

What are deep neural networks and what are they good for?

Kendrick Kay

<http://cvnlab.net>

Center for Magnetic Resonance Research (CMRR)

University of Minnesota, Twin Cities



UNIVERSITY
OF MINNESOTA

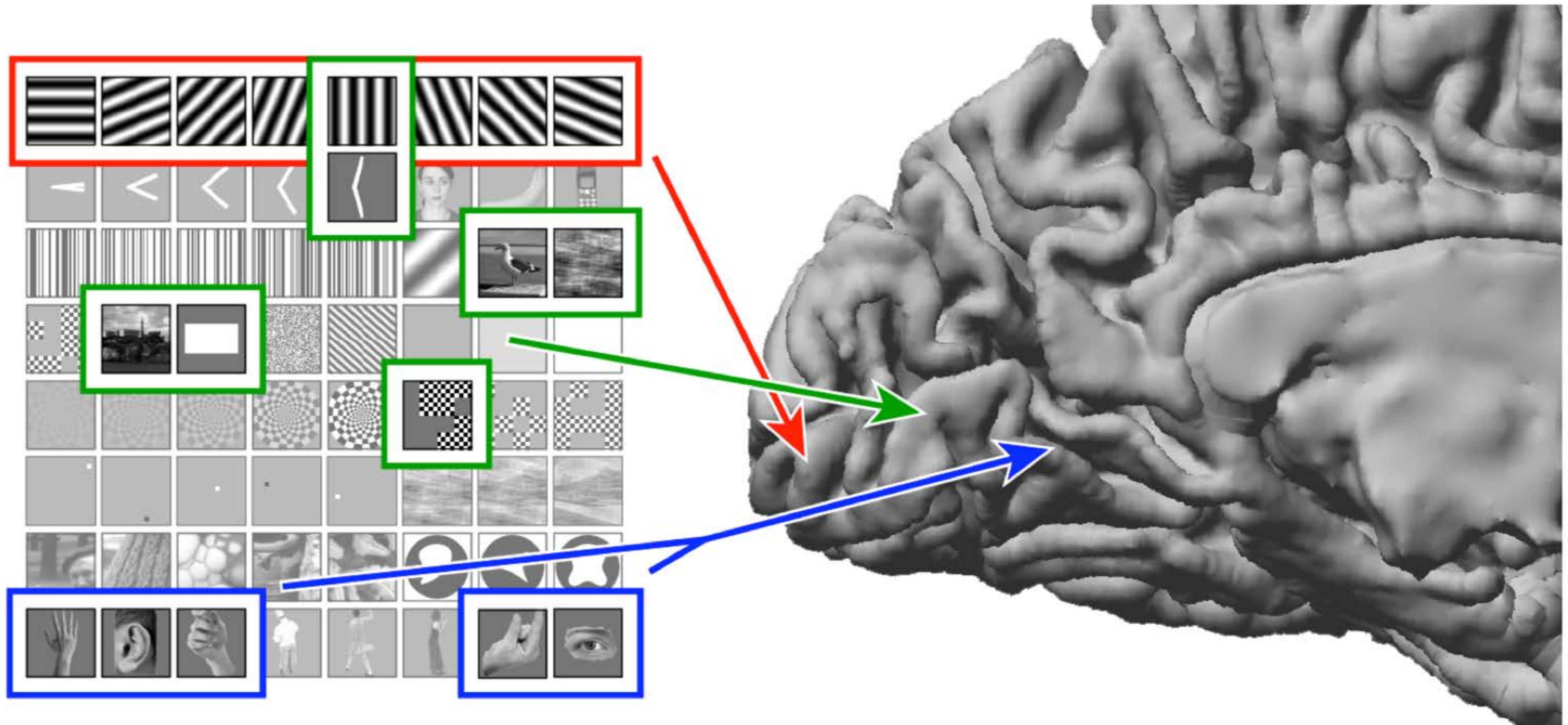
Outline

What are deep neural networks (DNNs)?

How should we evaluate DNNs?

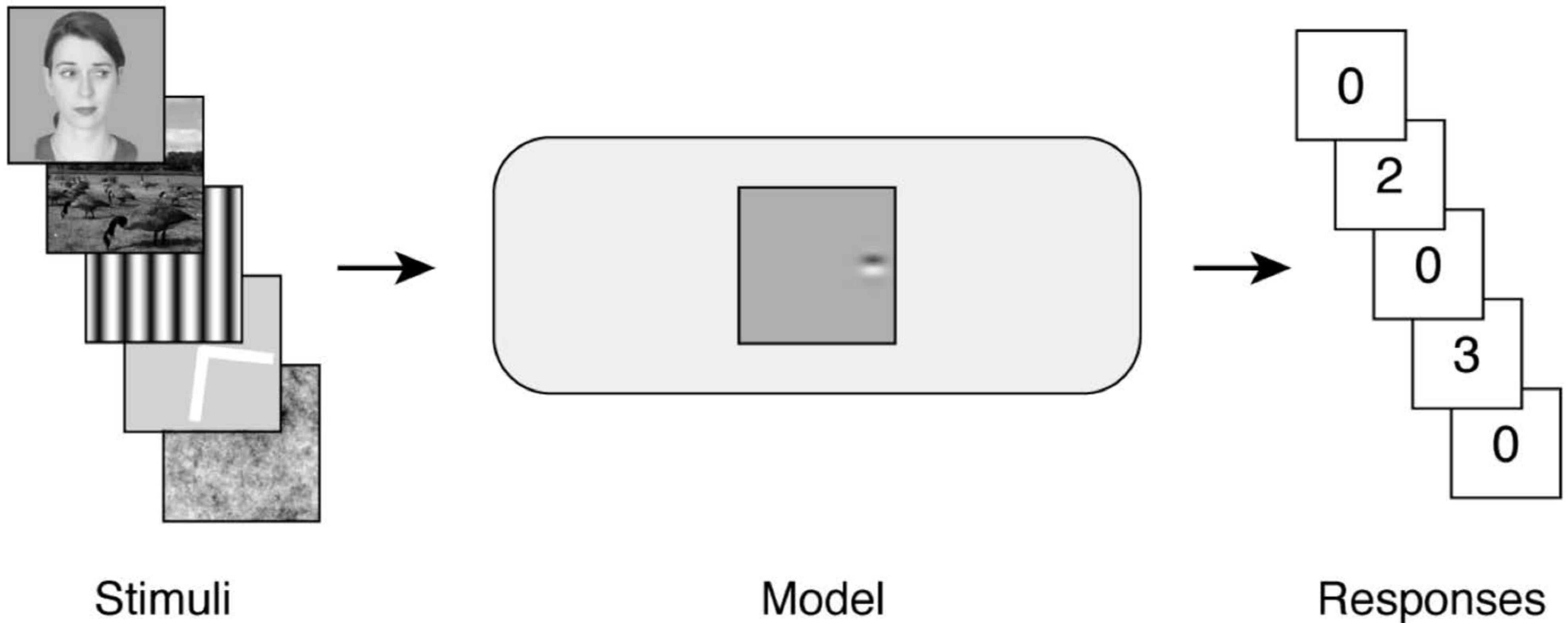
Context for DNNs

- **Observation:** visual areas show stimulus selectivity



Context for DNNs

- **Observation:** visual areas show stimulus selectivity
- **Goal:** develop image processing models
(receptive-field model, forward model, encoding model, representational model)



Context for DNNs

- **Observation:** visual areas show stimulus selectivity
- **Goal:** develop image processing models
(receptive-field model, forward model, encoding model, representational model)
- **Details:**
 - individual units or similarity matrices
 - experimental design, cross-validation, noise analysis

Brief overview of DNNs

- **History:**

- Neural networks
- Improvements in training procedures

More information:

- Yamins and DiCarlo, *Nat Neurosci*, 2016
- Kriegeskorte, *Annual Reviews*, 2015

- **Why exciting?**

- Powerful (very good performance on computer vision tasks)
- Possibly a good model of the brain?

Yamins et al., *PNAS*, 2014

Khaligh-Razavi and Kriegeskorte, *PLoS Comp Bio*, 2014

Cadiou et al., *PLoS Comp Bio*, 2014

Agrawal et al., *arXiv*, 2014

Güçlü and van Gerven, *J Neurosci*, 2015

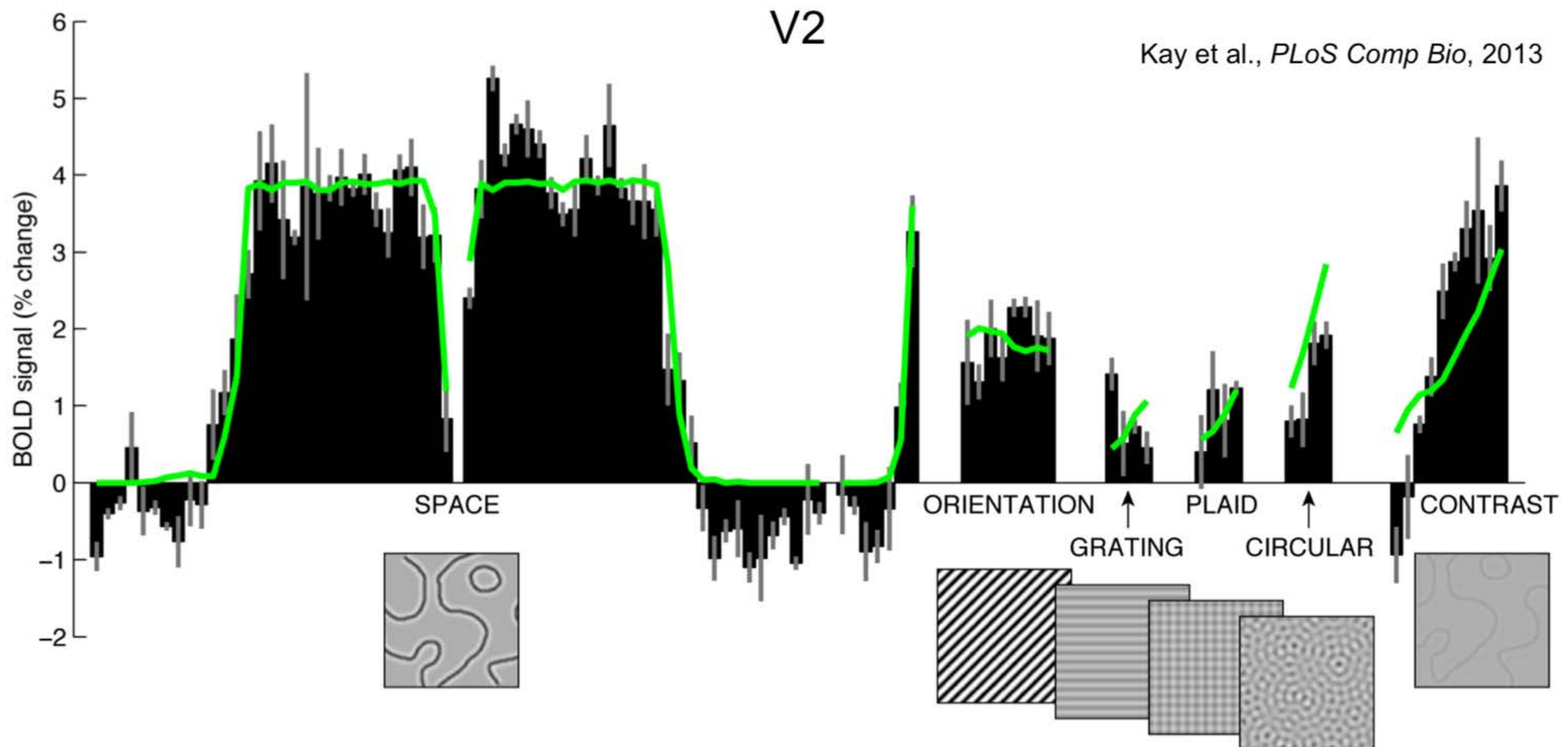
Kubilius et al., *PLoS Comp Bio*, 2016

Cichy et al., *arXiv*, 2016

What makes a good model?

- **Accuracy**

- Cross-validated predictions of experimental data



What makes a good model?

- **Accuracy**

- Cross-validated predictions of experimental data
- Caution: NOT circuits. We are just observing responses.

Inhibition, Spike Threshold, and Stimulus Selectivity in Primary Visual Cortex

Nicholas J. Priebe¹ and David Ferster^{2,*}

Neuron 57, February 28, 2008

What makes a good model?

- **Accuracy**
 - Cross-validated predictions of experimental data
 - Caution: NOT circuits. We are just observing responses.
- **Understanding**
 - "I have a model that perfectly simulates your data."



or



?

What does it mean to understand?

- **TUNING**

Do you know how the model behaves?

- **PARAMETERS**

Do you know what happens if you perturb the parameters?

- **ARCHITECTURE**

Have you done model surgery to identify important parts?

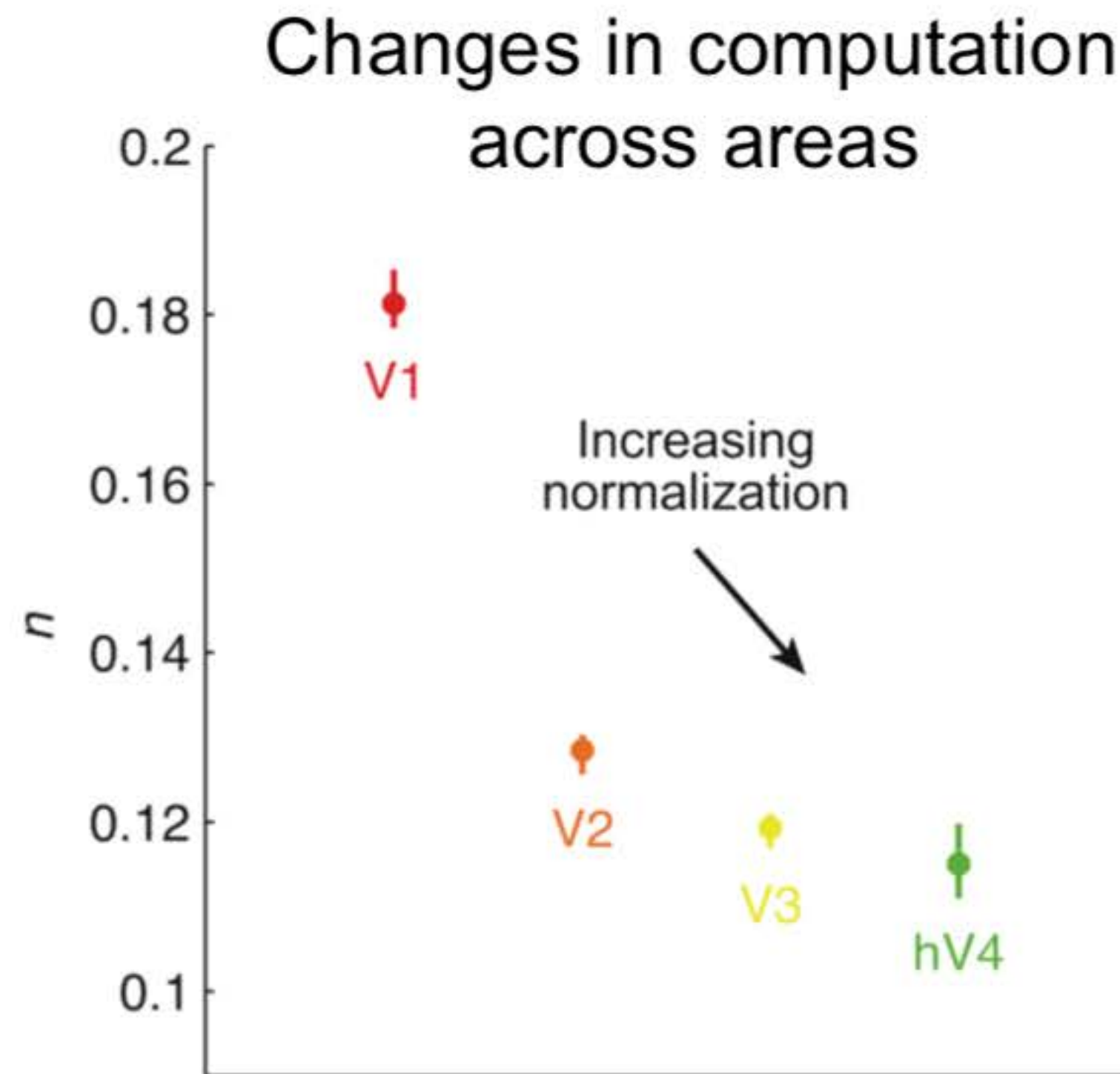
At a deeper level:

Can you predict responses in your head?

Can you implement the model from scratch?

Why do we want to understand?

- Understanding enables simplification
 - Smaller, more efficient models
 - More insightful comparison across visual areas, cognitive states, individuals, groups

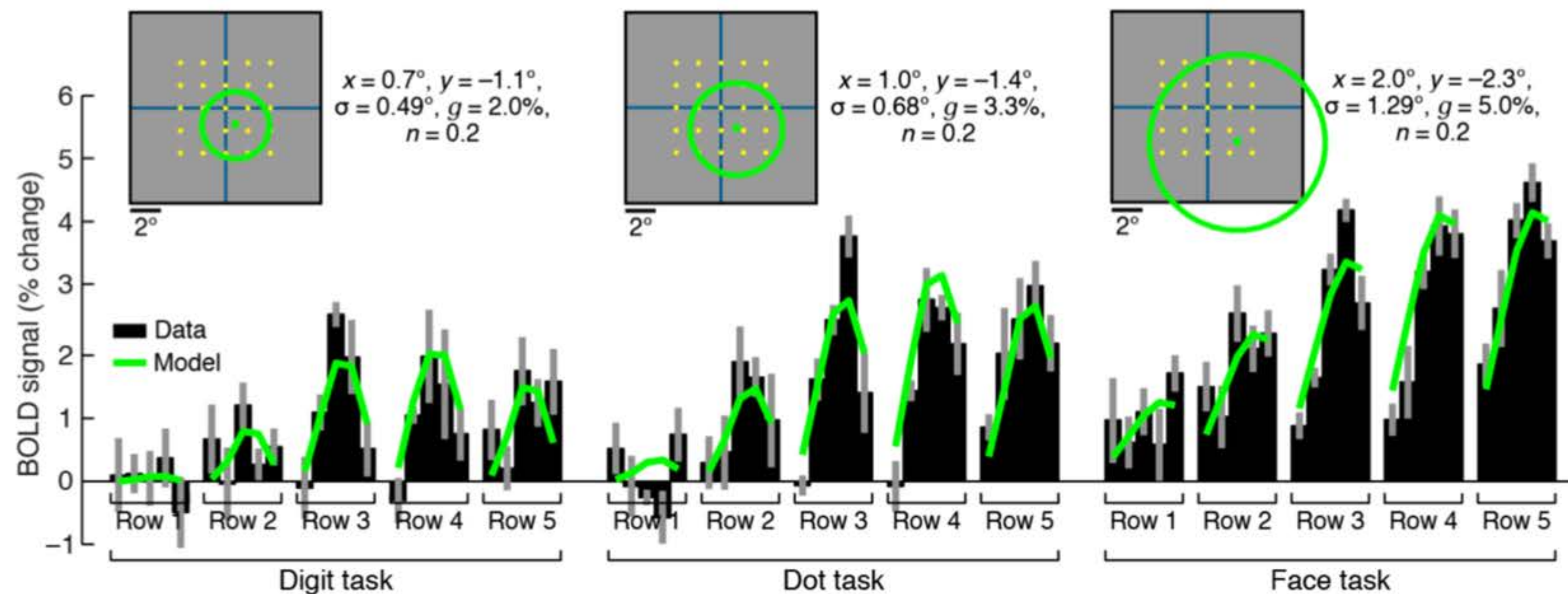


Kay et al., *PLoS Comp Bio*, 2013

Why do we want to understand?

- Understanding enables simplification
 - Smaller, more efficient models
 - More insightful comparison across visual areas, cognitive states, individuals, groups

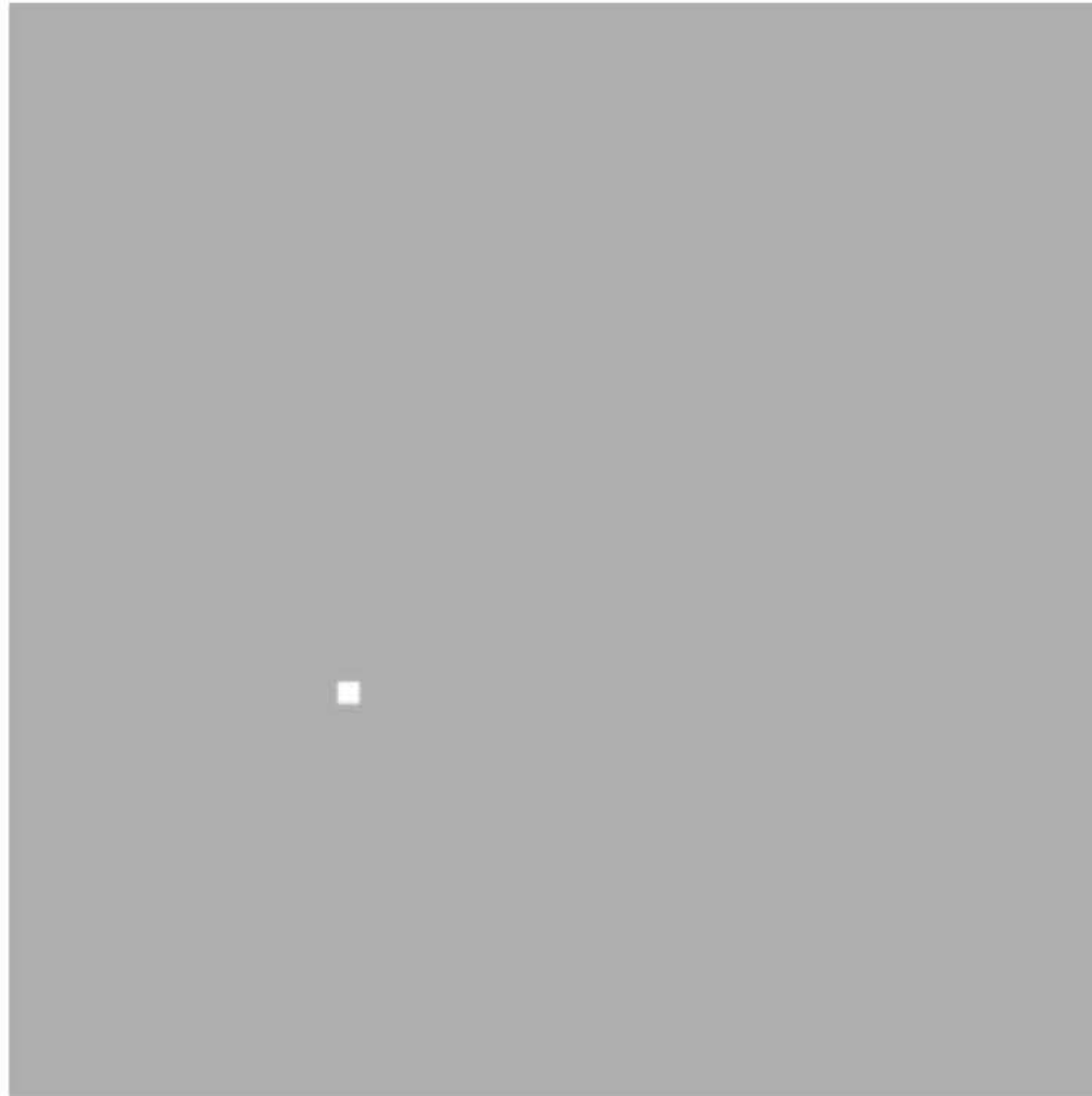
Changes in computation across cognitive states



Kay et al., *Curr Biol*, 2015

How do we understand?

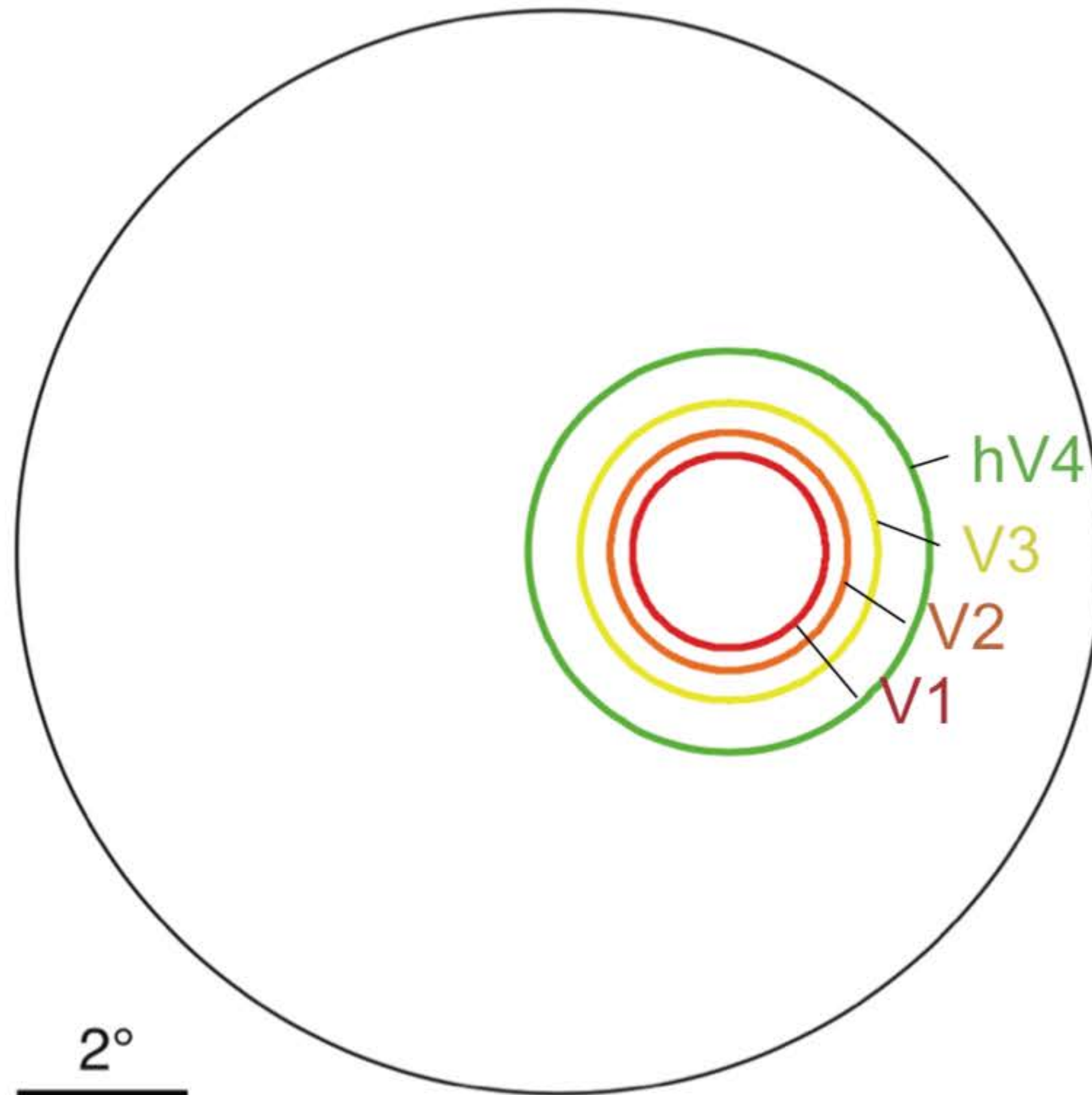
- **Observe it**



Kay et al., *PLoS Comp Bio*, 2013

How do we understand?

- **Observe it**



Kay et al., *PLoS Comp Bio*, 2013

How do we understand?

- **Observe it**
- **Manipulate it**

If we change

Architecture | Parameters

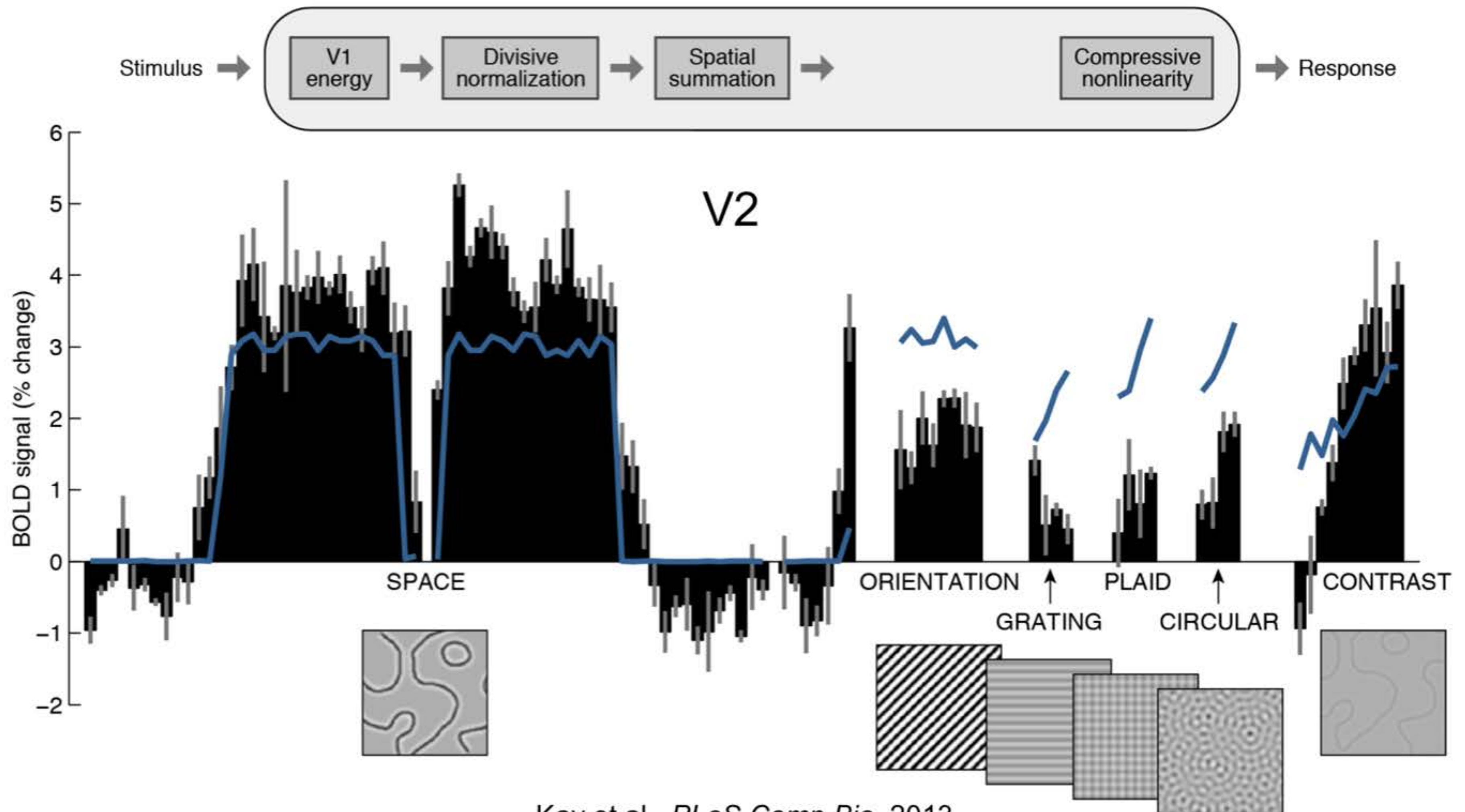


Is there an effect?

Accuracy | Tuning

How do we understand?

- Observe it
- Manipulate it

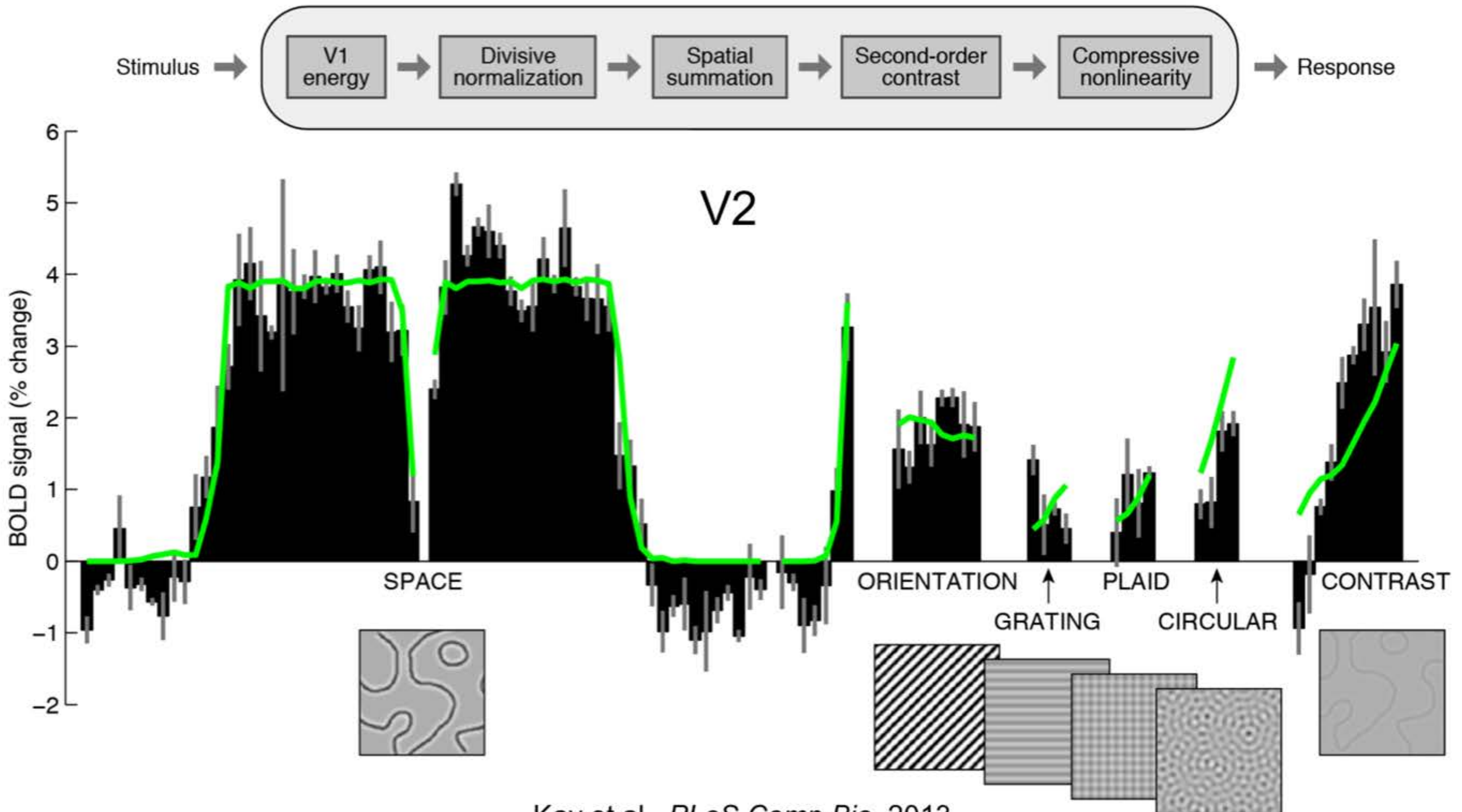


Kay et al., *PLoS Comp Bio*, 2013

How do we understand?

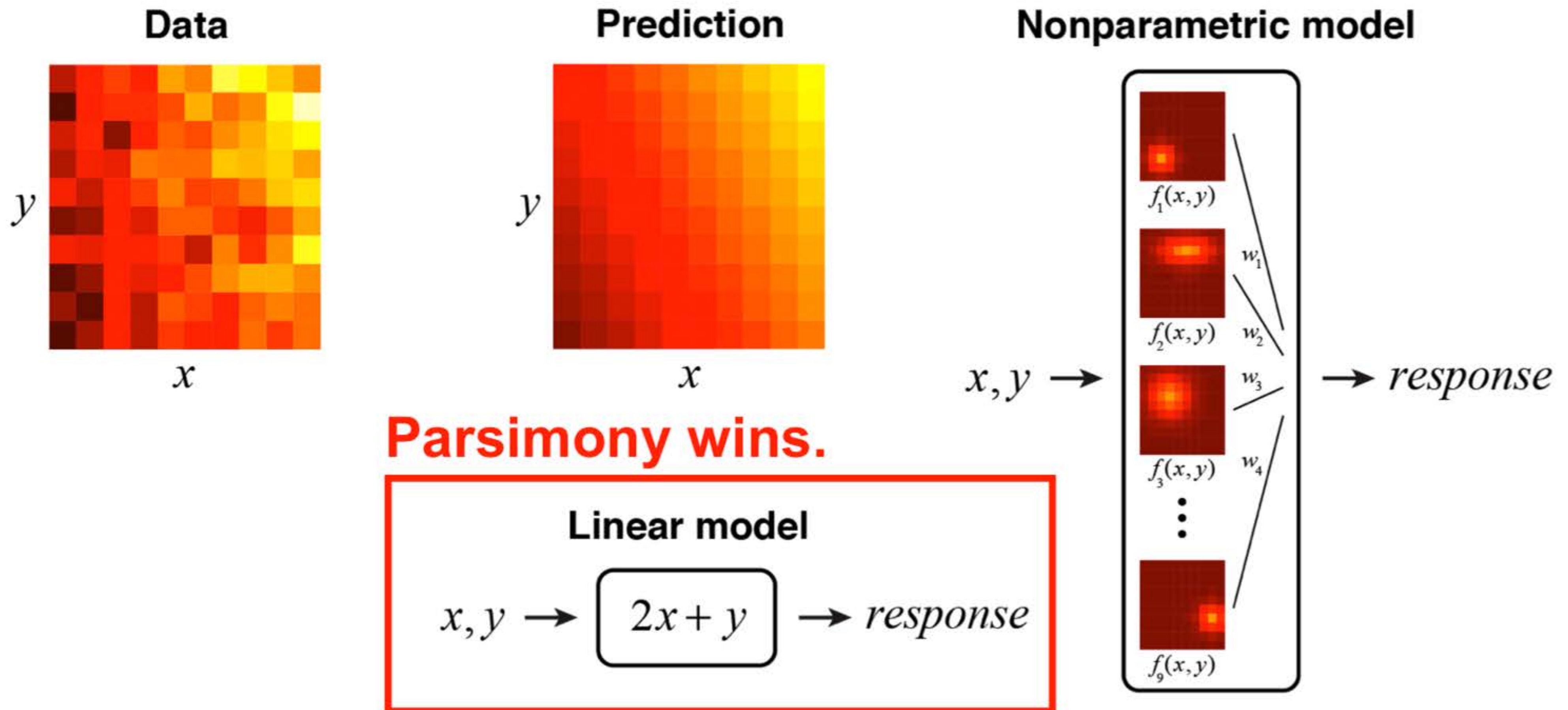
- **Observe it**
- **Manipulate it**

Also see Nishimoto and Gallant, *J Neurosci*, 2011



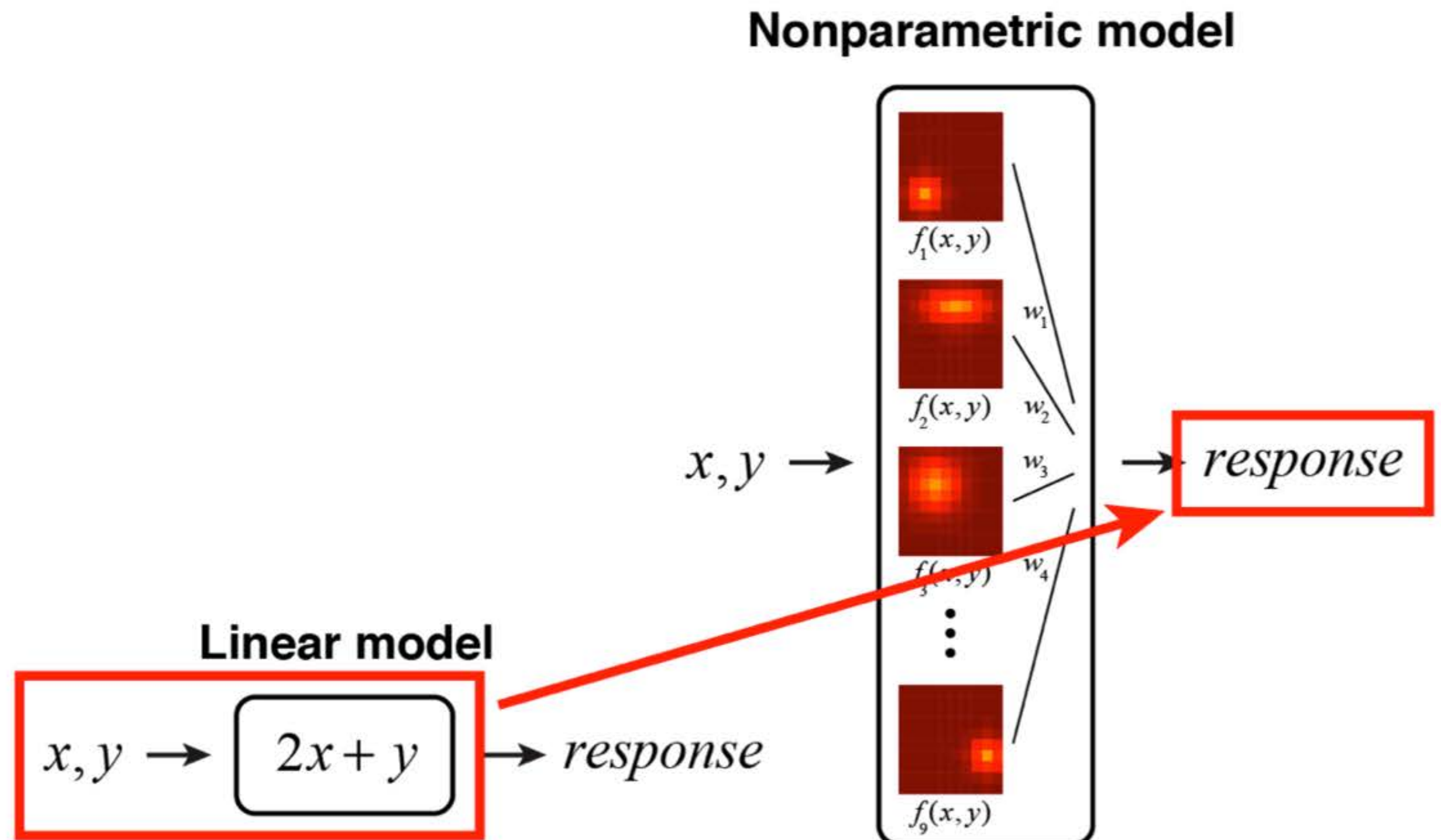
How do we understand?

- **Observe it**
- **Manipulate it**
- **Model it**



How do we understand?

- **Observe it**
- **Manipulate it**
- **Model it**



Take-home points

- Criteria:

- Accuracy ← **DNN strength**
- Understanding ← **DNN weakness**

See also:

- Sunday AM poster (33.4071)
- Sunday 8:30am talk (31.22)

- What we can do:

- Observe the model **TUNING**
- Manipulate the model **PARAMETERS, ARCHITECTURE**
- Model the model **PARSIMONY**

- Understand and simplify

- Which filter weights are actually important?
- Do you need all those layers?
- Are your effects just driven by receptive field size?

Take-home points

- Criteria:

- Accuracy ← **DNN strength**
- Understanding ← **DNN weakness**

See also:

- Sunday AM poster (33.4071)
- Sunday 8:30am talk (31.22)

- What we can do:

- Observe the model **TUNING**
- Manipulate the model **PARAMETERS, ARCHITECTURE**
- Model the model **PARSIMONY**

- Understand and simplify

In praise of shallow networks