The Recognition Moment: The Cognitive Dynamics of Pictorial Recognition

By Ron Gallagher

Abstract: The human visual system can identify the ‘spatial envelope’ of a scene, whether in a picture or real-life, in less than 100 milliseconds (one-tenth of a second). This is known as the ‘gist view’. In this initial recognition moment we roughly categorize what is in view, identify a few objects and start the pictorial interpretation process. The brain processes which are set in train in this moment are too fast and too complex to examine introspectively but recent breakthroughs in psychophysics and brain imaging technology have enabled researchers to dissect the recognition moment millisecond by millisecond. These techniques are providing insights into the cognitive dynamics which are in play in the first moment of looking at a picture. This recent research into ‘gist views’ and change-blindness eliminates the possibility that symbolism or resemblance is involved in the initial interpretation process and confirms a number of Flint Schier’s hypotheses about the cognitive dynamics of pictorial recognition.

Close your eyes and move your head so that you are facing in a new direction. Now open your eyes for half-a-second. That brief look has given you enough information to describe the scene in front of you and identify a number of objects.

The human visual system can identify the ‘spatial envelope’ of a scene and recognise a few objects in less than 100 milliseconds (one-tenth of a second). This is what vision researchers call the ‘gist view’. It has been found that if you flash pictures of objects onto a screen at the rate of ten-a-second, viewers easily identify the objects and scenes. Within the first 100 milliseconds, we categorise scenes as open-space/enclosed-space, and objects into categories such as animal/non-animal.[1]

Psychophysical and brain-scan experiments have found that in the 100-200 milliseconds after the initial categorization, the visual system refines that categorization and identifies specific objects in the scene. It is crucial to note that usually these experiments monitor the brain activity of subjects as they view *pictures*. The overriding view in vision science is that in the ‘recognition moment’ there is no significant difference between how the human brain responds to an actual object and to a picture of an object. This view is reflected in the methodology of vision research which now rarely uses three-dimensional objects to conduct research on visual recognition of objects. Instead, vision researchers use drawings, photographs and computer generated images.

Consequently the last 40 years of research into visual perception has, by default, been research into pictorial perception.

Recent breakthroughs in brain imaging and psychophysical technology now make it possible to track the timecourse of the brain processes of visual recognition millisecond by millisecond.[2] Such studies of visual perception and recognition have profound significance for the field of visual aesthetics because they are beginning to reveal the cognitive dynamics of how human beings recognize objects and pictures of objects.

The Visual System Trades Speed of Recognition for Visual Detail

The categorisation of a scene made in the initial few hundred milliseconds of an encounter is based on a visual stimulus which, due to the structure of the retina, is largely monochrome and blurred. The rod and cone arrangement of the human retina ensures that most of our visual field is effectively black and white and very coarse in its resolution. Only the fovea, located near the centre of the retina, can finely focus and resolve colour. Thus, in order for us to see colour, detail and actual objects, our eyes must scan the scene in a series of saccades. This takes time. Our eyes saccade around a scene at the rate of about 3 times a second, each time fixating on a point so that the fovea can focus. Studies confirm that the gist of a scene is available for purposes of categorisation of a scene on or before the first saccade – sometimes less than 50 milliseconds.[3]

According to the neurologist V.S. Ramachandran our visual apparatus evolved to enable our arboreal ancestors to make extremely fast judgements about movement and size which ensured their survival.[4] His point is that, from an evolutionary perspective, being able to see highly detailed coloured shapes in the ‘recognition moment’ is not a priority – being able to spot movement is. Our visual system is equipped to make an almost instant evaluation of whether a scene is threatening. Aude Oliva’s work at the Computational Visual Cognition Laboratory at MIT also supports the view that quick, rough and ready categorisation of a scene is the priority in the early stages of a recognition encounter. Her work suggests that before we identify an object in a picture we make a scene-based categorisation. That is, we identify a picture as a portrait, landscape or interior scene long before we identify an object in a scene.

The ‘recognition moment’ is a busy time for the visual system. Dozens of parallel mechanisms and object recognition modules dissect the stimulus for movement, edges, shapes, orientation, and other significant features. It is not completely understood how all these parallel analyses of the light stimulus come together to form the percept which we experience. However, one thing that is certain is that the processing of the various stimuli takes place at a different pace and using different brain modules and paths according to the nature of the stimulus. For example, it takes longer to process a colour image than a black and white image, and a colour photograph will stimulate mechanisms not stimulated by a black and white line drawing. A letter string always triggers activity in an area of the brain called the *fusiform gyrus* but pictures of objects do not.[5] Brain imaging experiments can correlate which areas of the brain are activated when we encounter different kinds of visual stimulus. By examining the brain scans of subjects it is possible to establish what kind of visual stimulus has generated the brain activity pattern. In 1998 the neurologist Semir Zeki conducted experiments in which he sought to establish which areas of the visual brain were active in processing different colours and shapes. He did this by presenting the subjects with ‘Mondrians’ (pictures featuring large geometric areas of colour) and simultaneously scanning their brain activity.[6] Similar experiments focussing on object recognition have also been conducted in which images are flashed onto a screen and the subjects’ brain activity monitored for neurological spikes indicating which areas of the brain are involved in the recognition process.[7]

Thousands of these recognition experiments have been conducted in the last few decades and researchers have found *no significant difference* in the time it takes to recognize an object and a picture of that object. Many of these experiments have been done using macaque monkeys and other primates such as chimpanzees. These higher primates have no difficulty recognizing drawings and photographs of familiar objects. The neural processes involved in the recognition process are very similar for human beings and macaques.[8] These findings are fatal for a theory which holds that pictures depict objects using a symbol system. It is inconceivable that macaque monkeys are interpreting a symbol system when they recognize a picture of a cat or a tree.[9] Like us, they are using natural visual recognition abilities to make an extremely quick assessment of the object in front of them. Nothing in recent vision research supports the idea that recognition of objects in pictures involves interpreting a symbol system. The evidence also suggests that resemblance theory which invokes property matching as a mechanism of recognition is at best simplistic and at worst unworkable. Recognition occurs so quickly that there is not enough time for the higher brain functions or memory to be accessed. Consequently no resemblance matching can take place in the crucial first few hundred milliseconds of recognition. The speed with which we obtain a ‘gist view’ precludes the possibility that our visual system is matching the shapes of objects in our visual field with stored memories of shapes we have encountered. Oliva observes:

> speed and accuracy in scene recognition are not affected by the quantity of objects in a scene, and recognition can be achieved equally well even when object information is degraded so much that objects cannot be locally recovered.[10]

Oliva’s work shows that the way the brain registers a scene and recognizes objects cannot be based on some kind of shape-matching. The gist view of the scene is not built up from its parts. In the recognition moment we take in the whole scene at once. Oliva argues that in the recognition moment the visual system makes a number of rough and ready judgments about what is in view based on rather abstract and generalized categories. In their paper ‘Building the Gist of a Scene: The Role of Global Image Features in Recognition’, Oliva and Torralba[11] argue that the visual system makes at least two distinct types of categorisation of the same scene – object-centred and scene-centred.

The gist view is typically identified using scene-centred categories such as large-space/small-space, man-made-scene/natural-scene, enclosed/unenclosed-space. We categorise paintings and photographs in a similar way – landscapes, interiors, street scene etc. In Oliva and Torralba’s opinion the gist view of a picture yields information about whether we are looking at a landscape or interior based on very abstract properties of a scene. In the recognition moment we are not evaluating objects, we are evaluating very coarse light features which indicate the presence of a horizon, or a wall, or a corner or a tree. Oliva and Torralba also found that they could refine the vocabulary for scene-description into measures of the volume of the space and the scene properties such as, depth range, openness, expansion, ruggedness, verticalness, naturalness, busyness and roughness. The theory is that we categorise the ‘spatial envelope’ using something like these properties before we go on to identify a definite object. This basic level categorisation enables us to recognise a street-scene, a forest, a highway, a panorama, an office or a living room in our first glance whether it be in a picture or in real life.

Schier's Two Stages of Pictorial Interpretation

In the recognition moment, a picture of a scene triggers many of the same recognition mechanisms which are triggered by a similar real-world scene. The mechanisms of recognition in human beings and macaques have evolved specifically to identify real-world objects.[12] In those initial 100 milliseconds of visual stimulus the visual system construes the light stimulus as coming from a real world object. It doesn't matter how abstract the stimulus is, the human visual system will attempt to recognize an object in it. It could be a stain on a wall, a couple of lines on a piece of paper or even a Jackson Pollock painting, but whatever it is, the first act of our visual system in the recognition moment is to try and categorise it and identify it as a real-world scene. This powerful drive of the human visual system explains many of the puzzles involving so-called optical illusions. The fascination with such phenomenon derives from the fact that when we look at something like the Ponzo illusion (fig. 1) our visual brain is conflicted; it is attempting to resolve two contradictory hypotheses about what is in view – a real world scene involving receding railway tracks or marks on a piece of paper. We cannot help seeing the marks on the paper as tracks receding into the distance. When we look at any picture our visual system reads the marks as 3D objects despite the fact that we know that we are looking at marks on a flat plane. This double awareness is the famous 'twofoldness' which has been the subject of much debate in the theory of depiction.[13] Each time we visually attend to the Ponzo drawing the 'gist' view of the receding tracks competes for our attention against the knowledge that we are looking at marks on a surface. After our initial encounter with a picture and as we scan around the picture picking out objects and features, the visual recognition module which categorized the scene based on the 'gist' view continues to broadcast its interpretation of the picture

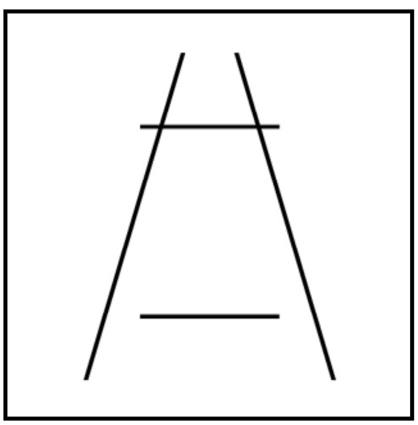

Figure 1 The Ponzo Illusion – the horizontal lines are of equal length

as a real-world scene. It seems that scene and object categorization in the 'gist view' is to some extent encapsulated from the awareness that we are looking at a picture. The 'gist view' impression is always the impression of a real-world scene or object. This is a very important finding and, to some extent, Flint Schier predicted it in his 1986 book *Deeper into Pictures.* Schier argues that in order to understand how we see objects in pictures we must first understand how we see objects in the world. This has become known as the 'recognition theory of depiction'. Lopes comments in 'Pictures and the Representational Mind':

> The recognition theory is inspired by the thought that some of the cognitive capacities engaged in perceiving things face-to-face are engaged in interpreting pictures of those things. Some version of this thought underlies much empirical research on pictures and vision. Only if the capacities engaged in picture interpretation are among those involved in perceiving things face-to-face can a study of pictures tell us anything about vision – or *vice versa*.[14]

Lopes argues that recognition theory is a better fit with empirical research on vision than resemblance and symbol theories simply because of its explanatory power.

In *Deeper into Pictures* Schier accounts for the phenomenon of 'twofoldness' by proposing a two-stage theory of picture interpretation. Schier uses the example of a picture of a peach. In the first stage we momentarily entertain the hypothesis that the picture of a peach (S) is a real peach, and then reject it because the cues for a flat picture-peach are more convincing. Schier comments:

> On this view, when S depicts a peach, the visual system processes cues from S, some of the cues being evidence for the presence of a round, ripe peach, and some of the cues intimating the presence of a flat, rectangular piece of canvas. For some reason the visual system prefers the evidence of S's being a flat, rectangular object to the evidence of its being a round peach. Nonetheless, the fact that one's visual system entertained the peach hypothesis affects one's perception of S: one sees S as a picture of a peach…… in the picture-peach case your visual system (or your visual homunculus if you like) entertains the conflicting 'peach' and 'flat' hypotheses; they are incompatible so it must reach a verdict and it decides that the

> ‘flat’ hypothesis is better grounded; the result is that you see a flat picture of a peach.[15]

In short, Schier’s theory suggests that S triggers object-recognising capacities even though that does not amount to recognising S as an object (O). He explains:

> Surely the ‘cash value’ of this claim must be equivalent to the claim that S induces me to entertain (albeit unconsciously) the O-hypothesis with respect to S.[16]

Schier concludes that the analysis of depiction necessarily involves a ‘cognitive theory of mental activity.’ [17] The primary reason Schier makes this claim is his conclusion that the first stage of the process of pictorial interpretation (which roughly coincides with what I call ’the recognition moment’) is not a conscious process and involves innate recognition abilities.

The Ability to See Scenes, Animals and Objects in Depictions is Innate

The ability to see pictures is a biological endowment. We see shapes in clouds, faces in stains on the wall, animals in shadows, and figures in the stars. Our ancestors 40,000 years ago saw these things as well and began augmenting with pigment the shapes of animals which they saw on the walls of caves.[18] This is, arguably, how depiction was born. It is a human invention, but it is built upon the innate ability to see animals and objects in every facet of our environment. It is an ability which relies on speed. There is a clear evolutionary advantage in being able to identify what we encounter very quickly and based on very rudimentary stimulus. When we encounter things in the dark we may react based on very little visual evidence. The conditions under which something is viewed vary radically according to light conditions, distance and whether we or the object are moving. The resources that the visual system uses to interpret the visual stimulus in that first few hundred milliseconds do not distinguish 2D from 3D and may not even discern whether a picture is a photograph or a line drawing.

We see objects in pictures because identifying objects is the primary objective of our visual system. As our eyes scan a picture each act of looking is underpinned by the object-seeking mechanisms which dominate visual perception. We cannot turn off this primal drive of the visual system. When we look at the Ponzo drawing it is almost impossible not to see the lines on the page as tracks receding into the distance. The recognition module of our visual system cannot help construing those lines as something real in the world. This recognition module is oblivious to style and technique. It doesn't register whether something is a photograph or a line drawing. It evolved to recognize and track real-world objects in spite of changing light conditions, poor visibility, camouflage, proximity, angle, movement and size. In short it is an object-seeking system which will lock onto a real-world interpretation of a scene regardless of the myriad distractions which may accompany the act of looking.

Experiments in vision science use photographs, drawings and computer generated images more-or-less interchangeably. Some of these pictures are very rudimentary.[19] After over 40 years of such experiments no evidence has emerged indicating that style or technique of depiction affects speed of recognition. We seem to perceive a far less densely populated and detailed image than we imagine. An artist can trigger recognition with a rough sketch as effectively as with a full-colour realist painting. And there is no evidence that the three-dimensionality of the object is a recognition factor. When we speculate on how the technique and style of a picture alters perception of its subject, we are analyzing post-recognition cognitive processes. In Schier's terms these are second-stage interpretations. These post-recognition mechanisms are not what kick-starts the interpretation and may not contribute to the recognition process. Perhaps one of the reasons aesthetics has been locked in a debate about whether pictures are a language-like symbol system or rely on resemblance is that philosophers and art critics have been focussing on the wrong interpretive processes. They have been analyzing recognition processes in the second and subsequent stages of visual processing whereas the crucial recognition event – the initial interpretation – takes place in that few hundred milliseconds which constitute the first stage. Composition, projection and style play a role in the way that we read a picture but they are not something that the human visual system evaluates within the crucial first half-a-second of looking at a picture

Recent Vision Research Supports Schier's Hypothesis

It is now possible, thirty years after Schier published his thesis, to analyse the 'unconscious' cognitive mechanisms which Schier suggests are at the root of pictorial interpretation. His hypothesis can be tested using the increasingly sophisticated techniques of brain imaging and psychophysics. It may be many years before the mechanisms of human perception are fully understood, but the progress that has been made in recent years reveals that these mechanisms are radically counter-intuitive. For example, it is tempting to imagine that a scene registers itself on our retinas as a kind of snapshot which then gets processed by our visual brain. Under this model of perception a scene more-or-less presents itself fully formed to our retinas and our perceptual mechanisms then set about analyzing and recognising the stimulus. This is clearly not what happens. There is no fixed image to analyse and, according to Daniel Dennett, it is misleading to imagine that there is a coherent analyzing perceptual system which deconstructs the stimulus and presents it to consciousness as a percept. Dennett points out that at the onset of visual stimuli dozens of separate processes are set in train in our visual system. These processes start at different times and yield results at different times. These findings support Schier's hypothesis that multiple stages of processing are involved in pictorial interpretation. Dennett observes that not only does our visual brain process the shapes and colours in a scene at radically different rates but these stimuli are processes in different neural pathways and do not 'come together' anywhere in the brain as a coherent percept. He describes the process thus:

> Visual stimuli evoke trains of events in the cortex that gradually yield discrimination of greater and greater specificity. At different times and different places, various "decisions" or "judgements" are made; more literally, parts of the brain caused to go into states that discriminate different features, e.g., first mere onset of stimulus, then location, then shape, later color (in a different pathway), later still (apparent) motion, and eventually object recognition. These localised, discriminatory states transmit effects to other places, contributing to further discriminations and so forth. The natural but naïve question to ask is: Where does it all come together? The answer is: Nowhere……There is no one place in the brain

> through which all these causal trains must pass in order to deposit their content "in consciousness."[20]

According to Dennett perception is like a vast film-editing job which never results in a finished piece. Our brain is constantly re-constructing the scene in front of us according to whichever sensory stimulus is currently dominant. The visual system is dozens of systems all working in parallel and linked to other perceptual mechanisms.[21] The speed and complexity of these mechanisms precludes analysis by introspection.

Recent vision research indicates another reason to doubt the introspective method. Change-blindness and attention-blindness experiments confirm that our subjective experience of visual perception is largely a post-recognition reconstruction of what we actually see.[22] These experiments indicate that our visual system registers very little in any given scene. What we do notice and recall is radically affected by what we expect, and what we are interested in seeing. The surprising finding in these studies has been that not only are subjects often mistaken about what they have just seen, but make up details which support their memory of the scene. In effect, our visual memory reconstructs history to fit what we think we have seen. This research confirms that our own account of our perceptual processes is unreliable as a means of identifying perceptual processes and our account of the details of the experience similarly unreliable. This is a fairly devastating outcome for philosophical accounts of pictorial perception, such as the 'experience resemblance' and 'seeing as' accounts, which rest on the phenomenology of the visual experience.[23] Research into visual metacognition[24] which has emerged from the attention-blindness studies confirms that people not only fail to notice or attend to most of a scene, but always overestimate their ability to remember what they have seen. That is, we massively overestimate our powers of perception, attention and memory, and we similarly overestimate other people's powers of visual cognition.

In his paper entitled 'Is the Visual World a Grand Illusion?' Alva Noë summarises the significance of the change-blindness work as follows:

> The fact of change blindness is widely thought to have several important consequences. First, perception is, in an important sense, attention-dependent. You only see that to which you attend. If something occurs outside the scope of attention, even if it's perfectly visible, you won't see it. In one study, perceivers

> are asked to watch a video tape of a basketball game and they are asked to count the number of times one team takes possession of the ball (Neisser, 1976; Simons & Chabris, 1999). During the film clip, which lasts a few minutes, a person in a gorilla suit strolls onto the centre of the court, turns and faces the audience and does a little jig. The gorilla then slowly walks off the court. The remarkable fact is that perceivers (including this author) do not notice the gorilla. This is an example of inattentional blindness. Second, perception is gist-dependent. Some changes, for example, in the features that affect the gist of the scene, are more likely to be noticed (Simons & Levin, 1997). Third, it seems that the brain does not build up detailed internal models of the scene; that is, it doesn't perform the integration of information across successive fixations, contrary to the assumption of traditional orthodoxy (Blackmore et al., 1995; Rensink et al., 1997; O'Regan et al., 1999; Rensink et al., 2000; Noë et al., 2000). Or if it does, we have little easy access to this detail. If we did, then presumably we'd keep track of change better than we do.[25]

The ‘basketball and gorilla’ experiment to which Noë refers has now become part of the popular folklore of change-blindness, and has been repeated on network television to the general delight and dismay of unprimed viewers. One major implication of this change-blindness work for theories which focus on the object-presenting *experience* of pictures (such as experience resemblance and ‘seeing-as’) is that pictorial experience itself has now become the subject of purely psychological interest and is not the key to unlocking the cognitive processes which enable us to recognise content in pictures. Attention blindness research indicates that we are radically mistaken about our visual phenomenology. It follows that intuition is not serving us well in our quest to understand how we see pictures, and the technique of introspection, as applied to these pictorial experiences, is suspect. Dominic Lopes comments:

> Unlike experiential theories, the recognition theory does not take the fact that pictures typically cause object-presenting experiences as primitive. It views this fact as in need of explanation. [26]

Lopes’ comment echoes Schier’s position on the subsidiary role the resemblance experience plays in pictorial interpretation. Schier argues ‘The respect in which S resembles its depictum O is this: there is an overlap between the recognitional abilities triggered by S and O.'[27] Schier is here saying that the resemblance experience is the *result* of recognition abilities being triggered. Resemblance does not trigger recognition and therefore has no role in the primary interpretation process. Understanding the phenomenology of the experience of pictorial recognition, such as the feeling of resemblance we sometimes feel when we look at a picture, is useful from a psychological perspective. However, that goal is subsidiary to, and dependent on, understanding the object-recognising cognitive processes which underlie pictorial perception.

The Alliance of Aesthetics and Cognitive Science

Research on change-blindness and 'gist' views is already uncovering surprising facts about how the visual system recognises scenes and objects. Rigorous experimentation using the methodologies of psychophysics and neuroscience can dissect the act of recognition and relate object-recognising behaviour to brain activity. The question posed by aesthetics is: 'How are objects recognised in pictures?' Clearly we are still looking for an answer to that question but the accumulation of new facts about vision and perception mean that the kind of answer we are looking for has changed. It is no longer enough to speculate that some kind of symbol interpretation is taking place when we look at a picture, because there is no evidence of symbol manipulation and ample evidence to the contrary. The onus is therefore on advocates of symbol-theory to provide empirical evidence of symbol interpretation in picture-object recognition process. Neither is it enough to speculate that resemblance plays a role in pictorial recognition. It is clear that resemblance is the output of prior mechanisms of recognition and it is these mechanisms which are doing the heavy lifting of pictorial recognition. The fundamental question is 'What are the mechanisms of recognition and how are they triggered?' Importantly, it looks as if the 'mechanisms' in question are the same whether we are inquiring about recognition of objects in the world or objects in pictures. To the extent that aesthetics is concerned with understanding fundamental human perceptual processes, as Gombrich and Goodman envisaged, the aims of research into visual cognition and the aims of those in the philosophy of art are one and the same. In order to understand how we see *pictures* we must first understand how we see *objects*.

Examining the process of pictorial interpretation using the timeframes and techniques of cognitive science immediately clarifies the issue of how we see objects in pictures. By dissecting the act of looking at a picture millisecond by millisecond it is possible to separate object and scene recognition from other acts of pictorial interpretation. The gist view which is obtained in the recognition moment – that first few hundred milliseconds – is the prime act of recognition and pictorial interpretation, and our subsequent analysis of the picture in front of us depends on that act of recognition. Knowing what a picture represents is crucial to all subsequent acts of pictorial interpretation. Rollins comments 'recognition theories can be taken to identify necessary conditions……for the interpretation of works of pictorial art.'[28]

The Two-way Street: the Changing Nature of Aesthetic Enquiry

When Lopes and Rollins say that aesthetics and cognitive science is a two-way street[29] they are really saying that the nature of explanation and what is being explained is changing. Lopes comments on the analogous case of the changing nature of enquiry in the philosophy of mind:

> The trend in philosophy of mind itself is towards a greater dialogue and interaction with the behavioural and brain sciences. Philosophers of mind have not simply acquainted themselves with the controversies, methods and results of the empirical sciences. Many of the more heated philosophical debates – about individualism for instance – concern how to interpret psychological and neuroscientific practice and what conceptual resources are needed for scientific explanation of mental phenomena.[30]

Much debate in the philosophy of mind now centres around the evidence about how the brain works which is emerging from cognitive science. This change in the nature and scope of debate is also happening in the philosophy of art. Cognitive science isn't just providing answers to old questions, it is posing new questions and opening up new debates. The old questions and debates are being superseded by new concerns as we begin to understand more about the cognitive dynamics of pictorial perception. The recognition moment is the foundation moment for every perception that follows. It is the foundation for further recognition as much as it is the foundation for aesthetic judgments about high level features of pictures, including style composition and meaning, because these judgments rest on the initial act of recognition. In this sense it is imperative to get a purchase on the cognitive mechanics of that recognition moment on which all subsequent discriminations hinge.

Close your eyes again. Now look towards the door. Open them for a split second and close them again. Try and recall the details of the door. What kind of door knob has it got? Are the hinges showing? What kind of mouldings are the architraves made of? Are the panels tapered? Yes, you recognized the door, just as you recognized the scene which you turned your head towards at the beginning of this essay. But how much detail did you perceive? Did that split second view provide enough visual evidence for you to pick out doorknobs and architraves or was the gist view more general? What are the elements of the gist view? The sciences of visual cognition are already exploring this question and in doing so they will no doubt uncover more surprising facts about how we perceive the world and pictures.

Notes

[1] See Aude Oliva, 'Gist of a Scene', in Laurent Itti, Geraint Rees, and John Tsotsos (eds.), *Neurobiology of Attention* (New York: Academic Press, 2004).

[2] See Rufin Vanrullen and Simon J. Thorpe, 'The Time Course of Visual Processing: From Early Perception to Decision-Making', *Journal of Cognitive Neuroscience,* 13/4 (May 15 2001), pp. 454-61. and Jeffrey S. Johnson and Bruno A. Olshausen, 'Timecourse of Neural Signatures of Object Recognition', *Journal of Vision,* 3/7 (September 2, 2003), pp. 499-512.

[3] Monica S. Castelhano and John M. Henderson, 'The Influence of Color on Perception of Scene Gist', *Journal of Vision,* 5/8 (September 23, 2005), pp. 68-68.

[4] V. S. Ramachandran, 'The Neurological Basis of Artistic Universals', *Art and Cognition,* (9th December 2002).

[5] See Anna C. Nobre, Truett Allison, and Gregory Mccarthy, 'Word Recognition in the Human Inferior Temporal Lobe', *Nature,* 372/6503 (1994), pp. 260-63., Alan J. Pegna et al., 'Visual Recognition of Faces, Objects, and Words Using Degraded Stimuli: Where and When It Occurs', *Human Brain Mapping,* 22/4 (2004), pp. 300-11. or B. Rossion et al., 'Early Lateralization and Orientation Tuning for Face, Word, and Object Processing in the Visual Cortex', *Neuroimage,* 20/3 (Nov 2003), pp. 1609-24.

[6] Semir Zeki, *Inner Vision: An Exploration of Art and the Brain* (Oxford: Oxford University Press, 1999).

[7] See, for example, R. Q. Quiroga et al., 'Invariant Visual Representation by Single Neurons in the Human Brain', *Nature,* 435/7045 (Jun 23 2005), pp. 1102-07.

[8] See K. Tanaka, 'Mechanisms of Visual Object Recognition: Monkey and Human Studies', *Current Opinion in Neurobiology,* 7/4 (Aug 1997), pp. 523-29. and N. Sigala, F. Gabbiani, and N. K. Logothetis, 'Visual Categorization and Object Representation in Monkeys and Humans', *Journal of Cognitive Neuroscience,* 14/2 (Feb 15 2002), pp. 187-98.

[9] See M. C. Booth and E. T. Rolls, 'View-Invariant Representations of Familiar Objects by Neurons in the Inferior Temporal Visual Cortex', *Cereb Cortex,* 8/6 (Sep 1998), pp. 510-23.
[10] Oliva, 'Gist of a Scene'.

[11] Aude Oliva and Antonio Torralba, 'Building the Gist of a Scene: The Role of Global Image Features in Recognition', <http://cvcl.mit.edu/Papers/OlivaTorralbaPBR2006.pdf>, accessed 15th May 2007.
[12] See Booth and Rolls, 'View-Invariant Representations of Familiar Objects by Neurons in the Inferior Temporal Visual Cortex', (P. Janssen, R. Vogels, and G. A. Orban, 'Macaque Inferior Temporal Neurons Are Selective for Disparity-Defined Three-Dimensional Shapes', *Proceedings of the National Academy of Sciences of the United States of America,* 96/14 (Jul 6 1999), pp. 8217-22. P. Janssen et al., 'Macaque Inferior Temporal Neurons Are Selective for Three-Dimensional Boundaries and Surfaces', *Journal of Neuroscience,* 21/23 (Dec 1 2001), pp. 9419-29. and G. Kovacs et al., 'Effects of Surface Cues on Macaque Inferior Temporal Cortical Responses', *Cerebral Cortex,* 13/2 (Feb 2003), pp. 178-88.
[13] See, for example, J. Dilworth, 'Three Depictive Views Defended', *British Journal of Aesthetics,* 42/3 (Jul 2002), 259-78. and Richard Wollheim, 'In Defense of Seeing-In', in Heiko Hecht, Robert Schwartz, and Margaret Atherton (eds.), *Looking into Pictures: An Interdisciplinary Approach to Pictorial Space* (Cambridge, Massachusetts: The MIT Press, 2003).
[14] Dominic M. Mciver Lopes, 'Pictures and the Representational Mind', *The Monist,* 86/4 (2003 2003), p. 642.
[15] Flint Schier, *Deeper into Pictures: An Essay on Pictorial Representation* (London: Cambridge University Press, 1986), p. 192.
[16] Ibid., p. 194.
[17] Ibid., p. 195.
[18] See Richard L. Gregory and E. H. Gombrich, *Illusion in Nature and Art* (London: Duckworth, 1973).
[19] See J. G. Snodgrass and M. Vanderwart, 'Standardized Set of 260 Pictures - Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity', *Journal of Experimental Psychology-Human Learning and Memory,* 6/2 (1980), pp. 174-215.
[20] Daniel C. Dennett, *Consciousness Explained* (London: Penguin, 1993), pp. 134-5.
[21] see G. A. Rousselet, S. J. Thorpe, and M. Fabre-Thorpe, 'How Parallel Is Visual Processing in the Ventral Pathway?' *Trends Cogn Sci,* 8/8 (Aug 2004), pp. 363-70.
[22] See Alva Noë (ed.), *Is the Visual World a Grand Illusion?* (Thorverton: Imprint Academic, 2002b). and D. J. Simons and M. S. Ambinder, 'Change Blindness - Theory and Consequences', *Current Directions in Psychological Science,* 14/1 (Feb 2005), pp. 44-48.
[23] See, for example, Robert Hopkins, 'El Greco's Eyesight: Interpreting Pictures and the Psychology of Vision', *Philosophical Quarterly,* (1997). and Christopher Peacocke, 'Depiction', *The Philosophical Review,* 96/3 (July 1987), pp. 383-410.
[24] Daniel T. Levin (ed.), *Thinking and Seeing : Visual Metacognition in Adults and Children* (Cambridge, Massachusetts: The MIT Press, 2004).
[25] Alva Noë, 'Is the Visual World a Grand Illusion?' in Alva Noe¨ (ed.), *Is the Visual World a Grand Illusion?* (Thorverton: Imprint Academic, 2002a), pp. 5-6.
[26] Lopes, 'Pictures and the Representational Mind', p. 650.
[27] Schier, *Deeper into Pictures: An Essay on Pictorial Representation,* pp. 186-87.
[28] Mark Rollins, 'The Mind in Pictures: Perceptual Strategies and the Interpretation of Visual Art', *Monist,* 86/4 (Oct 2003), p. 611.
[29] See Lopes, 'Pictures and the Representational Mind', and Mark Rollins, 'Special Symposium: Aesthetics and Cognitive Science - Introduction', *Philosophical Psychology,* 12/4 (Dec 1999), pp. 381-86.
[30] Lopes, 'Pictures and the Representational Mind', p. 633.