

“With Reference to Material Presented in the Semester, Speculate Whether Vision is Best Understood as a Top-Down or Bottom-Up Process.”

Vision has historically been attested to either be exclusively a bottom-up or top-down process (e.g. **Gregory, 1997; Gibson, 1986**). However, modern accounts are more appreciative of both processes, despite the overall preference for top-down explanations (**Ganis & Kosslyn, 2007**). This essay argues that, currently, vision is best understood in the context that visual information analysis is *not* enough to understand our environment; top-down mediation in the form of interpretation and prediction is needed to optimise human vision. That is, we can learn *more* by considering vision as an active, as opposed to passive, process. For example, cognitive neuroscientific research has recently discovered a neural-cognitive mechanism capable of fulfilling an efficient, predictive role in object recognition (**Kverga, Ghuman & Bar, 2007**). Change blindness research has also suggested the existence of internal visual representations (i.e. previously stored knowledge of the environment) to assist in change detection of real-world scenes (**Lamme, 2003; Simons & Rensink, 2005**). Top-down evidence also comes from psychophysical and neuropsychological domains (**Hill & Johnston, 2007; Schenk & McIntosh, 2010**). Recent advances in cognitive neuroscience have also reenergised the investigation of consciousness in vision (**Rees & Seth, 2010**). As a result, consciousness and awareness have also been mixed into the top-down/bottom-up debate (**Lamme, 2003**), however this may prove to become a fruitful endeavour. To elaborate: current top-down explanations, despite their popularity, still lack important components of a truly explanatory perspective of how knowledge facilitates vision. Prediction and representation views both posit the existence of environmentally derived ‘knowledge’ to pre-empt identity and change in the environment (**Kverga, Ghuman & Bar, 2007; Lamme, 2003**). However, they have yet to justify *how* such concepts are formed, what qualities they possess, and how they augment the qualitative nature of experiential vision. Cognitive neuroscientific research deciphering the nature of neural-cognitive consciousness will surely help further our understanding in how active cognitive processes assist in constructing our sense of reality, or ‘qualia’ (**Gregory, 1997**). Further, these insights will assist in reinforcing the currently untenable requirements of scope by **Simons & Rensink (2005)** to challenge purely bottom-up accounts.

Cognitive neuroscientific methods (particularly functional magnetic-resonance-imaging; fMRI), have been increasingly used to investigate the top-down cognitive-neural mechanisms involved in vision. A notable research domain is object recognition, which is seen as a means of facilitating efficient and predictive vision. Orbitofrontal cortex (OFC) is thought to be the executive of this process (**Bar, 2003; Kverga, Ghuman & Bar, 2007**). OFC theory suggests that there are two processing pathways in vision, one coarse and one refined (see **Figure: 1, next page**). The refined pathway refers to the selective and detailed (bottom-up) processing of high spatial-frequency (HSF) information (an object’s specific characteristics) that occurs from visual cortex onwards. The coarse pathway provides top-down support; low spatial-frequency (LSF)

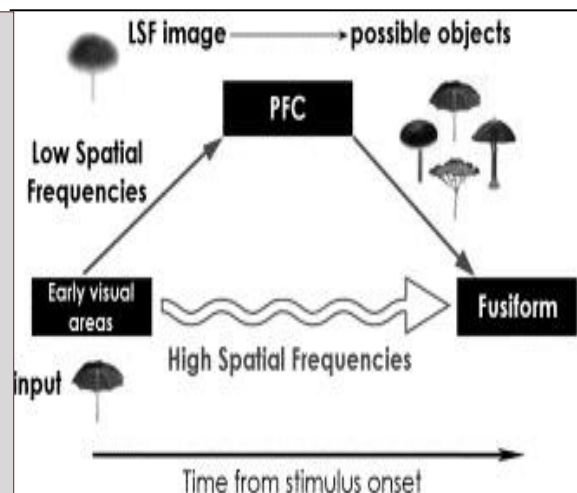
information (an object's global characteristics) is rapidly projected via the magnocellular pathway to the OFC region in the frontal lobe. This general information is sent to object recognition regions in inferior temporal lobe (IT) as a prediction of the objects' identity. Insights are gained by studying the coarse pathway, as it acknowledges the process of ascertaining the identity of familiar objects, especially those share global characteristics with others (e.g. a football and a balloon). However, an important component currently missing in this theory is the issue of how knowledge *becomes* familiar in the first place, and whether object knowledge is the only prediction that occurs – for example, is it the same for face perception? The issue of knowledge development is further discussed later.

Figure: 1

A Theory of (Familiar) Object Recognition

Some researchers assert that when a familiar object is seen, LSF information is sent to OFC via the *magnocellular pathway*, which is then evaluated and sent as a prediction to IT. It can be understood as a short-cut way of assessing familiar stimuli. Meanwhile, HSF information is sent to *visual cortex* for detailed, neurally selective analysis.

(Image from **Kverga, Ghuman & Bar, 2007; p.152**)



As a theory of object prediction, OFC theory has received noteworthy empirical support. **Kverga, Boshyman & Bar (2007)** using fMRI found significant activations in OFC, and differential activations between the magnocellular and parvocellular pathways as a result of visual stimuli manipulated by spatial frequency content. The magnocellular pathway reacted primarily to stimuli with LSF (global) information, and as predicted by the theory, increased activation within OFC for this information compared to HSF information. OFC theory also advances our understanding of top-down vision, as it provides an account of how we are able to perceive and identify our environment with such ease (indeed, without much subjective awareness of the effort either). Anecdotally, in the **Kverga Boshyman & Bar (2007)** study above, LSF information was perceived by participants to be easier to see, despite the higher saliency of HSF objects due to their characteristics. This suggests a neural preference in favour of the 'big picture' as opposed to specific details, which coincides with behavioural studies supporting a cognitive bias for holistic processing (**Navon, 1977**). Other research has found early activations of OFC in response to emotional stimuli, even before awareness occurs, suggesting an evolutionary role in quickly recognising and predicting dangerous events (**Carretie, Hinojosa, Mercado, & Tapia, 2006**). This also suggests that OFC activation may *precede* consciousness, insofar that it may act as an unconscious store of conceptual knowledge for object identity. Further research should consider OFC's role in consciousness, as it may be part of the process leading ultimately to our subjective awareness. Neuroimaging must be treated with caution, as a theoretical assumption of fMRI is that hemodynamic activity within these brain regions are directly involved in the cognitive task carried out. Further, although averaging methods are

generally accepted to reduce extraneous noise, there are statistical debates for analysing fMRI data (**Lindquist, 2008**). Despite this, neuroimaging provides us with a schematic, albeit arbitrary, capability of showing how top-down facilitation in vision may work on a neural-cognitive level.

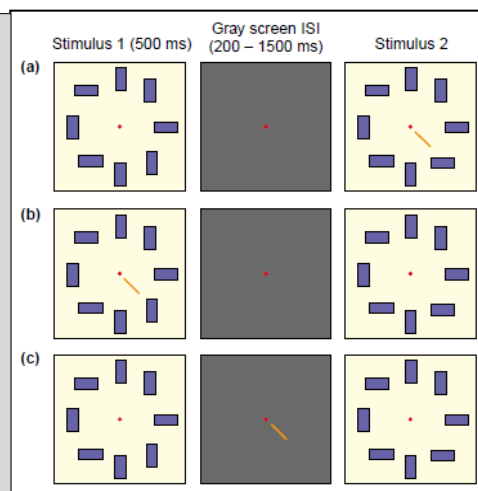
Change blindness is a behavioural paradigm showing how vision requires (top-down) awareness in order to perceive real world changes in the environment (**Lamme, 2003**). It is also a strong demonstration of the intermingled nature between vision, perception, attention and memory (**Lamme, 2003**). Change blindness typically incorporates a flicker paradigm (see **Figure: 2, below**). It occurs when a viewer fails to (efficiently) detect changes in a visual scene from one view to the next, when it is separated by a brief disruption, such as a blank interval (**Rensink, O'Regan & Clark, 1997**). It is thought that the blank interval negates what would be movement in the real world, and the grabbing of attention to the moving object. Eye movement research shows we are quite blind of high resolution details when using peripheral vision; we only 'see' with clarity and accuracy when we focus on objects (**Henderson, 2007**), so the existence of change blindness is unsurprising in this regard.

Figure: 2

The Flicker Paradigm

A visual array/scene is presented, in the case opposite for 500ms. A blank interval follows this, with variant duration. A modified version of the original array/scene then reappears. The task requires fast detection of the changes. When cued correctly, change detection can be improved; this can be viewed as evidence of a top-down mechanism internally representing the previous scene.

(Image from **Lamme, 2003, p.13**)



Top-down interpretations of this phenomenon posit the existence of an internal, short-lived cognitive representation of the original scene to support change detection. This has been substantiated in experimental formats, where change blindness susceptibility has been controlled by providing cues at different points in the flicker paradigm – cueing the changeable target before or during the interval both facilitate change detection, whereas cueing after the interval does not (**Lamme, 2003**). This cueing effect suggests that an attentional gate may provide the means to access this representation in order to aid change detection (**Simons & Rensink, 2005**). A typical Gibsonian, bottom-up account would struggle to account for change blindness – the blank interval removes the ecological advantage that movement would confer, and yet, reaction times dramatically improve when cued – this is because we seek knowledge of the previous array to inform our decision for the current array (**Lamme, 2003**). In addition, that viewers are able to *efficiently* seek out changes in scenes they have already seen (i.e. you don't miss the gorilla twice – **Simons & Levin, 1998**) suggests that our visual memory facilitates our capacity to spot changes in familiar scenes. Further, this would help explain why we are more efficient spotting changes within familiar, as opposed to novel, scenes. Representation views (e.g. **Lamme, 2003**)

convey a message similar to Working Memory accounts (e.g. **Baddeley, 2003**), that we are equipped with an innate cognitive capability to access and manipulate a limited store of fast-decaying information to understand and predict the ever-changing environment. Vision, like many cognitive abilities, is active; we do not passively process information (**Gregory, 1997**). Of course, the extent we are conscious of these processes is debateable (**Rees & Seth, 2010**).

Vision exists because it improves our actions (**Schenk & McIntosh, 2010**). However, analytic vision *without* perception would rely on always seeing clear-cut stimuli. We need perception to facilitate vision, because we often see unclear, occluded and ambiguous stimuli – we must make a decision on what we are viewing. Visual illusions are typically used to investigate this process; we can infer perceptual processes involved in understanding the world when we ‘see’ unclear or ambiguous images (**Gregory, 1997**). Consider the the Hollow Face: it shows how knowledge is used to interpret incoming visual information and ultimately construct our reality (**Gregory, 1997**). The illusion occurs when a hollow mask is rotated (see **Figure: 3, below**). The top images show its convex side, whilst the bottom images show its concave side. What we typically see though is a convex (i.e. typical) face *whichever* side we look upon. This occurs *despite* depth cues remaining constant, emphasising the overriding of knowledge (i.e. perception) on deficient visual information.

Figure: 3

The Hollow Face Illusion

This face has properties to appear convex on the front and concave on the back. However, what we tend to continuously see is a concave mask. This is typically interpreted as evidence of top-down processes biasing our vision in a situation where it is actually showing us the ‘wrong’ thing.

(Image from **Gregory, 1997, p.2**)



Further evidence of top-down effects contributing to this illusions' occurrence derive support from behavioural and developmental domains. Behavioural research shows that the occurrence and intensity of the illusion is variable. This research, by **Hill & Johnston (2007)**, found a stronger effect (measured in distance: larger distance = stronger illusion) for faces than for non-orientational objects (such as a circular jelly mould), suggesting we use typical characteristics of objects, which are internally stored as knowledge, to *predict* the likely objects' identity, which then interfere with the actual visual information. Therefore, we see the face as convex, because previous knowledge posits a face is not typically seen as concave. Secondly, developmental research shows differential effects of illusion occurrence when comparing adults (*circa.* 30-years-old) to infants (<1-year-old). This research, by **Tsuruhara and colleagues (2011)**, found that, using preferential looking and habituation paradigms, adults were more likely to look at the concave-illusory convex-face, whereas the older infants (7-8 months) preferred the convex-actually convex-face, while younger infants (5-6 months) had no preference. This

difference in preference may reflect the role learning in top-down vision. It is known that newborns are greatly susceptible to saliency (**Stechler & Latz, 1966**), so it may be that top-down vision becomes more prominent increasingly through life, as past knowledge, through conceptual and mnemonic acquisition, grow in size and potency. Further, it may be that saliency is a cognitive mechanism used to assist in the development of our top-down capability.

Motor research has also demonstrated distinctions between vision-for-perception and vision-for-action (**Schenk & McIntosh, 2010**). This may reflect the interaction of top-down (perception) and bottom-up (action) processes. Participants viewing movement targets within the Hollow Face were shown to *correctly* move their hands to the targets, despite still being under the illusory effect (**Króliczak and colleagues, 2006**). This suggests that our theoretical knowledge through *understanding* the stimulus, and applied knowledge through *acting upon* the stimulus, are both functioning independently. This emphasises the need to consider vision in the context of action; vision does *not* occur in a vacuum, it is present to aid our behaviour (**Beldaulf & Deubel, 2008**). It may also mean that vision and perception are functioning in a parallel and independent fashion – we are seeing (and correctly respond to) the correct image, however we are perceiving to see the wrong image. Experientially, participants ‘saw’ the illusion despite the correct movement – this demonstrates that knowledge is dominating our vision – which may be a neural trace of consciousness (**Rees & Seth, 2010**).

Although illusions can be used to expose the characteristics of visual processes, it cannot suggest anything towards *how* the process of perceiving the stimulus occurs. Other methodologies are needed: **Schmid & Bar (in prep – see Kverga Ghuman & Bar, 2007)** found greater OFC activation (using fMRI) when processing ‘ambiguous’ (as rated by previous participants) stimuli than ‘obvious’ stimuli. This may be the area responsible for illusory perception - future research could investigate how actual illusions are perceived by individuals with OFC damage, which may provide an insight into how neural activity contributes to the experience of ambiguous stimuli.

Neuropsychological evidence also shows how top-down processes seem to be required for efficient human behaviour. **Schenk & McIntosh (2010; p.53–55)** cite several studies investigating an individual, initialled DF, who had suffered from visual form agnosia as a result of an anoxic lesion, which destroyed the lateral occipital area bilaterally (specifically the ventral stream). While able to perform low-level gross and fine motor tasks, DF struggles with movements requiring high-level (top-down) input. For example, DF does not use visual cues to program her finger tip forces; she also makes semantic errors with objects, grabbing them in non-functional ways; she also is unable to place a plaque through a T-shaped target, despite being capable of placing it through a non-letter slot (**McIntosh, 2010; Carey and colleagues, 1996; Goodale and colleagues, 1994 – see Schenk & McIntosh, 2010**). These findings show us several things – firstly, it again emphasises the need to consider vision in its functional, action-based context. Secondly, it shows how top-down mediation is needed to access conceptual and

functional knowledge to guide our movements. Finally, it shows how a largely bottom-up explanation would struggle in explaining and potentially helping an individual with a brain lesion similar to DF – why can she place an object through a gap, but struggle when linguistic knowledge is needed to guide the movement? Also, given that her perceptual capabilities are intact, it seems that it is the integration of knowledge, vision and subsequent movement that is deficient (**Schenk & McIntosh, 2010**). Despite this evidence, we must be cautious when asserting brain function from case study data, as we cannot accurately ascertain the size and extent of damage, nor can we be sure if it is localised to one region. However, one can conclude with relative confidence that DF's motor deficits are related to visual brain areas, which emphasises the importance of multi-modal considerations in cognitive research (**Ganis & Kosslyn, 2007**).

It seems reasonable that object predictions and scene representations have both evolved as a means of 1) enhancing our understanding of the environment 2) actively pre-empting, as opposed to reacting, to changes in the environment 3) reducing the vulnerability that is evident in-between eye movements. Cognitive neuroscience has thus far identified potential brain areas for visual knowledge, such as OFC, and even predict when such activity occurs after the stimulus appears (**Kverga, Ghuman & Bar, 2007**). However, such explanations only go so far in aiding our understanding of vision – how does such knowledge form in the first place? What characteristics does a construct such as an internal representation possess? (In this case, do we have one master representation, or do we have many snapshot ‘pictures’ of the scene?) How do we maintain these representations? How does such a construct assist in building our subjective sense of vision? It may well be that such a mechanism may feed our confidence that we do in fact ‘see’ *everything* around us in complete clarity and form, despite the reality suggesting otherwise (**Rensink, O'Regan & Clark, 1997; Henderson, 2007**). In the case of a bottom-up theory, explanations to argue and predict the *non-existence* of such knowledge must also be included.

Top-down explanations that use perceptual knowledge to aid vision thus far only informs our understanding to the extent that we know how the brain acts as an information processor i.e. how specific inputs produce differential outputs. Outstanding questions include know how such capability develops, how it is maintained, and how it seems to construct our ‘objective’ visual reality. Therefore, it seems reasonable that *requirements of scope* are needed for *both* top-down and bottom-up accounts. **Simons & Rensink (2005)** outlined several unreasonable empirical requirements required to disprove a representational account of change blindness. The questions raised above are applicable to *both* top-down and bottom-up accounts of vision. That is not to say that both prediction and representation views do not account for the data they provide – they do – to an extent. The baby must not be thrown out of the bathwater (**Simons & Rensink, 2005**). These theories are sound foundations in the understanding of how vision *requires* perception, memory and attention to assist us in both adapting to and changing the environment (**Ganis &**

Kosslyn, 2007). However they must go further if they are to account for the the formation and maintenance of top-down knowledge, and its role in constructing our sense of ‘complete’ reality.

This essay has shown with different evidences that vision is a process best understood as top-down. To conceptualise vision as a predominantly bottom-up process *cannot* explain how we can identify objects in the visual scene solely based on global characteristics, how we can detect change *without* movement, and how we can see a typical face in an atypical stimulus. We rely more on knowledge-based neural-cognitive mechanisms to ‘fill in’ gaps in our visual capability. In regards to our sense of ‘objective’ reality, our perception is “...psychologically projected into external space and accepted as our reality” (Gregory, 1997; p.1121), *despite* the nature of our physiological limitations. It currently seems that knowledge-driven neural-cognitive mechanisms assist by helping to construct the 'objective' visual reality we perceive, however future research into the nature of consciousness in the brain is needed to verify this relationship. The extent is evident in LSF object, change blindness and illusion stimuli. We open our eyes, and it feels effortless to see – and yet, this is actually a very complicated process requiring knowledge. The extent to which we are conscious of vision, perception and other cognitive processes is debatable, but at this point it seems that the brain’s capacity to apply knowledge-based concepts of the world to incoming visual information is centre to our visual experience.

References

- Carretie, L., Hinojosa, J., A., Mercado, F., & Tapia, M., (2005). Cortical Response to Subjectively Unconscious Danger. *NeuroImage*, 24(3), 615–623.
- Baddeley, A., (2003). Is Working Memory Still Working? *European Psychologist*, 7(2), 85–97.
- Bar, M., (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition. *Journal of Cognitive Neuroscience*, 15(4), p.600–609.
- Beldauf, D., & Deubel, H., (2008). Visual Attention During the Preparation of Bimanual Movements. *Vision Research*, 48, 549–563.
- Ganis, G., & Kosslyn, S., (2007). Multiple Mechanisms of Top-Down Processing in Vision. In Funahashi, S., (Ed.), *Representation and Brain*, Tokyo, Japan: Springer-Verlag, p.21–45.
- Gregory, R., (1997). Knowledge in Perception and Illusion. *Philosophical Transactions (B) of the Royal Society*, London, 352, p.1121–1128.
- Gibson, J., (1986). *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates Inc., New Jersey: USA.
- Henderson, J., (2007). Regarding Scenes. *Current Directions in Psychological Science*, 16(4), 219–222.
- Hill, H., & Johnston, A., (2007). The Hollow Face Illusion: Object Specific Knowledge, General Assumptions or Properties of the Stimulus? *Perception*, 36, 199–223.
- Króliczak, G., Heard, P., Goodale, M., & Gregory, R., (2006). Dissociation of Perception and Action Unmasked by the Hollow Face Illusion. *Brain Research*, 1080(1), 9–16.

- Kverga, K., Boshyman, J., & Bar, M., (2007). Magnocellular Projections as the Trigger of Top-Down Facilitation in Recognition. *Journal of Neuroscience*, 27(48), 13232–13240.
- Kverga, K., Ghuman, A., & Bar, M., (2007). Top-Down Predictions in the Cognitive Brain. *Brain and Cognition*, 65, 145–168.
- Lamme, V., (2003). Why Visual Attention and Awareness are Different. *Trends in Cognitive Sciences (Opinion)*, 7(1), 12–18.
- Lindquist, M., (2008). The Statistical Analysis of fMRI Data. *Statistical Science*, 23(4), 439–464.
- Navon, D., (1977). Forest Before Trees: The Precedence of Global Features in Visual Perception. *Cognitive Psychology*, 9, 353–383.
- Rees, G. and Seth, A., (2010). The Cognitive Neuroscience of Consciousness. *Cognitive Neuroscience*, 1(3), p.153–154.
- Rensink, R., O'Regan, J., & Clark, J., (1997). To See or Not to See: The Need for Attention to Perceive Changes in Scenes. *Psychological Science*, 8(5), 368–73.
- Stechler, G., & Latz, E., (1966). Some Observations on Attention and Arousal in the Human Infant. *Journal of the American Academy of Child Psychology*, 5, 517–525.
- Schenk, T., & McIntosh, R., (2010). Do we have Independent Visual Streams for Perception and Action? *Cognitive Neuroscience*, 1(1), p. 52–78.
- Simons, D., & Levin, D., (1998). Failure to Detect Changes to People During a Real-World Interaction. *Psychonomic Bulletin & Review*, 5(4), 644–649.
- Simons, D., & Rensink, R., (2005). Change Blindness: Past, Present, and Future. *Trends in Cognitive Sciences (Opinion)*, 9(1), 16–20.
- Tsuruhara, A., Nakato, E., Otsuka, Y., Kanazawa, S., Yamaguchi, M., & Hill, H., (2011). The Hollow-Face Illusion in Infancy: do Infants see a Screen Based Rotating Hollow Mask as Hollow? *i-Perception*, 2, 418–427.