“A Critique of Pure Vision”

Patricia S. Churchland, V.S. Ramachandran, and Terrence J. Sejnowski

Presented by: Sawan Patel

Current Picture ('94)

- **Theory of Pure Vision**
    - Visual system serves to create a 'detailed replica' of one's environment with hierarchical organization that also operates independently of other 'sensory modalities' and 'previous learning, goals, motor planning and motor execution'
- Marr's Tenets
    - What we see is a fully elaborated diagram of a visual scene
        - Transforming two-dimensional data into a description of the three-dimensional spatio-temporal world (Tsotsos, 1987)
    - Hierarchical processing
    - Dependency relations
        - Higher levels in processing hierarchy depend on lower ones, but generally not vice versa

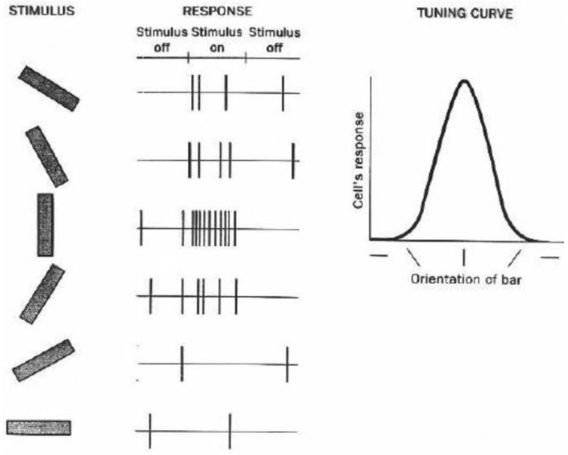


FIGURE 4.8 Response of a single cortical cell to bars presented at various orientations.

Example of V1 cell tuned to a particular line orientation (O'Reilly et al., 2020)

Discussion Question

How might a ***rich replica*** of surroundings be beneficial for practical/commercial applications? What would be some physical drawbacks to maintaining such a system?

(From Piazza)

- Beneficial: low-dimensional reinforcement learning applications, object detection
- Drawbacks: power load

Author Critiques

- Idea of pure vision is a stretch
    - Obscures most important computational strategies used by the brain
    - Term ‘impedes progress’ (e.g. ‘indivisible atom’)
- Question:
    - ‘Has research in vision now reached a stage where the orthodoxy no longer works to promote groundbreaking discovery?’
- Interactive Vision

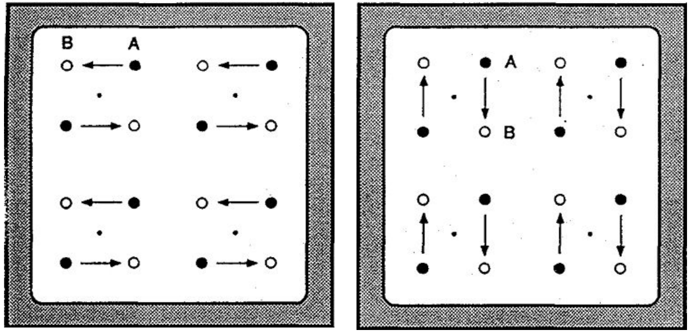
Interactive Vision

- Evolution of perceptual systems
    - Visual system and motor control system are intertwined
        - Visual system constructed to facilitate the four F's
- Visual 'semiworlds'
    - Only immediately relevant information is explicitly represented by our visual systems
        - Saccade/Attention system
            - Saccades every 200 msec
            - Amount of foveated area represented in detail in visual field depends on animal's general interests
- Interactive vision and predictive visual learning
    - Interactive ~ 'exploratory,' upkeep of predictive representations
    - Correlations between sensory modalities improve predictive representations in real world
        - Recognition improves if all perceptual apparatus are used to explore over time
- Motor System & Visual System
    - Motor assembling *starts* with minimal analysis of visual scene
        - 'Freezing' behavior
- Non-hierarchical organization
    - Real-world recognition (of visuomotor patterns) *depends* on recurrency
        - Recurrent connectivity is very evident in neuroanatomy (e.g. thalamus)
- Memory and Vision
- Pragmatics of research
    - Assuming the above, jointly studying visual system with other modalities is unavoidable

Is Perception Interactive?

Support from Visual Psychophysics

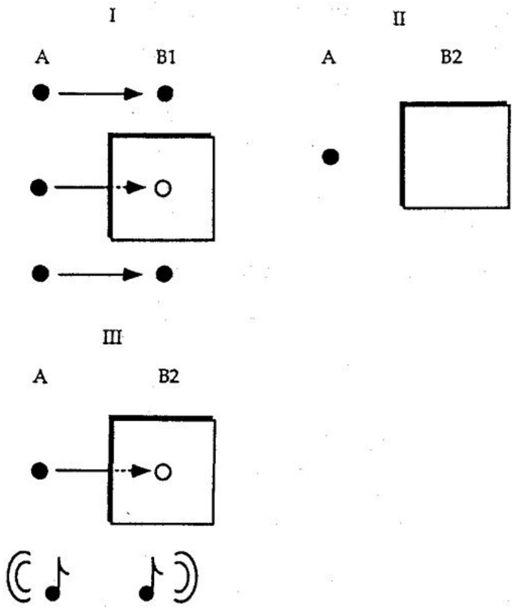
- Question: Do global factors play a role in the visual system's perceptions?
  - Context: Ullman ('79) proposed algorithm which solves problem of determining which features of earlier presentation go with which features at later presentation using only local information
- Bistable quartets



Perceived motion of all dots is uniform - either all 'move' vertically or all 'move' horizontally, never in between

Support from Visual Psychophysics

- Question: Do global factors play a role in the visual system's perceptions?
    - Context: Ullman ('79) proposed algorithm which solves problem of determining which features of earlier presentation go with which features at later presentation using only local information
- Occlusion Studies



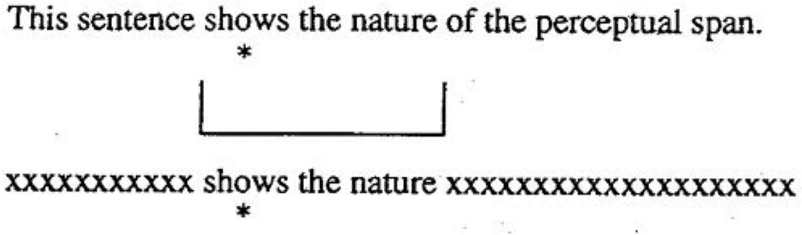
I. Dots in **A** blink to position in **B1** - middle dot appears to 'move' along with other dots
II. When dot in **A** blinks, motion is not perceived. Subjects detect a blinking dot on left and square on right in **B2**.
III. When dot in **A** blinks on, a tone is presented to left ear. When dot blinks off, a tone is presented to right ear. Motion is reported in this case, contrary to in (II).

Support from Visual Psychophysics

- Question: Can semantic categorization affect shape-from-shading?
- Masks
  - If concave mask is held 2m away from you, it is perceived as convex
    - Holds even if mask is illuminated from bottom *and* if subject is informed the direction of illumination!
  - Is this an effect of general assumption that most things are naturally convex?
  - **Experiment**:
    - Protocol: Two concave masks shown to people, one rightside up and other upside down. Subjects start at 0.5m and slowly walk backwards
    - 0.5m ~ subjects correctly identify both masks as concave
    - 1m ~ subjects see upright mask as convex
    - 2m ~ subjects see upside down mask as convex
  - **Results**:
    - Downstream processes (face categorization) affects upstream processes (shading/curvatures)
      - I.e. demonstrates existence of top-down correction

Visual Attention

- Conjecture: the undeniable feeling of having a whole scene visual representation is the result of:
  - Repeated visual visits to stimuli in scene
  - Short-term semantic memory on order of a few seconds that generally maintains ‘what is going on’
  - Brains objectification of sensory perception
  - Predictive dimension of pattern recognition
- Experiments:
  - Using online computer control to change what is visible on computer display as a function of subject's eye movements
  - Window text reading experiment
  - Attention precedes saccades
- Implications on relationship b/w perception, memory and attention:
  - If you are not attending to something, then you don't see it
  - If you are not attending to something and you don't see it, you do not have an iconic memory for it


This sentence shows the nature of the perceptual span.
*
xxxxxxxxxxxx shows the nature xxxxxxxxxxxxxxxxxxxxx
*

Found that maximum perceptual span is 2-3 character spaces left and 15 character spaces right

Discussion Question

Have you ever felt the ‘undeniable’ sense of having a representation of an entire scene in your head? If so, when?

Click HERE for selective attention video

Neuroanatomy Studies

- Backpropagations
  - Back-projections outnumber forward axon projections (e.g. V2 to V4)
- Diffuse ascending systems
  - Number of afferent systems that arise in small nuclei located in brainstem and basal forebrain
    - Not just thalamo-cortical (e.g. VTA sending da-ergic projections to frontal ctx)
- Corticothalamic connections
  - Sensory inputs from specific modalities project from modalities to middle layers of cortex
    - Reciprocal connections in deep layers project back to thalamus
- Connections from visual cortical areas to motor structures
  - 25 areas project to superior colliculus (saccades)
  - Nearly every area of mammalian ctx projects to striatum
    - Striatum lesions well-correlated with motor impairments

Neurophysiology Studies

- Connections from motor structures to visual ctx
    - Interactive effects shown even at early stages
        - Spontaneous V1 activity suppressed according to onset time of saccades (20-30 msec after saccade is initiated)
    - Neurons sensitive to eye position found in LGN, V1 and V3
        - Visual features encoded in egocentric coordinates via eye position information
- Dynamic mappings in exotropia
    - **Exotropia**: form of squint in which both eyes used when fixated on small objects close by, but when looking at distant objects, squinting eye deviates outward by as much as 60 degrees
        - Patient doesn't experience double vision - deviating eye's image suppressed, but unclear at what stage suppression occurs
        - Some claim that *binocular fusion* occurs in some patients (**anomalous retinal correspondence**), not popular among clinicians

Neurophysiology Studies

- Dynamic mappings in exotropia (cont'd)
  - Intermittent exotropia ~ patients appear to fuse images both during near vision and far vision
    - Experiment: found that disparities in locating source of a light as small as 20 min of arc could be perceived correctly even when anomalous eye deviated by as much as 12 degrees
      - Half-images of eyes were exciting non-corresponding retinal points separated by 12 degrees, but still showed small behavioral disparities
  - Claim: Binocular correspondence (and fusion) cannot be based exclusively on anatomical convergences of inputs in V1
    - Since binocular correspondence can change continuously in real time in a single individual depending on degree of exotroia
    - Relative displacement observed b/w two afterimages suggested local sign of retinal points must be continually updated as eye deviates outwards
  - Simple perceptual process like localization of an object in x/y coordinates is not strictly a bottom-up process

Computational Advantages of Interactive Vision

Does it make computational sense to have an interactive style rather than a hierarchical, modular, modality-pure and motorically unadulterated organization?

- Figure-ground segmentation and recognition are more efficiently achieved in tandem than strictly sequentially
    - Problem: resolution for V1 RF's is small and couldn't locally decide which pieces of an image belong together
    - Idea: use partial segmentation to help recognize and use partial recognition to help segment.
    - Performance of machine reading on numerals
        - Problem: efficient machine reading of zip codes on letters
            - No guarantee where zip code numerals will exactly be located (localization), which squiggles belong to which digits (segmentation) and whether a digit is a 0 or a 6 (recognition)
- Movement makes many visual computations more simple
    - Smooth pursuit
    - Optical flow
- Self-organization of model
    - Using nature to 'grow' vision system as is the case in nervous systems (via Hebbian schemes)
        - Correlation-based models show that properties such as ocularity, orientation and disparity can emerge from simple Hebbian mechanisms
        - Eye movement during development combined with hebbian plasticity can be capable of extracting higher order correlations from complex visual inputs
- Interactive perception simplifies learning problem
    - How does brain determine which features are relevant to reward and punishment (assuming maintenance of perfect visual scene)?
    - By narrowing number of visuomotor trajectories that count as salient, attention can bias choice of synapses to strengthen

Discussion Question

Have we since solved the figure-ground segmentation problem exemplified by the zip code case? What methodologies could be used and are they generalizable?

- (in-class) Yes! MNIST dataset has revolutionized handwritten digit classification problem and we now have models that achieve superhuman performance. Methodologies include: CNNs (some with recurrency), LSTMs, and more.

Learning To See

- Responses reinforced by a reward are likely to be produced again when relevantly similar conditions arise
    - Birds and bees, ‘two armed bandit’ conditions
- Reinforcement learning
    - Limits attached to RL should instead be attributed to *rich-replica* assumption of pure vision
    - Brain can create predictive sequences by rewarding behavior that leads to conditions that in turn permit a further response that will produce an external reward
        - Can then build a network replete with predictive representations that inform attention as to what is worth looking at given interests
            - Ex. Bees visiting flowers
                - Comparison between arithmetic mean of rewards and variance in distribution **or** comparison of received reward and predicted reward
    - Bees vs Humans
        - Human ‘increased intelligence’ compared to bees is really increased predictive-goal-relevant representational power
        - If some property of world is visually represented, it probably has **high utility** in predictive game

Concluding Remarks

- Pure Vision was critical to get the field to grow, but it must also be replaced should sufficient evidence be provided
    - The brain is only approximately hierarchical
    - Evolution suggests that having a very sophisticated, near-exact visual perception system does not guarantee survival
    - Motor processes can influence sensory processes (and vice versa!)

Piazza Discussions

- Reference biology when modeling (with caution)!
  - Adam Cheng Li
  - Following biology blindly is irresponsible as biology certainly has made mistakes (even in the human body)!
  - Also unfair to directly compare performance of robots and humans on vision tasks without recognizing that both have different tools to approach the same problem
    - Human eyes work differently in comparison to state-of-the-art sensors (both have their own advantages and limitations)
- Is collecting multi-modal data *worth it?*
  - Lance Bassett, Benjamin Steinig, Priya Thanneermalai
  - Storage/memory strain could be a huge issue with integrating multi-modal data into models, as is already the case with just high resolution visual percepts.
  - Sentiment prediction and transcription applications have benefitted from the use of multimodal learning
- Modern-day RNNs (LSTMs)
  - Jemuel Stanley Premkumar, Harikrishnan Seetharaman
  - Modern-day RNNs (LSTMs) enable flow of model to be backwards as well as forwards, allowing for recurrency from higher-level layers (which typically encode higher-order features) to lower-level layers

Citations

- Koch, Christof, and Joel L. Davis. *Large-Scale Neuronal Theories of the Brain*, MIT Press, Cambridge, MA, 1995.
- O'Reilly, Munakata, Hazy and Fank. *Oriented Edge Detectors in Primary Visual Cortex*, LibreTexts Medicine, 2020.