A new path to understanding vision
from the perspective of the primary visual cortex

Frontal brain areas

Visual cortices

Primary visual cortex (V1)

Retina

Li Zhaoping
July 16, 2018, presented at APCV 2018,
HangZhou, China
A new path to understanding vision

**Traditional paths to understanding vision**

1. Low level vision, mid-level vision, high-level vision
2. David Marr: primal sketch, 2.5 d sketch, 3-d model.

Li Zhaoping
July 16, 2018, presented at APCV 2018,
HangZhou, China
Talk outline

(1) The functional role of the primary visual cortex (V1)

(2) In light of V1’s role → a new plan to understanding vision

(3) A first example study in this new plan
The primary visual cortex (V1)

1953, Stephen Kuffler, retina, 1959-- Hubel and Wiesel, V1
The primary visual cortex (V1)

1953, Stephen Kuffler, retina, 1959-- Hubel and Wiesel, V1

Then …

Experimentally: V1 and beyond
Theoretical/modelling, Reichardt, Marr, etc.

2005: How close are we to understand V1? Olshausen and Field 2005
Do we really know what the early visual system does?

Carandini, Demb, Mante, Tolhurst, Dan, Olshausen, Gallant, Rust, 2005

**Standard models of V1 neural receptive field (combining filtering, rectification, squaring, normalization) captures only 15-35% of the variances in V1 responses.**

2012, David Hubel, in answer to “What Do You Feel Are the Next Big Questions in the Field?”

"We have some idea … for the retina, the lateral geniculate body, and the primary visual cortex, but that’s about it."

(Hubel & Wiesel 2012, Neuron)
The primary visual cortex (V1)

1953, Stephen Kuffler, retina, 1959-- Hubel and Wiesel, V1

Then ...

Experimentally: V1 and beyond
Theoretical/modelling, Reichardt, Marr, etc.

2005: How close are we to understand V1? Olshausen and Field 2005
Do we really know what the early visual system does?
Carandini, Demb, Mante, Tolhurst, Dan, Olshausen, Gallant, Rust, 2005

Standard models of V1 neural receptive field combining linear filtering, rectification and squaring, and response normalization captures only 15-35% of the variances in V1 responses.

2012, David Hubel, in answer to “What Do You Feel Are the Next Big Questions in the Field?”

“... We have some idea for the retina, the lateral geniculate body, and the primary visual cortex, but that's about it.”

(Hubel & Wiesel 2012, Neuron)

Questions:
Is a lack of understanding of V1 hindering our progress beyond V1?

Physiologically

Functionally (in behaviour)?
Information bottlenecks in the visual pathway:

10^7 bits/second
~ 10^6 neurons, 
~10 spikes/neuron
~1 bit/spike

10^9 bits/second (Kelly 1962)
~ 25 frames/second, 
2000x2000 pixels, 
1 byte/pixel

40 bits/second (Sziklai, 1956)

“To be or not to be, 
This is the question ..”

“To be or not to be, 
This is the question ..”
Information bottlenecks in the visual pathway:

We are nearly blind!

Vision ~ Looking (selecting) + Seeing

Task: find a uniquely oriented bar

top-down vs. bottom-up selection

10^7 bits/second ~ 10^6 neurons, ~10 spikes/neuron ~1 bit/spike

10^9 bits/second (Kelly 1962) ~ 25 frames/second, 2000x2000 pixels, 1 byte/pixel

"To be or not to be, This is the question .."
Information bottlenecks in the visual pathway:

Questions:
which brain areas are doing the bottom-up selection?

Frontal? Parietal?

Task: find a uniquely oriented bar

top-down vs. bottom-up selection

10^7 bits/second
~ 10^6 neurons
~ 10 spikes/neuron
~ 1 bit/spike

40 bits/second
(Sziklai, 1956)

"To be or not to be,
This is the question. .."
Information bottlenecks in the visual pathway:

**Questions:**
which brain areas are doing the bottom-up selection?

**Frontal? Parietal?**

Saliency regardless of visual features

Task: find a uniquely oriented bar

```
10^7 bits/second
~ 10^6 neurons
~10 spikes/neuron
~1 bit/spike
```

10^9 bits/second (Kelly 1962)
~ 25 frames/second,
2000x2000 pixels,
1 byte/pixel

“...to be or not to be, This is the question...”

Koch & Ullman 1985, Itti & Koch 2001, etc

**top-down vs. bottom-up selection**
The V1 Saliency Hypothesis: A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

- Retina inputs
- Saliency map
- V1 firing rates (highest at each location)

Winner-take-all
Superior colliculus
The V1 Saliency Hypothesis: A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

Retina inputs

V1 firing rates (highest at each location)

Winner-take-all
Superior colliculus
The V1 Saliency Hypothesis: A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

Retina inputs → V1 firing rates (highest at each location) → Winner-take-all

Superior colliculus

Attention auctioned here—no discrimination between your feature preferences, only spikes count!

A motion-tuned V1 cell → 1 spike

A color-tuned V1 cell → 3 spikes

An orientation-tuned V1 cell → 2 spikes

Hmm... I am feature blind anyway

Oh, no! He only cares about money

Neural activities as universal currency to bid for visual selection.
Retina inputs → V1 firing rates (highest at each location) → Winner-take-all Superior colliculus

The V1 Saliency Hypothesis:
A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

V1 isofeature suppression
The V1 Saliency Hypothesis:  
A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

Retina inputs → V1 firing rates (highest at each location) → Winner-take-all Superior colliculus

iso-feature suppression  
The V1 Saliency Hypothesis:
A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

- Iso-orientation suppression
- Iso-color suppression
- Iso-motion (direction) suppression

V1 firing rates (highest at each location)

Winner-take-all Superior colliculus
The V1 Saliency Hypothesis:
A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

V1 firing rates (highest at each location)

Winner-take-all
Superior colliculus

Left eye image
Fused perception
Right eye image

A fingerprint of V1
Escapes iso-eye-of-origin suppression

A surprising prediction: an invisible feature attract attention! (Zhaoping 2008, 2012)
Escapes iso-orientation suppression

Looking without seeing!
Replicated by multiple research groups since!
Testing the V1 theory on behaving monkeys --- Yan, Zhaoping, & Li, Submitted.

V1 neural responses to input stimulus (spikes/sec)

Saccade to an uniquely oriented bar ASAP

Receptive field

for trials with faster saccades

for trials with slower or failed saccades

Time since visual input onset (second)

Quantitative, zero-parameter, predictions from theory

Imply that higher visual areas are not involved

Koene & Zhaoping 2007
Zhaoping & Zhe 2015,

Solid curve --- Planck’s law
Squares: --- data points
Talk outline

(1) The functional role of the primary visual cortex (V1) ← Attentional selection

(2) In light of V1’s role →
   a new plan to understanding vision

(3) A first example study in this new plan
Information bottlenecks in the visual pathway:

10⁷ bits/second
~ 10⁶ neurons, 
~10 spikes/neuron
~1 bit/spike

10⁹ bits/second (Kelly 1962)
~ 25 frames/second, 
2000x2000 pixels, 
1 byte/pixel

40 bits/second (Sziklai, 1956)

“To be or not to be, 
This is the question ..”
A new path to understanding vision

Traditional paths to understanding vision

1. Low level vision, mid-level vision, high-level vision
2. David Marr: primal sketch, 2.5 d sketch, 3-d model.

40 bits/second

20 frames, 20 megabytes/second

40 bits/second (Kelly 1962) ~ 25 frames/second, 2000x2000 pixels, 1 byte/pixel

“To be or not to be, This is the question…”

(Sziklai, 1956)

Bottom-up selection

for motor action or cognitive decisions

(e.g., by retinal neural activities)

(attentional selection, often by saccade to selected location)

(e.g., face recognition)
Talk outline

(1) The functional role of the primary visual cortex (V1) ← Attentional selection

(2) In light of V1’s role →
                   a new plan to understanding vision
                   Encoding → Selection ←→ Decoding

(3) A first example study in this new plan
A new path to understanding vision

Visual inputs → Encoding → Selection → Decoding

20 frames, 20 megabytes/second
(e.g., by retinal neural activities)
(attentional selection, often by saccades to selected locations)
(e.g., face recognition)
40 bits/second

Looking
Seeing

center
periphery
A new path to understanding vision

Visual inputs → Encoding → Selection → Decoding

(e.g., by retinal neural activities)
(attentional selection, often by saccading to selected locations)
(e.g., face recognition)

20 frames, 20 megabytes/second

40 bits/second

Looking

Seeing

center

periphery

for motor action or cognitive decisions

© Jason Bradley
Looking and seeing — peripheral and central

Two separate processes

Demo: (Zhaoping & Guyader 2007)

V1: tuned to primitive bars

IT & higher areas: ‘X’ shape recognition, rotationally invariant

Add a horizontal bar or a vertical bar to each oblique bar
Looking and seeing --- peripheral and central

Two separate processes

Demo: (Zhaoping & Guyader 2007)

Add a horizontal bar or a vertical bar to each oblique bar

Add a horizontal bar or a vertical bar to each oblique bar
Display span 46x32 degrees in visual angle --- condition A
Gaze arrives at target after a few saccades

Looking --- mainly by the bottom up saliency of the unique orientation.

Then ...
Gaze dawdled around the target, then abandoned and returned.
A new path to understanding vision

Visual inputs → Encoding → Selection → Decoding

- 20 frames, 20 megabytes/second
- 40 bits/second

Looking (e.g., by retinal neural activities)
Selection (attentional selection, often by saccading to selected locations)
Decoding (e.g., face recognition)

Peripheral vision → Central vision

Must qualitatively differ in Seeing

Computational algorithms?
Focus on feedforward-feedback / V1

A demo of crowding in the periphery
Perception = ?

In V1, signals are efficiently encoded by these two de-correlated channels (Li & Atick 1994)

Subject task: report the perceived tilt.

$C_+ = C_-$

Fraction of trials seeing the summation tilt

Zhaoping 2017
Why does perception prefer ocular summation? (Zhaoping 2017)

For analysis-by-synthesis (c.f. predictive coding)

Feedforward, feedback, verify, and re-weight (FFVW)

If I perceive it, it is likely (prior) shown to both my left and my right eyes, so it should resemble the input in the sum channel!!!
Why does perception prefer ocular summation? (Zhaoping 2017)

Feedforward, feedback, verify, and re-weight (FFVW)

If I perceive it, it is likely (prior) shown to both my left and my right eyes, so it should resemble the input in the sum channel!!!

The Bayesian(?) monster

Fraction of trials seeing the summation tilt

Lower Peripheral 10 deg

n=13

Presentation duration (s)

0.1

Not because periphery cannot see tilt!

Little bias to Sum!

Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!
Why does perception prefer ocular summation? (Zhaoping 2017)

Feedforward, feedback, verify, and re-weight (FFVW)

If I perceive it, it is likely (prior) shown to both my left and my right eyes, so it should resemble the input in the sum channel!!!

The Bayesian(?) monster for analysis-by-synthesis (c.f. predictive coding)

Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!
A central disk (non-zero disparity) and a surrounding ring (zero disparity)

Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!
**Proposal:** Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!

A central disk (non-zero disparity) and a surrounding ring (zero disparity)

- Dots for the central disk correlated
- Dots for the central disk anti-correlated

Testing it in depth perception
Zhaoping & Ackermann, 2018

- No reversed depth percept
- V1’s report vetoed
- Expected disparity and binocular correlation not found in V1
- Top-down feedback to verify V1’s report
- V1 feeds forward reverse depth to higher brain areas!
Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!

A central disk (non-zero disparity) and a surrounding ring (zero disparity)

Dots for the central disk correlated

Dots for the central disk Anti-correlated

Testing it in depth perception
Zhaoping & Ackermann, 2018

No reversed depth percept

If peripheral vision has no feedback

V1 feeds forward reverse depth to higher brain areas!
Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!

A central disk (non-zero disparity) and a surrounding ring (zero disparity)

Dots for the central disk correlated

Dots for the central disk Anti-correlated

Testing it in depth perception
Zhaoping & Ackermann, 2018

Accuracy reporting disparity-defined depth

* p=0.003
* p=0.002
A new path to understanding vision

Visual inputs → Encoding → Selection → Decoding

- Visual inputs: (e.g., by retinal neural activities)
- Selection: (attentional selection, often by saccading to selected locations)
- Decoding: (e.g., face recognition)

40 bits/second for motor action or cognitive decisions

Falsifiable

Opening the window to our brain

Thanks to

**Funding source:** Gatsby Foundation, EPSRC, BBSRC.

**Collaborators:** Wu Li, Yin Yan, Ansgar Koene, Li Zhe, Joealle Ackermann, etc.