Journal of Consciousness Studies
Volume 9
Number 5/6, 2002
Articles
Temre N. Davies Donald D. Hoffman Antonio M. Rodriguez: Visual Worlds: Construction Or Reconstruction?

Temre N. Davies Donald D. Hoffman Antonio M. Rodriguez: *Visual Worlds: Construction Or Reconstruction?*

Page 72

Psychophysical studies of change blindness indicate that, at any instant, human observers are aware of detail in few parts of the visual field. Such results suggest, to some theorists, that human vision reconstructs only a few portions of the visual scene and that, to bridge the resulting representational gaps, it often lets physical objects serve as their own short-term memory. We propose that human vision reconstructs no portion of the visual scene, and that it never lets physical objects serve as their own short-term memory.

Introduction

Page 72

According to the standard account, vision is a process of reconstruction. From images at the eyes, human vision reconstructs those properties of the physical world that are useful to the viewer (Marr, 1982). The task of vision, on this account, is the inverse of the task of computer graphics.

Page 72

A graphics expert starts with a 3D specification of a scene, the positions and shapes of all its objects, the reflectance functions of all its surfaces, and the extent, position, and spectral composition of all its light sources. Then from any vantage point, and assuming any camera model, the expert can render an image of the scene using techniques such as ray tracing. Although rendering is computationally expensive, it enjoys the simplifying property that it is mathematically well-posed: a solution almost always exists, is unique, and varies continuously with changes in the scene or camera.

Page 72

The task of vision, on the standard account, is just the opposite. Vision starts with a rendered image, or a pair of rendered images, or even a dynamically changing pair of rendered images. The visual system must then reconstruct the physical scene, including the positions and shapes of all its objects, the reflectance functions of all its surfaces, and the extent, position, and spectral composition of all its light sources. As Yuille and Bülthoff (1996, p. 123) put it,

'We define vision as perceptual inference, the estimation of scene properties from an image or a sequence of images.' In particular, they are interested 'to model the individual visual cues for estimating the depth and material properties of objects...' (p. 124). So on the standard account, the scene consists of physical objects and their material properties, and the more accurately human vision can estimate depth and material properties from images, the better it can reconstruct the scene and its objects. The goal is reliable perception, i.e., to make the estimation as accurate as possible so that the objects and properties reconstructed by the visual system resemble as much as possible their physical counterparts in the scene.

To this end some vision researchers try to systematically measure the distribution of certain properties of the physical world. Once these distributions are accurately measured, they can be compared to the estimations computed by the visual system, to see how accurate the visual estimations are. Maloney and Wandell (1986), for instance, in a paper entitled 'Color constancy: a method for recovering surface spectral reflectance', propose a computational theory for color constancy. In their theory, color constancy is primarily the problem of estimating or recovering surface spectral reflectances. They justify their theory in part by appeals to objective physical measurements of natural terrain reflectances by Krinov (1947), and conclude that their theory can adequately reconstruct natural reflectances.

Page 73

Visual reconstruction is computationally expensive, and suffers the complicating property that it is mathematically ill-posed: solutions do not always exist, are almost never unique when they do exist, and need not vary continuously with changes in the images. This ill-posedness was well understood by Berkeley, who wrote

> It is, I think, agreed by all that distance, of itself and immediately, cannot be seen. For distance being a line

> directed endwise to the eye, it projects only one point in the fund of the eye, which point remains invariably the same, whether the distance be longer or shorter (Berkeley, 1709/1963, p. 19).

Berkeley's point is that for any given images at the eyes, there are countless distinct 3D worlds which could have projected to those images.

Page 73

What is true of depth is true of other properties such as shading, motion, reflectance, illumination, and even object identity: for any given images at the eyes there are countless distinct states of these properties in the world which could have generated those images. The task of vision is ill-posed everywhere you look.

Page 73

The standard account explains the success of vision, despite the ill-posedness of its task, by positing that the processes of visual reconstruction can be modeled as nondemonstrative inferences, typically unconscious, which exploit regularities of the physical world (Knill and Richards, 1996). This inferential explanation has a long and venerable history, dating back at least to the Islamic scholar Alhazen (965–1039 AD), who wrote:

> For the shape or size of a body, or the transparency of a transparent body, and such like properties of visible objects, are in most cases perceived extremely quickly, and

Page Break 74

> not immediately, since they are perceived by inference and discernment… (Translated by Sabra, 1978, p. 176).

An inferential account of vision was also elaborated by Hermann von Helmholtz (1821–1894) who wrote:

> The psychic activities that lead us to infer that there in front of us at a certain place there is a certain object of a certain character, are generally not conscious activities, but unconscious ones. In their result they are equivalent to a conclusion,…itmay be permissible to speak of the psychic acts of ordinary perception as unconscious conclusions, thereby making a distinction of some sort

> between them and the common so-called conscious conclusion (Helmholtz, 1910).

Page 74

Many regularities of the physical world have been studied, and shown to be in principle capable of leading to unique reconstructions of different aspects of the physical world. For instance, light sources tend to be overhead (Howard *et al*., 1990), and their spectra are usually linear combinations of just three basis vectors (Maloney, 1985). Surface reflectances tend to change abruptly (Land, 1977), and are often linear combinations of just a few basis vectors (Maloney and Wandell, 1986; Marimont and Wandell, 1992). Many objects move rigidly (Ullman, 1979; Bennett et al., 1989), quasi-rigidly (Ullman, 1984), or piecewise-rigidly (Hoffman and Flinchbaugh, 1982); and they intersect transversally to create parts of more complex objects (Hoffman and Richards, 1984).

Page 74

The standard account has proven powerful both in theory and in practice. Modeling vision as bayesian inference has led to the construction of numerous computer-vision systems with remarkable performance (Knill and Richards, 1996).

Change Blindness

Page 74

Although the standard account states that vision reconstructs the physical scene, it doesn't state how much, or which properties, of the scene are reconstructed at any one time. Marr himself was not explicit on this point, but a reasonable interpretation of his theory is that he intended the early stages of reconstruction, which he called the primal sketch and $2^{1}/_{2}$ D sketch, to simultaneously encompass the entire visual field. These stages reconstructed the edges, surfaces, reflectances, and viewer-centered depths of the visible world. The last stage, which Marr called the 3D model, reconstructed objects in an object-centered framework, and did not seem intended to encompass the entire visual field at once, but rather to proceed on a small number of objects at a time.

Page 74

Other theorists, however, have proposed that the reconstruction gives rise to stable and richly detailed representations of the entire visible world (Feldman, 1985; Trehub, 1991). Experimental studies of change blindness suggest that this is false (Rensink, 2000a). In a typical study using the 'flicker' paradigm, a subject is shown a picture of a scene for a few hundred milliseconds, followed by a blank screen for about one hundred milliseconds, followed by the original picture of the scene for a few hundred milliseconds, followed by a blank screen, and so on repeatedly until the subject responds or time runs out (Rensink, O'Regan and Clark, 1997, Rensink, 2000b; see also Phillips, 1974). The subject's task is to

decide if the two pictures of the scene are the same or different. The differences can include deleting objects, moving objects, or changing the colours of objects. Subjects typically find this task difficult, and can sometimes require several minutes to discover a major change, such as the deletion of a large object.

This result comes as a surprise to most subjects. It has also come as a surprise to many seasoned vision researchers. Several theories have been proposed to account for it (see Visual Cognition, 2000, volume 7, for several papers and a review). However the dominant theory is that human vision reconstructs certain primitive visual properties over the entire visual field, but that the reconstruction and storage of more advanced properties requires selective attention (Rensink, 2000b) and has a limited capacity of five or six items. Only those five or six items that have been reconstructed and stored in visual short term memory (vSTM) are available for change detection in the flicker paradigm, or for change detection across eye blinks or saccades. Normally the visual system relies on motion or brightness transients to draw its attention to image changes. However saccades, eye blinks, and the blank screens of the flicker paradigm all serve to interfere with the normal processing of transients, and force the visual system to rely only on the few items it has stored in vSTM to detect changes.

The recent literature on negative priming of visual objects indicates, however, that unattended objects can be recognized implicitly and that the visual system can detect changes to these unattended objects, as evidenced by implicit measures such as reaction times (Khurana et al., 2000). This suggests that current theories of change blindness might need to be modified somewhat to account for these negative priming results. In particular, attention is not required to hold object files in coherence over space and time, and unattended object files can persist for days or even weeks (DeShepper and Treisman, 1996). Again, these negative priming results are based on implicit measures of change detection; subjects cannot consciously report the changes that they implicitly report. The conscious reports appear to require attention.

The World As Short Term Memory

Page 75

If at most five or six items are stored in vSTM, and these are the only aspects of the visual scene that survive eye blinks and saccades, then it seems reasonable to suggest that the visual system relies on the physical world to serve as its own short term memory. Each time that information about the scene is needed, the visual system does not pull the details out of its own internal representation, instead it just looks to the right place in the scene and reconstructs whatever information is needed. The possibility that the world serves as its own short-term memory was suggested long ago by Stroud:

> In the case of vision for mammals, since our illumination is typically continuous sunlight and most of the scenery 'stays put,' the physical object can serve as its own 'short-term memory.' The way we 'remember things best' in the immediate visual present is to 'keep looking at them' (Stroud, 1955, p. 199).

It has also been suggested more recently by O'Regan:

Page Break 76

> seeing constitutes an active process of probing the external environment as though it were a continuously available external memory (O'Regan, 1992, p. 484).

For this approach to work, the visual system must reconstruct some stable properties of the entire visual scene as part of its early visual processing, so that this information can be used to direct attention to those parts of the scene which need to be reconstructed in greater detail. An architecture for doing this has been proposed by Rensink (2000a).

Page 76

These points are all easily accommodated by the standard theory. The key idea remains that vision reconstructs useful properties of the physical world. The only modification required is to note that vision is selective in what it reconstructs. Reconstruction proceeds primarily on an as-needed basis. And only five or six items are reconstructed in detail at any one time. We never feel bothered or limited by having details on such few items at a time, for the simple reason that we can quickly get details whenever and wherever we need them by simply redirecting our attention. As a result, as Noë (2000, p. 203) puts it, 'It seems to us as if all the detail is in the environment, which is where, in fact it is.'

Is It Reconstruction Or Just Construction?

One aspect of the standard account that deserves closer scrutiny is its claim that visual representations of a physical scene are, in whole or in part, reconstructions. It doesn't claim merely that they are constructions, but makes the stronger claim that they are reconstructions. A reconstruction, as most vision researchers use the term, means a construction with the further property of resemblance. To say that the visual system reconstructs the cats I see before me means that it constructs representations of cats that resemble, in relevant respects, the real physical cats. To say that the visual system reconstructs, or recovers, the 3D shape of the cats is to say that it constructs representations of the 3D shapes of the cats that correctly match, to within some useful tolerance, the real 3D shapes of the physical cats. To say that it recovers the colours of the cats is to say that it constructs representations of their colours that correctly match, to within some useful tolerance, the real colours.

Such claims of matching or resemblance are stronger than necessary, and stronger than is justified. First we will consider why they are stronger than necessary, and then why they are stronger than is justified.

Reconstruction Is Stronger Than Necessary

Consider the visual processes involved in watching a movie, such as The Score, on the big screen. The visual system of the viewer constructs a multitude of objects and people--the vault, the priceless sceptre, Marlon Brando, Robert Deniro, Ed Norton. But would it be correct to say that the visual system reconstructs these objects and people? It would seem not, since these objects and people are not literally in front of the viewer. Instead, the viewer looks at a white screen with changing coloured lights projected onto it.

Page Break 77
Page 77

Now one might maintain that, although the objects and people are not literally before the movie viewer, it still makes sense to say that the viewer reconstructs them. For at some point in the filming of the movie those objects and people were literally before the camera, and the camera served merely as a convenient surrogate for the viewer's own eye.

Page 77

Then consider a viewer watching a completely animated film, such as Shrek. In this case the objects and creatures that the viewer sees never were literally before the camera. The viewer constructs these objects and creatures, but could not reconstruct them because they don't literally exist.

Page 77

Take this one step further, and consider interactive virtual reality games, in which the viewer dons a helmet and body suit, and proceeds to fight virtual aliens, explore virtual worlds, or toss a virtual softball to the virtual image of another person who has also donned a helmet and body suit. Again all the objects and people that the viewer constructs are not reconstructions. There are no literal aliens that the viewer interacts with, and therefore no aliens to be reconstructed. What the viewer interacts with is a supercomputer and megabytes of software. The visual constructions of the viewer in no way resemble the diodes and resistors of the supercomputer that houses the software of the virtual-reality game. Nor do they resemble the lines of C++ and OpenGL code that constitute the software.

Page 77

And this is no problem at all for the viewer. To successfully fight virtual aliens, viewers don't need that their visual constructions are in fact reconstructions, all they need is for their visual constructions to be useful guides for their subsequent actions. One could in principle successfully fight those aliens by studying the C++ code of the virtual-reality game and then setting the correct values in the correct registers. But the visual constructions of the viewers allow them to bypass all the nasty code, even to be ignorant of the existence of code, and still successfully interact with that code in a way that lets them defeat the virtual aliens. Useful constructions are necessary for survival in virtual worlds or the real world; reconstructions are not.

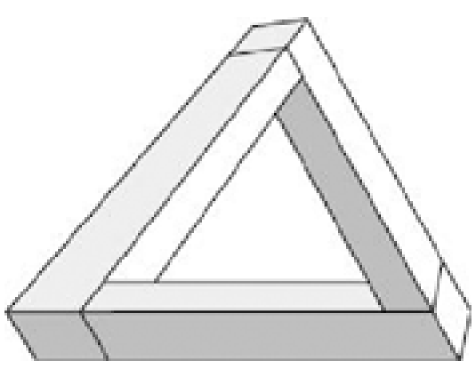
One might argue that the viewer's constructions in the virtual-reality game are reconstructions, not of physical objects directly before the viewer but of the objects that were in the mind of the game designer. And these in turn were the game designer's reconstructions of objects in the physical world. But suppose that the game maker did not directly code the objects in the game, but instead coded a genetic algorithm that probabilistically evolved the flora, fauna, and inanimate objects of the game. Then even the game maker would be surprised at the creatures that eventually evolved in the game, just as Richard Dawkins (1986) was surprised by the 'biomorph' shapes that evolved from his own genetic algorithm.

In this case of the blind game maker, the viewer is not reconstructing the objects in the mind of the game maker. Instead both the viewer and maker are constructing de novo the objects they experience when they don a helmet.

One might still argue that, although these novel visual objects are not reconstructions, nevertheless the parts of which they are made are not novel, and thus

these parts are reconstructions. The novel objects are merely novel combinations of familiar, and reconstructed, parts.

This argument is more difficult to counter. Examples of objects that cannot be reconstructions will not counter it. For instance, consider the 'devil's triangle' devised by Oscar Reutersvärd (1984) in 1934:



Page 78

It is fairly straightforward to build a 3D structure out of wood which, when photographed from exactly the right angle, gives rise to this image (Gregory, 1970, p. 56). However, the 3D 'object' that our visual systems construct when viewing this figure is physically impossible, i.e., it could not be built out of wood. Therefore it could not be a reconstruction. However, the individual parts of which it is made are perceived individually as 3D shapes that are physically possible. So one can argue that although the entire devil's triangle could not be a reconstruction, its parts could be.

Page 78

But our claim in this section is not that reconstruction is impossible, just that it is a stronger condition than is necessary: construction, without reconstruction, will suffice. For this purpose, the devil's triangle establishes that there are some perceptual objects, namely the 'impossible objects,' that cannot be reconstructions. They must simply be constructions. This demonstrates that at least some perceptual constructions are not reconstructions. Reconstruction is not a necessary property of visual constructions. Why require that object parts be reconstructions if the objects themselves are not reconstructions?

Page 78

Another case in which some constructions cannot be reconstructions occurs in colour perception. Some women are, genetically, tetrachromats rather than trichromats. They have four distinct colour pigments rather than the normal three. Careful psychophysical studies combined with genetic assays have found that these women perceive a richer world of colour than do the rest of us (Jameson, Highnote, and Wasserman, 2001). So if two observers, one a tetrachromat and one a trichromat, both see a peacock, the colours they construct are different. Therefore, assuming that the peacock has definite colours to be reconstructed, at least one of the two observers must not be reconstructing the colours of the peacock. And if at least one of them is merely constructing, not reconstructing, the colours of the peacock, it is surely in the cards that they both might be merely constructing, not reconstructing, these colours. Again, reconstruction is not necessary.

This same style of argument holds for photographic negation. If an image is shown in photographic negative, the patterns of shading that it displays are not, in general, depictions of physically possible illuminations of 3D objects. That is, there is no set of 3D objects and illuminations that could project to the given image. Therefore there are no physically possible worlds that could be reconstructions from that image. But observers nevertheless construct interpretations of these images as illuminated scenes of 3D objects (Subramaniam and Biederman, 1997), although negated faces may pose special problems (Liu, Collin, and Chaudhuri, 2000). Such constructions from negated images cannot be reconstructions. And once again this suggests that if some constructions are not reconstructions, then perhaps no constructions are attempts at reconstruction.

A further problem for reconstruction is synesthesia, an unusual mixing of the senses (Cytowic, 1993). Some synesthetes hear what they see, others see what they hear. One felt tastes with his hands. The taste of mint, for instance, felt to his hands as smooth, cool columns of glass. Every taste had its systematically associated feel, and he found this quite useful as an aid to creative cooking. However it would be quite a stretch to imagine that the feel of smooth columns of glass in response to eating mint is in any way a reconstruction. It is simply a construction that most of us don't make, but that happens to be quite useful to the one person who does make it.

Reconstruction Is Stronger Than Is Justified

It is one thing to argue that visual reconstruction is more than is necessary to account for our perceptions and for our survival, and that visual constructions will do just fine, but it is quite another to argue that in fact the constructions of vision are not justifiably called reconstructions.

We can use the powerful tools of the standard account itself to argue for this stronger claim. According to the standard account, vision is a process of inference in which the initial premises are images, I, and the conclusions are those properties of the visual scene, S, that the viewer constructs. The viewer determines the probabilities of various scene properties S given the images I, i.e., the viewer determines $P(S|I)$, and then selects those scene properties that satisfy some criterion such as maximizing probability or minimizing risk (Knill and Richards, 1996). The standard formulation of this inference uses Bayes rule:

$$P(S|I) = \frac{P(I|S)P(S)}{P(I)}$$

In this equation, the term $P(I|S)$ is a markovian kernel often called the 'likelihood function' by Bayesians and the 'rendering function' by vision theorists. It describes the probabilities that various possible images I would be rendered, given that the scene property is S. The computation of these probabilities just is the graphics rendering problem which, as we described earlier, is complex but well posed. The term $P(S)$ is a probability measure called the 'prior'. It describes the biases or assumptions that the viewer brings to the construction process, such as a bias toward rigid 3D objects or toward light sources that are overhead. The

Page Break 80
term $P(I)$ is simply a normalization factor which can be ignored unless it is zero, in which case continuous formulations of Bayes rule can be employed (Bennett et al., 1996). Finally the term $P(S|I)$ is a markovian kernel called the 'posterior distribution'. It describes, as mentioned above, the probabilities of various scene properties S given the images I.
Page 80
Now on the standard account the viewer's construction is some function f of the posterior distribution; that is, the construction is $f(P(S|I))$, where f is a function which optimizes some property such as risk. This account is intended to hold not only for vision, but for all perceptual modalities. In each modality the perceiver's constructions, and therefore the perceiver's perceptions, are a function of the relevant posterior distribution.
Page 80
The issue of reconstruction then becomes: Do the constructions $f(P(S|I))$ resemble or match the corresponding items in the physical world? To get empirical evidence to decide this question, we would need to compare the objective state of the items in the physical world against the

constructions $f(P(S|I))$. The problem is that the standard account allows us only one way to get information about the state of the physical world, namely via Bayes rule and $f(P(S|I))$. It does not allow non-inferential access to the objective state of the world. Every time we go to assess the state of the world, we are limited to seeing only what we construct (Bennett, Hoffman, and Prakash, 1989; Knill and Richards, 1996). This remains true even if we extend the range of our senses with various high-tech instruments. What we can perceive by means of those instruments, and of their readings, is limited to what our own senses can construct. It is true that these instruments can extend the range of our senses, e.g., from the visible electromagnetic spectrum to xrays and gamma rays. But they do not let us somehow bypass the inferential apparatus of the visual system and other perceptual systems, and indeed they often require, in addition, the more elaborate inferential apparatus of scientific theory building for their interpretation.

For this reason there is no way, on the standard account of perception, for the viewer to obtain the empirical evidence needed to justify the claim that perceptual constructions are in fact reconstructions.

One might ask how this can be so, since the standard account needs the formal model S before it can even get started. Where does S come from? And isn't S just the information required to determine if our perceptual constructions are in fact reconstructions? Unfortunately not. S does not represent the mind-independent external world. It represents the range of possible constructions available to the observer. Knowing S still leaves wide open the question of whether these constructions are reconstructions.

But doesn't natural selection guarantee that our perceptions are in fact reconstructions? Didn't those whose perceptions were more accurate reconstructions have a reproductive advantage over the rest, with the result, over aeons of evolution, that we are now a race that reconstructs the world quite accurately? Not at all. Natural selection promotes perceptions which guide useful behaviours. Roaches flee light, moths approach light. Neither species need accurately reconstruct the world in any sense; they just need perceptual constructions that

Page Break 81
usefully guide their behaviour. Arguments from natural selection do justify the claim that our perceptions are useful constructions; they don't justify the claim that they are reconstructions.

The Physical World Is Not A Short-Term Memory

Page 81
The idea that perceptual constructions are reconstructions, and the idea that the physical world can serve as its own short-term memory, are intimately linked. The reason the physical world could serve as a short-term memory is that the information that is effectively stored in the state of the physical world can be reconstructed as needed by the perceptual system of the observer. As long as the reconstruction can be triggered whenever needed, and as long as it proceeds quickly enough once it is triggered, then there is no reason to waste cortical resources to store what is already in the world. If, for instance, I am walking outside on a moonlit night, there is no need to store the moon in my short-term memory, since I can just look at the moon and quickly reconstruct it whenever I need.

Page 81
But we have just seen that perceptual constructions need not be reconstructions to be of use for survival, and that there are no empirical grounds to justify the claim that perceptual constructions are reconstructions. Thus there are no empirical grounds to justify the claim that the physical world serves as its own short-term memory. For, in order to justify this claim we would have to show that the constructions of the observer match, within allowed tolerances, the items in the physical world; without such a match, the world cannot serve as a high-fidelity memory. Now of course it is not true in general that the format of what is stored in a memory must match the format of what is ultimately retrieved from that memory. Good memories can use elaborate encoding schemes to improve efficiency in storage. But in the currently published accounts of the world as short-term memory, the assumption is made that vision recovers, i.e., faithfully reconstructs, the items that are in the world. And that assumption is not justified.

Page 81
Philosophical realists will of course object to this line of argument. A typical objection is that of Musgrave (1989):

> Indeed, some pretty mundane and well-entrenched results of science tell us that the moon (not some hyphenated moon, not even the Kantian moon-in-itself, just the moon) is objective and independent of us: it exists outside of our heads, it was not created by us, it existed before we did, and so forth (Musgrave, 1989, in Curd and Cover, 1998, p. 1221).

Whatever these results of science might be that are supposed to tell us this, they are surely not the results of physics, and especially quantum physics, which have told us instead that we should be extremely careful about our claims to knowledge of the world 'outside of our heads' (Albert, 1992; Barrett, 1999). Indeed the textbook interpretation of quantum theory, the so-called Copenhagen interpretation, maintains that observations of subatomic particles do not reconstruct dynamical physical properties of those particles because, between acts of observation, there are no values to reconstruct. The sciences that most directly bear on the issue of the relation between perception and the world 'outside' are the

cognitive science and neuroscience of perception. And pretty mundane and well-entrenched results of these sciences, as we have briefly discussed, tell us a quite different story about the moon (Hoffman, 1998).

The World As A Reliable Trigger

If we reject the concept of the physical world as a short-term memory, then how do we deal with the fact, revealed by change-blindness experiments, that visual short-term memory is limited to five or six items? If the memory is not in the head, and it's not in the physical world, where else can it be?

For an answer, we can consider again the example of the virtual-reality game. When the players have donned their helmets and body suits they find themselves immersed in some new visual world. Perhaps they are in a forest which is filled with various trees, rocks, sticks, leaves, and creatures. A player might look at a particular tree, then look away. If someone else asks what colour were its leaves or what branching structure characterized its limbs, then the player could look back at that same tree to obtain the answer. In this case the player is using the environment as a memory. But there is nothing in that environment that resembles the tree that the player observes. In this example, the environment is some supercomputer with many megabytes of software, but no trees. Yet this treeless environment effectively serves as a memory for the tree, because the player can act on that environment in such a way that the environment, in turn, triggers the player's visual system to construct the tree. The player acts on the environment by means of eye, head, and body movements, which are measured and transmitted to the supercomputer. The environment, in turn, triggers the player to construct the tree by having the supercomputer transmit a carefully crafted spray of photons to the player's helmet. The player is not reconstructing a tree that is in the environment; the player is instead constructing a tree in response to triggers from a treeless environment (Hoffman, 1998).

There is evidence as well from studies of eye movements that the environment triggers the visual system to construct, preattentively, a description of the 'gist' of an entire visual scene with each glance of the eye (De Graef, 1998). This gist describes the kind of visual scene it is, such as a garden scene or a gymnasium scene, and includes a parsing of the scene into objects, with a description of each object that is at least rich enough to determine if the object fits meaningfully into the scene, or if it is instead anomalous, such as a garden hose in a gymnasium. This preattentively constructed scene can then be used by the observer to guide eye movements and focus attention on specific objects in order to make more detailed constructions of these objects (Rensink, 2000b). Again these more detailed constructions are triggered by the interactions of the observer with the environment.

What is the nature of this environment? To investigate this question is a matter for scientific theory building, and beyond the scope of this paper. One theory is a form of naive realism which holds that, in many important respects, this environment is isomorphic to the constructions of observers, and so these constructions are in fact reconstructions of the environment. But as we have seen, this theory is at present not necessary and not justified.

Page 83

Nevertheless it is natural to ask: If perceptual constructions are just constructions   not reconstructions how can one account for the consistency of the visual world? I can inspect some portions of a table, turn away for a while, then come back and continue where I had left off. Doesn't this consistency of visual experience have everything to do with the environment serving as an external memory?

Page 83

Indeed it does, and this is no problem even if we reject reconstruction in favour of construction. One can still retain the notion of an external environment that is mind independent, in the sense that its existence does not depend on the mind or observations of any particular observer. In the virtual-reality example above, the supercomputer with its game software served in this role as the mind-independent environment. The key point is that this environment need not in any way resemble anything in our worlds of visual experience, just as the supercomputer and its software don't resemble the rocks, trees, and creatures of the virtual world that the helmet-laden observer experiences. Those worlds of experience can merely be a useful graphical user interface to the external environment, allowing us to interact effectively with that external environment.

Page 83

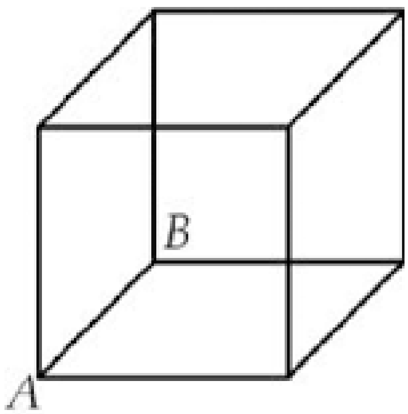
So the environment can serve as an external memory, without needing to resemble anything in our worlds of visual experience. As long as it is mind independent, and allows us to consistently interact with it in a manner that triggers us to create the same visual worlds, it can serve as a source of consistency for our visual worlds, and therefore as a reliable memory.

Page 83

Of course a mind-independent environment cannot be the only source for the consistency of our visual worlds. Another source is the rule-governed nature of our own constructive processes of perception. Systematic computational and psychophysical studies of human visual perception have uncovered dozens of interacting rules which guide the construction of our visual worlds. Some of these we mentioned earlier, such as trying to construct rigid objects, or modeling light sources and surface reflectances as linear combinations of a small number of basis functions. If the observer interacts with a consistent and mind-independent environment, and if in consequence of the interaction the observer is triggered to engage the same set of rules of construction, then the resulting worlds of visual experience will also be consistent. And all without need of any resemblance relation between the worlds of visual experience and the mind-independent environment.

Page 83

Even cases of inconsistency in visual experiences can be understood in this framework. Consider, for instance, the well-known Necker cube, published in 1832 by the Swiss naturalist Louis Albert Necker:



Page Break 84

Page 84

If you look at the figure, you will see a cube, perhaps with the corner labelled A in front. If you look away and then look back at the figure, you might again see a cube with corner A in front. That is an example of consistency of perception, and it can be explained as a result of your application of a consistent set of rules of construction: You see the same cube repeatedly because you engage the same rules repeatedly each time you look.

Page 84

However, sometimes when you look you might see a different cube, one with corner B in front. This demonstrates a failure of consistency in our worlds of visual experience. But it can also be explained in the construction framework. The rules of construction which create a cube are here being given a trigger which is a tad ambiguous. The rules result in two, rather than just one, 3D construction. Your visual system must pick one or the other construction, and sometimes it switches which one it picks. So the rules that help explain the consistency of visual experience also can account for its occasional multistability, viz., as a consequence of rules leading to multiple constructions. Notice again that the cubes being constructed here are not likely to be reconstructions, since it is highly improbable that there is a cube in a mind-independent physical world that is changing its shape each time your perception of the Necker cube reverses.

Page 84

After reading this section one might ask, 'What is the deep difference between an external stable trigger and an external memory?' Our answer is 'None'. We are not trying to distinguish between an external stable trigger and an external memory. We do distinguish between two theories of external memory. The first says that external memory is in fact a physical world whose contents resemble the contents of our perceptions. The second says that external memory need not in any way resemble our perceptions, any more than the software and hardware driving a virtual reality display resembles the perceptions of someone immersed in a virtual world. And we are endorsing the second theory.

So What?

Page 84

So what if perception is construction, and not reconstruction? This might be of interest to philosophers, but what difference does it make to practicing vision scientists? One difference is that the two theories make different empirical predictions. D'zmura, Colantoni, and Seyranian (2000) have created an immersive virtual world with four spatial dimensions and one temporal dimension. Users don a helmet and data glove, and set off exploring this 4D world, finding 4D objects, and chasing and shooting 4D aliens. The entertainment value of such a system is obvious. Once you've battled aliens in 4D, then 3D seems insipid by comparison. But an intellectual question was the primary force driving the creation of the 4D virtual world: Could human users learn to build visual maps of 4D worlds? It is too early to know the answer for sure, but initial results are encouraging.

Page 84

However, if perception is reconstruction, and if the physical world in fact has only three (uncurled) spatial dimensions, then there is no need to do perceptual research on whether human subjects can learn to perceive and visualize in 4D.

Page Break 85
Page 85

The answer must be that they cannot. Since they can only reconstruct, and since our best theories of physics tell us that there is no 4D world to reconstruct, they can never have 4D perceptions.

Page 85

But if perception is construction, and not restricted to being merely reconstruction, then it's an open possibility that some human observers might learn to perceive 4D worlds. This is a different empirical prediction than one obtains from the reconstruction theory. Thus, asking if perception is construction or reconstruction is not like asking how many angels can dance on the head of a pin. It makes another difference as well. Giving up the doctrine of reconstruction frees the theorist to consider a much wider range of possibilities for the relationships between perception and the world. One example comes from a practical problem faced by various intelligence agencies. They must daily comb through mountains of books, magazines, newspapers and other media looking for those rare tidbits of information that might prove critical to national security. At present this must ultimately be done by human readers, since only such readers can effectively do the job.

Page 85

But their job could be made much easier if an artificial intelligence program could first search all the articles for key words, and then organize them and present them graphically in such a way that the human readers would look first at the most interesting and most informative articles. So what is needed is a way to map abstract categories of information, say information about terrorists, bombs, weapons and oil, into a virtual visual world so that the human user can navigate through this virtual world and quickly find the important information. The virtual world might take the appearance of a 3D city with visual icons of oil cans and bearded hoodlums; or it might look like a forest populated with predatory animals and hostile plants; or it might assume whatever other appearance turns out to be an effective guide for the human users. Clearly this mapping need not, indeed in most cases cannot, be an isomorphism between the various important categories of abstract information and the virtual visual worlds that are used to display them; the abstract information might be twenty-dimensional, and the virtual visual world but three-dimensional, precluding an isomorphic map. That is no problem. The virtual worlds need not be reconstructions of the abstract worlds of information to be useful guides. On the contrary, the very usefulness of the virtual worlds derives from the fact that they are not reconstructions, but are instead well-chosen simplifications of some aspects, and exaggerations of others.

Conclusion

Page 85

The standard account of vision has the observer reconstructing those properties of the physical world that are useful. Change-blindness studies, and their discovery that vSTM is limited to five or six items, constrains the standard account by only allowing the observer to reconstruct five or six items at any one time. This limitation of the vSTM of the observer entails that the memory must be somewhere else, and the physical world seems to be the only candidate.

Page Break 86

Page 86

But the very notion of reconstruction on which this account relies is itself problematic, and careful examination suggests that it is not necessary and not justified. When we reject the notion of reconstruction, and replace it with the more conservative notion of construction, we give up nothing the observer needs for survival. We do give up the idea that the physical world serves as short-term memory which the observer can access when needed to reconstruct the desired items in that world. But we replace it with the more conservative idea that the observer can interact when needed with the environment in such a way that the environment in turn triggers the observer to construct the needed perceptual information.

Page 86

The move from reconstruction to construction is not a move to solipsism, but merely a move to more modest knowledge claims. To say that perception is reconstruction is to claim that the problem of the relationship between perception and the world is essentially solved, and that the relation is a particularly simple one: a rough isomorphism. To say, more modestly, that perception is construction is to recognize that the problem of the relationship between perception and the world is an open scientific problem with many possible solutions; isomorphism is just one solution, and perhaps not a likely one, given the variety of organisms and their perceptions.

Acknowledgements

Page 86

This work was supported in part by grant 009083 from the U.S. National Science Foundation to Don Hoffman. We thank Manish Singh and an anonymous reviewer for helpful comments.

References

Page 86
Albert, D.Z. (1992), *Quantum Mechanics and Experience* (Cambridge, MA: Harvard University Press).
Page 86
Barrett, J.A. (1999), *The Quantum Mechanics of Minds and Worlds* (New York: Oxford University Press).
Page 86
Bennett, B.M., Hoffman, D.D., Nicola, J.E. and Prakash, C. (1989), 'Structure from two orthographic views of rigid motion', *Journal of the Optical Society of America, A*, **6**, pp. 1052–69.
Page 86
Bennett, B.M., Hoffman, D.D. and Prakash, C. (1989), *Observer Mechanics: A formal theory of perception* (New York: Academic Press).
Page 86
Bennett, B.M., Hoffman, D.D., Prakash, C. and Richman, S. (1996), 'Observer theory, Bayes theory, and psychophysics', in Knill and Richards (1996).
Page 86
Berkeley, G. (1709/1963), *Works on Vision*, ed. C.M. Turbayne (New York: Bobbs-Merrill).
Page 86
Dawkins, R. (1986), *The Blind Watchmaker* (New York: Norton).
Page 86
DeShepper, B. and Treisman, A. (1996), 'Visual memory for novel shapes: Implicit coding without attention', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22**, pp. 27–47.
Page 86
Curd, M. and Cover, J.A. (ed. 1998), *Philosophy of Science: The central issues* (New York: Norton).
Page 86
Cytowic, R.E. (1993), *The Man Who Tasted Shapes* (New York: Putnam).
Page 86
De Graef, P. (1998), 'Prefixational object perception in scenes: Objects popping out of schemas', in *Eye Guidance In Reading and Scene Perception*, ed. G. Underwood (New York: Elsevier).
Page 86

D'Zmura, M., Colantoni, P. and Seyranian, G. (2000), 'Virtual environments with four or more spatial dimensions', *Presence: Teleoperators and Virtual Environments*, **9**, pp. 616–31.
Page 86
Feldman, J.A. (1985), 'Four frames suffice: a provisional model of vision and space', *Behavioral and Brain Sciences*, **8**, pp. 265–89.
Page 86
Gregory, R.L. (1970), *The Intelligent Eye* (New York: McGraw-Hill).

Page Break 87
Page 87
von Helmholtz, H.L.F. (1910), *Treatise On Physiological Optics*, trans. J. Southal, 1925, (New York: Dover).
Page 87
Hoffman, D.D. (1998), *Visual Intelligence: How we create what we see* (New York: Norton).
Page 87
Hoffman, D.D. and Flinchbaugh, B. (1982), 'The interpretation of biological motion', *Biological Cybernetics*, **42**, pp. 195–204.
Page 87
Hoffman, D.D. and Richards, W.A. (1984), 'Parts of recognition', *Cognition*, **18**, pp. 65–96.
Page 87
Howard, I., Bergstrom, S.S. and Ohmi, M. (1990), 'Shape from shading in different frames of reference', *Perception*, **19**, pp. 523–30.
Page 87
Jameson, K.A., Highnote, S.M. and Wasserman, L.M. (2001), 'Richer color experience in observers with multiple photopigment opsin genes', *Psychonomic Bulletin & Review*, **8**, pp. 244–61.
Page 87
Khurana, B., Smith, W.C. and Baker, M.T. (2000), 'Not to be and then to be: Visual representation of ignored unfamiliar faces', *Journal of Experimental Psychology: Human Perception & Performance*, **26**, pp. 246–63.
Page 87
Knill, D.C. and Richards, W.A. (ed. 1996), *Perception as Bayesian Inference* (Cambridge: Cambridge University Press).
Page 87
Krinov, E.L. (1947), 'Spectral reflectance properties of natural formations', *Technical Translation, National Research Council of Canada TT:439, 1947*.
Page 87

Land, E.H. (1977), 'The retinex theory of color vision', *Scientific American*, **237**, pp. 108–28.
Page 87
Liu, C.H., Collin, C.A., and Chaudhuri, A. (2000), 'Does face recognition rely on encoding of 3-D surface? Examining the role of shape-from-shading and shape-from-stereo', *Perception*, **29**, pp. 729–743.
Page 87
Maloney, L.T. (1985), 'Computational approaches to color constancy', Ph.D. Dissertation, Stanford University (Applied Psychological Laboratory Technical Report 1985-01).
Page 87
Maloney, L.T. and Wandell, B.A. (1986), 'Color constancy: a method for recovering surface spectral reflectance', *Journal of the Optical Society of America A*, **3**, pp. 29–33.
Page 87
Marimont, D. and Wandell, B.A. (1992), 'Linear models of surface and illuminant spectra' Journal of the Optical Society of America, A, 9, pp. 1905–1913.
Page 87
Marr, D. (1982), *Vision: A computational investigation into the human representation and processing of visual information* (San Francisco: Freeman).
Page 87
Musgrave, A. (1989), 'NOA's Ark   Fine for realism', *Philosophical Quarterly*, **39**, pp. 383–98.
Page 87
Noë, A. (2000), 'Beyond the grand illusion: What change blindness really teaches us about vision', *Visual Cognition*, **7**, pp. 93–106.
Page 87
O'Regan, J.K. (1992), 'Solving the “real” mysteries of visual perception: the world as an outside memory', *Canadian Journal of Psychology*, **46**, pp. 461–88.
Page 87
Phillips, W.A. (1974), 'On the distinction between sensory storage and short-term visual memory', *Perception & Psychophysics*, **16**, pp. 283–90.
Page 87
Rensink, R.A. (2000a), 'Visual search for change: A probe into the nature of attentional processing' *Visual Cognition*, **7**, pp. 345–76.
Page 87
Rensink, R.A. (2000b), 'Seeing, sensing, and scrutinizing', *Vision Research*, **40**, pp. 1469–87. Rensink, R.A., O'Regan, J.K. and Clark, J.J. (1997), 'To see or not to see: The need for attention to perceive changes in scenes', *Psychological*

Science, **8**, pp. 368–73.
Page 87
Reutersvärd, O. (1984), *Unmögliche Figuren: Vom Abenteuer Der Perspektiven* (Munich: Verlag Moos & Partner KG).
Page 87
Sabra, A.I. (1978), 'Sensation and inference in Alhazen's theory of visual perception', in *Studies in Perception: Interrelations in the history of philosophy and science* ed. P.K. Machamer and R.G. Turnbull (Columbus: Ohio State University Press).
Page 87
Stroud, J.M. (1955), 'The fine structure of psychological time', in *Information Theory in Psychology: Problems and methods*, ed. H. Quastler (Glencoe, IL: Free Press).
Page 87
Subramaniam, S. and Biederman, I. (1997), 'Does contrast reversal affect object identification?' *Investigative Ophthalmology & Visual Science*, **38**, pp. S998.
Page 87
Trehub, A. (1991), *The Cognitive Brain* (Cambridge, MA: MIT Press).
Page 87
Ullman, S. (1979), *The Interpretation of Visual Motion* (Cambridge, MA: MIT Press).
Page 87
Ullman, S. (1984), 'Maximizing rigidity: The incremental recovery of 3-D structure from rigid and rubbery motion', *Perception*, **13**, pp. 255–74.
Page 87
Yuille, A.L. and Bülthoff, H.H. (1996) 'Bayesian decision theory and psychophysics', in Knill and Richards (1996).

Page Break 88