

# What do you see, and how?

## The cognitive infrastructure of vision

Sunny Y. Auyang

Seeing a rose or hearing the doorbell is among the most common and immediate of experiences. Sense perceptions are also most fundamental and important; on them base all our factual knowledge and empirical science. Does their epistemological priority stem from their being unanalyzable primitives given to us? Do they have structures? If so, what are the structures and where do they come from? The importance of these questions extends beyond psychology to the justification of all knowledge and science.

After centuries of debate, philosophers are still widely divided on the nature of sense experiences. One school insists them to be unanalyzable primitives given to us. Its opponent argues that they have complicated structures resulting from active contributions.

Recent researches in cognitive psychology and neuroscience find our perceptual experiences to be anything but primitive. They are emergent properties resulting from complicated cognitive and neural processes, which constitute what I call the *infrastructure of mind*. We are unaware of them, but without them, we would not have the kind of perceptual experiences that we have. This talk examines the infrastructural processes underlying vision. They build many structures into our most immediate experiences. Among the structures is what reasonably interpreted as the concept of physical objects.

### **Experience and the basis of knowledge**

Postmodernists are waging what they call “Science Wars” in America and Britain, I do not know if it is as bad in Australia. Scientists accuse postmodernists for erroneous reporting and distorting scientific research and results. Postmodernists contend that their practice is justified because the notions of nature, reality, and truth are all illusions. With these notions outlawed, science becomes epistemologically no different from myth. Both are social constructions with no claim of validity beyond the culture in which they are constructed.

Postmodern writings are notorious for their vagueness and muddle thinking, as exposed by the Sokal hoax. Yet ideas similar to theirs have also been advanced by respectable philosophers who argue with logical clarity. For example, W. V. O. Quine wrote: “I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience. Physical objects are conceptually imported into the situation as convenient intermediaries – not by definition in terms of experience, but simply as irreducible posits,

comparable, epistemologically, to the gods of Homer. . . . In point of epistemological footing the physical objects and the gods differ only in degree and not in kind. Both sorts of entities enter our conception only as cultural posits.” [From a Logical Point of View, p. 44.]

Physical objects constitute nature and the real world. Electrons and planets, genes and brains are all physical objects. So are human beings, although they have peculiar mental capacities and enjoy special respect from others of their kind. The concept of physical objects is fundamental to natural science. When it sinks to the status of Homeric gods, one is on a slippery slope to postmodern nihilism: Anything goes! Quine did not slide all the way. His is an empiricist position with a long philosophical tradition tracing back to John Locke. Empiricism upholds a bedrock for knowledge: sense experiences.

Our senses of vision, audition, taction, olfaction, gustation, proprioception, and interoception bind our cognition to the physical world, including our own bodies. They are the bases of all knowledge, including empirical science. Physics, chemistry, biology, and other natural sciences are all empirical sciences. Their validity depends on experiments, which in turn depend on scientists observing experimental conditions and results.

Most people agree pure thinking alone is insufficient for knowledge and science, sense experiences are essential. In this they are at one with empiricism. However, I guess many would also balk at the Quine’s notion of physical object as “a convenient myth.” What in empiricism leads to this notion? The answer lies in its peculiar view of sense experience. On whether this view is correct Cognitive science can shed light.

### **Active and passive perception**

Many philosophical debates reduce to the question: Is sense perception active or passive? Do we passively receive what is given in the sense stimuli as experience? Or are our immediate sense experiences already structured by our active mental contributions?

The passive view of perception finds its major champions in empiricists, the active view in Kantians (Box 1). Both parties agree that all our knowledge is ultimately based on sense experiences. They disagree about the characteristic of our most immediate and direct perception, and hence its strength as the ground of knowledge.

For example, empiricists argue that you see roundness and redness, which are sense stimuli or sense data that are given to you. Kantians argue that your immediate visual experience is not merely roundness and redness; you directly see a tomato. Your experience has already involved certain notions, especially that of a persisting physical object, which come from you. Without the concept of object there is no experience in the sense we know it, for the raw sense stimuli are not intelligible at all.

## Box 1. Two view of sense experiences

### Passive perception (empiricists)

Experience as “sense data,” “sense impression,” “surface stimulations,” “ocular irradiation patterns,” the *given*.

“As an empiricist I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience. Physical objects are conceptually imported into the situation as convenient intermediaries.”

“The conceptual scheme of physical objects is a convenient myth.”

– W. V. O. Quine

### Active perception (Kantians)

“The employment of our ordinary, full-blooded concepts of physical objects is indispensable to a strict, and strictly veridical, account of our sensible experience. . . . I have argued that mature sensible experience (in general) presents itself as, in Kantian phrase, an *immediate* consciousness of the existence of things outside us.” – Peter Strawson

“Concepts of objects in general must underlie all empirical knowledge as its *a priori* condition.” – Immanuel Kant

Can you see physical things directly? Or do you see only sense data and later construct theories about physical objects? These are the central contention of the two views of perception. Sense impressions and physical things differ in many general ways. Here we note three; we will examine scientific evidence for them later:

- Sense impressions are private; I cannot share your experiences of redness. Physical objects are public; the tomato is there for both of us to see.
- Sense impressions are ephemeral; you blink and the redness and roundness are gone. Physical things endure when you look away; the tomato is there when you blink. Whereas an empiricist doctrine states: “To be is to be perceived,” a criterion for testing the reality of things is their persistence: to be is to be independent of being perceived.
- Redness is the same even when it appears in different sense impressions. Sense impressions are bundles of qualities. The tomatoes are two even if they are exactly alike in redness and roundness. Physical objects have qualities and their numerical identities, by which they are distinct from each other even if they have the same properties and cause the same pattern of stimuli on our sense organs.

Because of their endurance and numerical identity, physics objects cannot be merely given in sensory stimuli, which are transient and carry no lasting identity. Thus to say that you see a physical object is to say that you see more than meets the eyes, that your immediate visual experience has already involved certain contribution from your part.

The two views of perception entail two views of knowledge and two notions about the external world. In some way, passive perception provides more secure knowledge. Sense data are certain and infallible, because they are given to us and uncontaminated by our fallible processing. The

possibility of errors arises later, when you introduce concepts to organize your immediate visual experience of roundness and redness.

Unlike passive perception, active perception cannot claim certainty. Humans are fallible beings. Our contributions introduce the possibility of errors in our immediate perceptual experiences. You can see things wrong or you may be hallucinating. You can make more mistakes when you conceptually organize your experiences, but your experiences are already fallible.

On a closer look, the epistemological superiority of passive perception dwindles. Its certainty extends only to private sense impressions, but most valuable knowledge is about things and people in nature and the public world. For such knowledge, it is inferior to active perception; it does not allow cross checking and verification, active perception does.

If I saw a tomato and later wonder if I had mistaken an apple for it, I can go and check. If it is not there, I can ask people if they have seen it, or I can try to find empirical evidences of the causal traces it leaves behind. These evidences are closely connected to my original perception of a physical object, which serves as the nexus that integrates various perceptual instances. None of our experience and knowledge is absolutely certain, but through multiple cross checking we can increase our degree of confidence to near certainty. This is what empirical sciences achieve for a wide variety of topics.

In contrast, if I see merely redness and roundness, there is no way for me to return and check, they are gone for good. Nor can I ask people; they have no access to my sense impressions. If I introduce the concept of a tomato, it is something alien and unconnected to my sense impressions. Although I can be certain about my sense impressions, about other things I can hardly claim to have knowledge; I cannot perform test to verify them. A person who perceives only sense impressions is trapped beyond “a veil of perception” and cut off from physical world. This is why knowledge about nature reduces to a myth.

Which view of perception is correct? This is a factual question about human perception and factual questions are susceptible to scientific investigation. Fortunately, characteristics of sense perception have been intensively research topics in recent decades.

Let us examine vision, our most important sense that accounts for about 40 percent of our sense input. Scientific results are overwhelming in favor of the active view of perception. Recall Kant’s slogan: “Thoughts without contents are empty, intuitions without concepts are blind.” Kant’s “intuition” is the sensual aspect of perception. When the intuition is *visual* sensibility, intuition with concept is almost literally blind.

### **Mind and its infrastructure**

Research on sense perceptions is tricky; cognitive psychologists have to make sure that subjects report what they *immediately* perceive without additional concepts, which they would introduce in most descriptions. Fortunately, their research receives help from other branches of cognitive science, especially neuroscience.

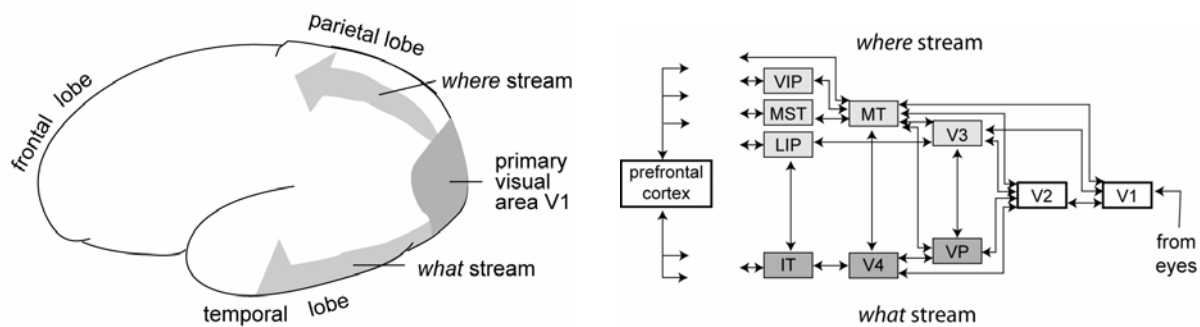
Our sense experiences are so immediate and effortless they do seem to be *given* to us. However, scientists have discovered in our brains a host of complex processes that intervene between the reception of stimuli at the sense organs and the emergence of conscious perceptual experiences. These processes, of which we are unconscious of, constitute what I call the *infrastructure of mind*. The cognitive infrastructure is the locus of current cognitive science.

Research on experiences involves at least two levels of phenomena, the *mental* level (or what I call the situated-personal-level, which highlights that mind belongs to a person consciously engaged in the physical and social world) and the *infrastructural* level. The two levels have radically different properties. Our mental life is conscious and meaningful. It consists of processes that are effortful, flexible, variable in scope, mostly learned, voluntarily controllable, and slow, with characteristic times ranging from seconds to hours or days. In contrast, the mental infrastructure is unconscious and mechanical in the widest sense. It consists of processes that are automatic, rigidly specialized, narrow in scope, mostly genetically determined, beyond voluntary control, and fast, with characteristic times of less than a tenth of a second.

Our mental infrastructure rests on neural substrates. However, infrastructural processes are seldom described in neural terms, because their operations usually involve millions of neurons that exhibit large-scale organizations. They make contact with their neural substrates via brain anatomy and networking. Take the visual infrastructure for example. In terms of brain anatomy, the infrastructure appears as a complicated network with thirty odd interactive brain areas (Box 2). The visual areas are distinguished more by functions than by neural structures, the interactions more by roles than by neural synapses. Usually, the brain anatomy for the visual infrastructure is depicted as an abstract circuit diagram. This is a grossly simplified version of it. It by itself conveys little information about whether the circuitry is neural, electronic, or constituting an “oil refinery,” as some neuropsychologists describe it.

Most infrastructural processes specialize to various functions, which are very specific. The conscious process of having a visual experience, for instance, is based on numerous visual infrastructural processes. Some are specifically sensitive to motion, others to color, and so on. Patients with damaged motion-sensitive processes see everything in slow motion. Those with damaged face-sensitive processes cannot recognize any face, not even their own. In short, a single conscious experience depends on many specialized infrastructural processes, which can malfunction individually and cause specific visual impairments.

We are unaware of the infrastructural processes and cannot control them voluntarily. Mental and psychological concepts such as perceiving and knowing are not applicable to processes in the mental infrastructure. A person sees and is informed by what he sees; his visual experiences are meaningful. His visual infrastructural processes respond to optical signals but do not see; they know no meaning and are not informed. When cognitive scientists talk about “computation,” they usually refer to the infrastructural process. Actually, so-called “computation” is just computer models for a process, not that different from computer models for the weather.



**Box 2. The cognitive infrastructure for vision.** Stimulated by light, photoreceptors in the eyes generate signals that, after some processing, enter the cerebral cortex at V1, the primary visual area at the back of the brain. If V1 is destroyed, a person is blind, perfect eyes notwithstanding.

From the primary visual area, the signals proceed toward the front of the brain via more than thirty secondary visual areas with specialized functions. These areas fall into two major parallel pathways, called the *where* and *what* streams according to their functions. Areas within a stream are densely connected, but only limited crosstalk exists between the streams. A simplified schematic of some visual areas is shown above at the right.

The *where* stream includes areas in the parietal lobe located in the brain's upper part. These areas, such as MT, respond selectively to spatial factors such as movement, speed, direction and vector of motion, and spatial relations among objects. They also contribute significantly to visuo-motor coordination such as tracking eye movement.

The *what* stream includes areas in the temporal lobe located in the brain's underside. These areas, such as V4, respond selectively to features and forms of objects such as color, contrast, orientation, and geometric configuration. In humans there are even areas specialized to faces and work form.

Secondary visual areas in both the *where* and *what* pathways project directly and individually to the prefrontal lobe, where large-scale integration of signals from various visual areas occurs. Most visual areas also have direct connections to brain areas for motor control.

Depending on their proximity to optical stimuli, areas in a pathway roughly organize into "earlier" and "later," "upstream" and "downstream." However, traffic on the pathway goes both ways. An area usually sends output both forward and backward, and it received input from both downstream and upstream areas. The backward projections ensure that our visual experiences are driven not only by optical stimuli but also by such high-level organizations as attention and concept. Because of the backward projection, your primary visual area is excited when you close your eyes and visualize something.

## Eye tracking experiments

The cognitive infrastructure supporting vision is briefly described in Box 2. Its sheer complexity is enough to cast doubt on the passive view that our visual experiences are identical to “ocular irradiation patterns,” as Quine put it. If vision is so simple, what are all these processes in the brain for? If they were redundant, would nature be so wasteful as to evolve them in eons of struggle for survival? The doubt is confirmed by many experiments, including eye-tracking experiments designed to investigate what exactly subjects see.

Let us start with what really are given in the optical stimuli. Our eye is not a simple camera. It is more like the combination of two cameras, one with high resolution but a narrow angle of view, the other low resolution but a wide angle of view. The fovea is the only area in our retina that supports high acuity color vision. It has a tiny visual field, which is about the size of a thumbnail at arms length. Sensitivity to details outside fovea’s visual field drops drastically, as extra-fovea area of the retina support crude panoramic vision. To compensate for the narrowness of the fovea view, our eyes flicker rapidly several times a second to aim the fovea at different spots. An average person’s eyes flicker more than 100,000 times a day. Each flicker, called a saccade, takes 10 to 80 milliseconds to execute. Between two saccades, the eyes fix on a spot for two to three hundred milliseconds to take in the details within the fovea field. Thus the optical input we receive are tiny disjoint bits and pieces, fleetingly and erratically succeeding one other. That is not the contents of our immediate visual experience at all. We see a stable and coherent world. How does our visual infrastructure produce this experience?

To produce what we see, our visual infrastructure must somehow integrate the bits and pieces received by our fovea. To integrate, it must retain something of the previous fixation. According to empiricists, we retain various colors and forms, which we then consciously organize, perhaps by employing the myth of physical objects. Experiments find that it is not the case.

Cognitive scientists design specially eye tracking machines that monitor the fixation and eye movement of a subject while he reads a text or examines a picture on a computer screen. The machine enable researchers to know exactly what optical pattern stimulates the subject’s eyes, specifically, what pattern irradiates his fovea with acute vision. Box 3 provides two sample displays. The dotted letter is the fixation point of the eyes. The field of the fovea is about twenty letters.

When the eye tracker detects the beginning of a saccade, it with electronic speed changes the display on the screen before the saccade completes. For instance, the display contains sentences written in a changing mixture of upper and lower cases, which are changed during a saccade and change back during the next saccade and so on. The irradiation patterns of “ThE” and “tHe” are very different. Instead of complaining about the difficult it causes in reading, however, subjects *do not see* the change at all. Many are surprised, some annoyed, that they can be the victim of such a trick. More interestingly, often subjects do not see the change even if they know about it. This happened to the scientist who designed the eye-tracking machine. He blamed the machine for failing to make the desired changes, while he himself was to be blamed for failing to see them. If people do not see the irradiation patterns on their fovea, what do they see?

### Box 3. Displays in eye tracking experiment

- indicates the position of the subject's eye fixation.

Subjects do not see the changes in the display as they read the text.

Fixation $n$	xx xxx notice anything straxxx xx xxxx xxxxxxxx •
Fixation $n+1$	xx xxx xxxxxx anything strange ix xxxx xxxxxxxx •
Fixation $n+2$	xx xxx xxxxxxx xxxxxxxx strange in this xxxxxxxx •
Fixation $n$	Do YoU nOtIcE aNyThInG StRaNgE iN tHiS dISpLaY? •
Fixation $n+1$	dO yOu NoTiCe AnYtHiNg sTrAnGe In ThIs DiSpLaY? •
Fixation $n+2$	Do YoU nOtIcE aNyThInG StRaNgE iN tHiS dISpLaY? •

What our visual infrastructure retains across saccades and integrates become the contents of our conscious experience. Eye tracking experiments show that we retain very little about colors and forms across saccades. The same happens when subjects look at pictures as when they read. Often a subject's eyes fix on a spot on a big red balloon, saccade elsewhere, and saccade back on the balloon, but the subject fails to see that the color of the balloon has changed from red to green. Experimental data are convincing that what we retain across eye flickers are not colors and forms but something more abstract: meaning in the case of reading, and persisting physical objects in pictures. These are what we see.

### Sense impressions and physical objects

As discussed earlier, physical objects are more complex than sense impressions in two ways: individuality and endurance. Sense impressions are bundles of qualities that exist only when being perceived. In addition to qualities, a physical object also has its numerical identity that individuates it and identifies it as an entity that endures through qualitative changes and persists when we are not looking. We say *this* yellow banana is the *same one* that was green yesterday. Numerical identity, absent in sense impressions, is a nuisance to empiricists. Quine advocated banning it in his regiment of language. Can he ban it from vision also?

Eye-tracking experiments have shown that we do not see merely bundles of qualities such as various colors and forms. What more do we see? Does numerical identity feature in the contents of our immediate visual experience?

The characteristics of our visual infrastructure offer a clue. Beyond the primary visual area, it divides into two pathways: the *what* and *where* streams. Areas in the *where* stream are responsible to tracking positions and motions. Spatial position and temporal endurance in motion are the essence of numerical identity.



The *what* stream contains areas that are sensitive to colors, edges, and simple shapes. These qualities are sufficient for sense impressions, but not for normal visual experiences. The quality areas are excited early on in processing signals generated by optical stimuli. Many experiments find that even when they are excited, the subject may still fail to anything – not shape, not color, not anything – if mishaps occur later in the pathway. Many things can go wrong. The one most interesting to us is the infrastructure’s failure to excite what cognitive psychologists call a *token* that binds the excited *types* to produce a visual experience. *Token* and *type* are *identity* and *quality* in our terminology. The necessity of token in vision is revealed in many phenomena, including repetitive blindness and attentional blink.

Perhaps you have noticed that line editors or proofreaders often overlook the second instance of a repeated word in a sentence. The weakness is not peculiar to their trade. *Repetitive blindness*, as psychologists call it, is a robust phenomenon observed in many laboratory experiments. When the words in the sentence “her jacket was red because red is conspicuous” are presented in rapid succession, most subjects miss the second “red.” They do not make this mistake when the first “red” is replaced by “pink.” Repetitive blindness is not limited to words; any repeated item is susceptible to selective omission. This effect is strong enough to override people’s sensitivity to meaning. People fail to see the second “red” even when its omission produces a meaningless word sequence “her jacket was red because is conspicuous.”

What cause repetitive blindness? After a variety of experiments to test various hypotheses, psychologists converge on the explanation in terms of failure to individuate particular objects. Each word in a sentence must be processed by the visual infrastructure as an individual object if it is to be consciously perceived. For a repeated word, the first occurrence is individuated and becomes contents of the subject’s visual experience. Objects usually persist through brief absence, and the infrastructural processes for them persist similarly. Thus the numerical identity of the first occurrence persists and assimilates the properties of the second occurrence. There is no excitation for a second numerical identity to bind the qualities registered from the second stimulus. Consequently, the subject does not see the second stimulus.

Repetitive blindness is related to other psychological phenomena that have been confirmed in many independent experiments, e.g., apparent motion, attentional blink, and object specific advantage. They all point to a similar conclusion. Normal subjects fail to see a stimulus if the infrastructural processes for *type* are excited but the processes for the *token* are derailed. It is the *token* or *numerical identity* of the object that binds various qualities into an entity in visual experience. Without the extra infrastructure for the numerical identity, an optical stimulus can incite a bundle of quality registrations but no visual experience. Without the notion numerical identity that binds various qualities into the perception of an individual object, we simply have no visual experience of the bundle of qualities.

All these experimental results suggest an answer to the question of what the visual infrastructure retains across eye saccades to produce coherent visual experiences. It retains not the impressions of color or shapes but the abstract concepts of *particular enduring objects*, which are often identified by their spatial relations. As Strawson argued, the employment of our ordinary, full-

blown notion of physical object is indispensable to a strict, and strictly veridical, account of our sensual experience. Kant was literally correct: intuitions without concepts are blind.

Quine advocated the “naturalization of epistemology.” Instead of arguing about generalities in a vacuum, philosophers should conduct epistemology as a science. Cognitive science seems to turn some empiricist doctrines on their head. Under normal conditions – not knocked on the head or under substance influence – we do not see bundles of qualities and then introduce the arbitrary notion of physical object to organize the received sense impressions. Thanks to the automatic and complex processing in our visual infrastructure, we immediately see physical objects, and if we please, later construct stories about our experiential contents as sense impressions or whatever culturally fashionable. Empiricists cannot salvage sense impressions by doing away with the visual infrastructure, because it is indispensable of visual experience. If its color areas are damaged, we are colorblind. If the gateway to it, V1, is destroyed, we are totally blind.

If naturalized epistemologists take scientific findings on our perceptual experiences seriously, what changes would they make to empiricist tenets?

Talk presented in the Department of Philosophy  
University of Sydney  
May 1999