Perceptual Principles and Computer Graphics

Jon May

Department of Psychology,
University of Sheffield,
Sheffield S10 2TP

Submitted to Computer Graphics Forum, May 2000

Perceptual Principles and Computer Graphics

Jon May

Department of Psychology, University of Sheffield, Sheffield S10 2TP

Abstract

Now that technology allows us to present photorealistic animations of scenically lit objects acting in real-time, the problem of computer graphics has changed from making displays recognisable, to ensuring that users notice what they are intended to see, without being distracted by irrelevant information. Worse than that, the use of veridical displays that are intended to be lifelike runs the risk of introducing unpredictable sources of information, that can lead users to infer all sorts of unwanted details. Traditional visual theory, based upon bottom-up models of feature extraction from the retinal image, cannot inform us about these aspects of perception. Broader based cognitive theories are required that integrate visual perception with attention, memory, emotion and inference. Theories such as Barnard’s Interacting Cognitive Subsystems enable phenomena such as change blindness and the craft principles of film editing to be interpreted within a common framework, supporting extrapolation to computer graphics.

Introduction

Until comparatively recently, the major problem with mass-market computer display technology was making anything recognisable at all. Eighty character-width displays, with eight or nine brilliant green lines per character, slow to respond and slow to decay, somehow enabled people to use their vast new computers with their kilobytes of memory. The pace of change should really astonish us, as we contemplate flat, bright and crisp LCD screens that require separate graphics processors and megabytes of video memory chips just to display our favourite desktop images. It now seems possible for our technological artefacts to display almost anything in as much detail as we would like, whether from a high resolution photographic image or, via skilfully implemented algorithms, by photorealistic rendering from data. In the course of this rapid development, the major problems have themselves changed: now we must ask ourselves what it means for our displays to be recognisable, and what is it in the display that needs to be recognised?

There is a role for psychology in answering these questions. The ultimate purpose of any computer display is for its content to be recognised by a human, and visual perception has been a cornerstone of the discipline since Wilhelm Wundt founded the first Laboratory of Experimental Psychology over a hundred years ago. The involvement of psychologists in human-computer interaction, and specifically in computer graphics, is not unheard of, of course. In the early days, choices of phosphor and of screen refresh times were driven not just by technical and manufacturing constraints, but also by detailed studies into phenomena such as critical flicker fusion frequency and contrast sensitivity. The introduction of colour within displays was (sometimes) backed up by usability studies showing that (sometimes) it improved performance. Compression algorithms were designed to take into account the discriminability of different levels of hue and saturation by the human visual system.

Contributions such as these have played an important guiding role within the development of computer graphics by providing principled and empirically justifiable ground rules. If the human visual system cannot see something, then you know there is no point displaying it like that. Now that the basic visual properties of displays have been determined, we know how to make displays that are readily perceivable and which are, to all extents and purposes, capable of presenting veridical scenes that are identical, in terms of their consequent monocular retinal image, to a natural scene filmed with a camera. If the perceptual processes that operate upon the retinal image can work exactly as they do with ‘real life’ scenes, seeing something upon a computer screen is now no more difficult than seeing it in any other representational medium: so why should psychology still have a role?

Psychology still has a role because it is about more than low level constraints upon the visual system. It is true that if you go into any psychology department and ask to speak to an expert in vision, you will find that they are concerned with detailed problems such as the perception of optical flow, or of binocular disparity in depth perception, or of texture discrimination. The visual stimuli that participants in their psychophysical experiments observe are dots, crosses, and lines of pure colours, not photorealistic or veridical images. Their research is, after all, still directed towards understanding the way by which information gathered by the retina is perceived at all. It is true that computer graphics has gone beyond this stage. There are other psychologists, though, who research later stages in perception, and it is this work that should now be of interest to the computer graphics community.

Does Computer Graphics need Psychology?

In providing a photorealistic image, the problem of making something recognisable has been solved, or at least, overcome. We are now faced with the problem of ensuring that the viewer sees what we intend them to see, rather than something else in the image. We have to ensure that the image is not ambiguous, and that the viewer will not interpret it as something other than we intended, and that the image is not so rich in information that the viewer is unsure about the relevant aspect. We have to be sure that the veridicality of the image does not lead the viewer to treat it as if the imaged scene were really present, and to respond to some channel of information that we have unwittingly introduced. These are the problems that can be investigated by cognitive psychologists, who are interested in the interplay of our thoughts, ideas, memories, emotions and perceptions.

Consider the addition of a realistic face to an audio signal that has been generated computationally, not spoken by a real human. Wonderful techniques have been developed for optimally rendering lip and face movements so that the generated face speaks visually in exactly the way a real person would, but the result is a rather eerie appearance. The absence of emotional components in the facial expression does not mean that we fail to attribute any emotional content to the message. Instead, we actively infer ‘absence of emotion’, which is a very definite and undesirable emotional state. To overcome this, cheek, eye and forehead components can be manipulated to form a basic grammar of affect, and these can be overlaid to add emotional tone: but now we need to know what emotional tones are appropriate, and how they interact with the emotional state of the viewer. Is an interrogative raised eyebrow always perceived as such, or does it appear condescending if the viewer is unsure of themselves?

One study indicates that it is not simple to predict the effects of introducing facial information into interfaces [1]. Students were asked to complete an online questionnaire about a computerised tutorial package that they had been using. A synthetic face was used to guide the students through the questionnaire, with its mouth movements synched to the text of the questions that were spoken. Half of the students saw a face with a ‘neutral’ expression, and half saw a ‘stern’ face that had a slight frown. Standard usability metrics indicated that the stern face improved participants’ performance, in that they took less time to complete the questionnaire, and answered more of the questions. Designers would be mistaken in inferring that all computerised avatars should adopt a stern demeanour in order to enhance usability, however. The actual content of the answers given on the questionnaire revealed that the participants who had been questioned by the stern face reported having enjoyed the tutorial package less, had felt more stressed by it, and had found it less usable, than had those who had been questioned by the neutral face. The emotional tone of the on-screen agent had migrated back into the participants’ assessments of the previous, agentless (and completely unemotional) interface; they were confusing the tenor of their interaction with the stern-faced agent with their assessment of the tutorial package: a classical misattribution effect.

Nothing in a theory of visual perception that is based entirely on the extraction and combination of features from the physical image projected onto the retina can account for such subtle emotional transfer effects. There must be a link between the perceptual processes and other, non-visual processes that identify aspects of the world and use stored knowledge to add in inferred information, allowing us to construct an internal narrative about the scenes that we observe. To design computer interfaces that include realistic, multivariate representations, we must understand the operation and consequences of such inferential processes.

Continuity, cutting and change blindness

Another problem that has attracted a large amount of research effort is that of the realistic rendering of motion through three dimensional space, and how to link it in a usable and ‘natural’ fashion with user interface actions. It has been a long time since our ancestors swung through the trees, after all, and while swooping through abstract cyberspaces may become second nature to our descendants, at the moment the best virtual environments still risk making people nauseous. Leaving aside the buzzword of ‘intuitive’, which often means no more than ‘usable by other people in my lab’, it is worth stepping back and asking whether it really is in the interests of the viewer to have every frame of their trip from A to B animated in front of their very eyes. Especially if they tend to close them to preserve their lunch. Is it worth using all those processor intensive routines to interpolate and blur, when film directors find it just as convenient to cut directly between camera positions? Pans and zooms in cinematography are limited by extensive conventions, that have been developed by a century of experience, and so film should have something to say for computer graphics.

The analogy between film editing and motion and animation is most obvious, and has been addressed by several researchers [2], but a broader argument can also be made: cutting from camera to camera is akin to opening windows on a display, since the 'director' changes the scene in front of the viewer just as instantaneously as the application does, but arguably with less disruptive effect upon the viewer's comprehension of the scene. Most film cuts go unnoticed (an ordinary 90 minute film can contain 1,500 cuts); most unnoticed new windows are those that lead to interaction errors when the user attempts to continue interacting with a different window. To understand how film cutting can help us create a smoother interaction style in computer interfaces, we need to do more than study the craft skills of cinematography: we need psychology to understand why films are easy to watch.

Our ability to perceive and comprehend film has long been recognised as a challenge for theories of visual perception. As early as 1916, Hugo Münsterberg compared the close-up shot to perceptual attention; flashbacks to acts of memory and mental imagery; and the sequencing of shots to the sequential direction of attention around a real-world visual scene [3]. Carroll [4] reports that another early film theorist, Pudovkin, described the role of the film editor as guiding the viewers' attention to certain elements of the scene, the laws of editing therefore being the same as those governing 'ordinary looking'. He and other analysts [5, 6] also discuss the use of close-up shots to magnify critical details to the exclusion of the surrounding scene, in the same way that a viewer in the real world can concentrate upon one part of the scene to the exclusion of the periphery of their gaze.

These early analyses of film concentrated on the relationships between film perception and the attentional capacities of the viewer, what might be thought of as a 'high level' view of perception. Lower-level analyses of film have also been informative. In his exhaustive review of the differences between the images available from filmed and real world scenes, Julian Hochberg [7] pointed out that the optical apparatus intervening between the real world and film, and between the film and the projected image, meant that many of the 'invariances' available in real world scenes were either distorted or no longer invariant in film. That we can still visually perceive objects and their relationships in film was, to him, evidence against Gibsonian ideas of 'direct perception' [e.g., 8], and in favour of the 'traditional' view following Helmholtz. Gibson had argued that our perceptual system has evolved and learnt to perceive certain 'invariances' or 'affordances' within the visual scene directly, without much processing or interpretation at all. Thus the motion of dots fixed to the joints of a dancer is readily seen as human, because they preserve some invariant relationships about limb lengths and motions [9]. The Helmholtzian view is that perception is an exploratory process in which visual sensory information is evaluated against the viewer's expectations about the scene; a top-down, rather than a bottom-up process.

Hochberg's paper had two main aims: a theoretical one, to contrast Helmholtzian and Gibsonian ideas about perception by examining the empirical evidence relating to the perception of dynamic visual scenes as represented in film and the real world; and a practical one, to advocate the application of psychological knowledge of perception to the then emerging technology of computer graphics. His argument was that in the absence of applicable psychological theory, film makers at least had the advantage of being able to point their cameras at real world events, and so many of the constraints upon object construction, appearance and behaviour that our visual systems might make use of were implicitly recorded in the resulting film, despite the optical interventions he went on to detail. Computer graphics, on the other hand, has no such constraints, and its scenes can portray anything, behaving in any fashion, at any level of veridicality, ranging from pixelated monochromatic wire-frame sketches to high-resolution, anti-aliased photographic renderings complete with multiple light sources, reflections, and receding surface textures.

While Hochberg's theoretical argument against Gibsonian perception led him to concentrate on the differences in continual motion available in film and the real world, he also pointed out the problems that film cutting raised, and that to understand how viewers could comprehend edited films:

> 'we can no longer act as though the physics of the pattern of stimulation and the action of direction-sensitive cells and other pattern-analyzing devices in the visual nervous system will suffice to explain the phenomenon at hand or predict the efficacy of the motion picture sequence. Such concepts as schematic maps, schematic events, and cognitive processing (and perhaps even a "linguistics of film") become necessary to any intelligent discussion of the problem.' [7, pp. 22-40]

Despite this assertion, there has been very little work on such 'high level' problems in visual perception in the years since Hochberg's paper, while research at the lower levels has proceeded apace. Perhaps because of this, there has also been little interdisciplinary work between psychologists and researchers and practitioners in Computer Graphics, where the dominant mode is the development of algorithms for rapidly rendering veridical, film-like imagery from underlying models of objects. While the research in this area is methodical and exhaustive, little consideration is given to the processes of perception, and psychological theories play no role in directing research.

The recent interest in the phenomenon of 'change blindness' illustrates the problem that follows from a focus upon low-level physical perceptual processes. Change blindness is a fascinating affront to our expectation that we should easily be able to spot the appearance or disappearance of whole objects within our field of view. The basic phenomenon is easily illustrated with the 'flicker paradigm', in which original and modified versions of a scene are alternated every second or so, with a 40-80 millisecond blank frame intervening [10]. Observers cannot report the nature of the modification without exhaustive serial inspection of the scene. The change does not 'draw attention' to itself, contrary to the expectations of those unfamiliar with the effect. Figure 1 shows a schematic of a typical flicker paradigm presentation.

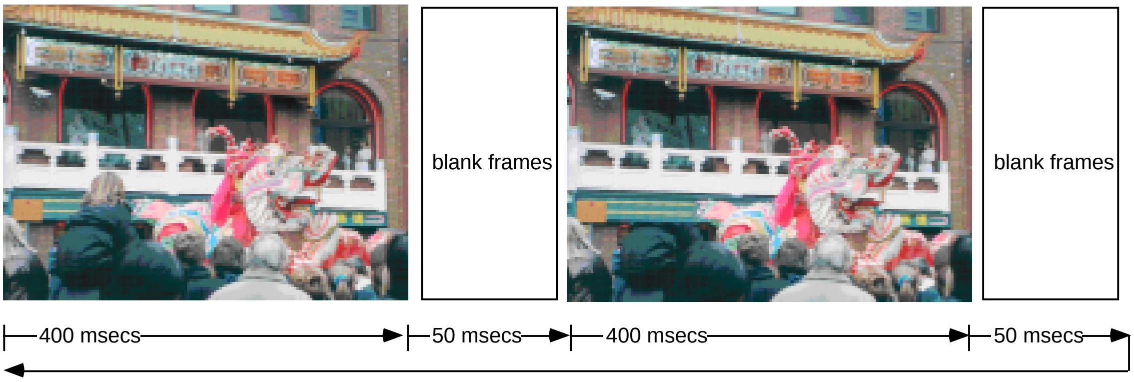

Figure 1: Change blindness can be induced by the ‘flicker paradigm’, in which two versions of a scene are shown in alternation, each presentation lasting around 400 milliseconds, with a blank screen of between 40-80 milliseconds between each presentation.

The two scenes can alternate directly, without an intervening blank frame, if small ‘mudsplashes’ (rectangular areas of visual ‘noise’ patterns) are superimposed on each image for 80 milliseconds following its onset [11]. Change blindness occurs for dynamic scenes, as well as for static pictures [12], with two-thirds of observers failing to report that the sole person in shot across a cut between two camera positions was played by two different, dissimilar actors [12]. Extending this to real world experiences, people who were stopped by an experimenter and asked for directions continued their explanation without hesitation after two confederates walked between them carrying a large door, despite the fact that while the experimenter was occluded he had been replaced by a confederate [13]. Clearly, a sound understanding of change blindness would be useful for interface designers, for it would allow them to predict which changes to a display would be noticeable, and so could serve as discrete alerts and pointers, and which would not be noticeable, and so could be used to update displays without distracting users from other, primary tasks.

A recent review of ‘high level scene perception’ [14] was restricted to the integration of views of a scene over very brief eye movements (saccades) and the effect of scene context upon object recognition. The startling series of change blindness experiments were interpreted solely in these terms. Levin & Simons [12], on the other hand, have concluded that “Our intuition that we richly represent the visual details of our environment is illusory”, and O’Regan et al [11] suggest that “We have the impression of simultaneously seeing everything, because any portion of the visual field that awakens our interest is immediately available for scrutiny through an unconscious flick of the eye or of attention”. According to this view, the visual scene acts as an ‘external memory’, and our internal representation is of a much less sensorial, object-based, and abstract nature. Such conclusions have implications far beyond the processing of visual images over saccades, and pose a challenge for a body of theory which is based upon the rapid extraction of as much detail from the visual image as is possible.

It may be more profitable to interpret change blindness effects as not being informative about low-level visual processes at all, but rather showing the dominance upon our perception of the high level cognitive processes. Our phenomenological sensation of a rich visual percept is not an illusion, but it is not the level of mental representation that we use to obtain information about object properties and relationships, nor about their meaning. In the rest of this paper I will describe a particular cognitive level description of the perception of dynamic scenes, which is based upon the idea that mental representations exist at varying levels of abstraction, each containing qualitatively different levels of information. Our mental tasks are driven by the properties of one or more of these levels of representation, depending upon the nature of the task, and the suitability of the information that is available for the performance of the task. It is not our purpose to deny that low level visual processes exist, or that they are important in perception: clearly, without such processes, the higher levels of mental representation would have no material to be abstracted from. Like Hochberg, we do regard it as limiting the application of perceptual psychology to focus upon the physical levels to the exclusion of cognitive levels of explanation.

Fragmentation of Knowledge and Unification of Theory

There are two real reasons why the practical knowledge of cinematographers has been difficult to integrate with computer graphics. First, the very pace of change mitigates against the systematic application of interdisciplinary knowledge. No-one can be expected to know about all of the research that has been conducted outside their domain of expertise, and those who do know about it cannot be expected to drop their own research to keep an eye on your field, on the off-chance that they will be able to help. Secondly, and more importantly, it is very hard to map knowledge or principles from one domain to another without some common theoretical framework. Film makers express their craft skills in terms of film-making situations that do not occur in computer graphics. Psychologists who are researching vision, or emotion, or spatial navigation, are doing so with their own theoretical concepts, and it is no easier for them to map these onto a computer graphics problem than it is for a computer graphics researcher to understand the psychologists. The theories within a domain are often too detailed and require too much specific input to be applicable to problems outside their native empirical paradigms. In fact, this is as much a problem within psychology as it is between psychology and other disciplines. Human behaviour has been partitioned into so many areas, at so many levels of analysis, that the mutual ignorance between researchers of vision, memory and emotion is astounding.

Fortunately there is an ongoing effort to develop integrative approaches within psychology that enable different aspects of behaviour to be linked at a less detailed level. Because they are not tied to any particular domain, such approaches also provide a way to communicate psychological research to non-psychologists, particularly those working in applied domains. One such technique that is becoming known within computer graphics is the Interacting Cognitive Subsystems model (ICS) that has been developed by Barnard and his colleagues [15]. The model is simple, in that it breaks cognition down into just nineteen ‘transformation processes’, between nine ‘subsystems’, each of which deals with a different level of mental representation, each with its own memory, or ‘image record’.

Although the model itself is simple, the mental behaviour it allows is complex, since some of these processes provide feedback, and there are sometimes two or more ‘routes’ between two subsystems.

One level of mental representation deals with ‘vision’ at a low level, where sensory attributes such as hue, brightness and motion are represented; another with the ‘objects’ that can be perceived within visually based scenes; another with ‘propositions’, semantic facts about objects and their relationships; and a fourth with ‘implications’, the real meanings that can be inferred from sets of propositions. Barnard’s approach is not limited to this linear or bottom-up process of recognition and comprehension, though. Implications feed back to influence the formation of propositions, and these feed back to influence the formation of object representations. The inclusion of internal feedback and top-down influences within the cognitive flow is the key to ICS. A detailed account is beyond the scope of this paper, and can be found in recent reviews in the human-computer interaction literature [16, 17]. The next few paragraphs attempt to give a brief introduction to the model, to show how it can help us understand the phenomena of film editing and change blindness.

Visual Perception and Comprehension in ICS

The overall architecture is illustrated in Figure 2, with the common generic structure of a subsystem shown in the inset. Visual perception is mediated by four different levels of representation, with the information flow shown by the arrows linking four of the subsystems. The visual subsystem uses its representations (richly sensorial but lacking any object based information or spatial relationships) to produce object representations (abstract, spatially structured objects) and implicational representations (qualitative, holistic interpretations of stimuli, such as their affective meaning). These are used by the respective subsystems to produce propositional representations (semantic, relational facts about the scene), which can in turn be used by the propositional subsystem to produce additional object and implicational representations. Note that a subsystem does not produce its own named representation, but receives them and turns them into other representations. At the same time as these transformations are being carried out, each subsystem copies the representations that it is receiving to its own ‘image record’, and it is this copying to memory that gives rise to the phenomenological sense of awareness of information. At each stage in the flow of information, the transformation processes within a subsystem may recruit this stored knowledge, derived from previous experiences, to enrich the incoming information.

The object level of representation can be thought of as the ‘mind’s eye’, where our awareness of a visual scene resides, but ICS includes a second route by which visual information can affect cognition. The visual level of representation is used to produce implicational meaning directly, in addition to the interpretative object and propositional route. A flashing red light, to take an extreme example, has an implicational meaning that is directly inferred from the sensory level. This meaning is, paradoxically, available to influence in a top-down manner the bottom-up structural interpretation. The same is true for sensory attributes that might not even be represented at the object level, such as aspects of facial expression, or of co-variation in movement of scene elements. The

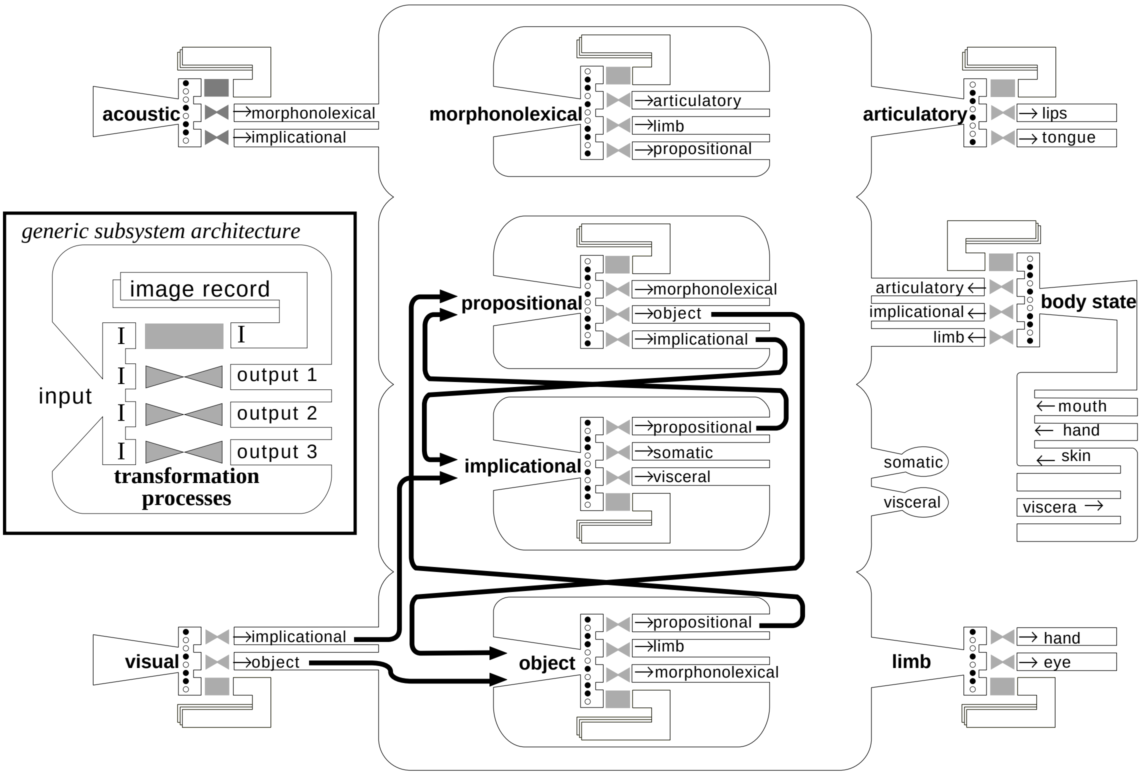

Figure 2: The overall ICS architecture, with the routes for visual perception and comprehension indicated by the arrows between subsystems, and the generic subsystem architecture (inset).

impact of such features of a display can occur despite our lack of awareness of their presence, and their consequent unreportability. As such, it is clearly dangerous to rely upon introspection or self-report assessments of display adequacy.

In this theory, visual representations correspond to the features produced by pre-attentive stages of processing. The object representations that the visual subsystem produces combine with the ‘top-down’ flow of representations being output by the propositional subsystem, to produce an integrated, object-based level of representation, which corresponds to the perceiver’s internal ‘mental image’. The overall phenomenological experience of a rich, meaningful visual percept is based upon information being copied to memory at several different levels of representation. The most sensorial visual level receives information only from the external world, via the eyes and pre-cognitive visual processes, and so has only a very limited temporal persistence and is limited in its fidelity by the scope of the sensory apparatus. The more abstract object level receives information both from the externally-derived visual level and the internal propositional level, and so can be maintained, or even constructed, in the absence of sensory input, or when sensory input is fragmentary, ambiguous, or flickering. The propositional and implicational levels contribute to semantic and holistic awareness of a scene.

At any particular moment, the quality of the phenomenological experience depends upon the locus of attention, which in Barnard's model, is governed by the competing demands of the various transformation processes. Any process can access stored knowledge within its subsystem, by revival of representations from an 'image record'. When the incoming stream is impoverished to the extent that the transformation cannot produce its output mapping continually, it can revive experiential records that have only just been copied to the image record, a mode of activity termed 'buffered processing'. This gives rise to the sensation of focal awareness, or attention. Because of the consequences that buffered processing has for the quality and rate of change of information in the overall flow of information through the whole cognitive system, only one process can be in this mode at a time, and so only one representation can be in focal awareness.

Within this conception of attention, the visual level of representation would be present and so providing a sense of diffuse awareness of the visual scene, even when the focus of cognitive processing might be upon the more abstract object or propositional levels. Change blindness could then occur because a changing object or feature might not be the focus of processing at the appropriate cognitive level, even though the information is available at another level, and the viewer is 'aware' of the object or of that area of the visual scene that contains it.

Applying cinematographic principles to interface design

May & Barnard [18] use ICS ideas to describe film editing techniques that are used to maintain the viewer's comprehension of a scene across cuts from one camera viewpoint to another. These cuts can be labelled as 'filmic' if they adhere to 'good practice', in which case they do not interrupt the viewer's comprehension (and are often not even noticed unless people are instructed to detect them), or 'unfilmic' if they do not follow filmmaker's craft experience and rules of thumb. Unfilmic cuts are perceived as 'jumps' and distract viewers by forcing them to attend to the structure of the visual scene rather than the narrative, in order to reorient themselves with relation to the objects and events that are being portrayed. By analysing the techniques film-makers use, we can reason about the psychological principles for their success, and use these principles to transfer their knowledge to other domains such as computer interface design.

One of the techniques identified is the common one of 'collocation' or a 'match cut', in which the object within the scene that the editor wants the viewer to attend to following a cut is placed in the same visual position as the object that they were likely to be watching immediately before the cut. Thus a gunman might raise and fire his gun, whereupon the scene cuts to the victim, who is shown in the same screen position as the gun. Collocation allows the cognitive processing that supports narrative comprehension to continue smoothly, while an unfilmic cut would require a different mode of processing to be briefly executed in order to scan the screen to find an object that makes narrative sense. Most cuts also represent a change in the point of view of the person who is watching the scene, and filmic cuts tend to be consistent with viewers' experience of the perceptual world, such as simply turning their head from side to side without physically changing position(although the consequences for the visual image are not optically identical [7]), or 'zooming in on' or 'pulling back from' a scene, which is analogous to changing their focus of attention.

ICS represents the 'narrative comprehension' mode of processing as one in which memory access and attention is located within the two 'central' cognitive subsystems. These process propositional and implicational representations of information, and the reciprocal processing between these two subsystems is the essential cognitive task in watching a film. While the propositional information about physical changes in the scene (derived from the object structure of the visual information) is consistent with the propositional information about the meaning of the scene (derived from the implicational interpretation of the scene), the propositional subsystem can blend and operate on its incoming representations without difficulty. Similarly, if any implications drawn directly from the visual and acoustic information are coherent with the implications drawn from the propositional representation, the implicational subsystem can operate on a coherent data stream. While both of these subsystems are operating on coherent data streams, memory access can be used to support the revival of records that elaborate and explain the actions and events that are portrayed, to anticipate the narrative, and to direct the viewer's attention around the scene. Changes in point of view that are consistent with the viewer's experience, or which present consistent changes in the relative positions of objects within the scene, can be tolerated at this level of processing and do not require cognitive work to reorient the viewer with relation to the scene.

Unfilmic cuts can disrupt this narrative comprehension mode by presenting structural information about the visual scene that does not cohere with the ongoing interpretation of the scene. Viewers will require memory access within the central subsystems to scan and evaluate the new scene, to locate an object that does fit with the interpretation, and even to assess the spatial relationships between objects within the new scene to determine how their own point of view has changed. Kraft [19, 20] has shown that directing the viewers' attention towards detection of cuts impedes their comprehension of a film's narrative, as would be expected if the focus of processing is shifting from the propositional and implicational levels of representation to the object and visual levels. Changes that are filmic pass undetected because they are not incongruent with processing at the object, propositional or implicational levels.

Although this explanation really requires a viewer to be engaged in active comprehension of the scene in order for unfilmic cuts to be distracting, a strong version of the argument would be that even a comparatively meaningless dynamic scene will show similar, if weaker, effects of unfilmic and filmic cutting. As long as some dynamic activity is occurring, the central subsystems will be engaged in following the events and attempting to predict future actions. Unfilmic cuts, such as those that present a change in the visual structure of the scene that is not consistent with the immediately prior visual structure, should be more distracting to this activity, and hence more 'noticeable', than filmic cuts, which might go unnoticed, or take longer to report. The increased detection within these experiments for jumps where the direction of motion altered, in addition to the position of the object, support this argument.

The same account can be applied to change blindness studies, whether using the flicker paradigm, mudsplatter patches, continuity errors in films, or real-world encounters. When the processing task is at a propositional or implicational level, such as in giving a stranger directions, gross changes to the object level of representation can pass by without impeding the flow of processing and drawing the individual's attention to the change. Even when the task is to look for changes, changes to the visual representation that are not part of the current topic of processing at the object level will be difficult to detect.

Conclusions and future directions

It is not necessary to conclude from change blindness research that the phenomenology of a rich visual percept is illusory, caused by the constant availability of the external world for re-inspection. The distinction within ICS between the visual, object and other levels of mental representation accords each equal reality, with the functional demands of an individual's momentary task determining which provides the quality of focal awareness, otherwise known as attention.

This approach to cognition allows a new set of perceptual principles to be added to the low-level constraints upon visibility of displays and the mysterious Gestalt Laws of perception. These new principles govern the requirement of congruency between the arrangement of scene elements and the viewer's expectations about the scene; about changes within the scene and thematic transitions within the viewer's comprehension of the 'narrative'; and about latent aspects of the interaction that can influence the viewer's interpretation of the scene. The traffic is not all one way, of course. By providing a framework for the modelling of cognition in complex tasks, ICS may enable psychologists to develop empirical paradigms that do not rely on highly reduced stimuli. The powerful graphics workstations in our laboratories that currently display red and green dots for hours on end may also be used to display photorealistic or rendered images, without the psychologists muttering about irrelevant complexity.

At the heart of ICS is its assertion that the meaning of an image can have as important a contribution to its perception as its physical structure. This is perhaps the holy grail of graphical rendering: to convey meaning as economically and accurately as is possible. Economy resides on both sides of the interaction: in terms of processing resources and hardware constraints on the system side, and in terms of attention, cognitive effort, and time on the user side. The solution will require an understanding of meaning, of the representation of meaning, and of the perception of meaning. The research path that is opening up requires the computer graphics community and cognitive psychologists to work together in a truly meaningful way.

References

1. Walker, J.H., L. Sproull, and R. Subramani. *Using a Human Face in an Interface*. in *CHI'94*. 1994: ACM.

2. Li-wei, H., M.F. Cohen, and D.H. Salesin, *The virtual cinematographer: a paradigm for automatic real-time camera control and directing*, in *SIGGRAPH'96: 23rd annual conference on Computer graphics*. 1996: New Orleans, LA USA. p. 217-224.

3. Münsterberg, H., *The Film: A Psychological Study (1916)*. 1970, New York.: Dover.

4. Carroll, J.M., *Toward a Structural Psychology of Cinema*. 1980, The Hague: Mouton.

5. Balázs, B., *Theory of Film (1945)*. 1970, New York: Dover.

6. Eisenstein, S.M., *Film Form.* 1949, New York: Harcourt Brace.

7. Hochberg, J., *Representation of motion and space in video and cinematic displays*, in *Handbook of Perception and Human Performance*, K.R. Boff, L. Kaufman, and J.P. Thomas, Editors. 1986, John Wiley and Sons: New York.

8. Gibson, J.J., *The ecological approach to visual perception*. 1979, Boston: Houghton Mifflin.

9. Johansson, G., *Visual perception of biological motion and a model for its analysis.* Perception and Psychophysics, 1973. **14**: p. 201-211.

10. Rensink, R.A., J.K. O'Regan, and J.J. Clark, *To see or not to see: The need for attention to perceive changes in scenes.* Psychological Science, 1997. **8**(5): p. 368-373.

11. O'Regan, J.K., R.A. Rensink, and J.J. Clark, *Change-blindness as a result of 'mudsplashes'.* Nature, 1999. **398**(6722): p. 34.

12. Levin, D.T. and D.J. Simons, *Failure to detect changes to attended objects in motion pictures.* Psychonomic Bulletin & Review, 1997. **4**(4): p. 501-506.

13. Simons, D.J. and D.T. Levin, *Failure to detect changes to people during a real-world interaction.* Psychonomic Bulletin & Review, 1998. **5**(4): p. 644-649.

14. Henderson, J.M. and A. Hollingworth, *High-level scene perception.* Annual Review of Psychology, 1999. **50**: p. 243-271.

15. Barnard, P.J. and J. May, *Interactions with Advanced Graphical Interfaces and the Deployment of Latent Human Knowledge*, in *Proceedings of the Eurographics Workshop on the Design, Specification and Verification of Interactive Systems*, F. Paterno', Editor. 1995, Springer Verlag: Berlin. p. 15-48.

16. Barnard, P.J., *Interacting Cognitive Subsystems: modelling working memory phenomena within a multiprocessor architecture*, in *Models of working memory: mechanisms of active maintenance and executive control*, A. Miyake and P. Shah, Editors. 1999, Cambridge University Press: Cambridge. p. 298-339.

17. Barnard, P.J. and J. May, *Representing cognitive activity in complex tasks.* Human-Computer Interaction, 1999. **14**(1&2): p. 93-158.

18. May, J. and P.J. Barnard, *Cinematography and interface design*, in *Human-Computer Interaction: Interact'95*, K. Nordby, *et al.*, Editors. 1995, Chapman and Hall. p. 26-31.

19. Kraft, R.N., *The role of cutting in the evaluation and retention of film.* Journal of Experimental Psychology: Learning Memory and Cognition, 1986. **12**(1): p. 155-163.

20. Kraft, R.N., *Rules and strategies of visual narratives.* Perceptual and Motor Skills, 1987. **64**: p. 3-14.