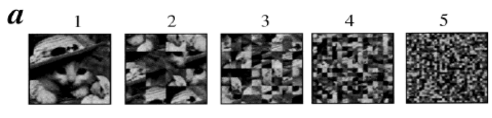
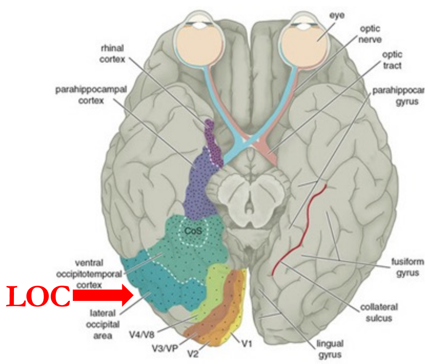
PSYC 365 Class 11:
Classifying Objects and Predicting Perception: Are you like me or not like me?

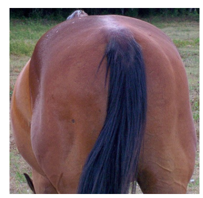
This week

- **Today**:
  - Back to Passingham Chap 2! (classifying objects)
  - Selecting attention
  - Stimulus similarity in the LOC
  - Reading: Egner, Monti & Summerfield, 2009
- **Tuesday, Feb. 25th**: Selecting and sustaining attention
  - Reading continued: Egner, Monti & Summerfield, 2009
  - Passingham Chapter 3
  - Get a head start on reading Rosenberg et al., 2016
  - In-class discussion on Thursday, Feb. 27th – submit your discussion questions before then (if you haven't already for this term!)

Review of Object Recognition in LOC







2/13/25 | PSYC 365: Class 11 | 3

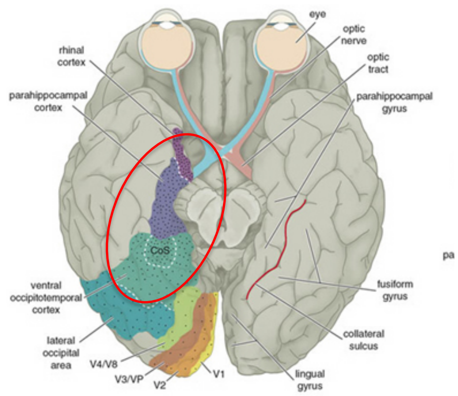
Previously, we looked at two classic fMRI studies that focused on LOC or lateral occipital complex – that patch on the outside or lateral section of the back of the brain, which comes right after primary and secondary visual cortex in the ventral visual stream. In the first study we saw that while voxels in early visual cortex showed the greatest bold response to the little bits of contrast and edge in the most scrambled images, voxels in LOC responded more to images of whole object, and regions in between responded to intermediate levels of scrambling. In the second study we saw that voxels in the anterior or forward part of the LOC showed adaptive suppression for objects of different sizes or viewpoints, indicating viewpoint invariance, where voxels in early visual cortex and other posterior regions of LOC did not.

Talk to your neighbour

- What is the difference between recognizing an object and making sense of it?
- How is classifying an object a form of making sense of it?

Introducing inferotemporal cortex (IT)

- Important for object categorization
  - Sensitive to semantic meaning



The ventral visual system categorizes objects at different levels

Superordinate	Animate	Inanimate
Basic	Face	House
Subordinate	Man	Mansion
Exemplar	President of the United States	White house

2/13/25 PSYC 365: Class 11 6

Once you get into higher level visual areas such as IT you get into populations of neurons that are selective to specific categories of objects. But within IT the pattern of categorization that neurons or voxels respond to ranges from very broad to quite specific. Grill-Spector has done a lot of research on how the ventral stream handles different levels of categorization. She has described 4 levels of categorization: Superordinate, Basic, Subordinate, Exemplar.
Superordinate: As an example, animacy is a very basic category and there are large patches of brain that are more sensitive to either animate or inanimate objects. Barack Obama, who was president in 2014 when the paper this information is from was written, is animate and the white house is inanimate. At the **basic** level, this animate object is a face and the inanimate object is a house. At the **Subordinate** level, more specifically the face is a man's face and the type of the house is a mansion. An exemplar is the individual example of the category above– in this case the man is the then president and the house is the white house.

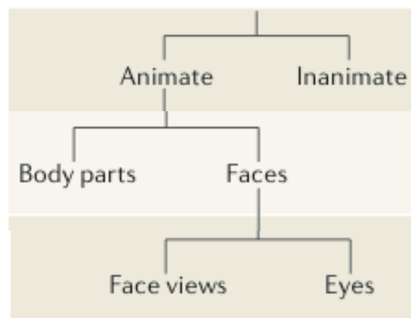
Encoding approaches: Divisions and Hierarchies in Ventral Visual Cortex (IT)

- The larger the category the bigger the amount of brain area.
- The smaller the category the smaller the amount of area – but only to a point

Superordinate > 1 cm
More abstract

Basic: Big component parts mm-cm, more concrete

Subordinate Complex Features < mm and distributed



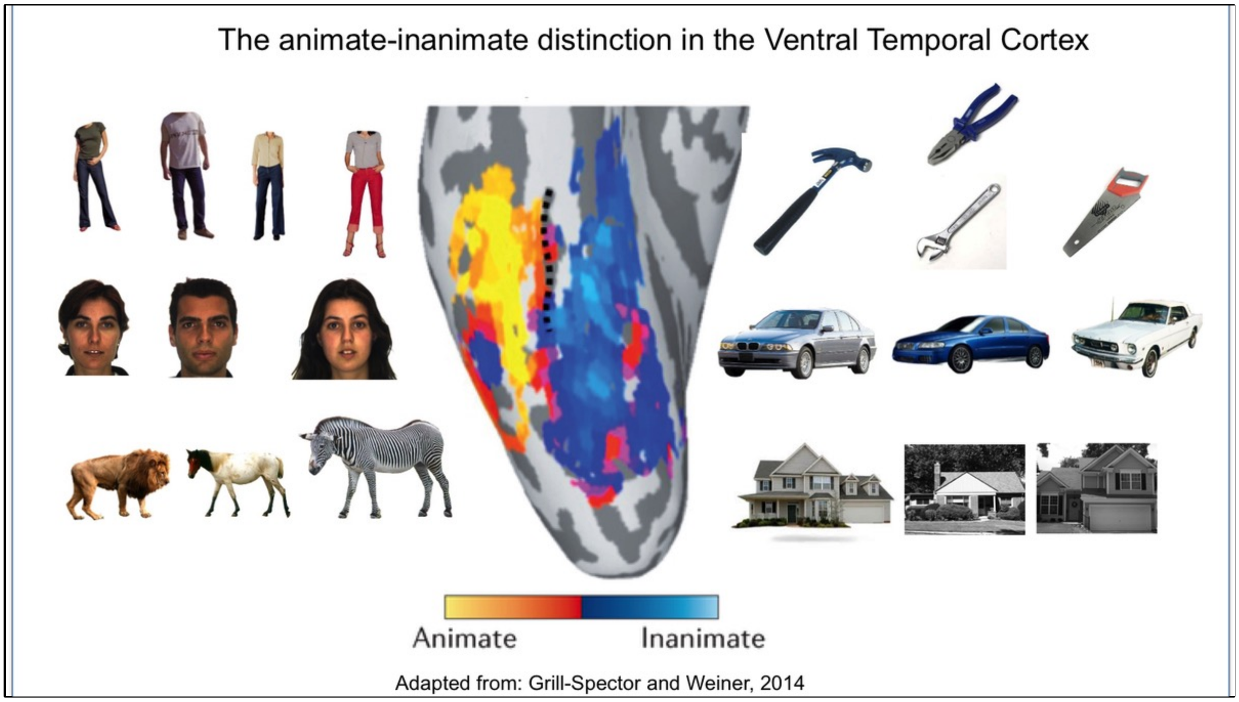
So one important area of reseach has looked at: How does the human VTC anatomically organize information and use it for efficient categorization? BBQ Ghost's work has shown us that **IT shows two patterns of categorization. In one pattern of organization** more abstract information is represented at a larger spatial scale and more-concrete information at a finer spatial scale. What's cool is that the more basic the category the more brain space – the more real estate is devoted to processing it. 1) The largest amount of space for the superordinate category of animate vs inanimate. As I said, the animate vs. inanimate distinction is a biggie for us. **Q: Why might that be?** In fact the ventral stream splits into 2 different pathways, one that prefers animate objects and one that prefers animate objects, pretty early on. 2) Next basic categories take up a bit less area – mm to cm. 3) Finally if it's a smaller feature, like the eyes of a face, it takes up a much smaller portion of real estate than processing an entire face, which in turn takes up less real estate than processing animate vs. inanimate objects.
4) So the smaller the category the smaller the amount of brain area devoted to it-- but this is true only to a point. In the study I'm about to describe, the authors point out that for categories more fine-grained than this it's unlikely that there will be identifiable anatomical landmarks. When we talked about decoding vs. encoding fMRI approaches I told you that brain activity that represents specific

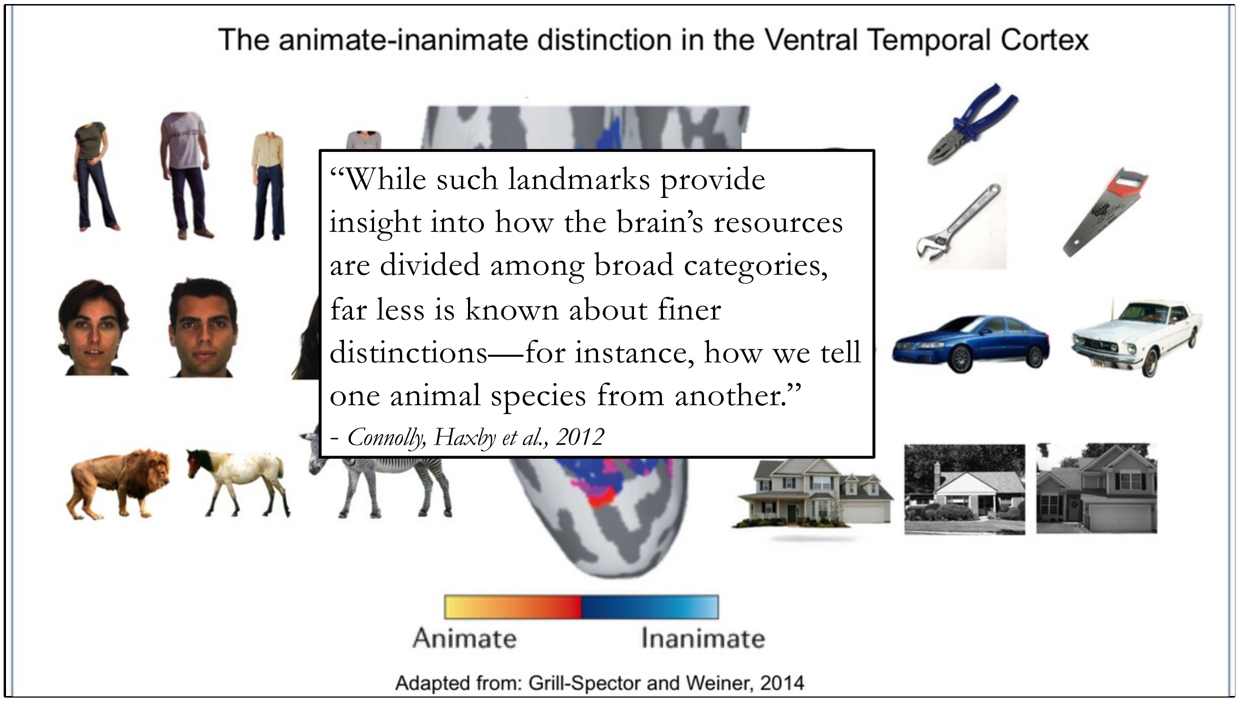
exemplars of faces or objects, for example, is more likely to be seen in the pattern of BOLD activity across voxels, not in the voxels of a particular area being more active for one than another. So beyond measuring broad geographical boundaries, for fine-graned categories you need a more sensitive measure of how the brain discriminates – how it represents fine-grained categories. In the next video we're going to enter the world of representational methods – of MVPA and RSA.

Classifying objects

"It is not enough simply to recognize what we see; we also have to make sense of it. This involves classifying things, whether animate or inanimate…"

-Passingham, Chapter 2


The animate-inanimate distinction in the Ventral Temporal Cortex
Animate
Inanimate
Adapted from: Grill-Spector and Weiner, 2014


The animate-inanimate distinction in the Ventral Temporal Cortex
"While such landmarks provide insight into how the brain's resources are divided among broad categories, far less is known about finer distinctions—for instance, how we tell one animal species from another."
- Connolly, Haxby et al., 2012
Animate
Inanimate
Adapted from: Grill-Spector and Weiner, 2014

Categorizing bugs, birds & mammals

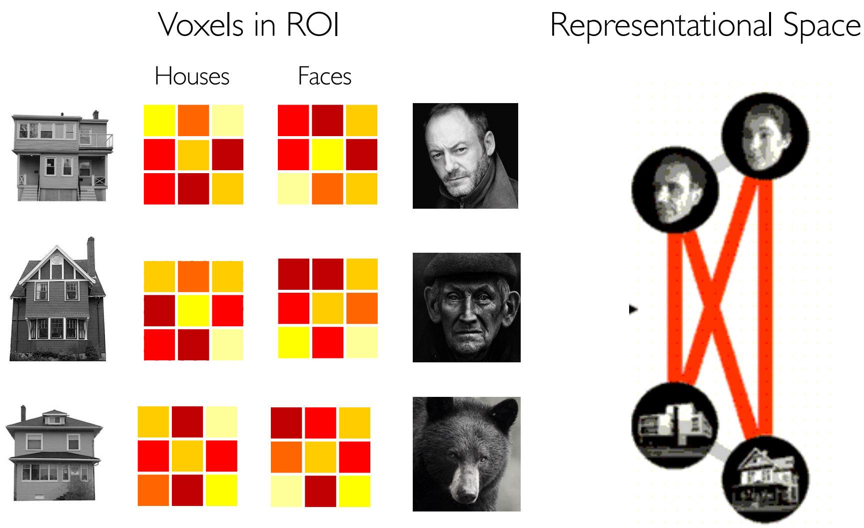
- How are finer grained categorical distinctions (subordinate classes) represented in the brain?
- Previous research: *Functional landmarks* for broad categories
  - e.g., animate vs. inanimate
- Less known about finer distinctions
  - e. g., different classes of animal
  - Goal: Look at patterns across voxels (RSA) to look for *similarity structure* sat the level of *biological class*

Connolly, Haxby et al., 2012

12

So the big-picture question guiding the Connolly & Haxby study that Passingham describes was [1]? As I've described, previous research using the encoding approach found that for broad categories such as animate vs. inanimate, Faces vs body parts, animals vs.artifacts you can find functional landmarks. That is, big hunks of real estate devoted to one broad category or another. [2] But before this study less was known about how the brain categorizes finer distinctions But at the level of finer grained distinctions its unlikely we would see a region specific for squirrels and another for raccoons. MAYBE though we can see these discriminations in the pattern across voxels in the ventral stream! This is where we turn to RSA. Before this study was done RSA had been used on monkeys to reveal a *finer-grained pattern* that reflected the *biological relationships* present among animal species. In this study, animal species were chosen to represent a simple natural hierarchy comprising three biological classes - insects, birds, and primates - as well as a higher-level grouping of invertebrates and warm-blooded vertebrates.

RSA: Decoding the representational space of object categorization



Redder voxel = More active voxel

13

Representational similarity analysis, or RSA, can be used for decoding the representational space of object categorizatoin. We decode this representational space by looking at the pattern of activation across voxels in the ventral stream. To do this we can look at the patterns of higher vs. lower activation across voxels for a single stimulus. [click] How similar or different is the pattern of voxels, illustrated here in the houses column next to this house, for this house to the pattern of voxels, illustrated here in the faces column, for this face. [click] How similar or different is the pattern for these to this house? [click] or this house? [click] or this face? [click] or this face? [clck' Then, based on whether it is more similar or more different from that pattern of activity for other face vs. house stimuli, we can read out the distance between one house and another house (short distances, which suggest stimuli belong to the same category and are close in representational space are illustrated by grey lines), or one face and another face, or one face and one house (long distances, illustrated by red lines, suggest stimuli belong to different categories and are distant in representational space)? The stimuli that cluster together in this representational space can then be said to belong to the same category. The stimuli that cluster far from each other can be said to belong in different categories. SO THE MORE ALIKE THE PATTERNS OF VOXEL, THE CLOSER IN

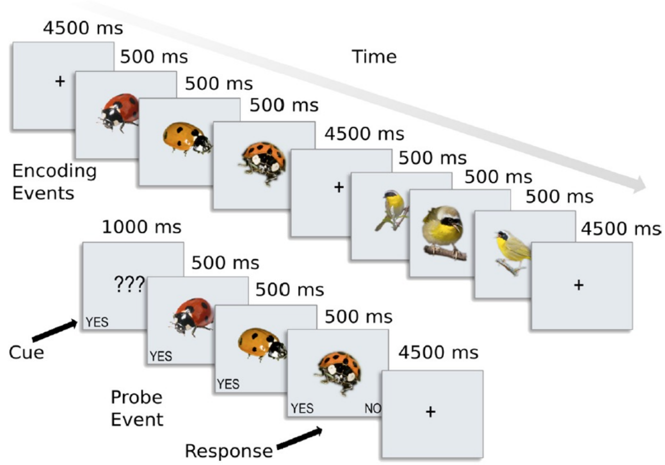
REPRESENTATIONAL SPACE, AND THE MORE CLOSELY CATEGORIZED AS THE SAME VS. DIFFERENT. But beyond telling us whether stimuli are in the same or different categories, RSA can RSA tells us how similar or dissimilar patterns of voxels are for one stimulus relative to another. So we can look at the degree to which two objects differ in representational space. The next slide is just a visual illustration of the concept of representational space.

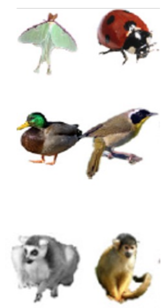
Representational similarity space

What is it?

fMRI Experimental Design

- Simple Recognition Memory Task
- Rate stimulus similarity





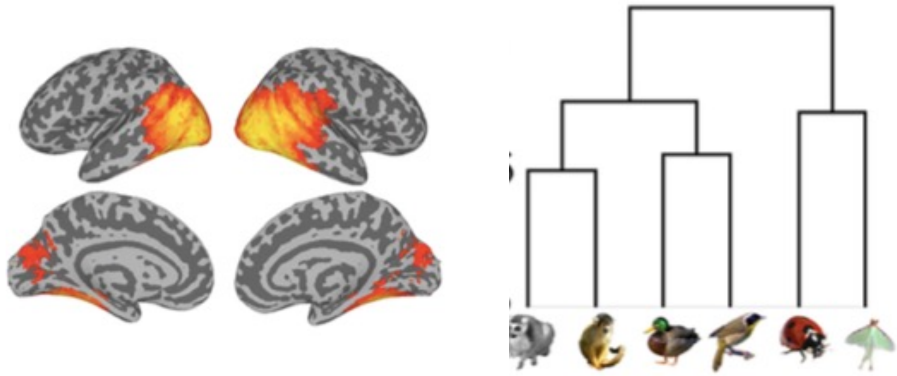
Connolly, Haxby et al., 2012

15

I will now give you more detail about the design of the study by Connolly and colleagues described in the classifying objects section of Chapter 2 of the textbook. 12 young adults participated in an fMRI study. As they lay there in the scanner in each trial they'd see 6 different series of 3 pictures of animals of the same class of animal -- either bugs or birds or primates. Then they would be shown one of those 6 series of 3 stimuli and would have to say if it was the same or different from what they saw in the previous trial. This probe question was just to keep the participants paying attention -- the experimenters didn't really care about how well they remembered the items. Here are some examples of the stimuli which were either bugs, birds & primates. Afterwards they had participants rate how similar all of the different stimuli were to each other and also do a task where they'd see three and choose the one that wasn't categorically like the others. In this case they did care about the participants' choices because they wanted to know how similar or different they thought the different stimuli were from each other.

Stimulus similarity

RSA Results: Ventral Stream (LOC & IT)



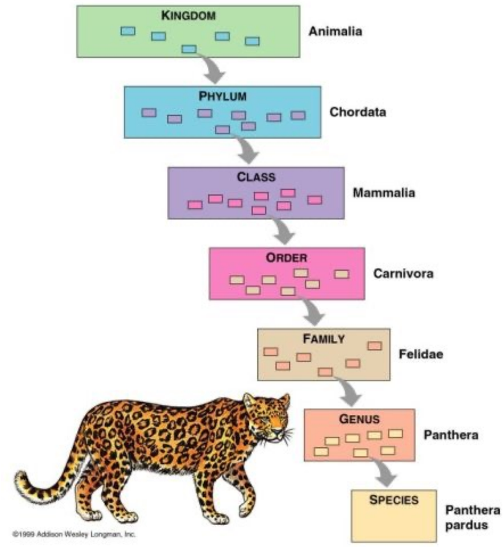
- Represents stimuli in a way that matches behavioural similarity ratings!! Maps onto biological class structure!

17

Now we are looking at a pattern represented by voxels in a large region that includes anterior LOC and inferotemporal cortex. [1] Now the lady bug and moth are very similar.   And the duck and the bird are very similar. And the monkey and monkey are also very similar. So here we see the semantic categories neatly reflected in the tree, with monkeys next to monkeys, birds next to birds, and bugs next to bugs. And when participants made judgments of the similarity their choices looked just like this, so the ventral stream activity reflected peoples' performance.   [2] It mapped onto biological class structure. Cold blooded invertebrates were furthest from warm blooded mammals. The authors of the study discussed about how this reflects categories learned in childhood, and they also discuss how LOC patterns depend on hierarchical organization of lower-level neurons in early visual cortex, as I illustrated earlier in the lecture.

Conclusion

Human neuroimaging reveals a "hierarchical category structure that mirrors biological class structure" in the ventral visual stream.

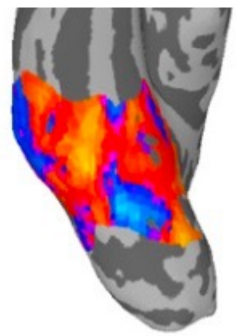


18

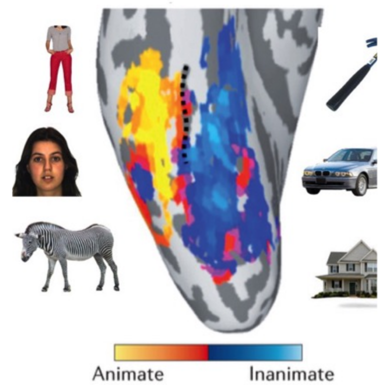
First human neuroimaging study to document category structure that reflects biological relations among species. Here is a textbook illustration -- Kingdom, phylum, class, order, etc. So our brains do it the way the textbooks do it (or vice versa). The authors conclude [1].

Using encoding for brain mapping in lateral occipital cortex (LOC)

- How does abstract representation of continuum from bug to primate map literally onto brain space?
- Does the map for primates vs. bugs looks like animate vs. inanimate?
- Lateral-medial organization



Primates vs. Bugs



19

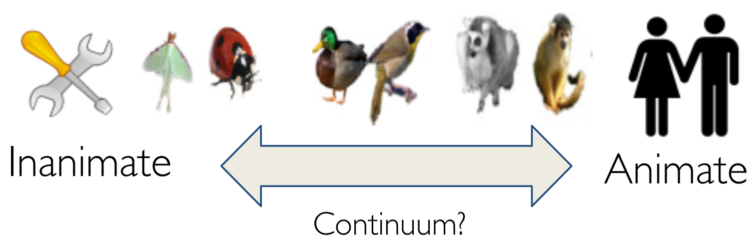
But the authors of this study weren't quite done. They had a further, more exploratory question: How does abstract representation of the continuum from but to primate map literally onto the landscape of the brain. How does this way of categorizing objects, which was based on the pattern across voxels across a big area of the ventral stream, map onto actual regions on the real estate of the brain? To examine this they went back and used an encoding approach. To remind you, in an encoding approach you compare 2 conditions and look for voxels where there’s a difference. [1] Here they compared BOLD response to primates with those of bugs and looked at the brain map of voxels that showed more activation to one than the other. They then compared the results to those from older older studies of brain mapping comparing animate with inanimate objects. Here blue blobs (more medial) illustrate greater activation for bugs, orange (more lateral) for primates. What they found was that the brain map for category differences between primates and bugs looked like the map dividing responses to animate vs. inanimate objects, with [2] a lateral-medial organization with animate objects/primates more lateral and inanimate objects/bugs more medial. The created a lateralization index and tested the accuracy of this claim that the animate to inanimate dimension was reflected in the cortical response to animal classes a number of ways.

The brain maps categories from inanimate to animate

- Animal categories are represented in LOC ranging from medial to lateral
- But this dimension only one of many ways of representing objects!

Using behavioral judgments as a target, we found semantic structure to be reflected strongly throughout the LOC

- Connolly, Haxby et al., 2012



20

[1]. In regions of the ventral stream, animal categories are represented from lateral to medial, from more animate to less animate. The "least animate" objects in this study—the bugs—pulled brain activity similar to that evoked by artifacts, whereas the most human-like—the primates— showed activity similar to that evoked by animate stimuli in previous studies. So more medial activations reflected more inanimate and more lateral more animate. This pattern was really reliably observed in all participants and accounted for a lot of the variance in LOC activation. However, it's important to point out this was not the only information the brain was using to classify these animals. It was using other information as well. so [2]The authors suggest that the animate–inanimate distinction in human cortex may also reflect a graded dimension among animate categories. We see animals that are more like us as more animate. We like bears and dogs more than bugs and reptiles. Monkeys more likely to: Have a mind, initiate actions, be aware of surroundings, have special rights. The authors also point out that the Animate vs.

Inanimate distinction is

- Formed during early development (Rakison & Poulin-Dubois, 2001)
- Kept into late dementia (Hodges et al., 1995)

Things we have learned from fMRI

- *Encoding Approach*: Responses to object parts and then whole objects along moving along occipital cortex from EV to LOC
- *Encoding Approach*: More invariant processing and more category specificity as we move along ventral stream.
- *Decoding approach:* shows continuum of finer-grained categories.
  - In the ventral stream it matches semantic judgments of animal class
  - In LOC it is organized along a spectrum of inanimate (not like me!) to very animate (a lot like me!)

21

So this just recaps what we previously learned from the Grill-Spector studies. We now also know [click] That as we go down the ventral stream we see a finer-grained pattern of category along an animate to inanimate spectrum represented in the pattern across voxels

Preview: predicting perception

2/13/25 PSYC 365: Class 11 22

End at 8:50 when we agree about our controlled hallucinations we call it reality

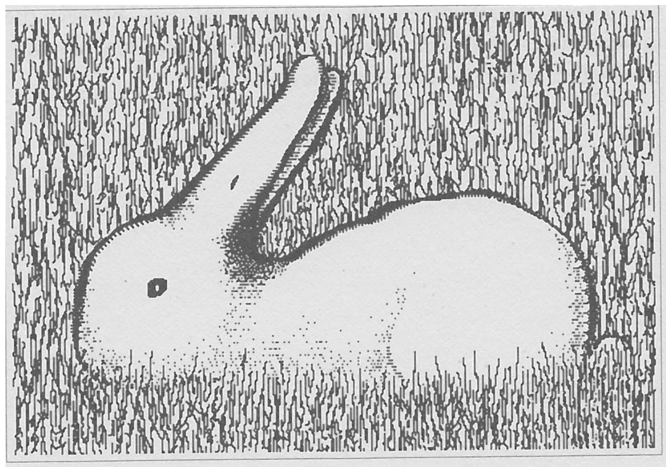
Learning Objectives: Predictive Coding

- Describe predictive coding models of perception
- Evaluate fMRI evidence from Egner et al. for predictive coding vs. bottom-up feature detection models
- Discuss what predictive coding views mean for our understanding of how our brains work and the grasp we can have on reality

How much of our perception is hallucination?

Now listen to this [play] Q: What do you think Grover is saying? NO he did not. He actually said "Yes, yes, that sounds like an excellent idea." Now listen to him again. [play]

How much of our perception is hallucination?



13/2/25 PSYC 365: Class 12 25

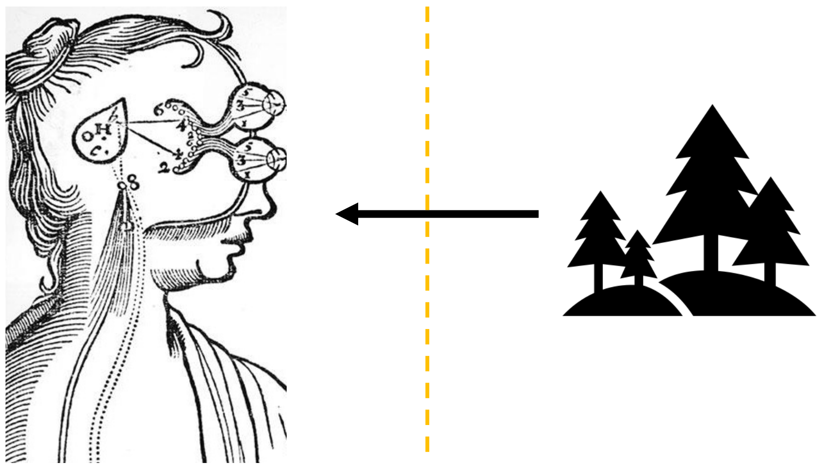
When you look at this picture, what do you see at first, a duck or a rabbit? What do you see now? Many of you have probably seen this before. It is used to illustrate the idea that we project our experience and expectations particularly when stimuli are ambiguous - when they could be one thing or another thing. We've been talking about feedforward volleys of information, and about perception as if our brains are fairly passively taking in what is out there in the world But in fact it is not just what is out there in the world but also what goes on inside our brains that determines what we see. In fact in in his TED talk Anil Seth suggested that perception is fundamentally hallucination.

Testing models in our head against sensations from the world

Thanks to Madeleine Ransom

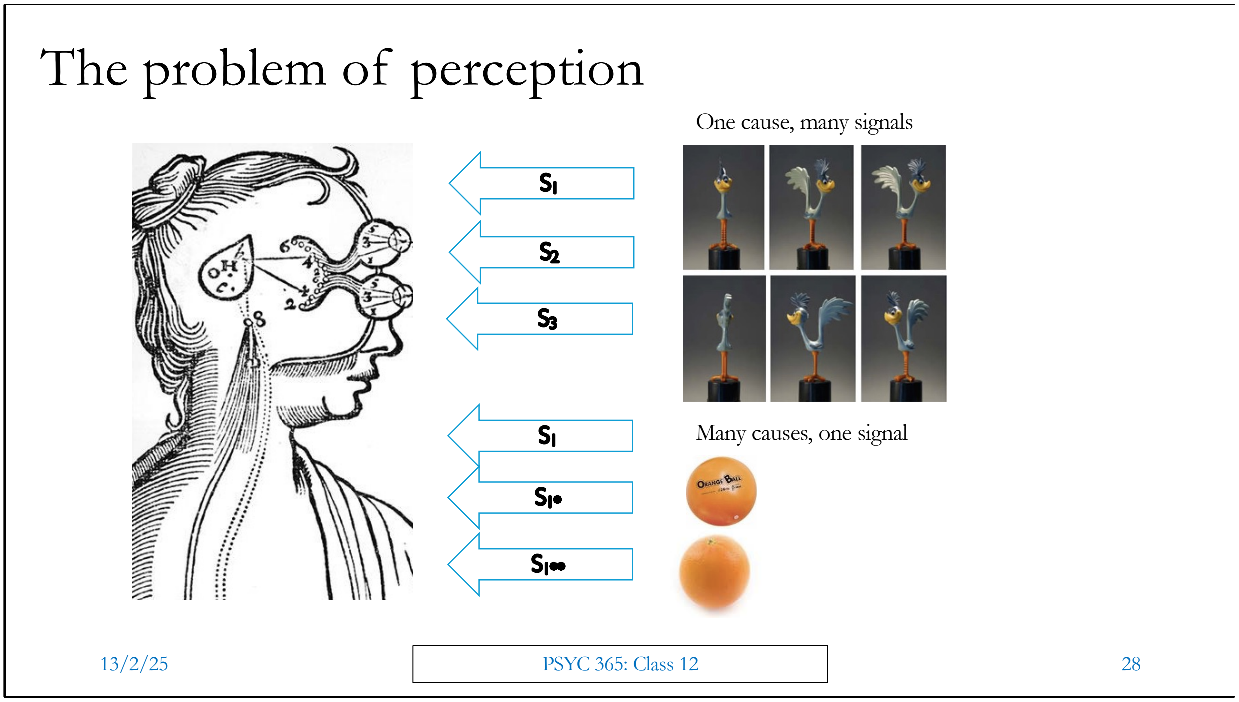
We've been talking a lot about bottom up processes by which we piece together information coming in from the world to make sense of it. But there are a lot of effects of experience those models don't account for. And they propose a view of the world that is quite passive. The world is there and we quite passively receive information through our senses. But there are many challenges to our sensory systems -- some of which we've described. For example, sometimes information is ambiguous. It could as easily be one thing as another. Humans have experience that allow them to make good guesses. Predictive coding models stress the idea that we don't have direct access to the world. We only get the signals that are interpreted by our sensory systems. And sometimes those could be interpreted in any number of ways. So we have these internal models or hypotheses about the way the world is -- these controlled hallucinations -- and we have to guess or predict what those sensory signals are given all the information that we have. We have to have what Bayesian statisticians call priors: Hypotheses about how likely things are in general and how likely they are to be true in the current situation. The Egner paper tests hypotheses from this view against the bottom-up feature-based models we've been discussing. But first let's review the basic ideas behind predictive coding models that Anil Seth introduced.

The problem of perception



13/2/25 PSYC 365: Class 12 27

Outside our heads, there is a world of things. But we don’t have direct access to these things [click], only their effects on our senses [click]. Light getting transduced into patterns on the retina that travel down the optic nerve. Sound waves that vibrate the hairs in our inner ears and get translated into frequencies we perceive as sound. And sometimes those are not exactly straightforward.



Complicating this picture is that there's not a one to one relation between causes in the world and their effects on our senses. One cause can produce many different effects, or signals. For example, the same object can look very different from viewpoints, as we've discussed, and as we see in the the image of birds on the upper left.. [click] At the same time different sources can produce almost the the same signal. such as this orange ball and this orange. How then, are we able to perceive these hidden causes out in the world? They are mysteries our brains need to solve.

Review: Feedforward Feature Based Models

- Signal or information gets passed forward from one node to the next
- What we call "Bottom-up" processing involves mostly feed-forward processes as information gets passed on from V1 forward
- Populations of neurons respond to features of objects at an increasingly large scale and higher levels of abstraction

VENTRAL STREAM

V1 → **V2** … → **LOC** … → **IT**

But here's the thing…

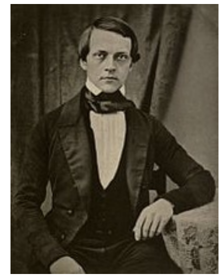
- Patterns of lower-level features detected can be interpreted in multiple different ways
- Need some way to decide between hypotheses: why choose H1 over H2 etc., since both account for the lower-level features equally well?
- Need additional, "top-down" information

[click]. For example, this photo of a dark lumpy shadow in the trees. It might be a tree stump. But it might be a bear. [click]. "Top-down" information refers to information your brain generates and applies to the world. It is your expectation or controlled hallucination. What are you expecting to see? Is it a bear or is it a tree stump? It's important to know!

Predictive coding models

Treatise on Physiological Optics, vol. 3 (von Helmhotz 1867)

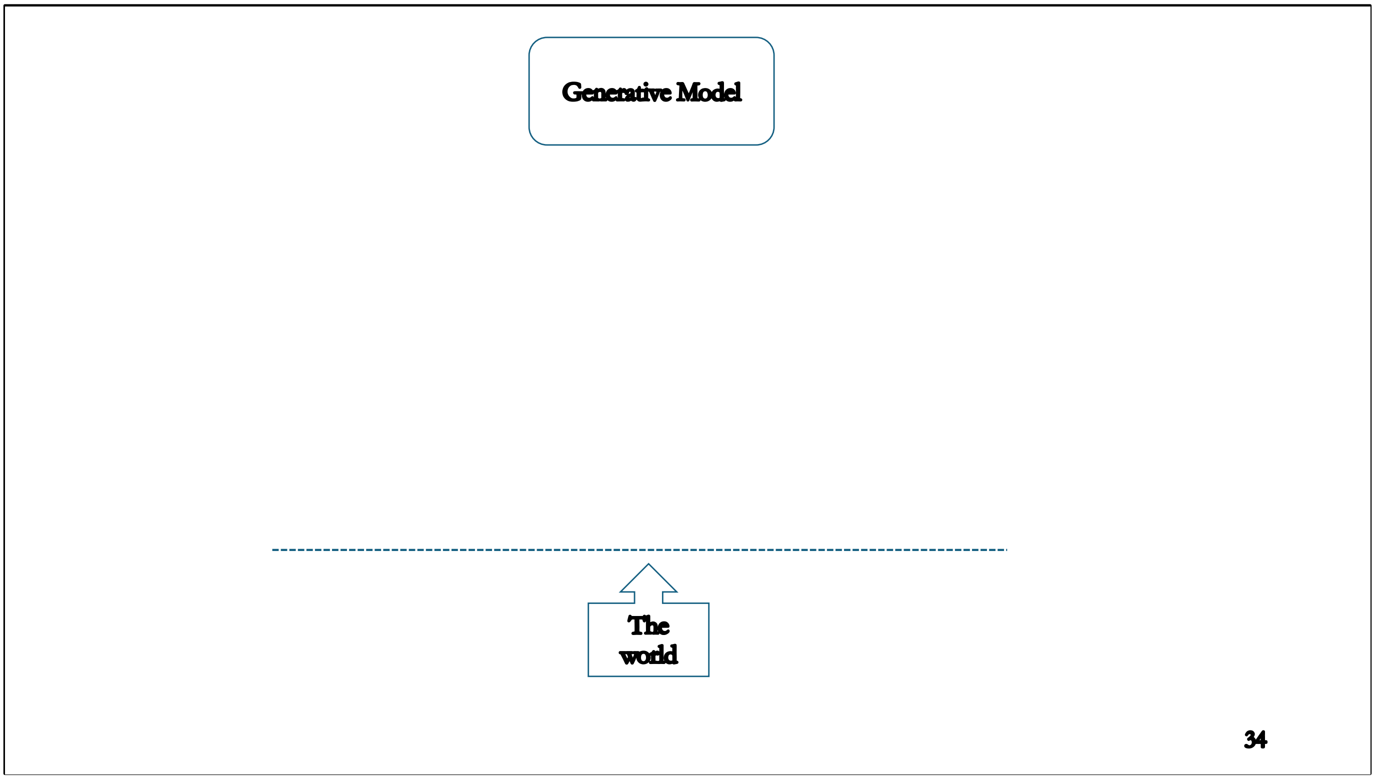
- The brain is a prediction machine
- Perception is just unconscious inference
  - we infer the cause of a sensation in the world via its effects
- Inference: idea or conclusion from evidence or reasoning
  - It's an educated guess



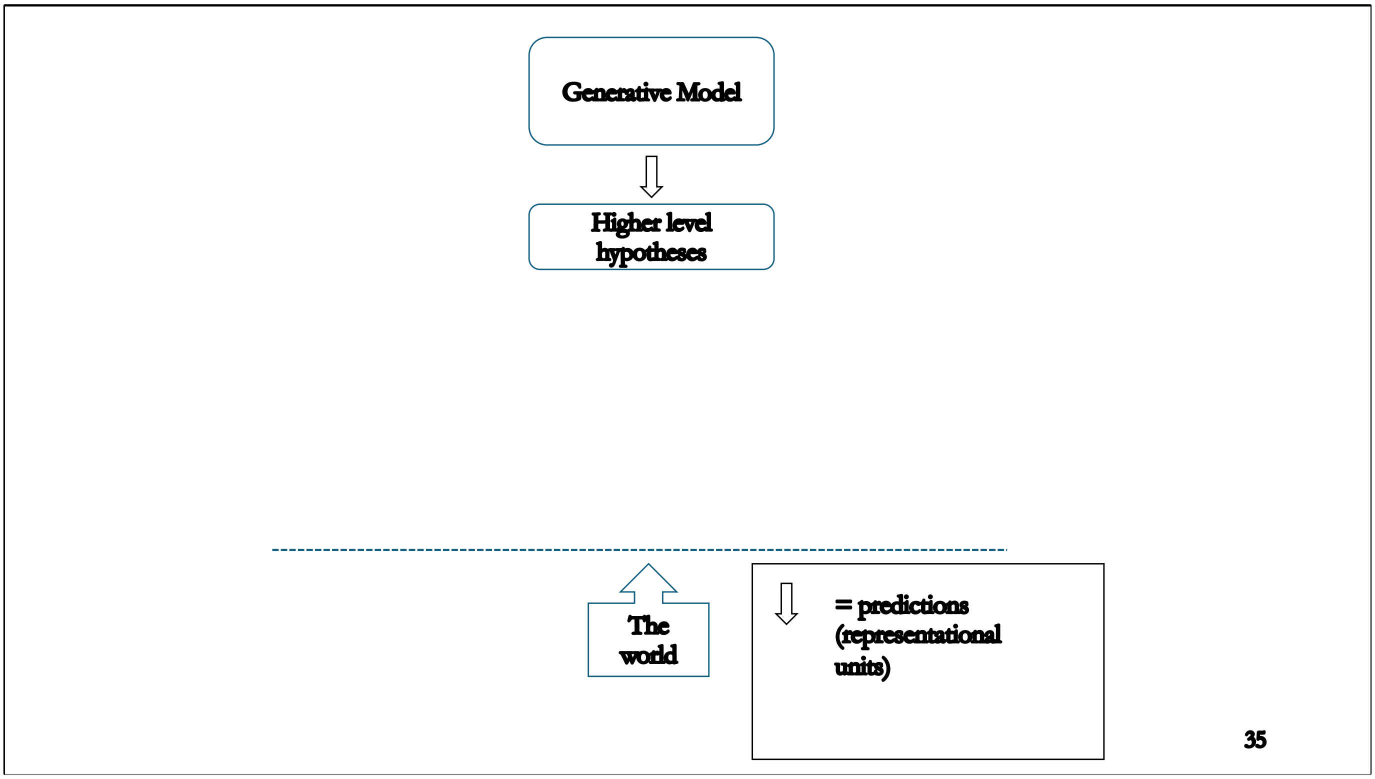
Well this is where PC models come in. And this idea is not new. von Helmholtz, top right here, famously viewed perception as the generation of a best guess (i.e., inference) about the state of the world, in view of the data. [1] [2] [3] In other words, the brain creates an internal generative model of the world that embodies a prediction of what will be observed next. We have these templates or models that reflect our expectations of what we're seeing and we project those down from – say our prefrontal cortices to our visual cortices. And what comes in from the world is the error between what we expect and the information we get. That is information from the world concerns whether our prediction is right or wrong. I'll illustrate this in more detail in the next slides

Questions?

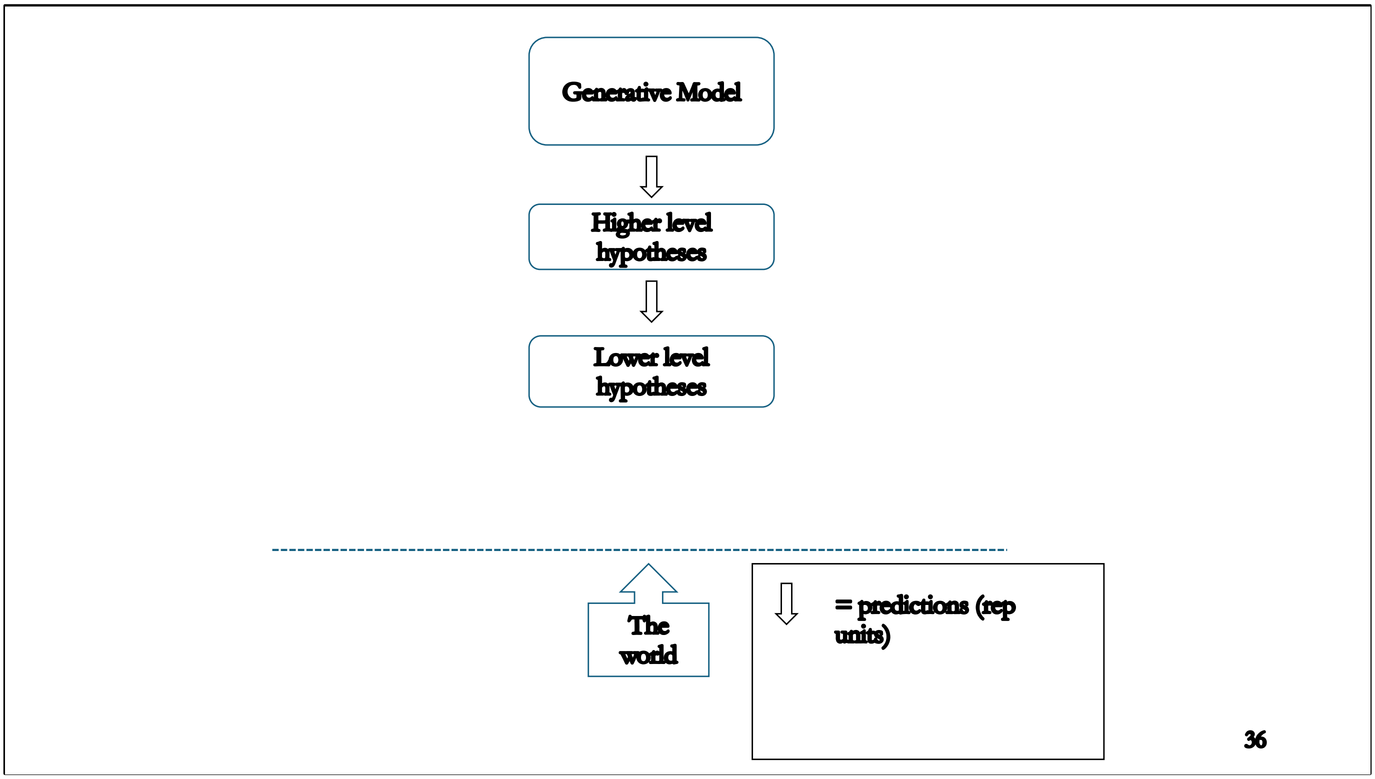
The predictive coding model, step by step



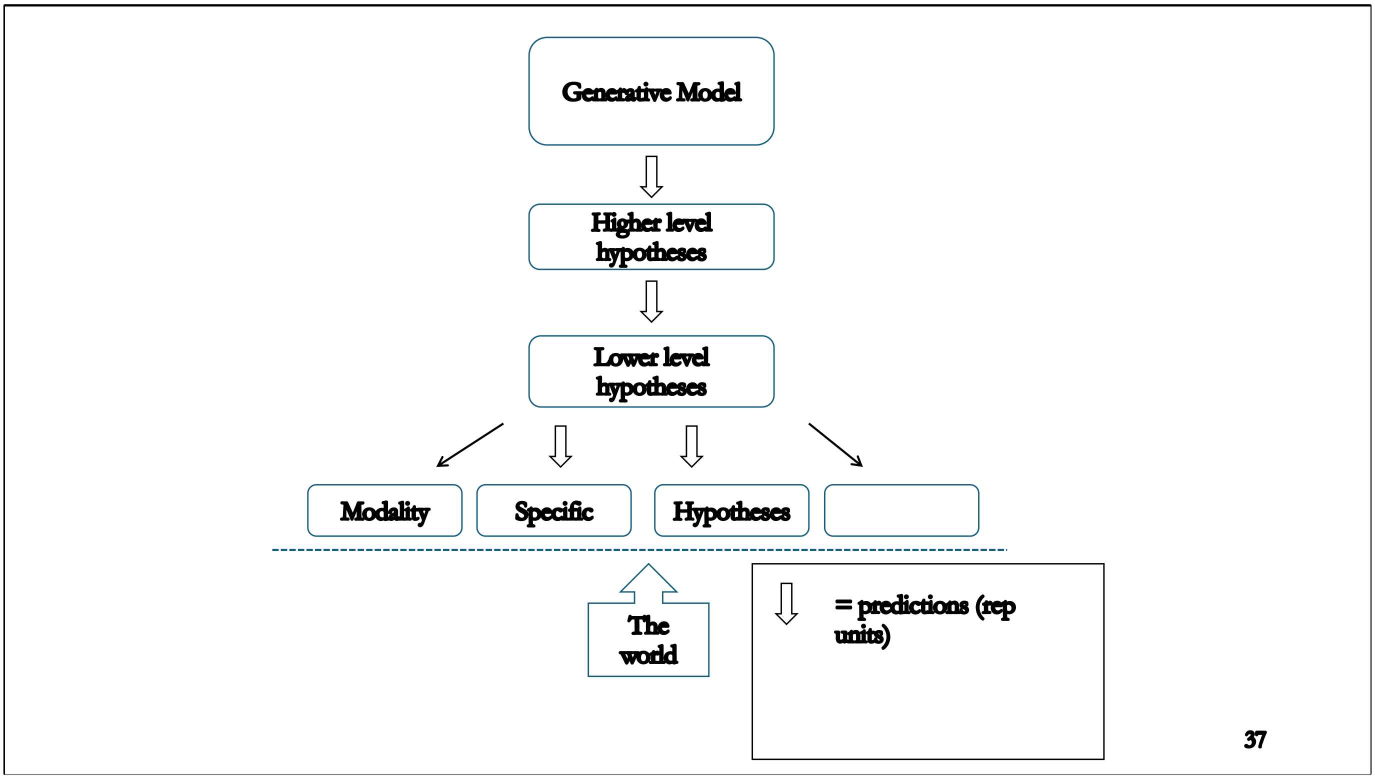
According to the predictive coding view, our job at any given point in time is to predict the hidden cause out in the world of what we're perceiving. To do so we begin with a generative model. That is a model of the world that our brain produces whose business is to use what we know about how the world works to generate predictions, or hypotheses, about what the object or scene in question is. You start with a generative model of where wild animals are found that estimates the probability that you are looking at a bear given a bunch of other things: Are you in the woods in the mountains? Are you in the Buchanan building at ubc?



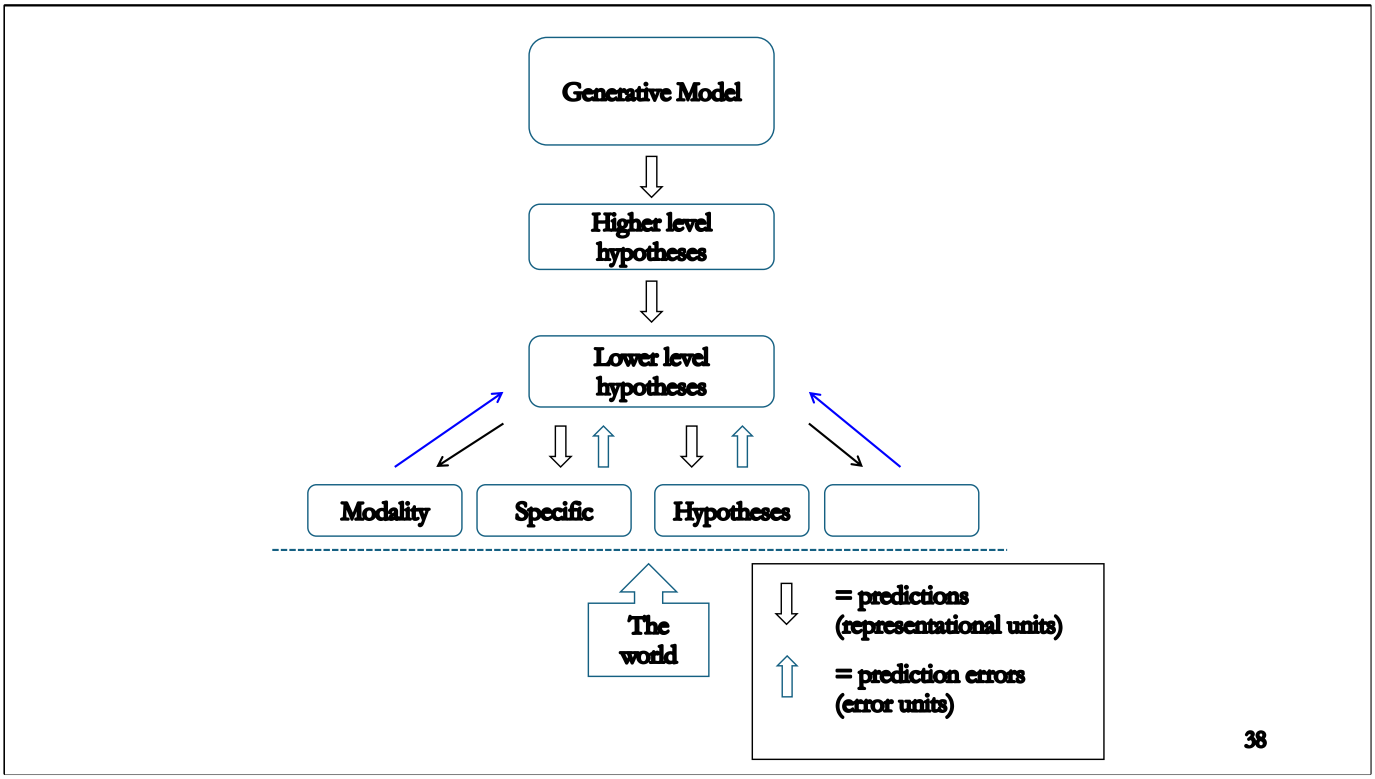
These models of the world in turn generate high level hypotheses such as "There are no bears in pacific spirit park" or "There are bears in the mountains ".



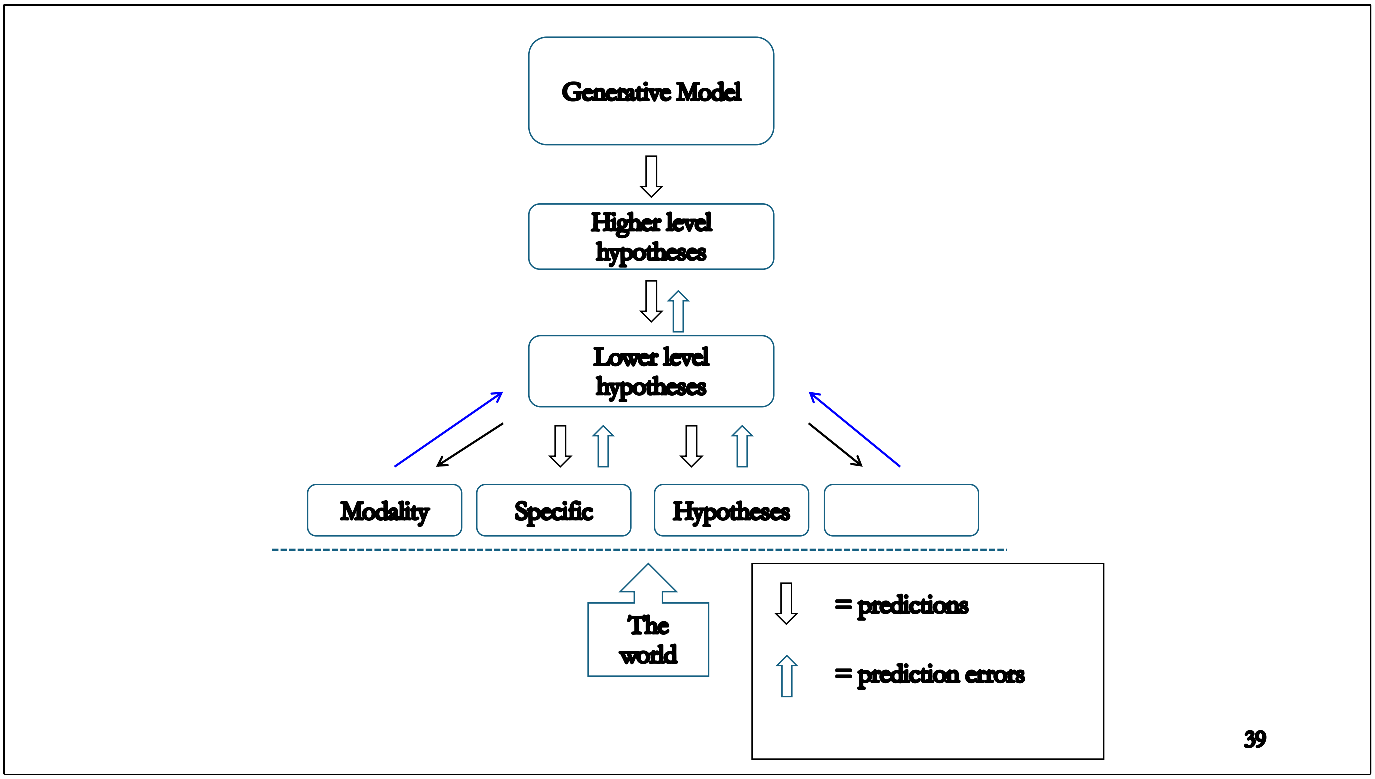
These in turn influence hypotheses at the lower levels, and so on, all the way down. Lower level hypotheses might be: If it's a bear there should be movement and sound (I should see movement and hear rustling). If it's a stump not much should change unless there is a gust of wind. And if that's the case I should hear/feel the wind at the same time.



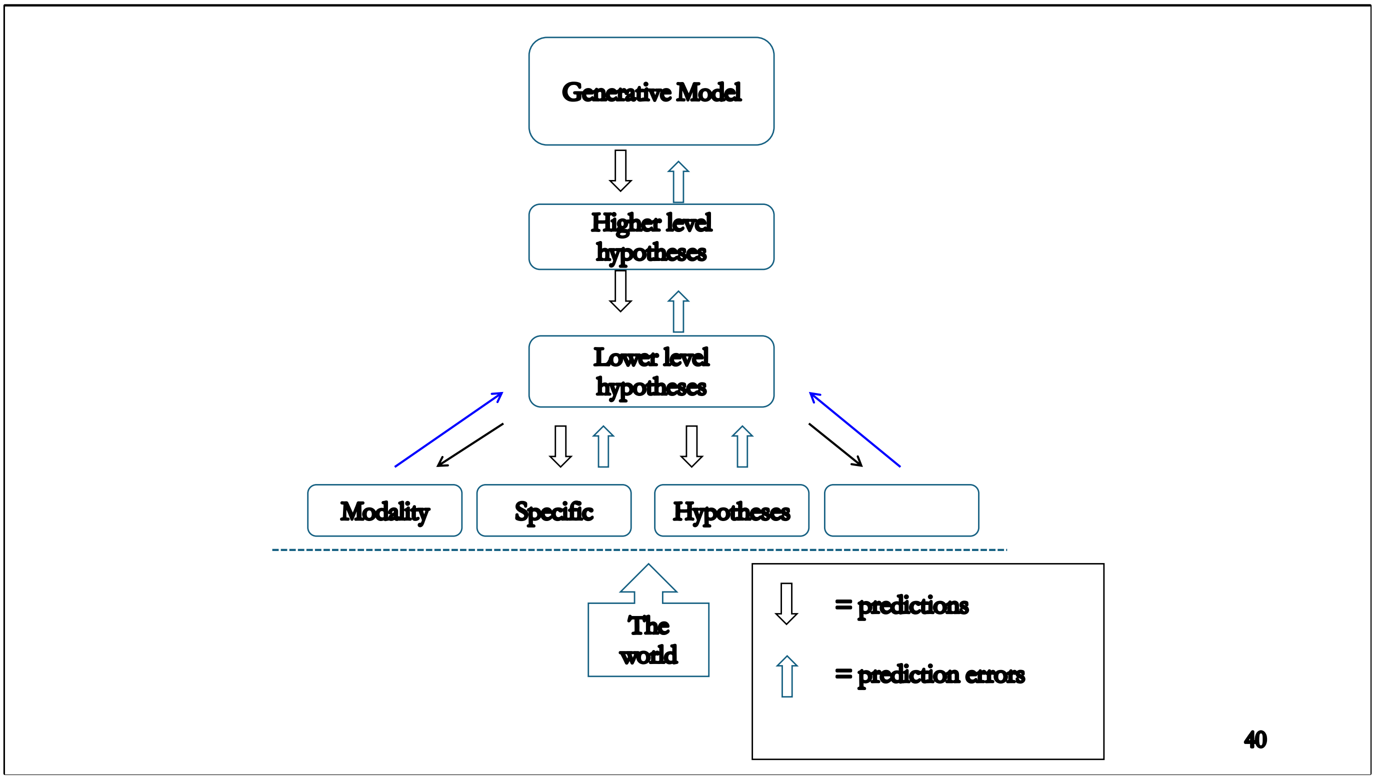
The lowest level hypotheses will be specific to each modality: taste, touch, smell, hearing, and vision. Now we're down to the primary visual cortex/lOC. We send our predictions down to all the specialized detectors of features. Here, something special happens – the hypotheses are compared with the incoming information from the senses. If it's a bear I should see a face, eyes, detecting eyes in the parts of visual cortex that pick up their contrast and shape. And at the part of the early visual cortices that like motion, such as MT, should detect motion. If it's a stump there should be no eyes, face, or signs of motion.



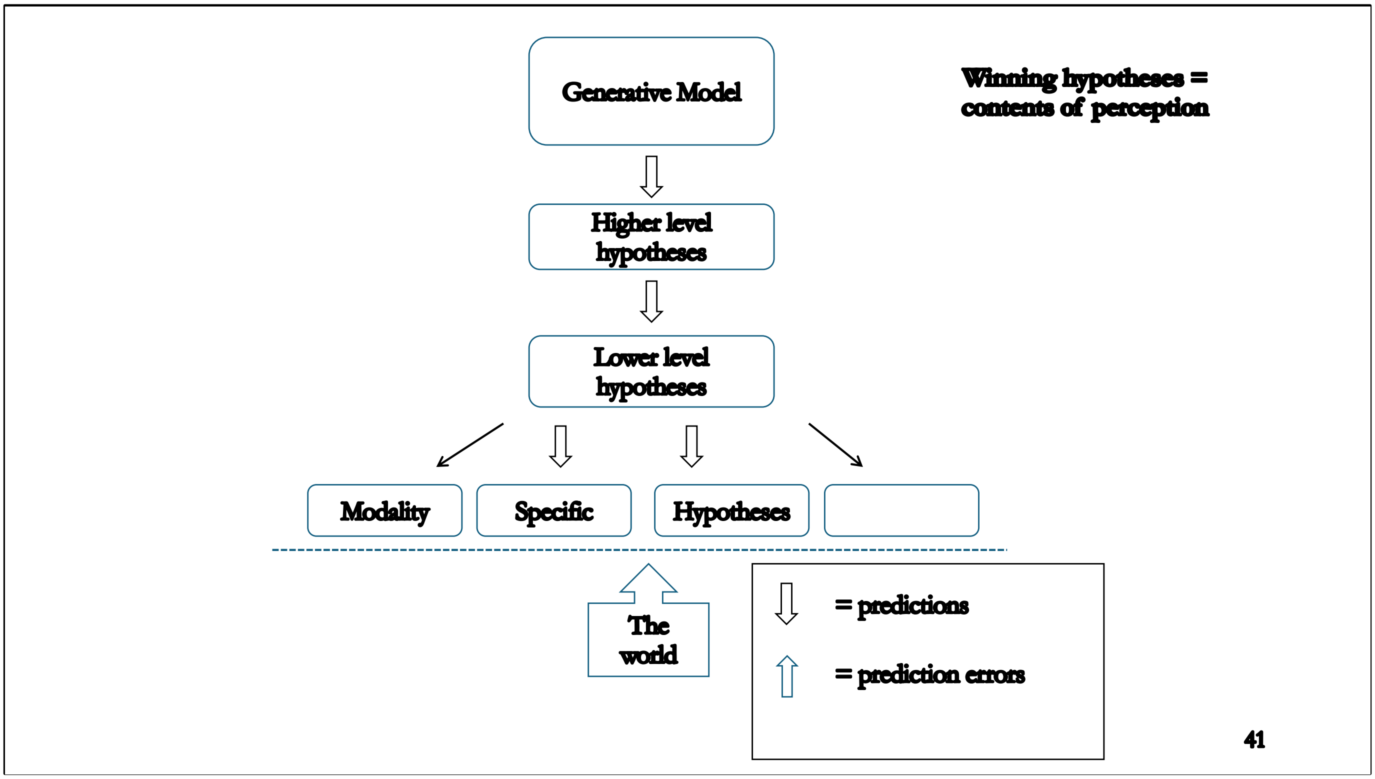
PREDICTION ERROR OR SURPRISE. This part is key to understanding the Egner paper. According to predictive coding, at each level there are *Representational units* that encode *expectation,* or the probability of a given stimulus under the circumstances (conditional probability). These send down their predictions of what they expect to receive from the lower level. And there are *error units that* encode or read *surprise,* or the mismatch between predictions and bottom up evidence from the senses. If I think I'm looking at a tree stump and I see motion because it's a bear that's a big important mismatch! These error signals are sent forward up to the next higher level where the expectations are adjusted. If there is a mismatch, then prediction error is generated, represented here by the blue arrows. The goal of all this is to minimize prediction error. To be right about what you're seeing in the world. Is this a bear? Or is this a tree stump? This has big advantages for survival. So to summarize, If there is a mismatch, then prediction error is generated. This prediction error is shunted up the hierarchy, causing the revision of the hypotheses at the level above. If the next level up can't minimize the prediction error,



then the prediction error gets pushed farther up the system.



The higher the level, the more substantial the revision in the hypothesis. The stump should stay still. A bear should move.



Perception happens when prediction error is minimized, with the winning hypotheses forming the contents of perception.

Questions?

Reading Question

- How do predictive coding models view the role of visual cortex neurons?

a. Feature detection – they respond to how well the stimulus features match the features the neuron is selective for

b. Predictions based on the probability the stimulus will have particular features

c. Error detection — they respond to a mismatch between predicted signal and actual signal

d. a & b

e. b & c

E

The Journal of Neuroscience, December 8, 2010 • 30(49):16601–16608 • **16601**

Behavioral/Systems/Cognitive

Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream

Tobias Egner,[1,2] Jim M. Monti,[3] and Christopher Summerfield[4]
[1]Department of Psychology and Neuroscience, and [2]Center for Cognitive Neuroscience, Duke University, Durham, North Carolina 27708, [3]Department of Psychology, University of Illinois, Beckman Institute, Urbana-Champaign, Illinois 61801, and [4]Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, United Kingdom

Tobias Egner

Chris Summerfield

This study is now 14 years old. It used an encoding approach to fMRI to empirically test predictions stemming from predictive coding models against feature-based models of object perception in the ventral stream.

Big picture question

- Do predictive coding models explain visual object recognition better than classic hierarchical feature-based models?
  - Examine by taking advantage of what we know about category selective voxels in fusiform face area (FFA) and parahippocampal place area (PPA)

[1] Now we're taking what we've learned from encoding approaches to brain mapping – that the FFA responds more to faces than houses and another region of the ventral stream, called the parahippocampal place area, responds more to houses. Then they are taking that brain mapping knowledge and leveraging it to answer a more focused question than just finding out what parts of the brain prefer certain categories of stimuli.

Background: 2 views of visual perception

Predictive Coding

- Perception is inference
- 2 processing units at every level of visual hierarchy
  - Representation ("conditional probability" or *expectation*)
  - Error ("mismatch between predictions and bottom-up evidence", or *surprise*)

Feature Detection

- Visual neurons just respond to features of an object
  - (e.g.,, FFA neurons respond to face features such as eyes, facial configuration etc.)

[1] Inference is a conclusion based on reasoning from the data. We don't really perceive the world directly. We're just guessing and testing our best guess. [2] As we saw, this theory says that at every level of the visual hierarchy there are representation units that are hypotheses about what you're seeing based on your internal model, and error or surprise units that record the mismatch between our internal representation and sensory information coming in from the world.

[2] You can think of processing units as populations of neurons - measured via voxels [3] This is a solely bottom up view. In this case feature detection view is the *alternative view* – the view they are pitting their predictions against.

Research questions & predictions

<u>Research Question</u>: Does BOLD activity in the FFA reflect responses to expectation + surprise? Or just face features?

- <u>General Hypothesis:</u> FFA activity will be an "additive function" of expectation and surprise.
- <u>Alternative hypothesis:</u> There will always be more FFA activation to faces
  - *Expectation and surprise will not matter!*

Reading question

- Who were the participants?
- How many of them were there?

After the reading break:

- More Egner, Monti & Summerfield, 2009
- Passingham Chapter 3
- Preparing for Rosenberg et al., 2016

Enjoy your break!