The Language of Vision

Patrick Cavanagh
Glendon College
Attention = selection

Selection by color

Selection by location
Multifocal Attention

Classically, single focus of attention
BUT most subjects can track 4 or 5 targets

Pylyshyn & Storm, 1988
Use moving attention to isolate and integrate brief events.

Example: Inspect the afterimage of a rapidly flickering target without attending to the target.
Vision exports descriptions to other modules in the brain
  – Jackendoff, Mandler, etc.

The format is a language
  – Fodor, Logan, Sereno
Language transmits conceptual representations from one person to another.
Central routing

“Attentive” Narrative System

Primary Visual System

Retinal Input

Selection

Scheduling Overhead

Auditory task

Vision

Daydreaming

Franconeri, Alvarez, & Cavanagh, TICS 2013

Other “low-level” channels

Motor System

Other “low-level” channels

eg

Planning

Emotions

Real Language

Declarative Memory

Other Senses
Both describe the properties of the same world
How does vision communicate with the rest of the brain?

Who does the talking
- pre-attentive vision
- visual attention

Who is listening
- other brain modules
How does vision communicate with the rest of the brain?

What is the message?
Pictures and movies?

Labels for events?

Formatted, language-like description?

(Gregory, Sereno, Fodor, Logan, Miller & Johnson-Laird)
Properties of Language

compositionality / productivity: combinations produce unbounded set of descriptions

arbitrariness: no link between symbol and referent

displacement: reference to something not present

recursion: the man’s lips; the man who was running fell
Why Language?

If not pictures, how about labels?

Label each event with a different symbol

A closed set

e.g., Monkey danger, leave, food calls

With more labels, probability of forgetting increases

Effective limit of Q labels
Event a combination of two labels: NV

Now describe \((Q/2)^2\) events

More efficient but requires grammar


“Language” not an option, it is the only choice
What is the visual narrative?

Perhaps only “conscious vision” is broadcast to other modules

Only attended content

Description sent out (to memory, language, planning) is then low-bandwidth

Evidence: Change blindness

(Rensink, Simon, O’Regan)
Language of Vision

• Nouns = objects

• Verbs = actions

• Prepositions = spatial, temporal relations
Language of Vision

Jim throws the rock at John

noun   noun   noun
verb   preposition

Red knocks green into black
Vision is knowing what is where by looking (Marr)

Many objects rapidly labeled by visual system.

Some objects identified rapidly, in parallel, pre-attentively


Klefner & Ramachandran, 1992
Visual “nouns” may be labelled pre-attentively.

They are labels for the objects not pictures of them.
Vision has verbs

Tenses: past (motion), present, and future (intentions)

Verbs (familiar actions, “sprites”) are identified by integrating patterns across time --> slow

Reusable: Mary walks, John walks, Dots walk

Not available in primary vision

Are they visual or cognitive inferences?
Familiar Action Units: Sprites
Cavanagh, Labianca, & Thornton, 2001

Flexible descriptions which embody object constraints
Online animation to fit changing image data

Reusable: Mary walks, dots walk
Choice of trajectory OR

OR
Like recognizing and following along with melodies

Once triggered, memory fills in any gaps

Sprites are the units of high-level motion, the “verbs of vision”

Rather than the left, right, up and down, etc of low-level motion

--> Roll, bounce, slap, break, flutter, glide
Can we process several “verbs” at once?

Visual search task:
Find rightward walker among leftward

Search is slow

Cavanagh, Labianca, & Thornton, 2001
Find walker among scrambled figures

In both cases, identifying walker requires attention.
Verb tense: present progressive

Perceiving motion includes a comparison between the present and the past

Present progressive?
Past tense

Terminated motion or action: past tense

Visual or cognitive?
Future

Visual or cognitive?
Goals guiding current and future action.

Cause and effect.

Tenses, intentions, and causality: are these visual computations, or cognitive deductions?

Most likely visual, at least for causality
(Rolfs, Dembacher, & Cavanagh, Curr Bio, 2013)
Visual adaption of the perception of causality


Causality

pass vs collision events
noncausal vs causal

How can we tell if this is visual?
Visual adaption of the perception of causality


Visual computation: show adaptation
Extended exposure
Effect is negative and local
Moves with eyes (retinotopic)

vs cognitive processes
Repetition: learning, priming, or boredom
Effects are global
Do not move with eyes
Visual adaption of the perception of causality


Causal adaptation stimuli, launching

Ambiguous tests

Various amount of overlap
Adaptation to causal events made tests appear less causal

Critically, only at the adapted location

Vision has verbs

Visual verbs have some tenses: past (motion), present, and future (intentions)

Verbs require attention: one at a time

These may be visual computations, not general cognitive inferences
Prepositions

Spatial and temporal relations between two selections.
Behind: Tokens in the visual description refer to an item that is not present, occluded
Prepositions

Behind: Tokens in the visual description refer to an item that is not present, occluded

Displacement: reference to an item not present

Michotte, 1954
Prepositions

Behind: Is what you expect to find behind

Visual or cognitive?
Actual events are continuous
But we see a start and end, an action
and the agents and objects
Event perception, Zacks et al., 2001
Assigning possible components is easy

Any evidence for grammar or syntax?
Language of Vision

Is there a grammar?

What would ungrammatical vision be like?

Impossible events, magic?

Deduce that it is impossible,

or see that it is impossible
Impossible, but we do not see this a visual error.

Physics of mirrors only roughly captured in visual grammar

Never get right description
Again, impossible but error not detected by vision.

Local syntax OK

“I am writing to you with my sword raised and a pistol in each hand.”
Aaron Burr
Basic constituent structure analysis of a sentence:

Sentence / Verb phrase

Noun Phrase

Prepositional phrase

Noun Phrase

Article Noun Verb Preposition Article Noun

The cat sat on the mat.

Visual syntax

Vision detects an error
Mistake in syntax
Man ?? Dog.

“Phrase structure” seems incomplete

Eventually get right description
Visual syntax

Which body goes with this head?
Visual syntax
Some structural errors only verified cognitively
A visual Language Acquisition System.

Picks up regularities in the visual input stream to determine objects, actions, spatial and temporal relations.

Infers nouns, verbs, prepositions

Regularities come from the physics of the world but system not specialized for physics but only regularities
Did the Regularity Acquisition System extend to the gestural and then spoken stream.

Is vision the “Ur” language? (Gregory, Sereno)

Look for similarities in acquisition system and development: fossils of visual structure
Asymmetry in Visual Search

Tilted in vertical easier than vertical in tilted

Long in short easier than short in long

Bright in dim easier than dim in bright
Asymmetry in Lexical Marking

Antonyms: bright/dim, long/short, wide/narrow

Often one is base term, it names the dimension

Long - short --> length

Bright - dim --> brightness

Tilted - vertical --> tilt

The other is marked: short = \{long\}^-

Takes longer to process in speech

Possible evidence of visual syntax retained in spoken language structure?
Powerful property of language

One sentence can be embedded in another
The man that jumps is happy

A noun phrase can be embedded in another
The man’s lips

Does vision allow multiple embedded messages?
The green disc that was hit by the red knocks the black disc.
Recursion

picture spaces embedded in picture spaces
The soda can that had been crushed is lying on the floor.
The stones that are under water.

See an object and its embedded history at the same time.

The soda can that had been crushed is lying on the floor.
The stones that are under water.
The train that failed to stop crashed to the pavement below

→ See the object and the processes that produced the deviation from its canonical form
Attention exports a description of visual events to the mind in a “language” format.

Language format more efficient.

Rules for acquisition of visual grammar may have been the seed for mechanisms of acquisition of other grammars.