Cognitive Science
Essay 2 - Oliver Lack

Explain the difference between digital and analog computational models of vision. Do the facts about mental imagery create a problem for digital models of vision?

The purpose of this essay is to overview digital and analog computational models of vision and illustrate differences between the two forms. It will show how mental imagery may influence the plausibility of digital models. Firstly, I will address the process of vision from the dominant perspective of constructivism. Secondly, I will propose the conventional digital computer as a model to illustrate the fundamental aspects of digital computation. I will use Biederman's (1987) Recognition by Components Theory (RBC) to exemplify a multi-stage algorithmic model of vision. Furthermore, I will contrast the differences to analog computation with regard to representation in a connectionist network and second-order resemblance. Additionally, I will use Shepard and Metzler's (1971) mental rotation experiment to demonstrate that mental imagery creates a problem for digital models. To conclude, I will give evidence from Brooks (1968) and Zeki (1992) suggesting mental imagery as part of visual cognition. I will describe the implications for digital models of vision and suggest that exploration into analog models may provide a more plausible account of visual cognition.

Vision is a process that takes retinal information from the external world and produces a useful description of the environment excluding irrelevant information. Marr (1982) describes vision as an active process that constructs perceived shapes, textures, motions, and entire objects. Marr founded his account of vision on the idea of unconscious inference proposed by Helmholtz (1867). Helmholtz recognised a gap between the optical information available directly from retinal stimulation and the perceptual knowledge derived from it. He contends that there must be a form of unconscious inference used to transform insufficient retinal stimulation into a useful perceptual model of the environment. He suggested that combining retinal stimulation and 'hidden assumptions' in the brain allows individuals to construct a visual perception. This laid the foundation for the dominant classical approach to vision: constructivism (Palmer, 1999, pp. 56). Digital theorists believe that visual construction is an active process that applies 'hidden assumptions' or rules of vision to retinal stimulation (Helmholtz, 1867; Biederman, 1987; Marr, 1982). In contrast, analog theorists believe visual construction results from a self-organised dynamic system powered by physical analogy (O'Brien, 1999; O'Brien & Opie, 2011). The differences between these models will be addressed below.

Digital computational models of vision take the popular view of the mind as an information processor similar to the processor of a digital computer. A digital computer partitions a continuous variable into discrete syntactic states. These states are assigned information (semantic properties) and represented using the binary numerals 0 and 1. These physically arbitrary units represent semantic properties. When syntactic states are assigned semantic properties, they create *symbols*. Programmed software (rules) sensitive only to the syntactic properties of the symbols manipulates them to produce output with respect to their semantic properties (Bermudez, 2014, pp. 43). Digital computation can be described as symbol manipulation. The computation bears an arbitrary correspondence to the representational content of the symbols; causal work is only performed on the syntactic properties of these symbols. The programmed software (rules) are designed to keep syntactic processing consistent with the intended semantic interpretation (O'Brien, 1999, pp. 5). In Haugeland's words, 'if you take care of the syntax, then the semantics will take care of itself' (Haugeland, 1985, pp. 106). A digital model of vision would suggest a similar form of symbol manipulation implemented in the brain.

A fundamental assumption of digital computation is the algorithm (rules) that governs the way in which representational vehicles (symbols) are manipulated. Without an algorithm, a digital model is blind. In contrast, a digital model that contains an algorithm behaves as if it were a semantic engine and so is capable of computational work. Digital theorists suggest that vision depends on a complex multi-stage algorithm. Biederman's (1987) Recognition by Components Theory (RBC) is an example of a digital model that uses different processing stages each with their own algorithm. Starting from the retinal image, input is extracted from the environment. It is then segmented into separate regions at points of deep concavity and discontinuities in curvature (Marr & Nishihara, 1978). This process is represented in the first three stages of figure 1, which are displayed in individual boxes. RBC assumes that each stage has its own algorithm that dictates how the input data is manipulated. Each segmented region is approximated to one of 36 fundamental components that Biederman called *geons* (geometrical ions). The hypothesised volumetric *geons* are combined in a separate stage to create a structural description and then matched to a stored categorical description of an object in the following stage. An example of these *geons* and how they combine to create structural descriptions of objects can be seen in figure 2. Although objects in the world can be complex and irregular, Biederman's digital model constructs components from retinal stimulation and manipulates them using a multi-stage algorithm. Combinations of components are matched to stored categorical descriptions of objects.

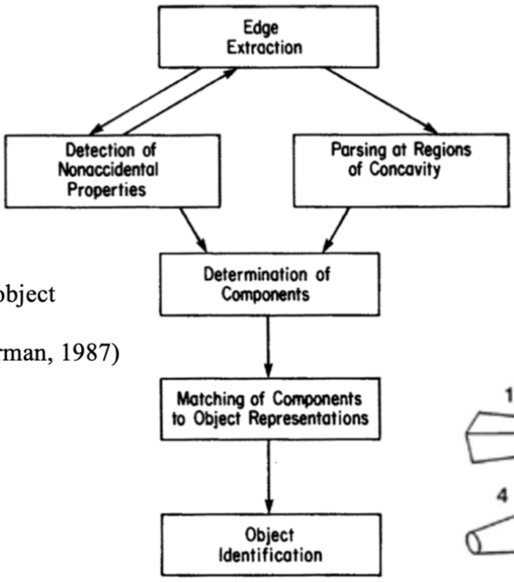


Fig. 1 Processing stages of object perception
(adapted from Fig. 2. Biederman, 1987)

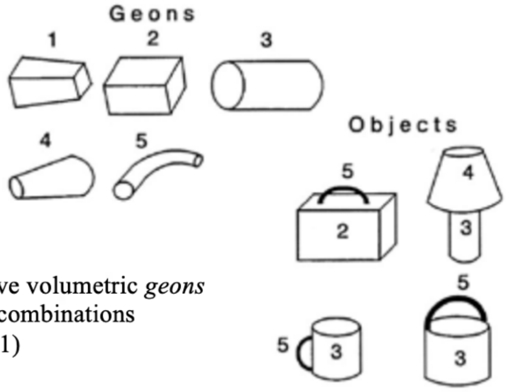
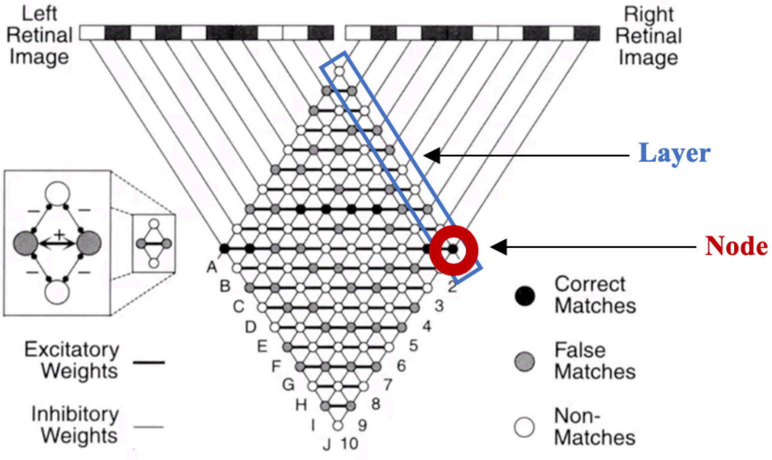


Fig. 2 Primitive volumetric *geons* and potential combinations
(Pujol, L. 2011)

Digital models assume that the localised symbols used in visual processing are the vehicles of mental representation. Analog models claim that the vehicle of mental representation is the state of the system as a whole (Cummins, 1989, pg. 157). Gestalt theorists of the early 1900s (Köhler, Koffka & Wertheimer) suggested that visual perceptions were holistic and different from the sum of their components. Köhler expanded connections between gestalt theory and biological mechanisms by proposing that the brain may be a dynamic physical system, converging towards an equilibrium state of minimum energy (Palmer, 1999, pp. 50-52). An example of this concept can be seen in a soap bubble. No matter the original shape of the bubble, it will always return to a sphere, its state of equilibrium. Particular analog models, known as connectionist networks, behave as dynamic physical systems. Connectionist networks are comprised of a large number of connected nodes capable of being excited or inhibited by input signals. These nodes are arranged into layers, usually consisting of input and output layers, that are mediated by multiple hidden layers. Each node has a weighted connection to each node in the following layer and a minimum threshold at which it will activate when given sufficient input from the sum of all attached connections. The input signals in the first layer of the network cause a pattern of activation to flow through the subsequent layers of the network. The activation is passed through each layer of the network until the activation is stabilised into a state of equilibrium (Sterelny, 1991, pg. 171). The stabilised state of the network's activation as a whole contributes towards the interpretation of one's visual perception. This is known as a distributed representation (Clark, 2001, pg. 66). Figure 3 shows an example of a connectionist network used to model stereo vision.

Fig. 3 Connectionist model of stereo vision (adapted from Opie, 2020)



The weight of each individual connection can increase (excite) or decrease (inhibit) the strength of the signal and is given by a real number adjustable by the system. These connections are used for transferring and altering the signal when passed through the network; they are the distributed information stores and processors of the system. As a consequence of distributed representation in analog connectionist models, information storage and processing are not separated; unlike digital computation which separates storage from processing (Clark, 1993, pg. 192).

Analog computation is powered by a physical analogy between the systems representing vehicles and the represented target domain. In contrast, digital computation is powered by the rules (algorithms) which govern the manipulation of the systems representing vehicles (symbols). Analog models bear a nonarbitrary correspondence to the represented target domain. The representing vehicles possess resemblance to the properties of the represented object; unlike the symbols in a digital model, which possess an arbitrary relationship to the represented object (O'Brien, 1999, pp. 7). The analog vehicles correspond to the represented target domain through a structural resemblance between the intrinsic properties of the representational vehicle and the represented object. O'Brien (1999) states, when the physical properties of a representational object are represented by equivalent physical properties of a representational vehicle, the relation is one of *first-order* structural isomorphism (first-order resemblance). This first-order resemblance of common physical properties is incompatible with what we know about properties of the brain (Block, 1983, pp. 511). However, the requirement of having equivalent physical properties may be relaxed; only systematically mirrored variations between properties of the representational vehicle and target domain is needed to form a *second-order* resemblance. For example, the relationship between the width of a tree's growth rings and seasonal rainfall; seasons with higher rainfall produce wide rings, whereas lower rainfall produces relatively narrow rings. If one maps the set of growth rings to seasons, variations in rainfall are reflected relative to the width of the rings (O'Brien & Opie, 2011, pp. 122). The relationship has a second order resemblance, as the intrinsic properties of each variable bear a nonarbitrary relationship. Evidence suggesting a second order resemblance between vehicles of visual processing and their represented target domain is addressed further in Shepard and Metzler's (1971) experimental results.

Shepard and Metzler (1971) developed an experiment using mental imagery, where two objects in a drawing are judged to be congruent (figure 5). The first two cases were the same object at different degrees of rotation. The third case had two incongruent images. The results found a direct linear relationship between the length of time that subjects took to solve the problem and the degree of rotation needed in each case. Subjects reported 'mentally rotating' visual images of the objects to judge congruency; the linear relationship between the time and degree of rotation supports this assertion if the speed of mental rotation is constant.

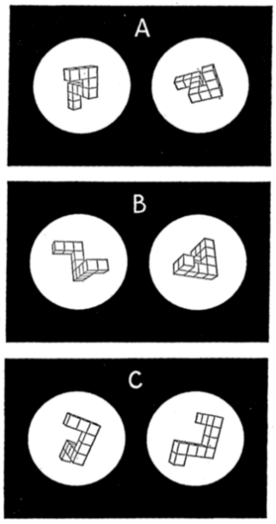


Fig. 5 Case examples of mental rotation of three-dimensional objects experiment (Shepard & Metzler, 1971)

An important feature of digital models is the processing time of information. This length of time is a function only of the quantity of information (Bermudez, 2014, pp.43). The type of information being processed should not matter, as the vehicles hold an arbitrary relationship to what they represent. Shepard and Metzler's (1971) experiments pose a problem for digital models. They suggest that the time taken for processing is dependent on the degree of rotation. The time is dictated by properties of the object, although the quantity of information remains the same. Kosslyn and colleagues (1978) demonstrate similar experimental results when scanning different distances on a memorised map. These experiments demonstrate a second-order resemblance between the processing of mental imagery and the represented object's intrinsic properties. Therefore, there exists a nonarbitrary relationship, implying that the processing of mental imagery may be powered by analog computation.

Mental imagery is believed to use the same cognitive system as visual perception. Internal visual representations used in conceptual and practical problem solving have shown to be part of one's cognitive visual system. A series of experiments by Brooks (1968) showed tasks which require mental imagery reduce the performance of simultaneous tasks which require visually guided responses. For example, Brooks had participants trace around the mental image of a blocked letter (figure 4) indicating whether each corner appears on the outside edge of the figure. Experimental groups consisted of a visually guided response requiring pointing to a written answer or an auditory response of 'yes' or 'no'. Performance of visually guided responses was degraded due to interference in visual cognitive processing. This interference suggests that visual perception and imagery use the same cognitive processes.

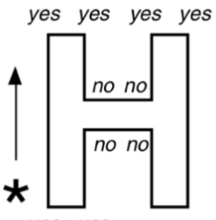


Fig. 4 Blocked letter task example (Brooks, 1968)

Furthermore, damage to specialised regions of the visual cortex in stroke patients has shown blindness to specific attributes of the world, such as colour, motion and form (Zeki, 1992, pp. 74). This damage also impairs patients' ability to form mental imagery. For example, when drawing an object, patients' artwork displayed the same modification shown in their visual impairment, such as incorrect colour placement or the form of line arrangements. If one uses mental imagery to draw, the correlation in impairment is likely to result from using the same cognitive system as visual perception.

In summary, the distributed representation of a dynamic physical system and the resemblance between the vehicle and represented domain, differentiate analog from digital models. In contrast, digital models separate processing from storage and have an arbitrary relationship between vehicles and their represented domain. Experiments from Shepard, Metzler (1971) and Kosslyn (1978) demonstrate a second-order resemblance between the mental rotation of objects and their spatial properties. If visual cognition uses digital computation, processing time should only be a function of the quantity of information. The experimental results contradict this assumption. Thus, the processing of mental imagery poses a problem for digital models. Evidence from Brooks (1968) and Zeki (1992) has shown that one's visual cognitive system accounts for both perception and mental imagery. Therefore, digital models may be unable to account for visual cognition. The nonarbitrary relationship between the degree of rotation and time is consistent with analog model conceptions. Investigation into analog computational models may result in a more plausible account of visual cognition.

References

Bermudez, J. L. 2014, *Cognitive Science*, 2nd Edition

Block, N. 1983, *Mental Pictures and Cognitive Science*

Brooks, L.R. 1968, *Spatial and verbal components in the act of recall*

Clark, A. 1993, *Associative Engines*

Clark, A. 2001, *Mindware*

Cummins, R. 1989, *Meaning and Mental Representation*

Kosslyn, S. Ball, T. & Reiser B. 1978, *Visual Images Preserve Metric Spatial Information: Evidence from Studies of Image Scanning,* Journal of Experimental Psychology: Human Perception and Performance Volume 4

Marr, D. 1982, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, MIT Press

Haugeland, J. 1985, *Artificial Intelligence: The Very Idea*

Helmholtz, H. 1867, *Treatise on Physiological Optics*

Marr, D. & Nishihara, H. K. 1978, *Representation and recognition of three-dimensional shapes*, Proceedings of the Royal Society of London, Series B, 200, 269-294.

O'Brien, G. 1999, *Connectionism, Analogicity and Mental Content*, Acta Analytica, Vol. 22, pp. 111-131

O'Brien, G. & Opie, J. 2011, *Knowledge and Representation, Representation in Analog Computation 2011 / Newen, A., Bartels, A., Jung, E. (ed./s), pp.109-127*

Opie, J. 2020, *Vision and Thought,* Cognitive Science: Minds, Brains and Computers

Palmer, S. 1999, *Vision Science*

Pujol, L. 2011, *Realism in Virtual Reality applications for Cultural Heritage*

Sterelny, K. 1991, *The Representational Theory of Mind*

Zeki, S. 1992, *The Visual Image in Mind and Brain,* Scientific American