Visual Intelligence; How We Create What We See by Donald D. Hoffman, W. W. Norton, New York, 2000

About the Book in General

Few days ago I was sitting on in the first compartment of a stationary train at a busy station. Looking through the windows I could see another stationary train on the adjacent platform. I could only see the train, nothing else, on that side. When I casually looked at the other train after few minutes, I felt very uneasy and almost dizzy. I was overwhelmed by the thought that my train was moving fast even though I couldn't feel it. My logical faculties were affirming me all along that it was the other train that was moving. But, it seemed like my body didn't want to believe it. My disbelief perplexed my thinking so much so that I didn't turn my head to the other side of the train I was sitting on. This is a very real experience about the dissociation between the vision and the signals from the rest of the body. It is also an example about reality challenged by 'virtual reality'. Something implicit in this example was a voice, however feeble it may be, against reductionism. It is the whole individual who would coax the brain for interpretation, not just the visual cues.

Even though Prof. Hoffman's book mainly focuses on the way we see the world, it is about the way we perceive the world in general. Prof. Hoffman fills the book with many examples from his wealth of experience, making it a tour de force in a moderately academic context. Thus, the book is not for the feeble minded who want to enjoy a quick and easy read. First six chapters of the book discuss the rules behind the way we see what we see and how we see the movement. With his experience in computing, Prof. Hoffman cannot be faulted for looking for rules to build algorithms usable in vision software. Given the sheer number of rules one wonders why the vision is so complex and overburdened by such a nexus of rules. The examples and rules remind one of the attempts of the proponents of Gestalt Theory to come up with laws to understand figure-ground issue. Here it is good to remind ourselves that figure-ground phenomenon is tied up with ambiguous figures.

The question I struggled with was the inefficiency that can be created by such a complex rule based system. One cannot stop wondering what sort of complex structure of reasoning our genes should construct to see what we see. On the other hand, if we look at the examples Prof. Hoffman gives in the book it is not difficult to see that almost all of the examples are about two dimensional projections of the three dimensional world. As he says on p.23 a two dimensional image showing depth "has countless interpretations in three dimensions." Should we, then, use such projections and build rules around why the eye interprets ambiguous two dimensional images the way it does? I believe this is more of a way to infuse algorithmic thinking into a process which is less complex in a more pragmatic sense.

As Prof. Hoffman reiterates, we live in a three dimensional world. Our eyes have evolved over long period of time to see the world in three dimensions. If we use the hackneyed argument from adaptationist viewpoint, any creature using vision to live on the earth should protect themselves in a three dimensional environment. Two dimensional images don't matter much as they occur in a three

dimensional background naturally as shadows and silhouettes. Thus, I believe many of the rules described in Prof. Hoffman's book can be merged to form a simpler structure for three dimensional vision unless we wish to develop computer algorithms.

1. Phenomenal world we live is three dimensional.

2. Irrespective of the two dimensional nature of retinal images, our visual systems have been shaped by Nature to live in a three dimensional phenomenal world.

3. Perception of hues, bundled here with brightness and saturation, and perspective is entrenched in such visual systems.

4. Over millions of years, eyes have been designed by Nature to look for three dimensional shapes and their defining features.

In my opinion, as an intruder into the realms trodden by experts and academics such as Prof. Hoffman, these are the basic rules which govern our vision. Human eye has not evolved to see the world in two dimensions. All the visual constructions Prof. Hoffman included in his book to show how we create what we see are two dimensional and hence, deceptive to the eye. Thus, I propose we need to be critical as to whether the arguments in the book about Visual Intelligence hold much water in the phenomenal world.

Let us have a look at Fig. 1 below showing white squares on white and black rectangular backgrounds. If you keep looking at them for a while you can see either a tunnel ending in a well-lit space or a flat-topped pyramid. Eye is struggling to create three dimensional visuals with ambiguous two dimensional pictures. Fig. 2 shows grey boxes of two different sizes on black background. Irrespective of what is obvious as figure and ground, both boxes can be seen either as a grey box or a space with two vertical walls and a floor. Why doesn't this happen with the "attached boxes" of Whitman Richards and Allan Jepson (p.30)? If we ignore the small box for a second we can see two walls attached to a ceiling instead of a floor. But this visual is not sustainable as the small box is not aligned to such a view. It cannot be seen as a similar space attached to a ceiling as the lines guiding the eye are not aligned for both boxes. One box is rectangular in shape while the other is more of a square shaped box.

All of this may tell us something intriguing about our vision. Given Prof. Chomsky's views about an innate grammar we are all born with, it is not hard to imagine the existence of a visual vocabulary. We all remember that Ancient Egyptians used a language based on pictograms. Modern Chinese still uses logograms representing visual cues. In contrast to these flat-world languages, we may have a built-in visual vocabulary in 3-D which will be called in whenever we see something. As the vocabulary is in three dimensions, finding a meaning for a two dimensional image with ambiguous three dimensional undertones is always difficult. Eye may stretch itself to find meaning in the image. It may change the hue, perspective or movement to try different interpretations.
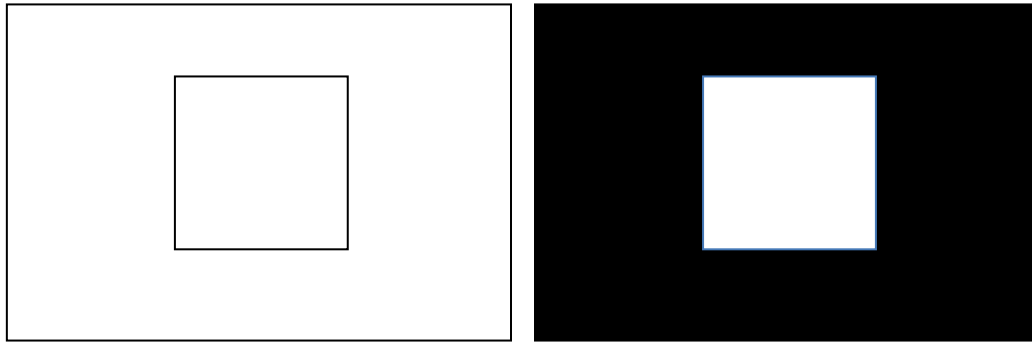
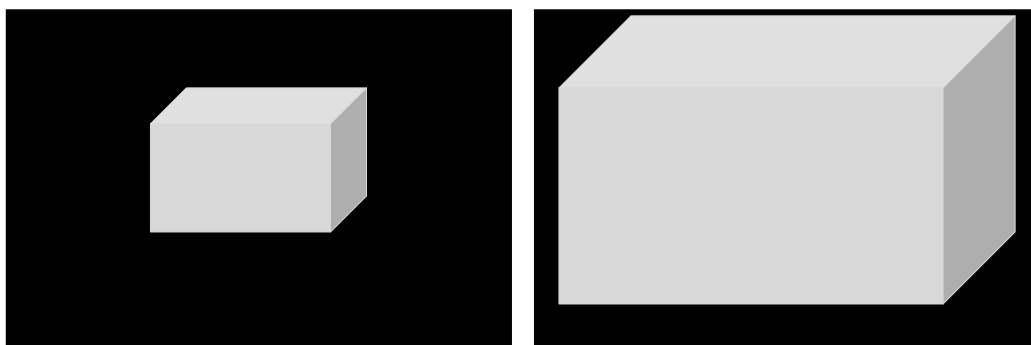Fig 1. White square within white and black rectangular backgrounds

Fig 2. Grey 3-D boxes in black background

Some of the ideas in the book are important in the selection of meaning of such ambiguous images. As Prof. Hoffman says in p.121, our 'visual intelligence tries to find the lowest cost solutions'. If it can be further interpreted this may mean the most energy efficient and effective solution to a visual problem. Unfortunately, some tricky images with no real existence presented to the eye can be costly and inefficient as the eye had evolved to work with our three dimensional world.

Philosophical Implications of Visual Intelligence and Frames of Reference

Now please forgive me for encroaching the philosophers' territory. To look at the next chapters of the book, it is necessary to invoke some philosophical musings. Some of the views expressed here are not in agreement with age-old philosophical traditions and thus, are invariably arguable. Unfortunately, these heretical views are required for the following discussion. Prof Hoffman discusses the virtual reality and brain research relevant to our perceptions such as synaesthesia and phantom pain. He also looks at the phenomenal brain and relational brain. Phenomenal brain constructs what we see. But it is present only when we perceive something. On the other hand, relational brain is the one which sustains the object when we are not aware of it. Berkeley attributed this to God who constantly perceives the material world. Prof. Hoffman says he carefully chose the word "construct" to describe the visual process (p.196-7) to avoid mixing the phenomenal and relational aspects. He believes if he uses the word "recover" or "reconstruct" it can mean recovering or reconstructing the forms of objects existing externally through our vision. In my opinion, this reservation arises due to lack of reference to the representational nature of our sensory inputs.

It is more fashionable in current times to explain our mind and our reality in terms of artificial intelligence. In a physicalist world everything is mechanistic. Our brain works like a computer consisting of myriad of binary circuits. Mind is simply physical processes arising from the central nervous system. This may not be far-fetched. But it can only be a science fiction until the immense gap between a modern super computer and the brain becomes more imaginable. No artificial intelligence system has so far passed the long form of imitation game. Even if such a system will pass the test one day, it might be doing it like a person in Searle's Chinese room.   What about virtual reality? Until we can call a robot a human or at least an early hominid, it may be far-fetched to imagine our reality in terms of virtual reality. Prof. Hoffman doesn't want to be a part of these arguments and he does this by avoiding the relational aspects of our sensory inputs.

Should or shouldn't we consider the relational aspect of our perceptions to describe the reality? I believe it is not quite right to say that we construct what we perceive. What we construct is only a 'representation'. That representation is a result of our senses and our interpretation of the sensory input. If the existential frame of reference tied to the sensory input is compromised by any physical defect, the representation can become 'particular' for the frame of reference rather than becoming 'universal'. Thus, the world and 'reality' are about sets of frames of reference. In a relativistic sense, a stationary observer will see something different to a moving observer. But the object that both see, is on its own frame of reference. Thus, all phenomenal and relational existences are relative to specific frames of reference. Reality exists in a relative sense. When I see a tree the representation of the tree I see is subject to a set of frames of reference comprising of two key types. We can use these two key types of frames, namely, *existential frame of reference* and *relational frame of reference* in space-time continuum to build a set of frames of reference. I think and thus, I am. This Cartesian view is defined within the co-ordinate system of my existential frame of reference. If I am colour blind, my interpretation of the image of the tree is of a different hue and will become a relational frame of reference that links me with the tree. If someone with normal vision sees the same tree at the same moment, that person will see the colour of tree which is 'universally' accepted and will become a relational frame of reference linking that observer with the tree. When none of us are watching, the tree still exists attached to the co-ordinate system determining its existential frame of reference. If the tree 'can sense' both of us, it may have two relational frames of reference about being observed. Thus, I believe we don't create what we see. What we do is constructing a representation of what we see with respect to a set of frames of reference. The objects giving rise to the representations exist independent of relational frames of reference.

Existential Frames of Reference and Berkley's God

 As the above discussion spurs us to think, the world we live in is full of frames of reference. A reality of any animate or inanimate agent is the existential frame of reference attached to its system of co-ordinates. No one can deny its reality as it is independent of any relational frame of reference. Virtual reality created at the Virtual Reality Exhibit described on p.185 has no existential reality independent of the existential and relational frames of reference to the super computer and software. It is like the experience I had on the train. As soon as the computer is turned off or the software running to create the virtual world is closed, the virtual reality ceases to exist. But if the above mentioned tree is moved to a different location, it still exists and the movement traces a series of existential frames of reference in a space-time continuum. Thus, our waking state or 'life' in general is not like a dream as Mahayana School of Buddhists would like us to believe in their concept

of *Vijnapthi Matrata*. A dream is more like a virtual reality that we earlier touched upon as the dream depends on the existential frame of reference of the dreamer. Thus, in my view, the reality or the objective world exists if we rein in our unconstrained philosophising. This helps us realise that we do not create what we see. We can replace Berkley's God with existential frames of reference and experience the existence independent of our sensory inputs.

Prof. Hoffman's book that I read with great enthusiasm was an interesting journey about ourselves. But it may not be for the travel-weary reader.

Darshi Arachige 5th Sept. 2017