Visual illusions and neurobiology

David M. Eagleman

The complex structure of the visual system is sometimes exposed by its illusions. The historical study of systematic misperceptions, combined with a recent explosion of techniques to measure and stimulate neural activity, has provided a rich source for guiding neurobiological frameworks and experiments.

The act of ‘seeing’ seems so effortless that it is difficult to appreciate the vastly sophisticated — and poorly understood — machinery that underlies the process. Illusions, often, are those stimuli that exist at the extremes of what our system has evolved to handle. Sometimes illusions stem from assumptions made by the visual system; at other times they represent an active recalibration. In all these cases, illusions serve as a powerful window into the neurobiology of vision, and have pointed towards new experimental techniques.

There is some difficulty in rigorously defining ‘illusion’, as there is a sense in which all of vision is an illusion. The resolution in our peripheral vision is roughly equivalent to looking through a frosted shower door, and yet we enjoy the illusion of seeing the periphery. Peripheral vision is an illusion. The resolution in our vision has limits; this is also an illusion of sorts. We depend on other illusions as a normal aspect of our lives: in the cinema, we pay money to watch a succession of flat, still images that appear to be rich with motion and depth. For our present purposes, I will review several categories of illusion that have been more traditionally explored in history and in modern neuroscience. The purpose of this article is to illustrate how such illusions have helped to guide neuroscience research. To that end, I will attempt to trace the growth of intellectual threads that directly led to insights into — and placed constraints on — the underlying neural mechanisms of vision. I will also attempt to highlight as much about our ignorance of illusions as about our understanding, in the hope of lighting the way to interesting new problems.

In the past, illusions were sometimes considered to be inappropriate objects of study. The nineteenth-century psychologist Oswald Kulpe expressed the intellectual climate of the era when he wrote that perceptual illusions are “subjective perversions of the contents of objective perception”. This is why Exner’s experiments on apparent motion2 in 1875 did not receive a great deal of attention, until Max Wertheimer, defining the Gestalt movement almost 40 years later, re-examined apparent motion in a climate in which the study of illusions had changed3.

We now appreciate that the systematic study of illusions provides important clues to the neural architecture and its constraints. This appreciation drives, and is driven by, a flood of physiological data from awake, behaving primates, both monkey and human.

Illusions from lateral interaction

In 1865, Ernst Mach reported illusory bands of bright and dark (Mach bands) at the edges of a luminance ramp differing in brightness: consonant Mach bands4. (FIG. 1a). Five years later, Ludimar Hermann was reading a book that contained a set of figures organized in a grid, and noticed between the figures grey spots that disappeared when he looked directly at them (Hermann grid; FIG. 1b). At the time of these reports, a rebellious teenager named Santiago Ramón y Cajal was apprenticed as a shoemaker; years later, in 1887, Ramón y Cajal began to experiment with Camillo Golgi’s technique of silver impregnation of neural tissue. Cajal began to view the nervous system as being made up of billions of separate nerve cells, instead of a continuous network. That idea, known as the Neuron Doctrine, established the heart of modern neuroscience. However, knowing that the brain is composed of billions of cells tells us little about how those cells encode information and the principles of their interaction. Illusions, from Mach’s to Hermann’s and scores of others, have contributed to our understanding of this interaction. In fact, illusions such as the Hermann grid and Mach bands led to one of the earliest neural theories: that of lateral...
interaction between cells. We now understand that from the retina onwards, neurons characteristically inhibit or excite their neighbours, depending on their connectivity. This allows the nervous system to enhance contrast between similar regions. Contrast enhancement is generally beneficial, but in some instances generates illusory percepts. In the Zollner illusion, for example, the visual cortex enhances orientation contrast by making similar orientations seem to tilt away from each other (FIG. 1c).

Classical receptive fields, defined as the area in which a visual stimulus evokes a change in the firing activity of a cell, can explain local perceptual effects such as the Hermann grid and Mach bands, but are not sufficient to explain global perceptual effects. Not long after the discovery of the classical receptive field, physiologists realized that the response of a neuron to a stimulus could be significantly affected by stimuli presented outside the receptive field. These surround effects indicate that individual neurons can integrate information over large areas of the visual cortex. Neural data on the surround effect has been informed, in part, by a long history of illusions. For example, in simultaneous contrast — described two millennia ago by Aristotle and again one millennium later by Alhazen of Cairo — surrounding colours or luminances influence the perception of a central target. In FIG. 1d, the grey patches in the middle of the coloured squares are physically identical, even though they appear to be different. Even in the absence of detailed knowledge of neural tissue, Mach correctly guessed that simultaneous contrast involved some sort

Figure 1 Illusions arising from lateral inhibition and excitation. a | Mach bands are the illusory bright and dark lines to the left and right of the luminescence gradient that connects the two uniform regions. Receptive fields in the uniform regions have a balance between their excitatory centres and inhibitory surrounds. However, a receptive field centred on the bright Mach band gives a stronger response because part of the surround is in the darker area (and so this field receives less inhibition from the surround). Conversely, the receptive field over the dark band receives more surround inhibition because part of the surround is in the brighter area. b | The Hermann grid illusion, in which illusory grey spots are perceived at the intersections, is seen because a retinal ganglion cell receptive field lying at the intersection of the cross has more light falling on its inhibitory surround than a receptive field that lies between two black squares. Consequently, its excitatory centre is suppressed to weaker activity. c | In the Zollner illusion, the parallel long lines appear non-parallel. Lateral interactions increase orientation contrast, giving the impression of an increased angle between the long line and the shorter lines that cross it. d | In simultaneous colour contrast, the colours of the background affect the way in which we perceive the centre grey patch. Although the two patches are identical, most observers see the left patch tinged with blue and the right patch tinged with yellow.
of neural comparison of sensations from neighbouring image regions. Ewald Hering shortly thereafter developed the related idea that reciprocal interactions in the ‘neural image’ determined much of surface colour appearance. Recently, physiological recordings have grounded these ideas. For example, if the receptive field of a V1 cell is covered with uniform grey illumination, changing the luminance of the surrounding regions will change the firing of a cell in a manner consistent with the psychophysics of simultaneous contrast. The neural and psychophysical data continue to inform each other as more work unravels how modulations caused by horizontal and feedback connections might reflect the integration of information that underlies perception. For example, the illusion that apparent motion produced by a sequence of collinear objects is perceived as faster than the same sequence of non-collinear objects seems to be consistent with properties of the horizontal axons that form the ‘silent’ surround of cortical receptive fields.

**Illusory contours**

The Kanizsa square in FIG. 2a shows that we can perceive the borders of an object even in regions of the image where there is no direct visual evidence for them. This is one example of the phenomenon of illusory or subjective contours. Illusory contours, and the concomitant brightness enhancement, have a rich history in psychology. Beginning in 1984, this perceptual phenomenon was reported to have a direct physiological correlate in macaque area V2, where some neurons were found to respond to an illusory contour moving across their receptive fields, and in 1993, responses were reported in V1 (REF. 20). The V1 responses show longer latencies than V2 responses, indicating that the brain fills in illusory contours on the basis of feedback from higher areas (FIG. 2b). A full understanding of how and why the visual system constructs these contours awaits further research.

**After-effects**

In the late 1870s, Hering noted that after an observer fixates steadily on a blue patch, he sees an illusory yellowish patch. To observe this simple colour after-effect, return to FIG. 1d and fixate the black dot between the two coloured squares for 30 seconds. Shifting your gaze to the black dot at the bottom will generate afterimages of complementary colours. Hering’s examination of these colour after-effects, and the distinct pattern of their colours, led him to propose an ‘opponent process’ theory of vision. In this framework, trichromatic signals from the cones (as postulated by Helmholtz) feed into subsequent neural stages and show opponent processing (red versus green, yellow versus blue, black versus white). Hurvich and Jameson later developed Hering’s theory, and in the 1960s, physiological recordings from primate lateral geniculate nucleus emerged to support this opponent model of colour processing. A natural consequence of opponent processing is the idea that competing neural populations exist in a balance of tonic activity. According to this view, a subpopulation can be ‘fatigued’, and another population can dominate the push–pull competition and briefly control the percept. This idea has been popular as an explanation for several kinds of after-effect. One example is the motion after-effect (also known as the waterfall illusion), which has enjoyed a particularly rich history of study, dating back at least to Aristotle. To experience this illusion, stare at a waterfall for a few minutes; after shifting your gaze, stationary objects such as the nearby rocks will briefly appear to crawl upward. In 1963, the fatigue explanation for after-effects was bolstered by physiological findings by Barlow and Hill. They showed that the firing rate in directionally selective neurons of the rabbit retina declines if a stimulus is continuously moved through the receptive field in the preferred direction; after the stimulation stops, the baseline firing rate remains suppressed for a short while. However, despite this early physiological support, it now seems clear that the fatigue of neuronal populations falls short as an explanation for after-effects. Even as early as 1911, Wohlgemuth showed that the motion after-effect can be stored: motion is viewed, the eyes are closed, and the after-effect remains when the eyes are later reopened. In addition, we do not see a motion after-effect resulting from driving a car, and under certain circumstances, non-moving stimuli can induce an after-effect. In 1965, Celeste McCollough reported the first contingent after-effect, in which prolonged viewing of a pattern consisting of, say, horizontal and vertical coloured stripes, results in the subsequent colour- and orientation-specific misperception of uncoloured stripes (FIG. 3). Even more strikingly than the motion after-effect, the McCollough effect can last overnight and sometimes for days. This seems less likely to represent fatigue than...
some sort of active recalibration. That is, perhaps the visual system seeks to eliminate (possibly spurious) correlations between colour and orientation by actively adjusting the perceived colour of particular forms. Whatever the case, the motion after-effect is still not fully understood.

In general, after-effects are observed not only in simple attributes such as colour and motion, but often seem to reveal different sorts of 'channels', or 'spatial filters', which can be exposed by selectively adapting them. For example, in 1969, Blakemore and Sutton showed that there are spatial-frequency-specific after-effects: when an observer adapts to a grating of a particular spatial frequency, a slightly lower frequency will then seem even lower, and a slightly higher frequency will seem even higher. This spatial frequency after-effect is often explained in terms of a shifted peak in the population activity of a distribution of filters. This explanation might prove to be unsatisfactory, however, as downstream mechanisms that only read a peak in population activity might imply that we would never see stimuli other than gratings. It is suspected that the channel story might not be borne out in the higher areas of cortex as the next generations continue their study, however, the study of the after-effects will continue to serve as a benchmark for which new theories must provide an explanation.

**Multistable stimuli**

Illusions such as the Necker cube and the face–vase illusion (FIG. 4a,b) are examples of multistable stimuli. Strictly speaking, all visual stimuli are ambiguous. For example, a distant large object or a nearer small object can cause a given projection on the retina. The interesting property of multistable stimuli, however, is that they can flip back and forth between different perceptions. Although nothing changes on the page, there is more than one way for the visual system to interpret the stimulus, and perceptual reversals occur. The perceptual reversal indicates that cortical processing is an active process that tries to make sense of incoming information.

Multistable stimuli are invaluable tools for the study of the neural basis of visual awareness, because they allow us to distinguish neural responses that correlate with basic sensory features from those that correlate with perception. With this in mind, a very useful example of bistability can be found in the phenomenon known as binocular rivalry (FIG. 4d). When different images are presented to the two eyes, one eye’s view dominates for several seconds and is then replaced by that of the other eye. What causes these perceptual changes in the absence of any change in the stimulus? Although the psychophysical study of binocular rivalry is reasonably old, and the effect was explored with electroencephalography (EEG) in the 1960s (REF. 35), experimenters have begun to explore the physiology of changing percepts in awake, behaving monkeys at the single-cell level only very recently. Recordings from neurons in areas V1, V2, V4 and MT indicate that only a small proportion of cells responds to the perception of the dominant stimulus, while the rest continue to respond to low-level features even while the image is suppressed. The proportion responding to the dominant stimulus grows at higher stages of processing until, remarkably, most of the cells in the visual areas of temporal cortex fire in a manner that is correlated with the stimulus that is dominating perception. Activity in these temporal lobe cells therefore seems to represent a stage of processing that is beyond the determination of low-level features, and instead represents something like the brain’s internal representation of objects. This physiological data from the rivalry studies seems to be incompatible with the previously popular idea that rivalry reflects a competition between the two eyes, and instead seems to reflect competition between top–down cortical explanations for the

Figure 3 | **After-effects and competing populations.** The McCollough effect. After observers are exposed for a few minutes to differently coloured gratings of different orientations, they perceive similarly oriented achromatic gratings as if they were tinted with complementary hues. This illusion can be experienced by viewing the coloured gratings to the left for a few minutes, and then shifting the gaze to the black-and-white gratings on the right.

Figure 4 | **Multistable stimuli and active perception.** a | The Necker cube is an ambiguous two-dimensional stimulus that lends itself to spontaneous depth reversal, as it is equally consistent with two different three-dimensional configurations. Note that the visual system chooses only a single interpretation at a time, never a mixture. b | The face–vase illusion is subject to flip-flopping interpretations of figure versus ground. c | Monkeys viewed binocularly rivalrous drifting gratings in a 1989 study by Logothetis and Schall. In single-unit recordings from superior temporal sulcus, the firing activity of some neurons strongly correlated with the monkey’s percept, even while the retinal stimulation remained unchanged. d | An example of binocular rivalry; the perceptual alternation can be experienced by cross-fusing the two images. The red and green stripes will alternate in perception. Panel a is adapted with permission from REF. 99.
inputs\(^{60,49}\). However, this point is still debated; new functional magnetic resonance imaging (fMRI) evidence supports interocular competition, indicating that rivalry can be fully resolved in monocular visual cortex\(^{41}\). In general, fMRI is becoming a popular technique to study rivalry; recent studies have shown that extrastriate regions such as the ventral visual pathway, parietal and frontal regions\(^{42}\), the fusiform face area and the parahippocampal place area participate in perceptual changes\(^{43}\).

**The timing of awareness**

Another area of active debate is how and whether the timing of neural signals leads to illusions of timing. This hypothesis has a long history, beginning at least as early as Pulfrich’s illusion of timing. This hypothesis has a long history, beginning at least as early as Pulfrich’s demonstration in 1922 that a pendulum seems to rotate in depth when a neutral density filter is placed in front of one eye. After a suggestion by Fertsch, Pulfrich proposed a timing difference between signals from the two eyes\(^{44}\). The idea of differences in physiologically measured ‘latencies’ translating directly into perceptual timing has recently re-emerged as a proposed explanation for certain illusions. For example, taking as a springboard the measured latency differences between colour- and motion-processing areas of the visual cortex, Moutoussis and Zeki showed that synchronous colour and motion changes can appear to be asynchronous\(^{45}\). However, we now know that this illusion occurs only when the stimuli are presented repeatedly, and not when they are presented only once\(^{46}\). This seems to rule against a simple model of signals racing like a rabbit and a tortoise to a ‘perceptual end-point’.

The idea that signal timing might equal perceptual timing has also been proposed recently as an explanation for the flash-lag illusion\(^{47,48}\); a phenomenon in which a flash and a moving object that appear in the same location are perceived to be displaced from one another\(^{49,50}\). However, there is opposing evidence that timing judgements between flashes and moving object are very accurate\(^{51}\), and instead that the flash-lag effect reflects something more fundamental about the timing of conscious awareness\(^{52,53}\). Specifically, the flash-lag illusion is consistent with the idea that the percept attributed to the time of an event is a function of events that happen in a small window of time after the event. This idea that consciousness is a retrospective reconstruction (or ‘postdiction’) has roots as early as William James\(^{54}\), and was expressed eloquently in the last decade by Dennett\(^{55}\). The reconstructionist framework fits well with the modern understanding that the visual system is not merely feedforward but also feedback, a concept nicely illustrated by recent transcranial magnetic stimulation (TMS) experiments in which feedback from MT to V1 was shown to participate in visual awareness\(^{56}\). Other classic illusions further support the idea that the visual system consults the ongoing input of information from the near future of an event before committing to a percept. In the illusion of backward masking\(^{57}\), a stimulus followed in rapid succession by a second stimulus can block or modify the perception of the first one. Recent experiments by Macknik and colleagues have found correlates of masking in monkey V1 using single-unit recordings\(^{58}\) and optical imaging\(^{59}\).

The nexus between temporal illusions and the timing of visual signals is likely to provide fertile ground for the future. It could be that the brain encodes time symbolically, such that stimulus time differences do not necessarily correspond to differences in physiologically measured latency. The latter idea, sometimes known as the ‘latency difference’ hypothesis, might be an example of experimental psychologists being misled by too simple an interpretation of physiological...
In more rare instances, biological knowledge lights the way to new illusions. One example is Edmund Mariotte’s 1668 chance discovery of the blind spot from his studies of the eye.\(^{25}\) The illusion of continuity in the visual field — despite the absence of photoreceptors in this sizeable patch of the retina — is so compelling that it is no surprise that the discovery was accidental.

More recently, the legacy of Hubel and Wiesel is an appreciation that the visual system generally processes different kinds of information in different processing streams\(^ {31,32}\). Consistent with anatomical and psychological evidence that colour and motion are analysed by separate channels in the visual system, Ramachandran and Gregory showed decreased ability to detect apparent motion with equiluminous coloured stimuli\(^ {41}\), verifying that the motion and colour pathways are perceptually separable. Two decades later, \(\text{La et al.}^{42}\) noted that a moving isoluminant colour grating can be perceived as standing still if all luminance and salience differences between the example in which human report would be useful. Human stimulation was first realized very early on\(^ {43}\), and is now carried out with TMS and implanted electrodes, again holding great promise for the future.

David M. Eagleman is at The Salk Institute for Biological Studies, 10310 North Torrey Pines Road, La Jolla, California 92037, USA. e-mail: eagleman@salk.edu


PERSPECTIVES

26. Sutherland, N. S. Figural after-effects and apparent size. Q. J. Psychol. 8, 222–228 (1961).
44. Puljic, C. Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie. (Voss, Hamburg, 1867).
PERSPECTIVES

55. James, W.
56. Dennett, D. C.
58. Bachmann, T.
60. McGurk, H. & MacDonald, J. Hearing lips and seeing voices.
91. Rubin, E. Synesthesia Pigey (Copenhagen, 1915).

Acknowledgements:
I thank my colleagues at the Salik Institute and the University of California at San Diego, in particular, G. Storer, A. Holcombe, B. Krekelberg, S. Anstis, M. van der Smagt and T. Sajnowski.

Online links
FURTHER INFORMATION
Amazing Optical Illusions: http://www.optillusions.com/
Illusionworks: http://www.Illusionworks.com/
high-level vision | illusions
SandlotScience: http://www.sandlotscience.com/
The Waterfall Illusions: http://www.psyp.un-kyoto-u.ac.jp/ashida/MWWW/home.html
Access to this Interactive links box is free online.