Script For Video Lecture On 'Seeing and Visualizing'
PSYCO 354, "Foundations of Cognitive Science"
Michael R.W. Dawson, University of Alberta
Prepared October 2013
To Be Used As Lecture For November 14, 2013

Slide 1: Title And Objectives

Welcome to Week 11, "Seeing and Visualizing". The purpose of this lecture is to explore an example hybrid theory in cognitive science. A hybrid theory includes elements of all three approaches to cognitive science: classical, connectionist, and embodied. The particular hybrid theory of interest in this class is Zenon Pylyshyn's recent theory of seeing and visualizing.

We will begin by considering how some classical theories of language are beginning to include elements of embodied cognitive science; they situate the symbolic!

We will then detail the properties of classical theories of visual perception, focusing in particular on problems of underdetermination and on Treisman's feature integration theory.

This will set the stage for our discussion of Pylyshyn's theory. The hybrid nature of this theory is an interesting reaction to classical models of seeing and visualizing.

Slide 2: From Symbolic To Situation

Classical cognitive science has launched a counter-reaction against embodied cognitive science. It claims that embodied cognitive science is merely recycled behaviorism. Consider Pylyshyn's position: "Some ideas are merely the perennial recycling of behaviorist ideology in psychology, which attempts to empty the organism of thought and replace it with increasingly complex reflexes". Ironically, many classical scholars turn to embodied ideas in order to generate workable theories!

Slide 3: On Beyond Zebra

For example, take Ray Jackendoff, a leading linguist. He is steeped in the classical tradition; he was trained by Noam Chomsky. But Jackendoff's theories require situation. He proposes strong links between the structure of linguistic representations and the structure of representations from other modalities. Consider his cognitive constraint: "There must be levels of mental representation at which the information conveyed by language is compatible with information with other peripheral systems such as vision, nonverbal audition, smell, kinesthesia, and so forth". In his book *Semantics and Cognition*, Jackendoff even abandons logicism. He argues that semantics does not involve assigning truth values to logical or linguistic expressions.

Slide 4: Grounding Language In Vision

A new generation of researchers takes Jackendoff's cognitive constraint very seriously. MIT's Deb Roy builds robots that learn language. His machines link phonetic features from recorded speech to information derived from their vision and action. Their semantics requires situation and embodiment. This is evident when we see a video segment of one of Roy's robots, Ripley, learning language concepts.

[Ripley video segment goes here. Source: MIT Media Labs, Cognitive Machines group]

Slide 5: Perception

To set the stage for Pylyshyn's hybrid theory, let us first consider classical theories of perception. According to classical theories, perception is used to provide information for constructing useful models or representations of the real world. Perception provides sense data for modeling, thinking, and planning in the classical sandwich.

Slide 6: Data-Driven Processing

One common theory about perception is that it is purely "bottom-up" or "data-driven". This means that perceptual processes detect visual features, and then combine these features into feature combinations that define objects. A famous version of this type of theory is Selfridge's pandemonium model of letter recognition – a model that is connectionist in nature!

Slide 7: Poverty Of The Stimulus

The problem with a purely data-driven perceptual system is that the visual information detected by the eye – the proximal stimulus – is not sufficient to uniquely specify the properties in the world that caused it. That is, one proximal stimulus is consistent with many different interpretations, only one of which is correct. Consider the Necker cube. If we watch it long enough, we will see that this proximal stimulus is ambiguous – it supports two very different three dimensional interpretations.

This situation is generally called "the poverty of the stimulus" or "the problem of underdetermination".

Slide 8: The Problem Of Underdetermination

The Ames chair is another example of the problem of underdetermination. From one perspective, the proximal stimulus gives rise to the perception of a normal chair. However, the physical arrangement that actually causes the proximal stimulus is an arrangement of parts that are not very chair-like! Clearly the proximal stimulus is consistent with two very different models of the world.

Slide 9: Top-Down Processing

Classical cognitive science has a standard solution to the problem of underdetermination. It assumes that perception is a form of cognition, a form of reasoning. We use our beliefs and expectations to provide information that is missing from the proximal stimulus. The information that we provide generates a unique interpretation of the visual world. Psychologist Richard Gregory argues that “Perception becomes a matter of suggesting and testing hypotheses” Seeing is believing!

Slide 10: Theory-Driven Processing

This classical view of perception is called “top-down processing” or “theory-driven processing”. To illustrate it, imagine that I have a proximal stimulus that delivers features like small, black-and-white, four legs, furry, two eyes, and nose. I use knowledge about where I am to create a sense of what I am seeing. If I am in my house, I expect to see my cat Phoebe, and this is my interpretation of the proximal stimulus. If I am in the ravine, I do not expect to see my (indoor) cat – but would not be surprised to see a skunk. A change in expectations results in a changed interpretation of the same proximal stimulus.

Slide 11: Top-Down Problems

A pure top-down theory has problems too. If we only see what we expect to see, then we will not see surprises. Unfortunately, the surprises in the world are what kill us! I don’t expect to see a tiger running around in Millcreek Ravine. However, I hope that in spite of this, if I did happen to encounter one – escaped from the zoo, perhaps – I would see it, and be able to avoid being eaten by it!

Slide 12: A Compromise View

Pure data-driven and pure theory-driven models of perception have problems. So, many modern theories are compromises that incorporate both. A modern theory would include data-driven modules that detect various features (low-level vision). A modern theory would also include theory-driven processes that link knowledge of the world to visual information, so that we can classify objects and know what they can be used for (high-level vision). A modern theory would also have a middle process – called visual cognition – that acts as a go between. That is, high-level vision can request visual cognition to invoke attentional processes called visual routines to return particular information (e.g. by pasting features together, or by computing the relationship between two objects (is one to the left of the other?)

Slide 13: Feature Integration Theory

An important example of a modern – yet classical – theory of visual processing is Anne Treisman’s feature integration theory. According to this theory, vision begins when low level processes separate stimuli into their component features. These features are represented as locations of activity on different feature maps. When required, a middle level of visual cognition can direct a spotlight of attention to a specific location in a master map. The attentional spotlight can ‘glue’ together all the features present at that location (features represented at that location in different feature maps). The feature conjunctions created by visual cognition can then be linked to object descriptions, called ‘object files’, for processing by higher-level process.

Slide 14: Treisman’s Visual Search Task

An important source of evidence for feature integration theory comes from experiments on visual search. A subject is presented a display in which there are a number of objects. In such a display, there may be an ‘odd man out’ – an object that is different from all of the other distractors. The task is to search through the display as quickly as possible to see if there is an odd man out target. The time required to determine if such a target is present is the dependent measure. Independent measures include how many distractor objects are present, as well as the nature of the objects.

Slide 15: Complex Objects

Complex objects are defined as combinations or conjunctions of simple features. The two objects shown here (a connected object and a disconnected object) are made from exactly the same features, but differ in how these features are combined. The following demonstration will give you the sense that it takes time to find complex objects in a visual search task. In contrast, simple objects defined only by the presence of a unique feature (like color or orientation) seem to pop out of the display. The differences between searching for feature conjunctions and searching for simple features are explained by feature integration theory.

[Visual Search Demonstration Video Goes Here – created by Dawson]

Slide 16: Feature Integration Theory

There are many different kinds of evidence that have been used to develop and support feature integration theory. For the purpose of this class, let us only consider this theory in the context of visual search. Feature integration theory can be used to explain why some targets pop out, while others do not.

Slide 17: Feature Integration Theory And Pop Out

First, how does this theory explain pop out? If a target pops out, then it is the only source of activity in a low level feature map. So, if there is only one active location in one of these maps, then the unique target can be detected without invoking any processes higher up in the model.

Slide 18: Feature Integration Theory And Pop Out (2)

For instance, if the target is the only red object in a display, it will be the only source of activity in the red color map. It can be detected immediately, without serial search from one location to another. The time to find it will not be related to the total number of distractors in the display.

Slide 19: Feature Integration Theory And Conjunctions

Feature integration theory has to deal with complex objects in a different way. These objects by definition are combinations of features. So, a unique complex object can only be found using attention. The spotlight of attention is moved to one location in the master map. It is then used to glue together any features that are present (in separate feature maps) at that location. Only after the features are combined by attention can the properties of the object be analyzed.

Slide 20: Serial Search

So, to find a complex odd man out, the attentional spotlight is directed to one location for processing.

Slide 21: Serial Search (2)

Then it is directed to another location to process the features there

Slide 22: Serial Search (3)

Then it is directed to another location. This search, from one distractor location to another, is called serial search. The greater the number of distractors, the greater the number of locations to search. This is why for complex objects reaction time increases with the number of distractors, as shown in the graph on the slide.

Slide 23: Pop Out And The Brain

Neuroscientists have found that there are independent pathways, responsible for detecting simple features, in early visual processing. It is these features – and only these features – that will produce pop out. Treisman feels that one of her main contributions was to propose this sort of model before neuroscientists came up with it, as can be seen in the short video that follows.

[Treisman National Medal of Science video goes here – source http://www.youtube.com/watch?v=AbwvmpANMi8]

Slide 24: Indexing, Not Feature Integration

Pylyshyn's theory of seeing and visualizing departs from feature integration theory in some radical ways. He replaces the single attentional spotlight with multiple attentional tags called FINSTs. These tags pick out objects, and 'stick' to objects, but do not deliver object features. In Pylyshyn's theory we detect objects before we detect object features!

Slide 25: Finger Instantiations (FINSTs)

FINST stands for 'finger instantiation'. It is an attentional tag that 'sticks' to an external object, just as if we were following that object by placing a finger on it – but not looking at the object. FINSTs track objects without delivering visual features. When we need featural information, we use the FINST to access the object for inspection.

Slide 26: Multiple Object Tracking

The original experimental support for FINSTs came from multiple object tracking experiments. Subjects fixate on the center of a display filled with identical objects. Some of the objects blink, identifying them as targets, and attracting FINSTs. Then the objects move independently and randomly. The task is to track the moving targets – with the mind, not the eye! After a period of time, movement stops, and an object is highlighted. The task is to say whether the object was a tracked target or not. Subjects can track 4 or 5 of these objects simultaneously – supporting the notion of FINSTs, and causing problems for theories that appear to an attentional spotlight. The following short video demonstrates the general paradigm.

[Multiple object tracking demonstration goes here – source http://ruccs.rutgers.edu/faculty/pylyshyn/DemoPage.html]

Slide 27: End of Part 1