

# Biomedical Coordinates of Human Perception

Dr. Ashoka Jahnavi Prasad

There are two essential features of this definition, the first being the distinction between perception and sensation. Sensation is the immediate neurobiological consequence of stimulating sensory transducers such as photoreceptors, mechanoreceptors, and chemoreceptors. Sensory events are ubiquitous and can affect behaviour directly – the spinal reflex of pulling your hand back from a hot surface is one simple example – but they are fleeting, discontinuous, and lacking semantic content. Perception enriches sensation by reference to other knowledge or experience. In the words of British associationist John Stuart Mill, “perception reflects the permanent possibilities of sensation” – things as they were or might be – and in doing so reclaims from evanescent sensory events the enduring structural and relational properties of the world. The philosopher and psychologist William James expressed a similar view in his discussion of ‘The Perception of Things’: “Perception thus differs from sensation by the consciousness of farther facts associated with the object of the sensation.” Or in the prescient words of nineteenth-century perceptual psychologist James Sully: the mind “supplements a sense impression by an accompaniment or escort of revived sensations, the whole aggregate of actual and revived sensations being solidified or ‘integrated’ into the form of a percept.” Building on the philosophical traditions from which the discipline of psychology was born, William James further stressed the need for associative neuronal processes to achieve this integration: “The chief cerebral conditions of perception are the paths of association irradiating from the sense-impression.”



**Dr. Ashoka Jahnavi Prasad** is identified as the most educationally qualified person in the world by The Polymath. He has a dynamic resume with a PhD in history of medicine from Cambridge, LLM from Harvard among other notable qualifications. Dr. Prasad has also worked as a consultant to the World Health Organization (WHO) and helped prepare two of their reports.

The second essential feature of neuroscientists’ definition of perception concerns the attributes of the thing (or things) to which sensation is referred. In particular, the referent is viewed as the ‘cause’ of sensation. This concept was developed by the eighteenth-century Scottish philosopher Thomas Reid, who argued that sensation “suggests to us” an object as its source: “We all know that a certain kind of sound suggests immediately to the mind a coach passing in the street.” As James later noted, perception is “the consciousness of particular material things present to sense.” This attribution to the source of sensation is profoundly important and meaningful for perception and behaviour. To fully appreciate this importance, it is useful to consider the larger function of sensory systems and the computational problems faced by them. For a variety of reasons – some technical, some conceptual, and some historical – this consideration is easiest to undertake for the visual modality, and that is the approach I will use herein. The principles of perceiving and the underlying neuronal mechanisms revealed by the study of vision nonetheless have broad relevance to other sensory modalities.

The biological problem, which is by far the more difficult of the two, thus involves reconstruction of the properties of the visual scene, given only the pattern of light present in the retinal image. This inverse problem of optics is an example of what is known formally as an ill-posed problem: it is a problem without a unique solution. Because of the dimensionality reduction that accompanies optical projection of the world onto the retina, there is simply not sufficient information present in the retinal image to uniquely identify its environmental causes. To put it bluntly, the number of visual scenes that could give rise to any specific retinal image is infinite.

In view of this intrinsic ambiguity, perhaps the most astonishing thing about visual perceptual experience is that we rarely have difficulty arriving at a unique solution for the environmental cause of the retinal image. Moreover, the solution we adopt is nearly always

the correct one (at least within margins of error allowable for our behavioural interactions with the world). Those cases where we arrive at the wrong solution are what we call 'illusions'. Vision is able to accomplish reliable disambiguation of the retinal image by virtue of context, which, broadly speaking, consists of other pieces of information that are either 1) co-present in the retinal image (spatial context); 2) learned based on the observer's prior experience with the world (temporal context); or 3) embedded in the computational machinery of the brain as a result of evolution in an environment that has consistent and well-defined properties (evolutionary context). Context is thus James' "farther facts associated with the object of sensation." Using the available context as clues, the process of disambiguation – the process of perceiving – is best characterized as probabilistic inference about the cause of sensation. Again according to James, "perception is of probable things." Or to use a familiar colloquialism, "we generally see what we expect to see."

Context plays an extended role in perception in that it also helps resolve symbolic form. Sensory inputs are replete with symbolism – often quite abstract and multifaceted – and to perceive is to grasp the meanings of the symbols. Perhaps the most dramatic example of this is human language, which is by definition an experience-dependent mapping of auditory and visual stimuli onto meaningful objects, actions, and concepts. To illustrate how context resolves this mapping and, by doing so, makes perception possible, William James offered the phrase *Pas de lieu Rhone que nous*. If the listener assumes that the spoken phrase is French, it is unintelligible. If, however, the listener is informed that the spoken phrase is English, the very same sounds are perceived as paddling your own canoe.

James further noted that "as we seize the English meaning the sound itself appears to change." In other words, the percept is both reconciled and subjectively qualified by the context provided. With these phenomenological aspects of perception in mind, research in recent years has focused on the mechanisms by which contextual cues modulate neuronal signals originating from the senses, thereby reflecting the thing perceived. In the case of vision, the question is: how do neuronal signals that initially reflect the properties of the retinal image become transformed (via context) such that they reflect instead the properties of the visual scene? Based on the discoveries presented below, I will argue that this transformation from an image-based representation to a scene-based representation is among the most basic and fundamental operations of the cerebral cor-

tex and is the neurobiological explanation for the fact that "perception reflects the permanent possibilities of sensation."

## Neuronal Basis of Perception

To provide a foundation for understanding the neuronal bases of perceiving, I will briefly review our present understanding of the organization of the primate visual system. Much of this information has come from the use of three key experimental approaches:

- 1) **Neuroanatomy:** These studies, which are made possible by methods for selective visual labelling of brain tissue, provide a picture of neurons – the basic cellular elements of the brain – as well as a picture of the wiring diagram of neuronal connectivity.
- 2) **Neurophysiology:** These studies reveal the types of sensory signals carried by neurons and the manner in which those signals are transformed at each computational stage in the signal processing hierarchy. In practice, this information is recorded physiologically by evaluating neuronal receptive field properties. The receptive field (RF) is a central concept in sensory neurobiology and is defined as the region of sensory space that, when stimulated, elicits a change in the activity of the recorded neuron (typically quantified as frequency of action potentials). Visual RF properties may include, in addition to spatial location, sensitivity to complex spatial, temporal, and/or chromatic properties of light. A more recent concept important for this discussion of perception is that of the RF surround (sometimes known as the 'non-classical RF'). A sensory stimulus that falls in a neuron's RF surround does not directly cause a change in activity (by definition), but rather has the ability to modulate activity driven by a stimulus in the RF. As we shall see, the RF surround provides a mechanism by which contextual cues may influence (and disambiguate) neuronal responses to sensory stimuli.
- 3) **Behavioural analysis:** This approach involves acquiring behavioural reports of subjective states, such as those associated with perceiving. These behavioural measures reveal correlations (and implied causal relationships) between a perceptual state and the neuronal signals recorded from specific anatomical circuit locations.

The first act of visual processing, which occurs in the retina, is phototransduction (the conversion of energy in the form of light into energy in the form of neuronal activity). Signals leaving the retina are carried

by the optic nerve/tract and convey information about the location of a stimulus in visual space, the wavelength composition of the stimulus, and about spatial and temporal contrast between different stimuli. Optic tract fibres synapse in a region of the thalamus (a form of sensory transponder located in the centre of the forebrain), known as the lateral geniculate nucleus, and the signals are passed on from there via a large fibre bundle to terminate in the primary visual cortex, or area V1. Area V1 is located in the occipital lobe (the most posterior region) of the cerebral cortex, which is the massive convoluted sheet of neuronal tissue that forms the exterior of the forebrain. Signals recorded physiologically reveal that V1 neurons extract a number of basic features of the pattern of light in the retinal image, including position in visual space, colour, contour orientation, motion direction, and distance from the observer. These signals ascend further to contribute to functionally specific processing – object recognition, spatial understanding, visual motor control – in a number of additional visual areas, which collectively make up approximately one-third of the human cerebral cortex.

Attempts to understand how visual cortical neurons account for perceptual experience have been largely reductionist in approach. These experiments, which spearheaded the field of sensory neurobiology 50 years ago, typically involve manipulation of a very simple stimulus along a single sensory dimension – a contour is varied in orientation, a patch of light is varied in colour, a moving texture is varied in directionality – and placed within the RF of a physiologically recorded neuron. This approach has revealed fundamental features of the ways in which basic attributes of a sensory stimulus are detected and encoded by neuronal activity. But in the end this approach has told us little about perception because the stimuli are devoid of meaning. They lack the context needed to identify environmental causes and are thus ambiguous in the most complete sense possible. The alternative experimental approach is one in which the context is varied, such that the percept (the inferred environmental cause) is independent of the parameters of the sensory stimulus. The real world presents a rich set of such conditions, of course, but there is an advantage to an intermediate experimental approach in which perceptual and neuronal responses to a simple well-defined stimulus – the ‘target’ – are evaluated in the presence of simple contextual manipulations. In our laboratory we have used both spatial and temporal context for this purpose. Spatial context is defined here as other

(non-target) features of the retinal image, such as the colour, pattern, or motion of spatial regions surrounding the target (the “surround”). There are many well-known perceptual phenomena in which experience of a simple target stimulus is markedly influenced by the surround. Perception of brightness and colour hue, for example, are heavily context dependent, often in complex and revealing ways. Neuroscientists have also discovered and explored a number of compelling effects of context on perceived motion.

### Current State of Biomedical Research Relating to Human Perception

In all of these cases, the percept varies because the spatial context leads to different plausible interpretations of the cause of the sensory stimulus. Gene Stoner (Salk Institute) has worked on context dependent visual motion perception – his work conclusively establishes this. Stoner designed a stimulus in which a diamond-shaped pattern of vertical stripes was viewed in the context of a textured surround. (He called this the “barber diamond” stimulus because it mimics features of the classic ‘barber pole’ illusion, in which rotating stripes appear to move downward along the axis of the pole.) Portions of the textured surround were manipulated (using binocular disparity cues) such that they were perceived in different planes of depth. He used two complementary depth configurations: 1) foreground (near) panels appear at the upper-right and lower-left corners of the stimulus and a background trough runs from the upper-left to lower-right; and 2) foreground panels appear at the lower-right and upper-left and the background trough runs from lower-left to upper-right. In both depth configurations, the diamond shaped pattern of stripes appeared at an intermediate depth plane (above trough and below foreground panels). For each of these two configurations we moved the pattern of stripes either leftward or rightward within the stationary diamond shaped window.

He predicted that the perceived direction of the stripes would vary depending on their depth relationships with the surround. To understand this prediction, note that the edges of the striped pattern that lie adjacent to a far panel are readily (probabilistically) perceived as edges of an object in the visual scene. He called these ‘intrinsic’ edges because they are intrinsic to the object that caused their presence in the retinal image. By contrast, edges of the striped pattern that lie adjacent to a near panel are readily perceived as accidents of occlusion of one object by another. He

called these 'extrinsic' edges because they are extrinsic to all objects in the visual scene. Thus, the moving lines that rake (as retinal stimuli) along an intrinsic edge should be perceived as the consequence of a striped object moving through space in that direction. Conversely, the moving lines that rake along an extrinsic edge should be perceived as bearing no reliable relationship to the motion of the striped object, for it is impossible to verify, based upon these retinal motions, the extent to which the object moves under or parallel to the foreground surface. As a consequence of these depth-dependent perceptual interpretations of the cause of the retinal stimulus, the physically invariant motions should be perceived as arising from an object in the visual scene that slides along one or the other diagonal, depending on the spatial context. This is precisely what he found to be true; indeed, the effect is universal and striking.

The configuration of these context dependent motion stimuli enabled us to place them within the RFs of cortical neurons that respond selectively to the direction of stimulus motion. Such neurons are abundant in a small mid-level cortical region known as the middle temporal visual area, or area MT. A typical MT neuron responds best to movement of a pattern in a specific direction – upward, for example – within the RF, and response wanes as direction deviates from this "preferred direction." Much evidence suggests that the activity of these neurons underlies the perception of motion, so he reasoned that the response of an MT neuron to the barber-diamond stimulus should vary depending on the direction perceived (as dictated by spatial context), even though the motion in the retinal image never changes.

The barber-diamond provides a dramatic demonstration of how an invariant retinal stimulus can be perceived differently depending on context. Another property of perception is that it often generalizes across different sensory attributes: widely varying retinal stimuli will be perceived as arising from a common source in the visual scene. These effects are termed perceptual constancies. Size constancy, for example, refers to the fact that an object viewed from different distances will vary markedly in retinal image size but will be perceived as having the same size in the visual scene. This is a case in which the probable cause of the retinal stimulus is inferred using contextual cues for distance, such as linear perspective or binocular disparity.

He also observed a type of perceptual constancy known as 'form-cue invariance', which occurs when

essential attributes of the visual scene, such as object form or motion, are extracted invariantly from the retinal image by generalizing across image attributes that distinguish object from background. The profile of Abraham Lincoln's head, for example, is readily perceived regardless of whether it is defined relative to background by black-on-white, white-on-black, or red-on-green. This form of constancy reflects contextual information (evolutionary context) that is 'built-in' the human visual system: to wit, we have evolved to operate in an environment in which the properties of light change over space and time, and objects are thus to varying degrees visible to us by different image cues, including brightness, colour, and texture. Reliable perception depends on the ability to generalize across these state changes and extract the meaningful properties of the world.

He studied the neuronal basis of form-cue invariance using simple elongated rectangles that were defined by differences in brightness, texture, or flicker. When these different stimuli are moved, the percept of motion is invariant, despite the fact that the sensory attributes of the image vary markedly. When he presented motion sensitive MT neurons with these form-cue invariant motion stimuli, he discovered that the neuronal responses also generalized across the different form-cues. It was concluded that these responses underlie the invariant perceptual experience by representing common motions of objects in the visual scene, to the exclusion of information about the sensory features that distinguish the objects from the background.

In the examples considered thus far, perception does not require reference to the observer's prior personal experience with the world. Much of the time, however, the way we perceive a sensory stimulus depends heavily on what we have seen before. This temporal context is manifested perceptually both as the ability to identify the cause of a sensory stimulus and by the ability to identify its meaning (as symbolic form). Without prior experience with this particular pattern of retinal stimulation, it is difficult to identify a probable environmental cause (other than a surface with an apparently random pattern of light and dark regions) or to grasp what the sensory features symbolize. Similarly, the visual perceptual experience of *Devanagari* (Hindi script) is markedly different before and after learning the written language. These temporal context effects are, in fact, ubiquitous and fundamental to perceiving. They are rooted in the phenomena of associative learning and memory retrieval: as an observer learns the relationship between a sensory stimulus and the "farther

facts associated with the object of sensation," the stimulus alone is capable of eliciting retrieval of those 'farther facts', which become incorporated into the percept.

To understand the neuronal events that underlie the perceptual effects of temporal context, Stoner explored both the learning of sensory associations and the mechanisms by which associative recall influences perception. Sensory associative learning is the most common form of learning: it inevitably results when stimuli appear together in time, and particularly so in the presence of reinforcement; it can occur without awareness ('classical conditioning') or with ('instrumental conditioning'). The product of associative learning is that presentation of one stimulus elicits retrieval of its associate and all of the experience that retrieval entails. For Pavlov's dog, for example, the learned association between the sound of the bell and the taste of food elicited in the dog the physiological manifestations of eating (such as salivating) when the bell alone was struck.

He began by examining how the learning of sensory associations is implemented by neuronal plasticity by hypothesizing – in accordance with James's conjecture ("When two elementary brain-processes have been active together or in immediate succession, one of them, on reoccurring, tends to propagate its excitement into the other") – that associations are established through the formation or strengthening of neuronal connections between the independent neuronal representations of the paired stimuli. To test this hypothesis, he trained subjects to learn associations between pairs of visual stimuli that consisted of meaningless patterns ('clipart' figures). As the subjects acquired the associations, he monitored activity from neurons in the inferior temporal (IT) cortex, a region that lies at the pinnacle of the visual processing hierarchy and is known to be critical for object recognition, which relies on memory of prior associations. Neurons in the IT cortex respond selectively to complex objects, such as the visual patterns used for this experiment. We predicted that as subjects learned new associations between these stimuli, connectivity would increase between neurons that were initially selective for one or the other member of a pair. The result of this change in connectivity, then, should be expressed physiologically as a convergence of the magnitude of the neuronal responses to the paired stimuli. In other words, as the stimuli become symbols for one another by association, the neuronal representations of those stimuli become less distinguishable. He believed that these neuronal changes are the physical manifestations of the newly learned

associative memories.

Memories consolidated in this manner over the course of a lifetime provide the store of "farther facts associated with the object of sensation": the temporal context needed to interpret sensory events. Stoner thus hypothesized that if the activity of a neuron in the visual cortex underlies perception (as opposed to sensation), that activity should be influenced by the retinal stimulus as well as by information retrieved from the memory store (the cellular locus of stored information in the brain). To test this hypothesis, he first trained subjects to associate pairs of stimuli and then evaluated the responses to individual members of each pair. He predicted that the new 'meaning' given to a visual stimulus by the learned association would be reflected in a new form of neuronal selectivity for the stimulus. The stimuli used for this experiment were moving patterns and stationary arrows. Subjects learned that the direction of motion of each pattern was associated with the direction of the arrow. Upward motion, for example, was associated with an upwardly pointing arrow, and so on. Moreover, the selectivity seen was entirely predicted by the learned association. That is, if a cell responded best to upward motion, the arrow direction that elicited the largest response was also upward.

In these experiments, the associative training causes the arrow and motion to serve as symbols for one another in the same sense that the graphical pattern moment serves as a symbol for a specific person. The perceptual experience elicited by sensing the arrow naturally includes the things that the arrow symbolizes, which in this case is the motion with which it has been paired. A corollary of James's axiom about "farther facts associated with the object of sensation" is that perceptual experience includes tangible 'images' of the things recalled – images seen in the 'mind's eye'. Indeed, James defined his "general law of perception" as follows: "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." Stoner thus inferred that the perceptual experience elicited in our experiments by the arrow includes visual imagery of the motion, which is retrieved from the memory store. He furthermore concluded that the selective pattern of activity exhibited by MT neurons upon viewing the stationary arrow underlies imagery of the motion recalled by association

The conditions of the experiment described above elicit a form of visual imagery in which the thing imagined (motion) differs markedly from the sensory stim-

ulus (arrow). This is the same form of imagery that occurs when we explicitly conjure the face of a friend, the characters and places in a novel, or imagine how the couch would look if we moved it to the other side of the room. But there is another form of imagery that is ubiquitous and largely non-volitional, in which the thing (or things) recalled by association ‘matches’ the sensory stimulus and is in fact a probabilistic inference about the cause of that stimulus. Harking back to our earlier discussion of the ambiguity of sensory events and the role of context in resolving that ambiguity, we can see that this latter form of imagery is a critical part of the process. This is particularly true under conditions in which the sensory stimulus is impoverished by noise or incompleteness

In fact, simple consideration of the objects that surround you – a chair that is partially obscured by a table, a glass blurred by glare, scratchy notes on a pad of paper – suggests that your clear and complete perceptual experience of these objects is the result of sensations that have been complemented – fleshed out, if you will – by information provided by memory. As these experiments have shown, that information comes in the form of selective feedback into the visual cortex, which unites retinal signals with imagery signals to yield perceptual experience of probable things.

Given the ubiquitous, powerful, and implicit nature of this form of imagery, it is perhaps unsurprising that it can be readily manipulated to hijack perception. Indeed, a major category of performance magic relies on priming as a form of temporal context. The result is an illusion in which the viewer’s percept of an equivocal sensory event is not the ‘correct’ solution. Similar unconscious priming effects occur under normal (non-magical) circumstances as well: perceptual biases in eyewitness reports or in the interpretation of woolly data (such as x-ray and forensic fingerprint examination and proofreading) are well documented. There are also priming effects on perception that hap-

pen with full awareness. Pareidolia is the phenomenon in which we perceive coherent and meaningful patterns in response to (recognizably) random sensations, such as clouds that look like animals or foods that resemble Jesus. Finally, there are genres of art, such as impressionism, in which the image rendered – the sensory stimulus – is left intentionally vague in order to allow the perceptual experience to be completed by the viewer’s prior experiences (which art historian E H Gombrich evocatively termed “the beholder’s share”).

In this column, I have defined perceiving as a process by which the fundamental ambiguity of sensation is resolved through the use of contextual cues, which enable identification of the causes of sensation and attributions of meaning. From these discoveries, it is natural to conclude that a fundamental and generic computation in the cerebral cortex is the transformation from sensory attributes (the retinal image) to attributes of the external environment (the visual scene). While we know that context is used in that computation, and we can identify neuronal signals that reflect the outcome, we currently know very little about the neuronal mechanisms that give rise to these signals. How, for example, is information from memory selectively and dynamically routed back to the visual cortex in a context-dependent manner to complement information arising from the retina? How do neurons that represent visual motion incorporate information from memory selectively and dynamically routed back to the visual cortex in a context-dependent manner to complement information arising from the retina? How do neurons that represent visual motion incorporate information about the spatial layout of a visual scene? Identifying these processes presents formidable challenges, to say the least; but a variety of new experimental techniques—many from the fields of molecular biology and engineering—provide much promise for a future mechanistic understanding of perception.

+