European Review, Vol. 27, No. 3, 378–410 © 2019 Academia Europæa
doi:10.1017/S1062798719000061

From Unconscious Inference to the Beholder's Share: Predictive Perception and Human Experience

ANIL K. SETH

Sackler Centre for Consciousness Science, Department of Informatics, University of Sussex, Brighton BN1 9QJ, UK. Email: a.k.seth@sussex.ac.uk

Science and art have long recognized that perceptual experience depends on the involvement of the experiencer. In art history, this idea is captured by Ernst Gombrich's 'beholder's share'. In neuroscience, it traces to Helmholtz's concept of 'perception as inference', which is enjoying renewed prominence in the guise of 'prediction error minimization' (PEM) or the 'Bayesian brain'. The shared idea is that our perceptual experience – whether of the world, of ourselves, or of an artwork – depends on the active 'top-down' interpretation of sensory input. Perception becomes a generative act, in which perceptual, cognitive, affective, and sociocultural expectations conspire to shape the brain's 'best guess' of the causes of sensory signals. In this article, I explore the parallels between the Bayesian brain and the beholders' share, illustrated, somewhat informally, with examples from Impressionist, Expressionist, and Cubist art. By connecting phenomenological insights from these traditions with the cognitive neuroscience of predictive perception, I outline a reciprocal relationship in which art reveals phenomenological targets for neurocognitive accounts of subjectivity, while the concepts of predictive perception may in turn help make mechanistic sense of the beholder's share. This is not standard neuroaesthetics – the attempt to discover the brain basis of aesthetic experience – nor is it any kind of neuro-fangled 'theory of art'. It is instead an examination of one way in which art and brain science can be equal partners in revealing deep truths about human experience.

1. Introduction

> It is the power of expectation rather than the power of conceptual knowledge that moulds what we see in life no less than in art.[1]

How do we see? How can the complex operations of brains within bodies within worlds account for the rich phenomenology of visual experience? And are there

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

common principles uniting perception and phenomenology across modalities – and with experiences of selfhood?

These are challenges for science and art alike. A common intuition is that the function of perception is to recover an accurate – veridical – representation of some external state of affairs, of, for instance, a world full of objects of different shapes, sizes, and colours. This intuition is of course misleading. Perception can be 'for' a variety of purposes,[2] and both scientists and artists have long recognized that perception and phenomenology depend as much on the observer as on what is observed. William James, the father of modern psychology, said 'whilst part of what we perceive comes through our senses from the object before us, another part (and it may be the larger part) always comes out of our own head'.[3] For Georges Braque, 'objects don't exist … except insofar as a rapport exists between them, and between them and myself'.[4]

Despite these common ideas, neuroscience and artistic practice have followed largely separate trajectories regarding investigations of perception and phenomenology.[5] But now, a renewed enthusiasm for some old ideas about brain function, and a healthy scepticism about some narrow views on how science and art should relate, are opening new opportunities.

In terms of brain function, Hermann von Helmholtz's nineteenth-century notion of perception as unconscious inference has found new momentum in modern formulations of 'predictive processing', the Bayesian brain, and the 'free energy principle'.[6–8] On these formulations, which can all be subsumed under the general term 'prediction error minimization', [8] perceptual content is the result of probabilistic inference on the hidden causes of sensory signals. Generative models encoded by brain structure and dynamics make predictions about sensory inputs, based on prior expectations. Action (e.g. body movements) and perception conspire to reduce sensory prediction errors, giving rise to perceptual content and – perhaps more importantly – to predictive *regulation* of sensory variables.[9] More informally, what we perceive is the brain's 'best guess' of what's out there (or 'in here', for perceptions of body and self). This emerging perspective has found many applications in perception, action, learning, and cognition, and claims for its wider explanatory scope are extremely ambitious.[7,10] Despite this widespread attention, the implications of prediction error minimization (PEM) for human *experience* remain underexplored.

At the same time, there are suspicions that neuroscience is becoming increasingly colonial, with neuroscientific 'explanations' now offered in a host of contexts beyond the study of mind and brain. One of the most popular and most frequently criticized instance of colonial neuro-X is neuroaesthetics, which is usually understood as the study of the neurological basis of the creation and contemplation of works of art.[11–13] Considered as an extension of empirical aesthetics, there may well be value in characterizing neural mechanisms implicated in experiences of beauty, awe, or other subjective responses to artworks – insofar as these concepts can be adequately defined.[11,14] However, the notion that the subject matter of aesthetics will eventually yield to a reductionist story narrated in the language of brain networks has been justifiably resisted.[15–17] One reason for this resistance is a core assumption of

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

neuroaesthetics, sometimes implicit, that art and our responses to it are phenomena in need of explanation.

Another view is that art and (neuro)science are engaged in the same enterprise: the attempt to understand human experience. That they are complementary and should form a federal rather than colonial relationship. As Maurice Merleau-Ponty said, decades ago, the painter investigates *through painting* the means by which an object makes itself visible to our eyes.[18] More recently, Alva Noë, reviewing Eric Kandel's short book on art and brain science,[19] said 'Works of art do not merely cause experience; they also figure into the artist's ongoing effort, like that of science, to understand ourselves'.[20]

There are a number of examples of this federal approach, although the connections have usually been made with hindsight. For example, Patrick Cavanagh describes how visual artists developed insights into the perception of complex visual features such as shadows and transparency, noting how far artists could depart from the physical laws governing the relevant optics, while still preserving (or indeed enhancing) their perceptual impact.[21] Similarly, the powerful 'shimmering' of some Impressionist paintings has been attributed to the separate processing of colour and form in the visual system.[11,22,23] These examples, which stand for many others, illustrate the idea that the artists, through painting, had discovered 'neuroscientific' principles of human visual perception.

In this article, I develop this approach by exploring a striking parallel between two concepts: Helmholtz's notion of predictive perception, and the art-historical concept of *the beholder's share*, first introduced by Alois Riegl in the early twentieth century (as 'the beholder's involvement') and elaborated and popularized by Ernst Gombrich in his influential book *Art and Illusion*.[1]

Although the beholder's share has long been a staple of art history, a concise definition is hard to locate. For present purposes, we may set out a few properties which hew closely to the discussions in Ref. 1, and which apply principally to paintings. First, there is the core notion that the experience of viewing art depends on 'completion' of the image through the active involvement of the perceptual (i.e. not only conceptual) apparatus of the viewer. Second, this involvement:

> draws [the beholder] into the magic circle of creation and allows him to experience something of the thrill of 'making' which had once been the privilege of the artist. (Ref. 1, p. 202)

This marks a contrast with perception in general, which need not involve any experiential recapitulation of the creation of the causes that give rise to sensory impressions (e.g. the *making* of a painting). Third, there is a process of *guided projection* in which an artwork leads the viewer to contribute their perceptual expectations and memories in a particular way, while leaving sufficient ambiguity in the image for these projections to be assimilated into the resulting experience. Fourth and last, there is a proactive dimension in which the viewer (is guided to) test and update perceptual hypotheses:

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

> And what we call reading an image may perhaps better described as testing it for its potentialities, trying out what fits. (Ref. 1, p. 190)

This point emphasizes the iterative nature of perceptual inference, which is elaborated by Gombrich into expectations about the sensory consequences of actions such as head or eye movements; I will have more to say about this later on. Altogether, the beholder's share describes a process of active completion of an image through a guided process in which the viewer's perceptual expectations and memories are projected onto and into an image.

Taken together, these properties of the beholder's share raise immediate questions about which processes are specific to the viewing of works of art, and which are common to perception in general. That the former can shed light on the latter was recognized long ago, even by Helmholtz himself, who said in an 1871 lecture:

> We must look upon artists as persons whose observation of sensuous impressions is particularly vivid and accurate, and whose memory for these images is particularly true … The study of works of art will throw great light on the question as to which elements and relations of our visual impressions are most predominant in determining our conception of what is seen.[24]

In this quote we already see the seeds of the idea that art can investigate (e.g. through painting) the inferential mechanisms by which human perceptual systems transform noisy and ambiguous sensory signals into phenomenologically rich perceptual scenes. Of course, this idea only makes sense when considering perception in Helmholtzian terms – as the result of inference on the causes of sensory signals. It is striking that the above quote from Helmholtz appears in a critique of neuroaesthetics in which the author laments that 'Most visual scientists have abandoned Helmholtz's theory of vision' (Ref. 16, p. 25). Whether that was ever true is debatable, but it is certainly not the case now. The PEM view of cognition and perception is resurgent, and has been especially influential in research on visual perception.[6,25–27]

The re-emergence of generative Helmholtzian accounts of perception, in light of the insights into visual experience provided by the beholder's share, therefore offers a unique opportunity to develop an enriched understanding of perceptual experience. In what follows I take up this challenge, building on some previous accounts,[9,19,25,28,29] with the specific aim of exploring the value brought by considering perception as an intrinsically 'generative' or 'top-down' process, as compared with looking for parallels between art and neuroscience from a more standard 'bottom up' perspective on perception.

I will start by outlining some key concepts of the PEM view of perception and cognition, while calling on the artistic movement of Impressionism in 'reverse engineering' the visual system and focusing on conscious perception and phenomenology rather than unconscious inference. I then generalize the discussion to predictive perception of body and self, through the concept of *interoceptive inference*,[30] relating this

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

to Expressionist (and Abstract Expressionist) art which emphasizes emotional and embodied responses. I next consider how *active* perceptual inference – in which sensory prediction errors are minimized in part through action – shape the phenomenology of 'objecthood', calling on 'sensorimotor contingency theory' and its development in the context of PEM.[31,32] This discussion links to Cubism and the emergence of more abstract forms of visual art. I will then contextualize the discussion by reconsidering issues of epistemic gain, neuroaesthetics and of cultural influences on perception and art. Finally, I will return to the beholder's share itself, to consider what it now means in light of the insights provided by the cognitive neuroscience of predictive perception.

2. Predictive Perception and Impressionism

2.1. Predictive Coding and Generative Perception

The basic principle underlying (Helmholtzian) predictive perception is simple. It is that perception is in the business of inferring the causes of the inherently noisy and ambiguous signals that continually impinge on our various sensory surfaces: our eyes and our ears, but also our other senses including those originating from within the body (interoception). The many-to-many mappings between sensory signals and their hidden causes entails a Bayesian process in which prior beliefs or expectations about these causes are combined with representations of sensory data (the likelihood) to furnish a 'best guess' – or Bayesian posterior – specifying the most likely causes of current sensory inputs. These causes are referred to as hidden because they are never directly available to the brain, and always have to be inferred from their sensory consequences. Importantly, while we may be conscious of the results of this inferential process, we are not typically conscious of the process itself – hence Helmholtz's emphasis on *unconscious* inference. Also important is that the function of perception, on this view, is not necessarily to recover the most accurate representation of an external reality, but rather to enable the organism to engage with its environment (an environment that includes its own body) in ways which best support its behavioural goals.[33]

Operationally, predictive perception is usually considered in terms of predictive coding, which provides a process theory with a biologically plausible implementation and accumulating empirical support.[26,27,34–36] In predictive coding schemes (see Figure 1), neuronal representations in higher or deeper levels of neuronal hierarchies generate predictions about representations in lower levels.[37] These descending predictions are compared with lower-level representations to form a prediction error, which is passed back up the hierarchy to update higher-level representations. The recurrent exchange of signals between adjacent levels resolves prediction error at each and every level, resulting in a hierarchically deep explanation for sensory inputs (Figure 1A). In computational terms, the activity of neuronal populations is assumed to encode Bayesian beliefs or probability distributions over states in the world that cause sensations (e.g. my visual sensations are caused by a *face*). The relative

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

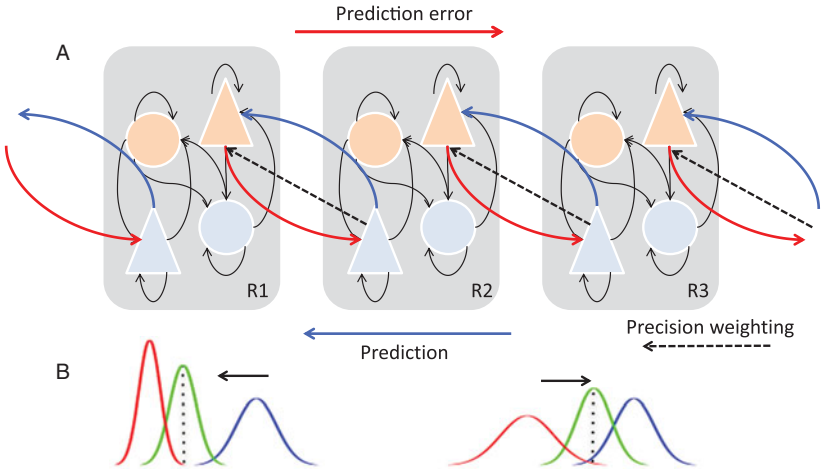


Figure 1. Predictive coding. (A) A schematic of hierarchical predictive coding across three cortical regions; the 'lowest' on the left (R1) and the 'highest' on the right (R3). Bottom-up (left-to-right) projections (red) originate from 'error units' (light orange) in superficial cortical layers and terminate on 'state units' (light blue) in the deep layers of their targets, whereas top-down (right-to-left) projections (dark blue) that convey predictions originate in deep layers and project to superficial layers of their targets. Prediction errors are associated with precisions, which determine the relative influence of bottom-up and top-down signal flow. Triangles represent pyramidal (projection) neurons; circles represent inhibitory interneurons. Solid black lines depict local circuit interactions wherein descending predictions are resolved with ascending prediction errors. (B) The influence of precisions on Bayesian inference and predictive coding. The curves represent probability distributions over the value of a sensory signal (*x*-axis). On the left, high precision-weighting of sensory signals (red, sharper distribution) enhances their influence on the posterior (green) and expectation (dotted line) as compared with the prior (blue). On the right, low precision-weighting (broader distribution) of sensory signals has the opposite effect on posteriors and expectations.[41] To view this figure in colour please see the online version of this journal.

influence of sensory signals on perception is governed by their (expected) precision. Precision is just a technical term for the reliability or confidence afforded to information – in this instance prediction errors. Mathematically, it is the inverse variance or dispersion (Figure 1B). Sensory data that are expected to be more reliable have sharper precision and therefore draw the Bayesian posterior further in their direction. The process of precision weighting has been associated with the role of attention in perception.[38]

A central assumption of the predictive coding architecture is that perceptual content is conveyed by top-down (or inside-out) connections, while bottom-up (outside-in) connections convey only (or largely) the prediction errors. This view of perception inverts classical frameworks in which perceptual content is assumed to depend on the hierarchical elaboration of bottom-up signals, with top-down signals providing at most some form of modulatory influence.[39,40]

In short, predictive coding provides a biologically plausible scheme for updating (typically implicit or unconscious) beliefs about the world, based on sensory signals.

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

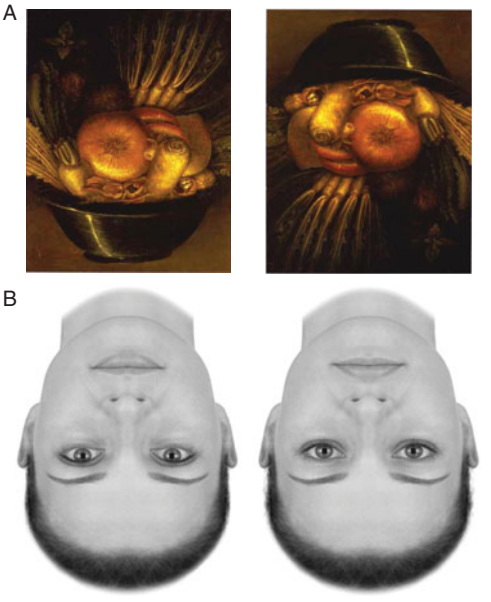


Figure 2. (A) Giuseppe Arcimboldo, *The Vegetable Gardener* (c.1590). Our percepts are constrained by what we expect to see. Arcimboldo, 'a 16th century Milanese artist who was a favourite of the Viennese, illustrates this dramatically by using fruits and vegetables to create faces in his paintings. When viewed right side up, the paintings are readily recognisable faces' (Ref. 28, p. 288). (B) Faces are probably one of the most important (hidden) causes of our sensations. While, in Arcimboldo's image, viewing right side up is needed for the configuration of features to appear as a face, when images are already recognizably faces, viewing right side up (by rotating the page) reveals that these faces might in fact be more different than they appear (this is the so-called 'Thatcher illusion' – though the face in this image is certainly not the former iron lady). These examples illustrate the complex interplay between prior expectations and stimulus features that shape perceptual content.[42]

In this view, cortical neuroanatomy and neurophysiology can be regarded as a distillation of statistical or causal structure in the environment that is revealed through sensation. The resulting anatomy and dynamics encode a *generative model* – generating predictions of sensations that can be compared with actual sensory samples. A powerful example of the influence of expectations on perception is given in Figure 2.

2.2. Impressionism and the Beholder's Share

The Helmholtzian notion that perceptual content is constitutively shaped by (explicit and implicit) expectations finds a natural complement in Gombrich's beholder's share. Eric Kandel, in his insightful *The Age of Insight*, puts it this way:

> 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. (Ref. 28, p. 204)

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press



Figure 3. *Hoarfrost at Ennery*, Camille Pissarro. Oil on canvas, 1873, Musee D'Orsay, Paris. Source: Wikimedia Commons.

This emphasis on the viewer's contribution to perception is especially evident in Impressionist art, which explores the idea that the painted image provides, not a detailed pictorial representation of some external situation, but the raw material to ignite perceptual and associative representations. In other words, the Impressionist painter attempts to

> create the conditions that enable the viewer to charge the percept, to complete the picture, based on his/her unique prior experiences. (Ref. 25, p. 241)

In Camille Pissarro's *Hoarfrost at Ennery* (Figure 3), 'pallete scrapings on a dirty canvas' are sufficient to provide the powerful perceptual impression of a sharply frosted field.[43] In one sense, Impressionism portrays an ideal of objectivity: to remove the artist from the act of painting; to recover the 'innocent eye'. In practice, this requires a deep understanding of the visual system along the lines advanced by Helmholtz. Pissarro's talent lies in 'reverse engineering' the visual system, to recover the afferent sensory signals that trigger a particular cascade of perceptual inference, rather than depicting the outcome of this process. Impressionist painting can therefore be understood as a series of experiments into the inferential operations of the visual system and – more broadly – into the nature of the subjective experiences entailed by these operations. These artistic 'experiments' complement contemporary neuroscientific attempts to reveal how top-down perceptual predictions underpin visual experience within the framework of PEM. Gombrich himself put it this way:

> When we say that the blots and brushstrokes of the Impressionist canvas 'suddenly come to life', we mean we have been led to project a landscape into these dabs of pigment. (Ref. 1, p. 170)

Of course, the idea of art as experiment neither begins nor ends with Impressionism. For example, earlier figurative art can be understood as

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

investigating the physical laws of optics – as exemplified by the beautifully detailed paintings of the seventeenth century Dutch artist Vermeer.[44] And as we will see, succeeding movements in modernist art such as Expressionism, Cubism, and various forms of abstraction, trace a distinctive path in the empirical investigation of visual phenomenology (in what is, of course, a highly complex history).[45] Nevertheless, the resonance between Helmholtzian inference and the beholder's share is perhaps sharpest for Impressionism. This resonance is exemplified by the insight that top-down influences do not simply 'fill in' details that might be 'missing' from sensory input;[19] rather, perceptual experience as a whole depends on reciprocal interactions between counter-flowing bottom-up and top-down signals across the entire visual field and across multiple hierarchical levels of the visual system.[46]

2.3. Perceiving the Expected

The subjective impact of Impressionism, seen through the lens of the beholder's share, raises a fundamental question for cognitive neuroscience: what aspects of predictive perception constitutively underpin visual experience? Addressing this question amounts to examining whether and how perceptual expectations determine conscious perceptual content. This turns out to be surprisingly difficult to do, since it is challenging to disentangle expectations from other cognitive processes such as attention and memory, and expectations can be of many different kinds – both implicit and explicit.

Controlling for these factors as far as possible, a number of studies now indicate that we are more likely to consciously see that which we expect, compared with that which violates our expectations. In one recent example, we used 'continuous flash suppression' to render an image (a house or a face) subjectively invisible for a period of time (Figure 4).[47] We found that if the image corresponded to a perceptual expectation it 'broke through' into consciousness faster than an unexpected image.[48] In other experiments, by orthogonally manipulating expectations and attention, we found that valid expectations enhance perceptual metacognition (the correspondence between accuracy and confidence when making a perceptual decision, which is often taken as a proxy for conscious perception).[49] Furthermore, we found that perceptual expectations exert their influence on perceptual content at preferred phases of the occipital alpha rhythm (a ~10 Hz brain oscillation that is particularly prominent over the visual cortex).[50]

These and other studies (e.g. Refs. 27, 35, 51–53) are collectively revealing the neurocognitive mechanisms by which perceptual expectations shape conscious visual content. However, the phenomenologically simplistic paradigms mandated by tight experimental control undersell the richness of real-world phenomenology. When held alongside the immersive experiential power of Impressionism, their paucity becomes even more apparent.

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

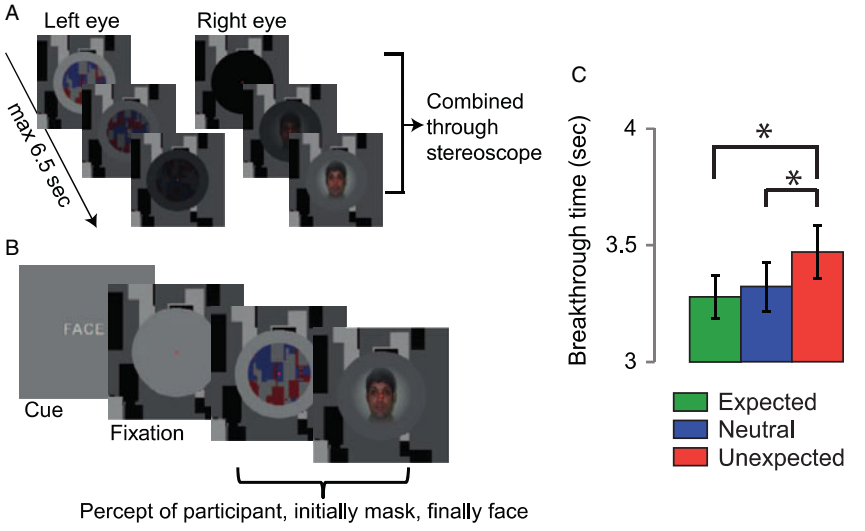


Figure 4. Perceptual expectations accelerate conscious access. (A) Continuous flash suppression design. A 'Mondrian' mask was presented to one eye, and a target image (face or house) to the other eye. Over time the contrast of the mask was gradually reduced and that of the target image was increased, so that the percept at some point transitioned from the mask to the target. (B) Participants were expecting either to see either a face or a house (cue) and were asked to press a button when they detected or were able to identify the target image. (C) Target images 'broke through' into consciousness faster for expected than for unexpected stimuli. Figure adapted from Ref. 48.

2.4. Peripheral Vision and Statistical Experience

Consider peripheral vision. Most laboratory experiments on predictive perception employ relatively familiar images (e.g. houses, faces) or rather nondescript psychophysical stimuli (e.g. Gabor patches) presented foveally (i.e. to central vision) at a fixed distance. Yet our visual experience encompasses a large and subjectively borderless peripheral area, outside the foveal/central region, which has a distinctive phenomenological character. Paul Cezanne's *Pine Tree near Aix* (Figure 5) provides a powerful Impressionistic analysis of peripheral phenomenology. Part of its power comes from the fact that the portion of the image representing the periphery is likely to be observed using central vision, under normal viewing conditions. The combination of hard contours and blended colour patches towards the edges of the painting may also contribute to the subjective effect.

It is well known that basic visual sensitivity in the periphery is altered, with a much lower density of light sensitive cells than found in the fovea. And within the peripheral population, there are many more rods than cones, compared with the foveal ratio.[54] These facts help explain some basic observations about peripheral vision, for example the greater sensitivity to motion and low light conditions (where rods are more sensitive and effective). However, not all features of peripheral phenomenology can be explained in this way.

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press



Figure 5. *Pine Tree Near Aix*, Paul Cezanne, 1890 s. State Hermitage, Saint Petersburg. Another powerful rendering of the peripheral phenomenology can be found in J.M.W. Turner's *Snowstorm: Steamboat Off a Harbour's Mouth* (1842).

One striking feature of peripheral phenomenology is its perceptual under-representation. Although foveal vision occupies a very small angular portion of the visual field, it seems in our phenomenology to occupy a much more extensive territory. Peripheral under-representation starts to make sense when considering the *cortical magnification* factor – the amount of cortex devoted to each degree of visual angle[54,55] – which is much larger for central vision than for the periphery. Figure 6 shows a recent artistic rendition of the phenomenology of peripheral under-representation, when paying attention to a particular object within a scene. As the artist notes, the resulting image is very different from a photographic



Figure 6. Robert Pepperell's depictions of global visual phenomenology, comparing foveal to peripheral experience: 'In each case I have tried to capture as faithfully as possible the actual appearance of the scene before me when fixating on a particular object within it'.[56] Left panel: *Still Life*, Robert Pepperell studio painting. Oil on shaped canvas, 2011. Right panel: *Self Portrait*, Robert Pepperell studio painting. Oil on shaped canvas, 2011. Images courtesy of the artist.

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

representation, in which the relative scaling of the various objects would be very different.[56]

Differences in cortical magnification may also account for other apparently surprising properties of peripheral vision. For example, typical visual phenomenology appears coloured throughout the visual field; at least, we do not experience the *absence* of colour in the periphery, despite the relative scarcity of colour-sensitive cone cells. However, when stimuli are scaled by the appropriate cortical magnification factor, peripheral colour experience may be more-or-less equivalent to central colour experience.[57,58]

The same strategy could be used to counter the common intuition that peripheral vision is a blurry or low-resolution version of central vision; perhaps, once cortical magnification is taken into account, it is again equivalent. However, rather than seeming blurry, peripheral phenomenology is perhaps better described as more textured or locally disordered; as 'statistical' in nature in comparison with foveal vision.[59,60] Recent studies using so-called 'perceptual metamers' seem to support this view.[61] These are images that are physically different but which look the same (when viewed from a certain distance with central fixation; see Figure 7). Rigorous psychophysical studies are needed to determine how perceptual metamerism relates to cortical magnification, and whether metamers reveal specific statistical features that are represented in peripheral phenomenology.[62]

From a beholder's share or Helmholtzian perspective, one might wonder whether peripheral phenomenology could result from top-down predictions informed by central vision. A simple example illustrating this possibility is given by the Cornsweet effect, in which peripheral regions of a simple colour-field stimulus are perceived as having different brightness levels, despite being physically isoluminant.[63] By assuming some simple prior expectations about the spatial organization of reflectance and



Figure 7. Perceptual metamers. (A) Undistorted image. (B) 'Metamerized' image. When viewed with central fixation (and at the appropriate distance) the images are subjectively indistinguishable, despite the metamerized image incorporating large distortions. Figure reprinted from Ref. 32. Original images provided courtesy of J. Freeman and E. Simoncelli.

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

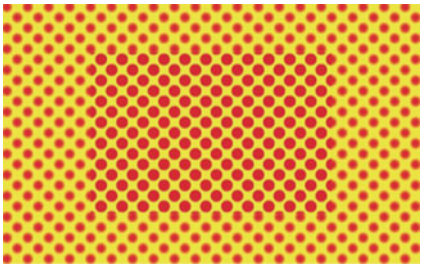


Figure 8. An example of the 'uniformity illusion' in which the periphery is blurrier than the centre. Fixating the centre of the image for a prolonged period should result in the peripheral region taking on the properties of the central region. From Ref. 65. Many more examples are available at www.uniformillusion.com.

luminance, it can be shown using computational modelling that the Cornsweet effect (and secondary phenomena such as Mach banding) emerge as a Bayes-optimal inference on the likely causes of sensory inputs.[64]

The Cornsweet effect is, perceptually speaking, extremely simple. A phenomenologically richer example is provided by the recently described 'uniformity illusion'.[65] In this illusion (see Figure 8), related but different features are presented in a central region of an image, and in the periphery. After fixating for several seconds, perception of the periphery takes on the properties of central vision, leading to a uniform phenomenological field. This is a powerful and robust effect that can be demonstrated across a broad range of stimulus features: luminance, orientation, texture, motion, and so on. While the neural processes underlying this illusion are not yet known, the Cornsweet example suggests that a Bayesian approach may be fruitfully applied. In particular, the subjective impression of the uniformity illusion suggests a predictive process in which imprecise peripheral predictions are replaced – over time – by more precise foveal predictions.

Overall, peripheral vision provides a fertile territory in which refined phenomenological intuitions can guide neurocognitive investigations to better elucidate the neural mechanisms that shape visual experience. Artworks, through their freedom from laboratory constraints, may catalyse this process through sharpening these phenomenological intuitions.

3. Interoceptive Inference and Expressionism

3.1. Interoceptive Inference and Emotional Experience

The classical exteroceptive modalities such as vision, audition and touch only account for part of the sensory input that continually streams into the brain. Two other major sources of sensory input are *proprioception*, which reports body position, posture and movement, and *interoception*, which collectively refers to a raft of sensory inputs signalling the internal physiological condition of the body (Figure 9).

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

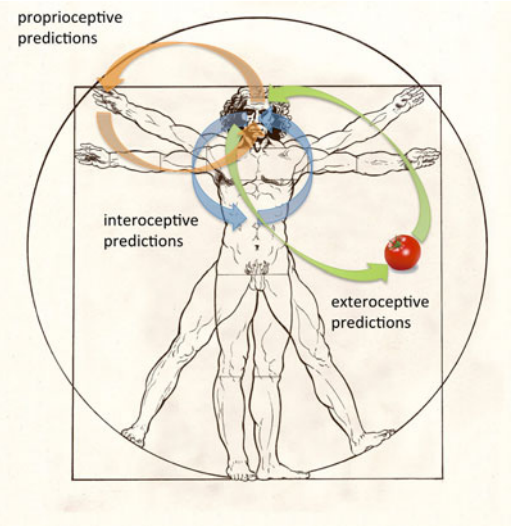


Figure 9. Inference and perception across different modalities. Green arrows represent exteroceptive predictions and prediction errors underlying perception of the external world. Orange arrows represent proprioceptive predictions (and prediction errors) generating actions. Blue arrows represent interoceptive predictions (and prediction errors) underlying emotional processing. From Ref. 9. To view this figure in colour please see the online version of this journal.

Recent extensions of PEM have sought to accommodate these additional modalities. Bodily actions have been proposed to result from the fulfilment of proprioceptive predictions,[66,67] and emotional experience and physiological regulation from interoceptive predictions.[9,30,68] As with predictive vision, the notion is that the corresponding perception (and/or behaviour) depends on the content of the descending predictions, rather than on the ascending prediction errors.

Applied to interoception, the specific idea is that emotional experience – and experiences of embodied selfhood generally – depend on the brain's 'best guess' of the hidden causes of interoceptive sensory signals. Informally, this is a simple generalization of earlier 'appraisal' theories. These earlier theories proposed that emotional responses depend on perceptions of changes in bodily physiology,[69] shaped by the cognitive context in which the changes occur.[70] Interoceptive inference brings several new dimensions to appraisal theories.[9,30] First, as with predictive vision, the neurocomputational vehicles of emotional experience are located in descending (top-down) predictions, rather than in the elaboration of bottom-up sensory input. Second, the sharp distinction between 'perception' and 'cognition' in theories of emotion is dissolved in favour of a smooth continuum of predictions and prediction errors at multiple hierarchical levels. Third, and related: the deeply hierarchical nature of perceptual inference entails that perceptual content is shaped not only by modality-specific predictions, but also by multimodal and amodal predictions that recruit and generalize across many perceptual modalities. In the setting of interoceptive inference this means that emotional experiences – and to a larger extent experiences of embodied selfhood generally – are

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

influenced by predictions about exteroceptive and proprioceptive signals, in addition to interoceptive predictions.[29,71]

The flip side of all this is that *visual* predictions will also be shaped by *interoceptive* (and possibly postural or proprioceptive) signalling. What this means, for the present discussion, is that the overall subjective experience of viewing an artwork will depend on the joint content of multimodal and amodal predictions that are elicited by the artwork together with the context in which it is being viewed. This trivially echoes the importance of context in the contemplation of art, but it does more than that. It provides a novel perspective on artistic traditions that are motivated by eliciting emotional responses to enhance or alter an artwork's subjective impact.

3.2. Affective Predictions and Expressionism

Expressionism originated in the late nineteenth and early twentieth century, primarily in Germany. While difficult to define precisely, a characteristic trait of Expressionism (at least in the visual arts) is that the artist seeks to elicit and give meaning to the emotions and responses that objects and events arouse within an observer.[72] This can be cast in terms of the beholder's share by recognizing that the emotional responses to an artwork depend on the prior expectations of the observer, where these expectations are now interoceptive and proprioceptive, as well as visual.

In Egon Schiele's *Prone Young Woman with Black Stocking* (Figure 10), the combination of angular lines, emphases and distortions of bodily shape, evoke in the viewer powerful emotional and bodily responses. This is a good example of Kandel's interpretation of the power of Expressionism:

> By exaggerating salient features [faces, hands, etc] and by showing the viewer how they produced these exaggerations, they invoke emotional primitives … and in this



Figure 10. *Prone Young Woman with Black Stocking*, Egon Schiele, 1913. Gouache, watercolour, and pencil on paper. Private collection.

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

> way uncover and bring to our conscious awareness the unconscious markers of our instinctive emotional systems. (Ref. 28, p. 447; Ref. 73)

A subtle distinction arises here. In one view of the relation between vision and emotion, affective responses are built into vision from the 'ground up,' and help shape perceptual inference, but emotional or affective content itself is not seen as being constituted by an inferential process. In this view, responses signalling an object's salience, relevance or value aid visual object recognition from the very moment that visual stimulation begins,[74] rather than after visual object recognition is completed. The current view encompasses these ideas but generalizes them to propose that all kinds of perceptual content – emotional and visual alike – are constituted by inference within and across modalities. This is a fully reciprocal, circular-causal relationship. Visual experience is shaped by interoceptive predictions, and in turn emotional responses in the viewer are shaped by visual predictions that are themselves informed by interoceptive inference.[9,30]

Admitting a constitutive role for interoceptive (and proprioceptive) predictions in shaping visual experience may constitute an extension of concept of the beholder's share, at least as it appears in Gombrich[1] where the focus is primarily on the relationship between perceptual expectations, memory and conceptual knowledge (see the Introduction). Interestingly, despite this focus, more recent references to the beholder's share have assumed a role for emotional responses. For example, Kandel describes Alois Riegl's first formulation of the 'the beholder's involvement' in the following terms:

> *[art] is incomplete without the perceptual and emotional involvement of the viewer.* (Ref. 28, p.189; emphasis in original)

Nonetheless, the current view underlines not only the importance of emotional associations and memories, but the constitutive nature of affective predictions in shaping visual experience.

Developing this argument further, Expressionist art uses representations of the human form to induce emotional responses, capitalizing on deep perceptual hierarchies that infer not just the immediate physical causes of sensations, but also the intentions and emotional states of other minds (and bodies) that underlie their physical appearance across time. This is made possible because the same inferential machinery that enables the brain to represent and control its own embodied state, can be deployed to infer the causes of similar reactions in others.[29,75] This idea has a great deal in common with concepts of 'mirror neurons' or other more general 'simulation' models in which our understanding of others' motivations, desires and actions arises from instantiating similar states in our own brains and bodies.[76,77]

Of course, artworks can evoke powerful emotional responses without any definite object-based representations at all – let alone representations of bodily forms. In Abstract Expressionism, 'action painters' such as Jackson Pollock and Willem de Kooning achieve this by working in a spontaneous, improvisatory manner.[45] In a second grouping, the vast abstract colour-field canvases of Mark Rothko, when

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press



Figure 11. *No. 61 (Rust and Blue)*, Mark Rothko, 1953. Museum of Contemporary Art, Los Angeles, Reproduced under fair use licence.

viewed in suitable gallery contexts (perhaps allowing sufficient engagement of peripheral perception), are capable of evoking overwhelming feelings of awe and spirituality (Figure 11). For Rothko, this depended on the breaking down of the finite and familiar associations that he considered to increasingly dominate society.[19,78] In its (historically complex) relations to Expressionism, the absence of bodily forms of any sort might entail that interoceptive and embodied predictions become necessarily less specific, eliciting emotional responses (awe, mysticism) that are themselves less embodied.[79]

Informally, Expressionism is sometimes portrayed as a response to, or reaction against, Impressionism. Instead of attempting to recover the 'innocent eye' by removing the artist from the act of painting, Expressionist art was meant to come from *within* the artist, to represent the swirl of emotional embodied feelings that define being an experiencing subject. The perspective advanced here instead suggests a continuity, in which Impressionism's reverse engineering of visual inference is generalized in Expressionism (and later, in the different forms of Abstract Expressionism) to embodied and interoceptive inference.

4. From Counterfactual Inference to Cubism and Abstraction

4.1. Counterfactual Predictions and the Phenomenology of Presence

Rene Magritte's ubiquitous Surrealist painting *The Treachery of Images* (1928–1929) makes plain the distinction between an object and an image of an object: in this case a pipe. When encountering a real pipe our visual experience is characterized

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

by the phenomenology of 'objecthood': we experience the pipe as really existing – as having *perceptual presence*. The phenomenology of presence is very common in our everyday experience and presents a puzzle for theories of perception. As Alva Noë puts it:

> How can it be true, as I think it is, that we are perceptually aware, when we look at a tomato, of the part of the tomato which, strictly speaking, we do not perceive. This is the puzzle of perceptual presence.[80]

Noë's preferred solution comes in the form of 'sensorimotor contingency theory',[31] which inherits from Gibsonian affordances,[2] and from 'enactive' approaches in cognitive science.[81] In this theory, perceiving is a skilful activity, which depends on the mastery of relevant laws of 'sensorimotor contingency'. For example, the phenomenology of redness is given by a 'practical mastery' of the regularities (sensorimotor contingencies) governing how red things behave under a variety of situations and actions. In this way, sensorimotor contingency theory can account for phenomenological differences among different modalities, since each modality will induce a distinct set of sensorimotor contingencies.

Sensorimotor contingency theory accounts for perceptual presence as follows. We experience presence or objecthood when we master the set of sensorimotor contingencies relevant to the property of 'being an object'. In Noë's example, we experience the (real) tomato as an object thanks to an implicit knowledge that certain sorts of actions would lead predictably to specific sensory inputs. For instance, we implicitly know that rotating a tomato will reveal a different part of it to our sensory surfaces. It is in precisely this sense that we become 'perceptually aware … of the part of the tomato which, strictly speaking, we do not perceive' (Ref. 80, p. 414).

This is a compelling intuition which provides a phenomenological gloss on the concept of form constancy in perception. However, sensorimotor contingency theory – at least within its original enactivist context – lacks any well-specified process theory or neurocognitive implementation. Fortunately, it is relatively easy to operationalize the core concepts of the theory in the framework of PEM.[32] This involves recognizing that sensory prediction errors can be reduced either by updating predictions or by performing actions to change the sensory data – which is termed *active inference*.[82] Critically, active inference implies the ability to predict the sensory consequences of different actions; in other words, the ability to make *conditional* or *counterfactual* predictions. As I have argued elsewhere,[32,83] the encoding, in brain structure and dynamics, of a repertoire of counterfactual predictions, linking potential actions to their sensory consequences, is equivalent to the mastery of a sensorimotor contingency in the enactivist tradition. In this account, the phenomenology of presence depends on the 'richness' or 'depth' (informally, the size of repertoire) of the counterfactual predictions relating to some specific inferred hidden causes of the sensorium.

To return to Noë's example: we experience the (real) tomato as a *present object* because our brain encodes a rich repertoire of predictions about how sensory signals would change under a variety of possible actions. This in turn implies a generative

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

model that has extracted, from sensory data, hidden causes that are invariant under these actions – i.e. a model that is hierarchically deep. A clear prediction from this theory is that changing the repertoire of action-related counterfactual predictions related to an object should change the level of experienced presence or 'objecthood' associated with that object. Some everyday observations are consistent with this: for instance, the perceptual experience of a uniform blue sky lacks a strong phenomenology of objecthood, and also lacks a rich dependency on any actions an observer might make (besides closing one's eyes).[83] Having said this, blue skies and tomatoes differ in many other ways and it stands as a challenge to manipulate the richness of action-related predictions, while preserving all other aspects of the perceptual encounter.[84]

4.2. Counterfactual Predictions and Cubism

When we view Magritte's *The Treachery of Images*, our visual experience encompasses the fact that it is a painting. We experience it along with its context, either in a gallery (with a particular frame, on a wall with such-and-such properties), or more likely on a particular computer screen in a room or an office. In these settings the picture (or computer monitor) may have the phenomenology of objecthood, while the depicted object does not – at least not to the same degree. Noë again: 'it is one thing to see a thing and another to see its depiction' (Ref. 17, p. 127).

The tension between the depiction and the depicted, with respect to objecthood, has been systematically explored in art ever since the formulation of laws of perspective hundreds of years ago (less systematic treatments of course extend much further back). Impressionist painters such as Cezanne began to experiment with these laws, in order to influence an observer's visual experience in ways that exceeded the capabilities of purely figurative art (as well as the emerging practice of photography). But it is with the emergence of Cubism in the early twentieth century, which was at least partly inspired by Cezanne's later works, that the phenomenology of objecthood is tackled head-on.[85]

Cubism investigated, through art, the representation and experience of three-dimensional forms – of objects. Pablo Picasso's *Still Life With Guitar* (1942) is a powerful example (Figure 12). In this image, objects are broken up, reassembled, and depicted from multiple points of view at the same time. The result is a challenging examination of the phenomenology of objecthood, given in a way that simultaneously emphasizes the flatness of the canvas itself. From a beholder's share perspective, Cubist art investigates not only the tension between the depiction and the depicted, but also the proactive and iterative character by which the viewer is guided to 'complete' the image through projection. The fragmented forms on the canvas continually challenge the beholder to develop and test new inferences that might provide a coherent 'explanation' of the sensory data – and perhaps also of the creative impulses that comprise their historically deep hidden causes. As with the bistable images so common in psychology textbooks, the impossibility of any such resolution is part of its power.

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

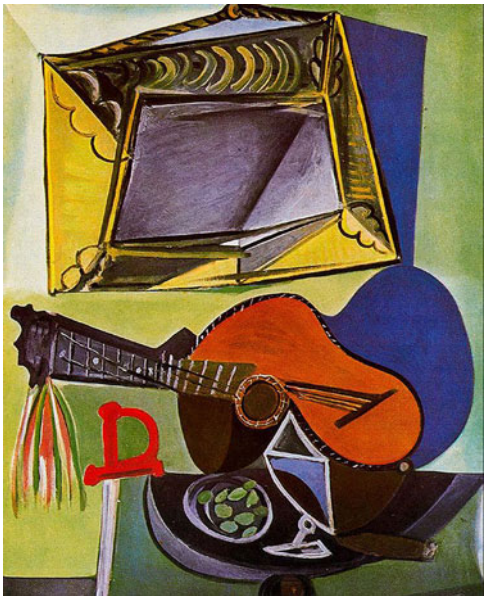


Figure 12. *Still Life With Guitar*, Pablo Picasso, 1942. Oil on canvas. Private collection. Image reproduced under fair use licence.

Such proactive and iterative testing of perceptual hypotheses strongly echoes the counterfactual aspects of perceptual inference that underpin the phenomenology of objecthood. Gombrich places this insight within a developmental context:

> Learning to 'see' may have much to do with the acquisition of expectations of serial orders, the sequence of shapes a chair or table will project on our retina as we move our head. (Ref. 1, p. 232)

This quote elegantly anticipates the importance of active inference in vision. The modern cognitive neuroscience perspective might only add that the acquisition and deployment of such expectations is not limited to the learning phase.

4.3. Epistemic Gain and Abstraction

Throughout the twentieth century, practices such as Cubism evolved into various forms of abstract art in which figurative or representational aspects were completely abandoned. In doing this, the artist further expands the scope of the beholder's share (as defined in the Introduction), going far beyond perceptual completion into a large territory of associations. In not depicting objects or scenes, the subjective power of abstract art rests even more strongly on the associations – perceptual, affective, and conceptual – evoked by the image. And as suggested earlier, in relation to Expressionism and interoceptive inference, the novelty brought by a Helmholtzian perspective is that the multifarious associations elicited by an abstract canvas may

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

not merely be subsequent to the visual perception. Instead, through the cascade of inside-out predictions, they may actually shape the visual experience itself.

It is a truism that individual reactions to abstract art vary wildly. One obvious reason for this is that each of us will bring different associations and memories to the act of beholding. And these differences may well be more diverse than, for example, individual variation in the functional neuroanatomy of the visual system that might underlie more basic aspects of vision, such as form perception.

Such a large variance across individual responses affords an opportunity to touch on the relationship between predictive perception and the pleasure elicited by beholding complex images. Kandel argues that the particular pleasure that one can derive from viewing abstract art reflects what James[3] called 'the victorious assimilation of the new': pleasure arises from the 'coherent perceptual experience of something we have never quite seen before – by its association with familiar things' (Ref. 19, p. 115). This ties the pleasure elicited by viewing very closely to the operation of the beholder's share. Indeed, Gombrich, identifies the pleasure of beholding with the cognitive effort required for (successful) interpretation:

> [W]e enjoy nothing more than the demand made on us to exercise our own 'imitative faculty', our imagination, and thus to share in the creative adventure of the artist. (Ref. 1, p. 236)

Considering these issues returns us to the territory of neuroaesthetics as the attempt to understand the brain basis of aesthetic experience. What can be said about aesthetic experiences from the perspective of the predictive brain? In their 'tentative prediction error account' of visual art, Van der Cruys and Wagemans[4] associate aesthetic pleasure with a transition from a state of uncertainty to a state of increased predictability, with regard to the top-down predictions brought to bear in the act of observing (see Kesner[86] for an interesting discussion and a detailed case study). This is closely related to theoretical proposals associating valence and affect with the rate of change of prediction error, in general formulations of the predictive brain,[87] and may further connect to ideas about the role of PEM in maintaining physiological homeostasis through predictive regulation.[9] Many interesting issues arise hereabouts – for example the role of negative affect in aesthetic experience (think of Goya's *Saturn Devouring His Son*), and the importance of engendering a certain inferential opacity (or 'obstinate obstruction', following Van der Cruys and Wagemans[4]) in the creation of an artwork, which the observer then attempts to overcome in the processing of beholding.

5. Discussion

This paper has not proposed any new theory of art, nor of neuroaesthetics. Its objective has instead been to explore how some forms of art, and some emerging (though simultaneously rather old) perspectives in cognitive neuroscience, can work together to elucidate generative contributions to human perception and

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

phenomenology. The basic idea is captured by the beholder's share in art, and by prediction error minimization (PEM) in neuroscience. The parallels between these concepts help bring into focus both relevant phenomenological explanatory targets, as well as potential neurocognitive mechanisms underlying these phenomenological properties.

The successive movements of figurative art, Impressionism, Expressionism and Cubism highlight a number of phenomenologically relevant features. There is the fundamentally generative nature of visual experience as emphasized in Impressionism: what we see is shaped by, but is not a direct representation of, some (hidden) state of affairs that are the distal causes of sensory inputs. Impressionism helps reveal the raw materials of natural visual experience and brings new impetus to investigations of challenging issues such as peripheral perception and phenomenology. Expressionism stresses the emotional and embodied involvement of the observer. Emotional responses do not follow or even simply accompany visual perceptual inference; rather, the act of beholding involves a coordinated engagement of exteroceptive (e.g. visual) and interoceptive (affective) predictions in shaping an overall subjective response. It is the whole organism, embodied and embedded, that engages in beholding – not the brain alone. Finally, Cubism emphasizes the phenomenology of objecthood and the tension between an artistic depiction of a thing, and the thing-in-itself, as well as the proactive and iterative nature of perceptual hypothesis testing in resolving ambiguity. This relates to emerging ideas about counterfactual or conditional sensorimotor predictions, which provide a neurocognitive operationalization of key concepts in sensorimotor contingency theory.

The emphasis on generative or top-down processes, both in the beholder's share and in PEM, contrasts with accounts linking neuroscience and art that focus on how artworks may elicit specific kinds of bottom-up activity in the visual system. (For example, see Zeki's discussion of Piet Mondrian's work in terms of functional specialization within the primary visual cortex.[88]) Instead, I have tried to address the challenge set out by Kesner (but not answered there), namely: 'to recast Gombrich's approach in the explanatory terms of predictive processing' (Ref. 86, p. 2). Doing this hopefully avoids some of the stronger criticisms of other approaches to linking art and brain science.[15–17] It is deliberately self-limiting in not proposing any general theory of art across formats, cultures or time. The focus is on art and neuroscience together exploring phenomenology, and not on a neuroscience *of* art or *of* aesthetics. This in turn means that art can be 'brought back into view' as a contextualized encounter between an observer and an art object, rather than being seen simply as a trigger for aesthetic experiences.[17,89] In addition, the rich neurocognitive framework of PEM helps avoid a tendency to propose overly reductionist neuroscientific explanations, which tend to emphasize bottom-up processing,[88] sometimes even while acknowledging the importance of top-down predictions.[19,28] Here, it is important to recognize the radical nature of full-fledged PEM, in which top-down signals are taken to *constitute* perceptual content, rather than merely modulating the activity of bottom-up sensory processing.

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

5.1. The Beholder's Share Revisited

Can the concept of the beholder's share, and its relation to Helmholtzian perception, be revised in light of the preceding discussion? The core of the concept, as identified in the Introduction, holds as strongly as ever: beholding involves active completion of an image through a guided process in which the viewer's perceptual expectations and memories are projected onto and into an image. Other aspects of Gombrich's concept also find clear resonances with contemporary expressions of Helmholtzian perception, in particular the proactive character of (possibly counterfactual) perceptual hypothesis testing, and its relation to learning to see and to the phenomenology of objecthood.

Does the Helmholtzian view enrich the beholder's share? One contribution is of course to provide a general neurocognitive operationalization of the concept in terms of PEM. The emphasis on affective (interoceptive) and counterfactual predictions may also be noteworthy. More specifically, both theory and experiment suggest that visual experience may not merely be *accompanied by* specific emotional and conceptual consequences: instead, top-down predictions that incorporate affective, embodied, and conceptual components may *constitutively shape* the corresponding visual experience. What is (visually) perceived, under the PEM view, depends on the brain's best guess of the causes of sensory inputs, in which guesses of all kinds can play a part.

The beholder's share, in turn, adds richness to the Helmholtzian view by drawing attention to the differences between perception in general and perception of artworks in particular. For Gombrich, the beholder's share draws the viewer into the process of creation of the artwork itself. This generalizes PEM to inference about the *historical* causes of current sensory inputs: in other words, one's perceptual experience of an artwork becomes partly constituted by predictions about the artist's motives and methods. Similarly, the notion of *guided projection* adds to Helmholtzian accounts of natural perception the artifice needed to manipulate the beholder and to enable their expectations or projections to find a foothold in the image. This notion may well have more general explanatory power, for example in understanding how joint attention shapes visual perception.[90]

5.2. Art Theory and the 'Period Eye'

Adopting a PEM perspective does not mean that contributions to theories of art or aesthetics are forever out of scope. Some tentative proposals, at least for aesthetic responses, have already been advanced.[86,91] And a deeper knowledge of the neurobiology of predictive perception no doubt helps understand subjective responses to specific artworks, including perhaps more general features of the traditions they represent.

Predictive perception may also shed light on the basis and reach of what has been called the 'period eye' in art theory;[92] broadly, the influence of social, cultural, and historical context on the perception and experience of artworks. While the concept

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

remains controversial,[93] it can be related to the concept of 'cognitive penetration'[94] in cognitive neuroscience by the question: can high-level beliefs and expectations (e.g. social or cultural stereotypes) drill down all the way to shape perceptual experience? In principle, the hierarchical structure of PEM admits different levels of abstraction without requiring any bright line separating perception from cognition. However, as an empirical issue, cognitive penetration itself is controversial, with early studies plagued by methodological concerns.[95] Despite these reservations, there now exists a core of compelling evidence that perceptual experience can indeed be influenced by high-level beliefs and cognitions,[96] including social expectations (e.g. stereotypes).[97] The PEM framework thus encompasses and operationalizes the 'period eye' at least as to the extent that it concerns influences of high-level cultural or social predictions about the causes of sensory signals.

5.3. Phenomenological Reverse Engineering

One exciting opportunity for future research may lie in emerging machine learning methods that recapitulate at least some aspects of the functional architecture of the visual system. So-called 'deep convolutional neuronal networks' (DCNNs) are now approaching or exceeding human-level performance in certain image classification tasks.[98,99] Despite utilizing feedforward architectures (i.e. lacking explicit generative models or top-down signalling), and although designed for engineering goals, some DCNNs demonstrate striking similarities with aspects of human visual processing. For example, deeper layers of trained DCNNs are sensitive to features that selectively activate deep layers of the visual ventral stream, when compared in terms of the representations these systems encode.[100,101] This line of research is rapidly developing computational models of visual perception that are increasingly faithful to the underlying neurobiology. Possibly, these models could be analysed in terms of those aspects supporting specific aspects of perceptual phenomenology.

One way to do this would be to use these models to interpret neuroimaging data[102] collected while perceptual experience is manipulated in phenomenologically relevant ways, perhaps by using stimuli designed with reference to works of art and presented immersively (e.g. in virtual reality). Another option is to run the models 'in reverse' so that an input image is modified according to the perceptual 'expectations' encoded in its networks. Using DCNNs 'in reverse' corresponds, intuitively, to generating the sensory impressions that would have been generated by the representations or expectations at higher (deeper) levels of the deep network. This inversion of 'recognition dynamics' highlights the fact that although they are remarkably good at classifying incoming stimuli, feedforward deep learning schemes struggle to reproduce the context sensitive and dynamic capabilities of human vision – precisely because they lack explicit generative models.

Inversion of recognition for standard feed-forward DCNNs has already been accomplished by so-called 'deep dream' algorithms. These algorithms enhance input patterns according to preferred features at user-selected layers in a DCNN

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

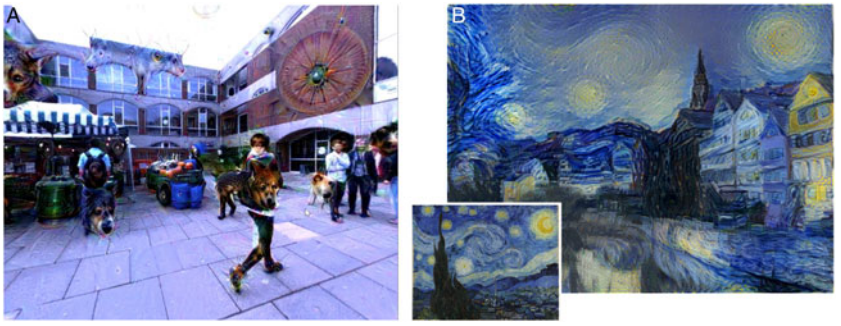


Figure 13. (a) A single frame from panoramic video taken on Sussex University campus and processed through a 'deep dream' algorithm, so that each frame is altered by the network's expectation that the image will contain a dog. Image courtesy of Keisuke Suzuki. (b) An example of how a DCNN can extract elements of artistic 'style' and reapply this 'style' to another image (in this case a photograph of buildings on the bank of a river).[106] Image courtesy of Leon Gatys.

(see Figure 13(a)).[103] Although the 'deep dream' procedure was initially developed to aid visualization of network behaviour, the imagery produced is rather striking and has been likened to a kind of hallucinogenic experience, which itself has been interpreted in terms of overactive perceptual expectations.[104,105]

The promise of this approach, as I see it, is not so much to develop 'computer art', but to advance an algorithmic understanding of the top-down processes that underlie specific phenomenological properties and, by extension, perhaps even distinct artistic styles (see Figure 13(b)).[106] Making good on this promise may be helped along by emerging types of deep neural network which are incorporating explicit generative elements,[107] recalling the innovative Helmholtz machines that preceded the DCNN revolution.[108,109]

5.4. Summary

In summary, Gombrich's beholder's share and Helmholtzian perception as inference both enhance our understanding of visual perception and phenomenology through an emphasis on top-down generative processes. In this view, art and brain science do not inhabit entirely different domains; nor is one subordinate to the other. Rather, both can work together to reveal deep truths about the nature of human experience. The phenomenological insights of art can provide new explanatory targets for cognitive neuroscience, while also highlighting the limitations of standard laboratory paradigms which substantially undersell the range and depth of human subjective experience.[110] At the same time, a deeper understanding of the neurocognitive basis of non-trivial phenomenological properties (objecthood, emotion, and so on) offers new opportunities for analysing the contextualized subjective responses to art that necessarily escape overly 'reductionist' accounts based – for example – on functional

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

specialization in the visual system. In this way, art and brain science can bring each other into proper focus, to the benefit of both.

Acknowledgements

I am grateful to the Dr Mortimer and Theresa Sackler Foundation, which supports the Sackler Centre for Consciousness Science, and also to the Wellcome Trust (Engagement Fellowship) and the Canadian Institute for Advanced Research (Azrieli Programme on Mind, Brain, and Consciousness). Many thanks to Urban Strandberg and Christian Munthe for inviting me to a forum 'Reconsidering Humanity' in Gothenburg (2015) from which this paper eventually emerged. I am also grateful to Tim Satterthwaite and the Courtauld Institute, to Margaret Boden for helpful comments, to Gary Hatfield and Ian Verstegen at the University of Pennsylvania for opportunities for discussion, and to my reviewers for their constructive comments.

References and Notes

1. E.H. Gombrich (1961) *Art and Illusion: A Study in the Psychology of Pictorial Representation* (Ewing, NJ: Princeton University Press), p. 188.
2. JJ. Gibson (1979) *The Ecological Approach to Visual Perception* (Hillsdale, NJ: Lawrence Erlbaum).
3. W. James (1890) *The Principles of Psychology* (New York: Henry Holt).
4. J. Richardson (Ed.) (1964) *Georges Braque: An American Tribute* (New York: Public Education Association). There are many other examples of this sentiment. John Dewey writes 'The product of art … is not the work of art. The work takes place when a human being cooperates with the product' – J. Dewey (2005) Art and Experience (New York: Penguin), p. 222. And Henri Matisse said 'Je ne peins pas les choses, je ne peins que les différences entre les choses' ['I do not paint things: I only paint the differences between things'] (L. Aragon (1971) Henri Matisse: Roman (Paris: Gallimard), p. 140). These quotes feature in S. Van der Cruys and J. Wagemans (2011) Putting reward in art: a tentative prediction error account of visual art. *Iperception*, 2, pp. 1035–1062, from where the translation is also taken.
5. C.P. Snow (1959) *The Two Cultures* (Cambridge: Cambridge University Press).
6. A. Clark (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, **36**(3), pp. 181–204. doi: 10.1017/S0140525X12000477.
7. K.J. Friston (2010) The free-energy principle: a unified brain theory? *Nature Reviews Neurosciences*, **11**(2), pp. 127–138. doi: nrn2787.
8. J. Hohwy (2013) *The Predictive Mind* (Oxford: Oxford University Press).
9. A.K. Seth and K.J. Friston (2016) Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **371**(1708). doi: ARTN 2016000710.1098/rstb.2016.0007.
10. K.J. Friston (2013) Life as we know it. *Journal of the Royal Society, Interface*, **10**(86), 20130475. doi: 10.1098/rsif.2013.0475

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

11. A. Chatterjee and O. Vartanian (2014) Neuroaesthetics. *Trends in Cognitive Science*, **18**(7), pp. 370–375. doi: 10.1016/j.tics.2014.03.003
12. V.S. Ramachandran and W. Hirstein (1999) The science of art: a neurological theory of aesthetic experience. *Journal of Consciousness Studies*, **6**(7), pp. 15–51.
13. S. Zeki (1999) Art and the brain. *Journal of Consciousness Studies*, **6**, pp. 76–96.
14. S. Zeki (2014) Neurobiology and the humanities. *Neuron*, **84**(1), 12–14. doi: 10.1016/j.neuron.2014.09.016
15. B.R. Conway and A. Rehding (2013) Neuroaesthetics and the trouble with beauty. *PLoS Biology*, **11**(3), e1001504. doi: 10.1371/journal.pbio.1001504
16. J. Hyman (2010) Art and neuroscience. In: R. Frigg and M. Hunter (Eds), *Beyond Mimesis and Convention* (The Netherlands: Springer Netherlands), vol. **262**, pp. 245–261.
17. A. Noë (2016) *Strange Tools: Art and Human Nature* (New York: Hill and Wang).
18. M. Merleau-Ponty (1964) Eye and mind. In: J.E. Edie (Ed.), *The Primacy of Perception* (Evanston, IL: Northwestern University Press), pp. 159–190.
19. E.R. Kandel (2016). *Reductionism in Art and Brain Science: Bridging the Two Cultures* (New York: Columbia University Press).
20. A. Noë (2016). Scientist's guide to modern art. *Science*, **353**(6305), p. 1215.
21. P. Cavanagh (2005) The artist as neuroscientist. *Nature*, **434**(7031), pp. 301–307. doi: 10.1038/434301a
22. For example, *Monet's Impression at Sunrise* (1872).
23. M. Livingstone (2002) *Vision and Art: The Biology of Seeing* (New York: Abrams).
24. H. Helmholtz (1995) On the relation of optics to painting. In: D. Cahan (Ed.), *Science and Culture: Popular and Philosophical Essays* (Chicago, IL: University of Chicago Press), pp. 279–308; p. 280.
25. T.D. Albright (2012) On the perception of probable things: neural substrates of associative memory, imagery, and perception. *Neuron*, **74**(2), 227–245. doi: 10.1016/j.neuron.2012.04.001.
26. R.P. Rao and D.H. Ballard (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, **2**(1), 79–87. doi: 10.1038/4580.
27. V. Weilnhammer, H. Stuke, G. Hesselmann, P. Sterzer and K. Schmack (2017) A predictive coding account of bistable perception - a model-based fMRI study. *PLoS Computational Biology*, **13**(5), e1005536. doi: 10.1371/journal.pcbi.1005536).
28. E.R. Kandel (2012) *The Age of Insight: The Quest to Understand the Unconscious in Art, Mind, and Brain, from Vienna 1900 to the Present* (New York: Random House).
29. K.J. Friston (2013). Book review: the fantastic organ. *Brain*, **136**(4), pp. 1328–1332.
30. A.K. Seth (2013) Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Science*, **17**(11), pp. 565–573. doi: 10.1016/j.tics.2013.09.007.
31. J.K. O'Regan and A. Noë (2001) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, **24**(5), pp. 939–973; discussion, pp. 973–1031.
32. A.K. Seth (2014) A predictive processing theory of sensorimotor contingencies: explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, **5**(2), pp. 97–118. doi: 10.1080/17588928.2013.877880.
33. A. Clark (2016) *Surfing Uncertainty* (Oxford: Oxford University Press).

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

34. A.M. Bastos, J. Vezoli, C.A. Bosman, J.M. Schoffelen, R. Oostenveld, J.R. Dowdall, … P. Fries (2015) Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron*, **85**(2), pp. 390–401. doi: 10.1016/j.neuron.2014.12.018.
35. N. Gordon, R. Koenig-Robert, N. Tsuchiya, J.J. van Boxtel and J. Hohwy (2017) Neural markers of predictive coding under perceptual uncertainty revealed with Hierarchical Frequency Tagging. *Elife*, **6**. doi: 10.7554/eLife.22749.
36. T.S. Lee and D. Mumford (2003) Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, **20**(7), pp. 1434–1448.
37. This description of predictive coding is closely adapted from A.K. Seth and K.J. Friston (2016) Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **371**(1708). doi: ARTN 2016000710.1098/rstb.2016.0007.
38. H. Feldman and K.J. Friston (2010) Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, **4**, p. 215. doi: 10.3389/fnhum.2010.00215.
39. D.J. Felleman and D.C. Van Essen (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, **1**(1), pp. 1–47.
40. D. Marr (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (New York: Freeman).
41. Adapted from A.K. Seth (2013) Interoceptive inference, emotion, and the embodied self. *Trends in cognitive Science*, **17**(11), pp. 565–573. doi: 10.1016/j.tics.2013.09.007.
42. Image from A.K. Seth and K.J. Friston (2016) Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **371**(1708). doi: ARTN 2016000710.1098/rstb.2016.0007.
43. The example is from T.D. Albright (2012) On the perception of probable things: neural substrates of associative memory, imagery, and perception. *Neuron*, **74**(2), 227–245. doi: 10.1016/j.neuron.2012.04.001.
44. L.J. Snyder (2015) *Eye of the Beholder: Johannes Vermeer, Antoni van Leewenhoek and the Reinvention of Seeing* (New York: Norton & Company).
45. H.H. Amason and E.C. Mansfield (2012) *History of Modern Art* (7th edn) (London: Pearson).
46. Kandel argues that the function of top-down processing is to resolve the remaining ambiguities that cannot be resolved by bottom-up processing alone. See E.R. Kandel (2016). *Reductionism in Art and Brain Science: Bridging the Two Cultures* (New York: Columbia University Press), p. 22. I think instead that top-down processing is constitutively involved in all perceptual experience.
47. N. Tsuchiya and C. Koch (2005) Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, **8**(8), pp. 1096–1101. doi: 10.1038/nn1500
48. Y. Pinto, S. van Gaal, F.P. de Lange, V.A. Lamme and A.K. Seth (2015) Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, **15**(8), p. 13. doi: 10.1167/15.8.13.
49. M.T. Sherman, A.K. Seth, A.B. Barrett and R. Kanai (2015) Prior expectations facilitate metacognition for perceptual decision. *Consciousness and Cognition*, **35**, pp. 53–65. doi: 10.1016/j.concog.2015.04.015.
50. M.T. Sherman, R. Kanai, A.K. Seth and R. VanRullen (2016) Rhythmic influence of top-down perceptual priors in the phase of prestimulus occipital alpha oscillations. *Journal of Cognitive Neuroscience*, **28**(9), pp. 1318–1330. doi: 10.1162/jocn_a_00973.

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

51. L. Melloni, C.M. Schwiedrzik, N. Muller, E. Rodriguez and W. Singer (2011) Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. *Journal of Neuroscience*, **31**(4), pp. 1386–1396. doi:31/4/1386/10.1523/JNEUROSCI.4570-10.2011.
52. L. Muckli, F. De Martino, L. Vizioli, L.S. Petro, F.W. Smith, K. Ugurbil, … E. Yacoub (2015) Contextual feedback to superficial layers of V1. *Current Biology*, **25**(20), pp. 2690–2695. doi: 10.1016/j.cub.2015.08.057.
53. A. Pajani, P. Kok, S. Kouider F.P. de Lange (2015) Spontaneous activity patterns in primary visual cortex predispose to visual hallucinations. *Journal of Neuroscience*, **35**(37), pp. 12947–12953. doi: 10.1523/JNEUROSCI.1520-15.2015.
54. H. Strasburger, I. Rentschler and M. Juttner (2011) Peripheral vision and pattern recognition: a review. *Journal of Vision*, **11**(5), p. 13. doi: 10.1167/11.5.13.
55. S. Anstis (1998) Picturing peripheral acuity. *Perception*, **27**(7), pp. 817–825. doi: 10.1068/p270817; H. Strasburger, I. Rentschler and M. Juttner (2011) Peripheral vision and pattern recognition: a review. *Journal of Vision*, **11**(5), p. 13. doi: 10.1167/11.5.13.
56. R. Pepperell (2012) The perception of art and the science of perception. *Proceedings of SPIE, 8291, Human Vision and Electronic Imaging XVII* (p. 829113). Cortical magnification is one reason why so many sunrise or sunset photographs are disappointing. When we experience a real sunset, we typically fixate on the setting sun so that it benefits from cortical magnification. When we look at a photo, the whole photo is usually within foveal vision and so the whole photo is cortically magnified (but not the real world-scene surrounding the photo). The relative magnification of the sun within the frame of the photo is therefore missing. Zoom lenses only partially fix the problem since they uniformly reduce the span of the captured part of the visual field, rather than adjusting its relative scaling.
57. A.M. Haun, G. Tononi, C. Koch and N. Tsuchiya (2017) Are we underestimating the richness of visual experiences? *Neuroscience of Consciousness*, **3**(1), pp. 1–4.
58. C.W. Tyler (2015) Peripheral color demo. *Iperception*, **6**, pp. 1–5.
59. B. Balas, L. Nakano and R. Rosenholtz (2009) A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, **9**(12), p. 13, pp. 11–18. doi: 10.1167/9.12.13.
60. J.Y. Lettvin (1976) On seeing sidelong. *The Sciences*, **16**, pp. 10–20.
61. J. Freeman and E.P. Simoncelli (2011) Metamers of the ventral stream. *Nature Neuroscience*, **14**(9), pp. 1195–1201. doi: 10.1038/nn.2889.
62. T.S. Wallis, M. Bethge and F.A. Wichmann (2016) Testing models of peripheral encoding using metamerism in an oddity paradigm. *Journal of Vision*, **16**(2), p. 4. doi: 10.1167/16.2.4.
63. T. Cornsweet (1970) *Visual Perception* (New York: Academic Press).
64. H. Brown and K.J. Friston (2012) Free-energy and illusions: the Cornsweet effect. *Frontiers in Psychology*, **3**, p. 43. doi: 10.3389/fpsyg.2012.00043.
65. M. Otten, Y. Pinto, C.L. Paffen, A.K. Seth and R. Kanai (2017) The uniformity illusion. *Psychological Science*, **28**(1), pp. 56–68. doi: 10.1177/0956797616672270. The uniformity illusion recalls and substantially generalizes the earlier 'healing grid' illusion, see: http://illusionoftheyear.com/2005/08/healing-grid/.
66. R.A. Adams, S. Shipp and K.J. Friston (2013) Predictions not commands: active inference in the motor system. *Brain Structure and Function*, **218**(3), pp. 611–643. doi: 10.1007/s00429-012-0475-5.

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

67. K.J. Friston, J. Daunizeau, J. Kilner and S.J. Kiebel (2010) Action and behavior: a free-energy formulation. *Biological Cybernetics*, **102**(3), pp. 227–260. doi: 10.1007/s00422-010-0364-z.
68. L.F. Barrett and W.K. Simmons (2015) Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, **16**(7), pp. 419–429. doi: 10.1038/nrn3950.
69. W. James (1894) The physical basis of emotion. *Psychological Review*, **1**, pp. 516–529.
70. S. Schachter and J.E. Singer (1962) Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, **69**, pp. 379–399.
71. A.K. Seth (2015) The cybernetic Bayesian brain: from interoceptive inference to sensorimotor contingencies. In: J.M. Windt and T. Metzinger (Eds), *Open MIND* (Frankfurt A.M.: MIND Group), pp. 1–24. An important aspect of interoceptive predictions is that they are likely to be geared towards control of (the causes of) physiological signals, in order to ensure adaptive homeostatic regulation for the organism. The phenomenological consequences of such control-oriented inference are discussed, provisionally, in these works and A.K. Seth and K.J. Friston (2016) Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 371(1708). doi:ARTN 2016000710.1098/rstb.2016.0007. Understanding how these consequences may play into the perception of affectively laden art offers an interesting opportunity for future research.
72. J. Lloyd (1991) *German Expressionism: Primitivism and Modernity* (New Haven, CT: Yale University Press).
73. Saying this in no way condones Ramachandran's 'peak shift' theory, see V.S. Ramachandran and W. Hirstein (1999) The science of art: a neurological theory of aesthetic experience. *Journal of Consciousness Studies*, **6**(7), pp. 15–51. For a powerful critique of this theory see S. Zeki (2014) Neurobiology and the humanities. Neuron, 84(1), 12–14. doi:10.1016/j.neuron.2014.09.016; B.R. Conway and A. Rehding (2013) Neuroaesthetics and the trouble with beauty. PLoS Biology, 11(3), e1001504. doi:10.1371/journal.pbio.1001504; J. Hyman (2010) Art and neuroscience. In: R. Frigg and M. Hunter (Eds), *Beyond Mimesis and Convention* (The Netherlands: Springer Netherlands), vol. 262, pp. 245–261.
74. L.F. Barrett and M. Bar (2009) See it with feeling: affective predictions during object perception. *Philosophical Transactions of the Royal Society of London B, Biological Science*, **364**(1521), pp. 1325–1334. doi: 364/1521/1325/10.1098/rstb.2008.0312.
75. K.J. Friston, J. Mattout and J. Kilner (2011) Action understanding and active inference. *Biological Cybernetics*, **104**(1-2), pp. 137–160. doi: 10.1007/s00422-011-0424-z.
76. J.M. Kilner, K.J. Friston and C.D. Frith (2007) Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, **8**(3), pp. 159–166. doi: 10.1007/s10339-007-0170-2.
77. An interesting complication here is that when predictive models are used to infer the properties of others, the corresponding prediction errors need to be suppressed. For example, if proprioceptive prediction errors remain unsuppressed when proprioceptive predictions are deployed to infer the hidden causes of another's actions, the result will be mimicry (maybe even echopraxia), rather than action understanding. The same principle applies to (interpersonal) interoceptive inference. This speaks to a fundamental role for neuromodulation in attenuating the expected precision of prediction errors, with failures in this

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

process potentially underlying unusual social responses (possibly also unusual responses to art) and, at extremes, a variety of neuropsychiatric conditions. See J.M. Kilner, K.J. Friston and C.D. Frith (2007) Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, **8**(3), pp. 159–166. doi: 10.1007/s10339-007-0170-2.

78. W. Spies (2011) *The Eye and the World: Collected Writings on Art and Literature (Vol. 8: Between Action Painting and Pop Art)* (New York: Abrams).
79. Rothko's early work (ca. 1930s) was in fact characterized by a blend of Expressionism and Surrealism. See for example his Couple Kissing (1934).
80. A. Noë (2006) Experience without the head. In T. Gendler and A. Hawthorne (Eds), *Perceptual Experience* (New York: Clarendon/Oxford University Press), pp. 411–434, p. 414.
81. E. Thompson (2007) *Mind in Life: Biology, Phenomenology, and the Sciences of Mind* (Cambridge, MA: Harvard University Press).
82. K.J. Friston, R.A. Adams, L. Perrinet and M. Breakspear (2012) Perceptions as hypotheses: saccades as experiments. *Frontiers in Psychology*, **3**, p. 151. doi: 10.3389/fpsyg.2012.00151; K.J. Friston, J. Daunizeau, J. Kilner and S.J. Kiebel (2010) Action and behavior: a free-energy formulation. *Biological Cybernetics*, **102**(3), pp. 227–260. doi: 10.1007/s00422-010-0364-z.
83. A.K. Seth (2015). Presence, objecthood, and the phenomenology of predictive perception. *Cognitive Neuroscience*, **7**, pp. 1–7.
84. We are currently addressing this challenge using methods from augmented reality, in which virtual objects can be programmed to behave in different ways during interactions with an observer.
85. J. Golding (1988) *Cubism: A History and an Analysis, 1907-1914* (3rd edn) (Cambridge, MA: Harvard University Press).
86. L. Kesner (2014) The predictive mind and the experience of visual art work. *Frontiers in Psychology*, **5**, p. 1417. doi: 10.3389/fpsyg.2014.01417.
87. M. Joffily and G. Coricelli (2013) Emotional valence and the free-energy principle. *PLoS Computational Biology*, **9**(6), e1003094. doi: 10.1371/journal.pcbi.1003094.
88. S. Zeki (1999). *Inner Vision: An Exploration of Art and the Brain* (Oxford: Oxford University Press).
89. J. Dewey (2005) *Art and Experience* (New York: Penguin), p. 222.
90. A. Frischen, A.P. Bayliss and S.P. Tipper (2007) Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin*, **133**(4), pp. 694–724. doi: 10.1037/0033-2909.133.4.694.
91. S. Van der Cruys and J. Wagemans (2011) Putting reward in art: a tentative prediction error account of visual art. *Iperception*, **2**, pp. 1035–1062
92. M. Baxandall (1972) *Painting and Experience in the Fifteenth Century: A Primer in the Social History of Pictorial Style* (Oxford: Oxford University Press).
93. N. Carroll (2001) Modernity and the plasticity of perception. *The Journal of Aesthetics and Art Criticism*, **59**(1), pp. 11–17.
94. Z. Pylyshyn (1980) Computation and cognition: issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, **3**(1), pp. 111–132.
95. C. Firestone and B.J. Scholl (2016) Cognition does not affect perception: evaluating the evidence for 'top-down' effects. *Behavioral and Brain Sciences*, **39**, p. e229. doi: 10.1017/S0140525X15000965.
96. See collected papers in A. Newen, F. Marchi and P. Brossel (2017) Introduction – cognitive penetration and predictive coding. Pushing the

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

debate forward with the recent achievements of cognitive science. *Consciousness and Cognition*, **47**, pp. 1–5. doi: 10.1016/j.concog.2016.12.001.
97. M. Otten, A.K. Seth and Y. Pinto (2016) A social Bayesian brain: how social knowledge can shape visual perception. *Brain and Cognition*, **12**, pp. 69–77. doi: 10.1016/j.bandc.2016.05.002.
98. Y. LeCun, Y. Bengio and G. Hinton (2015) Deep learning. *Nature*, **521**(7553), pp. 436–444. doi: 10.1038/nature14539.
99. T. Horikawa and Y. Kamitani (2017) Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications*, **8**, 15037. doi: 10.1038/ncomms15037.
100. S.M. Khaligh-Razavi and N. Kriegeskorte (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, **10**(11), e1003915. doi: 10.1371/journal.pcbi.1003915.
101. N. Kriegeskorte (2015) Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, **1**(1), 594.
102. Following S.M. Khaligh-Razavi and N. Kriegeskorte (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, **10**(11), e1003915. doi: 10.1371/journal.pcbi.1003915.
103. See Figure 13a in A. Mordvintsev, C. Olah and M. Tyka (2015) Deep dream – a code example for visualizing neural networks. Retrieved from https://research.googleblog.com/2015/07/deepdream-code-example-for-visualizing.html
104. P.C. Fletcher and C.D. Frith (2009) Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews. Neuroscience*, **10**(1), pp. 48–58. doi: nrn2536/10.1038/nrn2536;.
105. C. Teufel, N. Subramaniam, V. Dobler, J. Perez, J. Finnemann, P.R. Mehta, … P.C. Fletcher (2015) Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Science USA*, **112**(43), pp. 13401–13406. doi: 10.1073/pnas.1503916112.
106. A. Gatys, A.S. Ecker and M. Bethge (2015) A neural algorithm of artistic style. Retrieved from https://arxiv.org/abs/1508.06576.
107. For instance, A. Radford, L. Metz and S. Chintala (2016) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv, 1511.06434v06432.
108. P. Dayan, G.E. Hinton, R.M. Neal and R.S. Zemel (1995) The Helmholtz machine. *Neural Computing*, **7**(5), pp. 889–904.
109. G.E. Hinton and P. Dayan (1996) Varieties of Helmholtz machine. *Neural Networks*, **9**(8), pp. 1385–1403. doi: S0893608096000093 [pii].
110. C.W. Tyler (2015) Peripheral color demo. *Iperception*, **6**, pp. 1–5; E. Podvalny, E. Yeagle, P. Megevand, N. Sarid, M. Harel, G. Chechik, … R. Malach (2017) Invariant temporal dynamics underlie perceptual stability in human visual cortex. *Current Biology*, **27**(2), pp. 155–165. doi: 10.1016/j.cub.2016.11.024.

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press

About the Author

Anil Seth is Professor of Cognitive and Computational Neuroscience at the University of Sussex and Founding Co-Director of the Sackler Centre for Consciousness Science. He is also a Wellcome Trust Engagement Fellow, a Senior Fellow of the Canadian Institute for Advanced Research, and Editor-in-Chief of *Neuroscience of Consciousness* (Oxford University Press). In his work, Anil seeks to understand the biological basis of consciousness by bringing together research across neuroscience, mathematics, artificial intelligence, computer science, psychology, philosophy and psychiatry. He has published more than 120 academic papers in a variety of fields, and he holds degrees in Natural Sciences (MA, Cambridge, 1994), Knowledge-Based Systems (MSc, Sussex, 1996) and Computer Science and Artificial Intelligence (DPhil, Sussex, 2000). He was also a Postdoctoral and then Associate Fellow at the Neurosciences Institute in San Diego, California (2001–2006).

https://doi.org/10.1017/S1062798719000061 Published online by Cambridge University Press