PSYC 365 Class 12:
Predicting perception and discussing attention


Are you on HSP?
Are you an IMMIGRANT?
Are you a MAN?
Yes to ALL 3?
Participate in our Study: “Exploring Men’s Cultural Experiences and Sexuality”
SWell
Scan to Participate on HSP!
BREB#: H24-0154

Coming up

- Today (Tuesday): Egner et al., 2010: Predicting Perception: Expectation and surprise in the ventral visual stream
- Thursday (Feb. 27th): Selecting attention.
  - Read Passingham Chapter 3 and Sustaining Attention: Rosenberg et al.
  - Class discussion – submit your question before then!

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

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

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

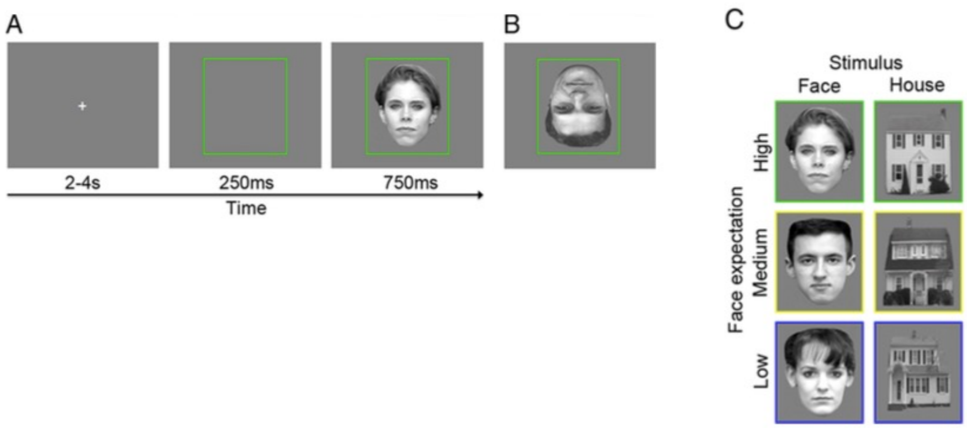
Research Question: Does BOLD activity in the FFA reflect responses to expectation + surprise? Or just face features?

- General Hypothesis: FFA activity will be an "additive function" of expectation and surprise.
- Alternative hypothesis: 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?

Experimental design: fMRI study

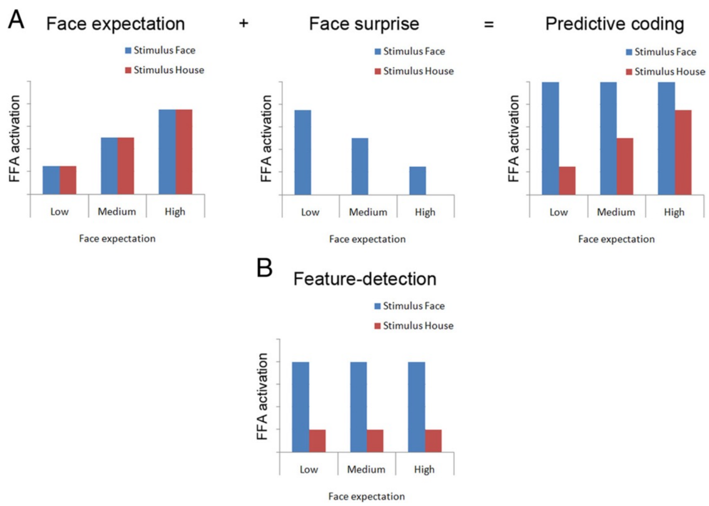


Variables

Independent **Dependent**

I want you to build skills in identifying these yourselves. **Independent variables**: Stimulus probability (% of time face vs. house); Stimulus feature (face/house); target vs. nontarget. Dependent variables. Reaction time. BOLD response in FFA and PPA.

Predicted results



Specific hypotheses

<u>Predictive Coding</u>

FFA responses to faces and houses should be MOST different when face expectation is low (25%)

<u>Feature Detection</u>

FFA responses to faces should always be greater than houses regardless of the level of expectation

Questions?

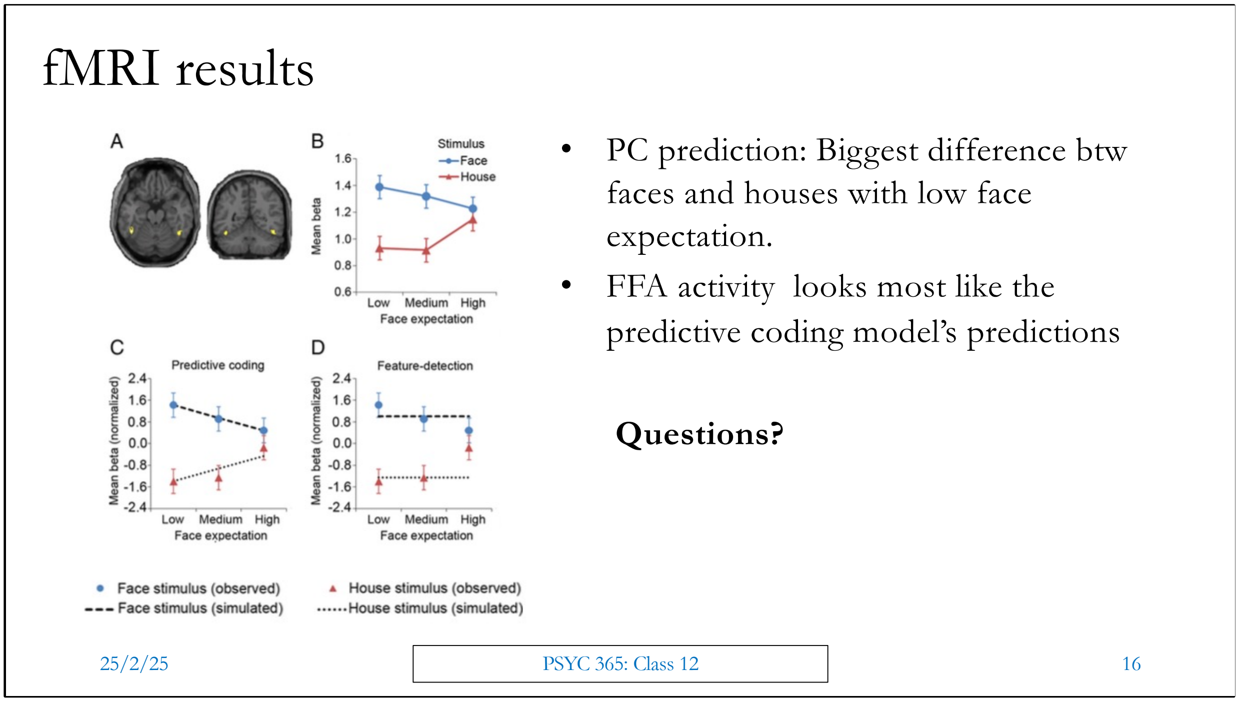
QUESTIONS??? Why? A. For PC model under low expectation there is a lot of surprise if you see a face. No surprise if you see a house. So there is large activation due strictly to surprise. And whether the stimulus in the end is a face or a house has nothing to do with face expectation. [click' B. Why? For FD model the FFA just likes the features of faces, doesn't care what you're expecting.

Actual results

Reading question

- What were the behavioural (reaction time) results?
- How did the authors interpret them?
- What did they see when they looked at accuracy in target detection?

Main effect: Faster to identify upside down faces than houses. They did not vary by expectation condition and there was no interaction between stimulus type and expectation condition. Conclusion: If participants were paying more attention to face features in the high face expectation condition then their RTs should be faster in that condition. Since they aren't faster, the expectation manipulation probably isn't influencing attention. ACCURACY was at ceiling. They were almost always correct. It was an easy task.



QUESTIONS??A. illustrates the fusiform face area seen from below and from behind. B shows the pattern of BOLD response they actually found. Once again, blue is for faces, red is for houses. On the horizontal axis is face expectation, high medium and low. On the Y axis mean BETA is a measure of the bold activation level. Here you see greater FFA responses for faces than houses in the low face expectation condition and the medium face expectation condition but not in the high face expectation condition where FFA responses to face and house stimuli were statistically indistinguishable. This doesn't look EXACTLY like the predicted pattern but it supports the prediction of greater differences between faces and houses for low vs. high expectation. [Click] C & D. These are predictions from the main computational models they used to predict the pattern of brain activity. They used a PC model with parameters representing expectation and surprise and a Feature Detection model with parameters = predicted FFA response to face and house features. C. and D. They did statistical tests (MLE) to compare how well each model fit the data. What you want to look at is how well do the lines, which are the model predictions, fit the observations (blue circles and red triangles)? As you see for the PC model they fit prettywell, not perfetly. And for the feature detection model they don't fit well. You can see that the predictive coding model looks a little bit different from the predictions of the last slide, which just assumed that expectation and surprise would have 50:50 influence. The PC model that fit best assumed a 1:2 ratio where surprise contributed 2x as much as expectation and that is what the data reflect. [click]

A. They tested additional attention models (feature + attention): 1) **Baseline shift model, additive**: FFA activation is enhanced or suppressed depending on expectation of faces for faces AND houses. 2) **Contrast-gain model, multiplicative**, FFA activation is enhanced or suppressed depending on expectation of faces for faces

Reading question

- Why did they also analyze fMRI data in the parahippocampal place area (PPA) as a test of whether the FFA results generalized?
- What did they find?

The mirror pattern of results for the FFA. That is the pattern they saw in the FFA for face expectation and surprise they also saw in the PPA for house expectation and surprise.

Egner & Summerfield conclusions

The authors pitted two views of how visual object recognition works in ventral stream

- "bottom up" feature detection model
  - Neurons respond to features that match preferred stimulus
- Predictive coding model
  - Representation units code expected features
  - Error units code mismatch between expected signal and actual signal from the world

Talk to your neighbour!

- Critiques?
- Are you buying their story?

Key – we know more about relative contributions of prediction and error units. Other things in discussion – there might be an interaction with attention if that was relevant to the task. And we don't know how much the BOLD response reflects top down vs. bottom up inputs.

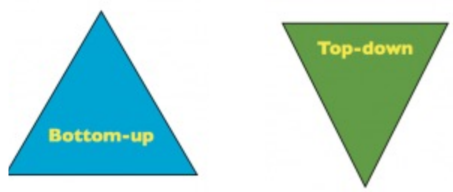
Egner & Summerfield Conclusions

- The pattern of results in FFA was consistent with a response that added expectation and surprise and with predictions of computational models based on predictive coding assumptions
  - They conclude: Prediction coding models describe the process of visual inference better than feature detection models
  - Encoding prediction and error is a general characteristic of how the brain works.

Summary: Pitting 2 models

- **Feature Detection Models**
- Feedforward volley of sensory information
- Hierarchical
- Feature Focused

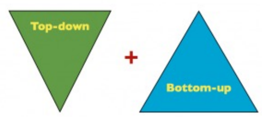
- **PC Model**
- Expectations/prediction
  - Based on memory/experience



PC Models: The Brain's Job is to Minimize Prediction Error!

- Top-down Processes
- Your *representation* or *model* of the world
- Generates predictions at every level of the visual hierarchy
- Tries to "explain away" sensory signal

- Bottom-up signal
- ONLY prediction error gets passed forward – not actual signal
- Propagated upward based on match between model and sensory information



Next up: Attention!

You have my undivided attention!

Learning objectives: Selecting attention

- Identify the difference between covert and overt attention
- Define an attentional set and explain an underlying neuronal mechanism
- Explain classic models of *selective attention* and underlying brain systems

Everyone knows what this quote says

Everyone knows what attention is. It is the taking possession by the mind in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought…

- *William James, 1890*

25/2/25 PSYC 365: Class 13 26

But there are actually many forms of attention and what has happened is that different groups of researchers have studied their particular flavor of attention so that it is very hard for anyone to say overall what attention is. In the next two classes we're going to focus on two very important forms of attention. Selective attention, which determines what we perceive of the world -- what we are actually aware of in our environment or our thoughts, and sustained attention, which keeps us on task, and which is challenging for all the many people who struggle with ADHD.

http://www.youtube.com/watch?v=IGQmdoK_ZfY

25/2/25 | PSYC 365: Class 13 | 27

The gorilla is a famous example of **inattentional blindness**, or the failure to see fairly major changes to a visual scene when you're attending to something else. It's a classic example of **selective attention**. The curtain changing is an example of **change blindness**, or failure to see gradual changes when they are not the centre of focal attention. This is an important concept for understanding **sustained attention**. Both are very classic illustrations of the limitations of attention. And that is one idea that is broadly agreed on -- that attention is limited.

Some forms of attention

1. Overt vs. Covert Attention
2. Selective Attention
   1. Identifying targets
   2. Classic model: Top-down and bottom-up
      1. Attentional sets in top-down attention
   3. Dorsal and ventral attention networks
3. Sustained Attention

Be careful – stop right here!

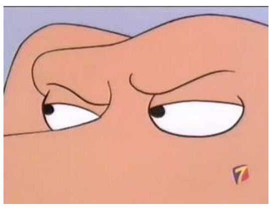
- "*Top-down*" and "*bottom up*" in attention aren't the same thing as in object recognition
- And *dorsal* and *ventral* **attention systems** are NOT the same as *dorsal* and *ventral* **visual streams**
- Same words, different systems!

[click]This causes a lot of confusion because the terms are the same but what they're referring to are not. Top down and bottom up are more general concepts that can be applied in different ways to different cognitive processes and the brain systems that underlie them. [click] Dorsal and ventral are directions so any number of brain systems are divided into dorsal and ventral divisions. [click] So for now forget everything you've learned about these terms in relation to object recognition and start fresh.

Overt vs. Covert attention

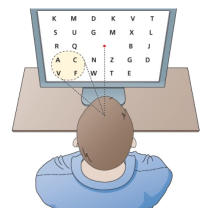
Overt attention

- The eye moves to focus on the object of attention
- **With the eyes**



Covert attention

- You attend to an area of space but the eye does not move
- Object of attention is in your peripheral vision
- **With the mind**



Overt vs. Covert

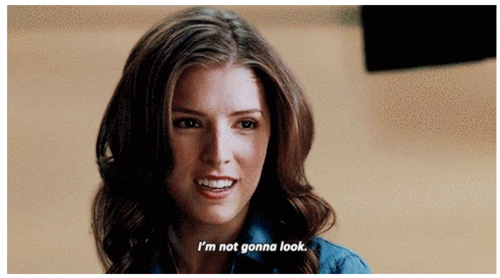
Overt attention

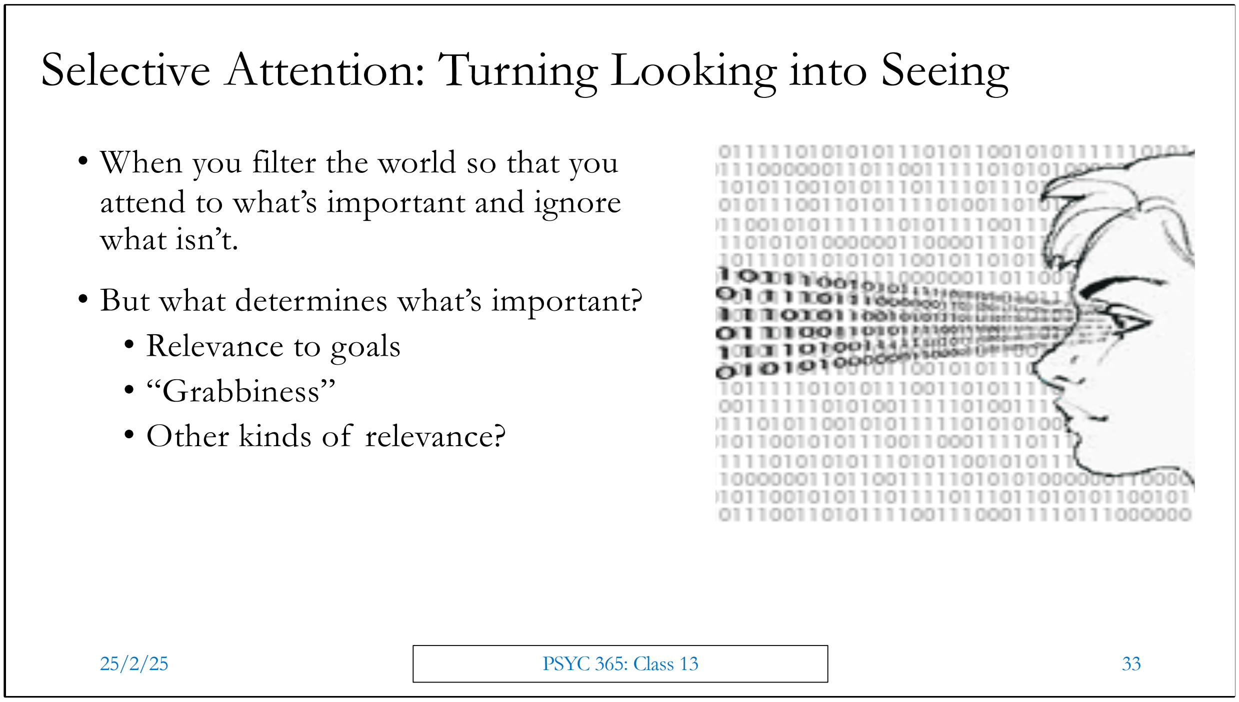
Covert attention

Overt vs. Covert (just one more example)

Overt attention

Covert attention





A good way of describing selective attention is that it is what turns looking into seeing. It's when you filter the world so you attend to what's important and ignore what is not. Why do we do this? [click]Well, millions of bits of information hit your retina. You are blissfully unaware of most of it. But only a couple of hundred bits of information reach the parts of the ventral visual stream involved in high level object recognition. As you know from our classes on object recognition as you you walk down a city street, the millions of pixels falling your retina are in your brain reduced down to chunks like "person", "sign", "building", "tree" and "car speeding towards me" by the time they reach IT in the ventral visual stream. But what process allows for that reduction? The study of selective attention is the study of how we filter the visual information down so that we see what is important and don't see what is not. But then[click] -- what determines what's important? [click]. Is something relevant to my current goals? I'm looking for my coat and hat and car keys so I can go out and run some errands. What features do I attend in my cluttered front hallway? [click]. Is it grabby? Is it shiny, bright, loud or moving? AsI look for my keys I see a flashing light and hear sirens out my window which pull my attention away from my search. [click]. And you can imagine that there are other reasons things could become relevant as well -- see if you can think of some off the top of your head.

Identifying targets

25/2/25 PSYC 365: Class 13 34

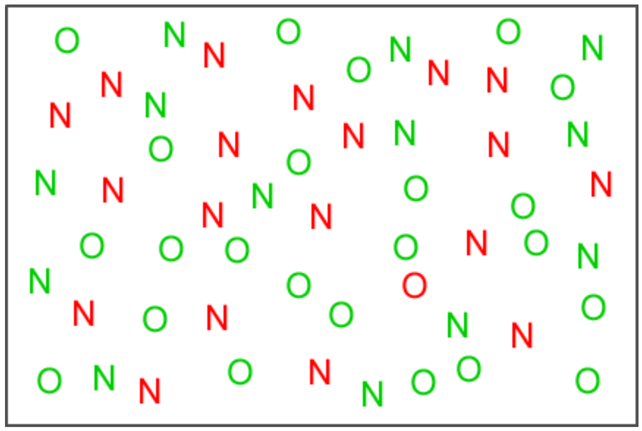
Passingham describes how one thing we use attention for a lot is to identify something a looking for. so looking for your keys on a crowded desk. Looking for your friend in a crowded room. Your favourite brand of beer in the crowded liquor store shelves. Visual search experiments are a type of experiment that are used to create that type of situation in the lab. And they have been used to identify different attentional systems involved in selective attention. I'm going to show you an example.

Identifying targets

25/2/25 PSYC 365: Class 13 35

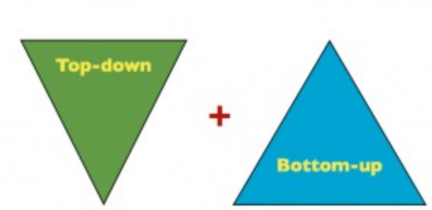
So when the image comes on the screen I want you to identify two unique things. Raise one hand when you've found the first. Your other hand when you've found the second. [click]… one more

Identifying targets



Classic model: 2 attention systems

- Top Down (controlled)
  - Deliberate
  - Conscious
  - Goal-directed *(task-based)*
  - *Attentional set*
- Bottom Up (automatic)
  - Involuntary
  - Captures attention
  - Despite our goals
  - Feature-based



25/2/25 PSYC 365: Class 13 37

- [click] As when you looked for the red O, top-down attention involves control. [click] it's deliberate. We do it on purpose. [click]. It's conscious. We're aware we're doing it. [click]. It's goal directed – we are doing it for a very specific reason. We are masters of directing top-down attention, and we do it all the time . We direct it to the lecture. We direct it to the road when we're driving in bad weather. Finally, top-down attention involves maintaining an attentional set [click] Passingham talks about maintaining a set. That means we maintain a template in our mind of what matters. The neurons and regions that are relevant to it fire up in anticipation. The neurons and regions that are tuned to distracting information are suppressed. It is important for deliberately focusing attention on what is relevant. And again, the key thing is that it is conscious and obeys our will. [click]
- But --"squirrel!" Sometimes our attention is captured by something, despite ourselves. [click]. It's involuntary. It may not help us with our current goals and so it can feel as if attention is captured against our will. So we say our attention is [click] captured – that it captures our attention.[click] despite our goals. Finally [click] it is feature based. That is, it depends on the low level features of the thing kind we discussed in relation to object perception – colour, motion, brightess, loudness, pitch… A siren catches our attention. A

sudden motion does . A bright flash of colour. Gradual change doesn't have any of those qualities – so we didn't see the change in the orange curtain in the gorilla movie**. Q: What might be the evolutionary advantage of each of these?**

"Top-down": Attentional sets

- Mental *templates* that allow us to selectively attend to a certain category of stimulus before it appears
- Involve holding in mind features or location of the object you're expecting

An attentional set is a mental template of either a location or features that you hold in mind so that you can pick the thing that you care about from its environment. [click]... And what is important for you to selectively filter IN is related to your goals.
A "real world" example that's often used is looking for your keys on a cluttered desk. In order to find them more efficiently you hold a *mental template* in your mind of the features of keys (shiny metal, key-shaped) so that you look for things that match the template and can ignore things that don't. That mental template is an *attentional set*. Passingham also describes a series of MVPA experiments that showed that representations of targets could be decoded even before the image is represented, suggesting the brain is keeping alive an attentional set for the features of those targets in advance.

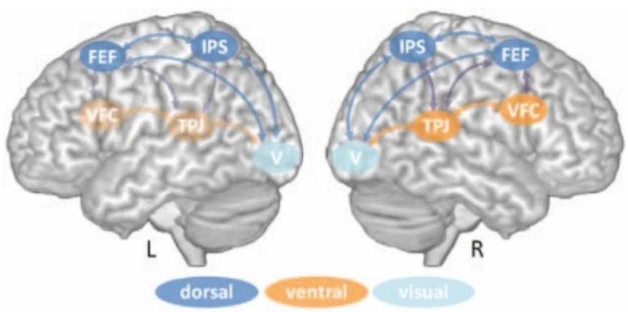
Examples of attentional sets

Mental Template

Q what are the features of Waldo you use to try to locate him? The avocadoes. What about the six pack of blue buck ale?

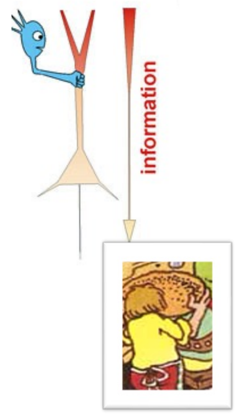
Dorsal (DAN) and Ventral (VAN) Attentional Networks

DAN – top down **VAN – bottom up**

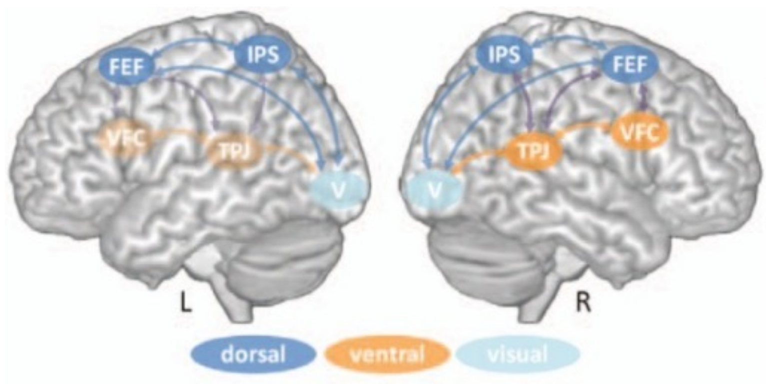


- FEF= Frontal eye fields
- IPS = Intraparietal Sulcus
- VFC = Ventral frontal cortex
- TPJ = temporoparietal junction
- V = Visual cortex

Biased competition: A neural mechanism for selective attention



DAN & VAN: Frontal and Parietal regions *modulate* activity in visual cortex



Interim summary

- *Selective attention* is the form of attention that allows us to select what is relevant and ignore what isn't.
- In classic models, attention can be *feature-based* (bottom-up), or *task-based* (top down)
- These types of attention are (mostly) mediated by VAN and DAN

Sustaining attention

Sustained attention

What is sustained attention?

- Staying on task, even when boring
- Often measured by the Sustained Attention to Response Task (SART)
- Individual differences associated with ADHD, impulsivity

Difficulties in sustained attention

- Easily distracted by non-relevant stimuli
- Often forgetful in daily activities
- Difficulty sustaining attention during activities
- Difficulty following instructions/failing to complete tasks
- Missing details/making mistakes
- Avoidance of activities that require sustained mental effort

Next class: Read Rosenberg et al.!

25/2/25 | PSYC 365: Class 13 | 47

b. Going beyond comparing 2 experimental conditions and “reading” the brain to see what a person is looking at. O