Solving “The Visual Expertise Mystery” with Models

Garrison W. Cottrell
Gary's Unbelievable Research Unit (GURU)
The Perceptual Expertise Network
The Temporal Dynamics of Learning Center
Computer Science and Engineering Department, UCSD

Joint work with Carrie Joyce, Maki Sugimoto, and Matt Tong
Why model?

- Models rush in where theories fear to tread.
- Models can be manipulated in ways people cannot.
- Models can be analyzed in ways people cannot.
Models rush in where theories fear to tread

- Theories are high level descriptions of the processes underlying behavior.
  - They are often not explicit about the processes involved.
  - They are difficult to reason about if no mechanisms are explicit -- they may be too high level to make explicit predictions.
- Using machine learning techniques, one can often build a **working model** of a task for which we have no theories or algorithms (e.g., expression recognition).
- A working model provides an “intuition pump” for how things might work, especially if they are “neurally plausible”
Models can be manipulated in ways people cannot

- We can see the effects of variations in cortical architecture (e.g., split (hemispheric) vs. non-split models (Shillcock and Monaghan word perception model)).
- We can see the effects of variations in processing resources (e.g., variations in number of hidden units in Plaut et al. models).
- We can see the effects of variations in environment (e.g., what if our parents were cans, cups or books instead of humans? I.e., is there something special about face expertise versus visual expertise in general? (Sugimoto and Cottrell, Joyce and Cottrell, Tong & Cottrell)).
Models can be analyzed in ways people cannot

In the following, I specifically refer to neural network models.

- We can do single unit recordings.
- We can selectively ablate and restore parts of the network, even down to the single unit level, to assess the contribution to processing.
- We can measure the individual connections -- e.g., the receptive and projective fields of a unit.
How (I like) to build Cognitive Models

- I like to build them in domains where there is a lot of data and a controversy about it.
- I like to be able to relate them to the brain, so “neurally plausible” models are preferred -- neural nets.
- The model should be a working model of the actual task, rather than a cartoon version of it.
- Of course, the model should nevertheless be simplifying (i.e. it should be constrained to the essential features of the problem at hand):
- Then, take the model “as is” and fit the experimental data: no fitting parameters is preferred over 1, 2, or 3.
The *other* way (I like) to build Cognitive Models

- Same as above, except:
- Use them as *exploratory* models -- in domains where there is *little* direct data (e.g. no single cell recordings in infants or undergraduates) to suggest what we might find if we *could* get the data. These models can then serve as “intuition pumps.”
- Examples:
  - Why we might get specialized face processors
  - Why those face processors get recruited for other tasks
The other way (I like) to build Cognitive Models

- Same as above, except:
- Use them as exploratory models -- in domains where there is little direct data (e.g. no single cell recordings in infants or undergraduates) to suggest what we might find if we could get the data. These models can then serve as “intuition pumps.”
- Examples:
  - Why we might get specialized face processors
  - Why those face processors get recruited for other tasks
A Good Cognitive Model Should:

- Be *psychologically relevant* (i.e. it should be in an area with a lot of real, interesting psychological data).
- Actually be *implemented*.
- If possible, perform the *actual task* of interest rather than a cartoon version of it.
- Be *simplifying* (i.e. it should be constrained to the essential features of the problem at hand).
- *Fit* the experimental data.
- Make new *predictions* to guide psychological research.
A Neurocomputational Model for Visual Recognition
(a.k.a. “The Model” (TM))

- Gabor Filtering
- PCA
- FFA
- LOC
- Expert Level Classifier
  - Book
  - Can
  - Cup
  - Face
  - Bob
  - Sue
  - Jane
- Basic Level Classifier
  - Book
  - Can
  - Cup
  - Face

Levels:
- Pixel (Retina) Level
- Perceptual (V1) Level
- Gestalt Level
- Hidden Layer
- Category Level
The Gabor Filter Layer

- Basic feature: the 2-D Gabor wavelet filter (Daugman, 85):

- These model the processing in early visual areas

Subsample in a 29x36 grid
Principal Components Analysis

- The Gabor filters give us 40,600 numbers
- We use PCA to reduce this to 50 numbers
- PCA is like Factor Analysis: It finds the underlying directions of Maximum Variance
- PCA can be computed in a neural network through a competitive Hebbian learning mechanism
- Hence this is also a *biologically plausible* processing step
- We suggest this leads to representations similar to those in Inferior Temporal cortex
How to do PCA with a neural network
(Cottrell, Munro & Zipser, 1987; Cottrell & Fleming 1990; Cottrell & Metcalfe 1990; O’Toole et al. 1991)

A self-organizing network that learns whole-object representations
(features, Principal Components, Holons, eigenfaces)

Holons
(Gestalt layer)

Input from Perceptual Layer

[Diagram of a neural network with holons and perceptual layers]
How to do PCA with a neural network
(Cottrell, Munro & Zipser, 1987; Cottrell & Fleming 1990; Cottrell & Metcalfe 1990; O’Toole et al. 1991)

A self-organizing network that learns whole-object representations
(features, Principal Components, Holons, eigenfaces)

Holons
(Gestalt layer)

Input from Perceptual Layer
How to do PCA with a neural network
(Cottrell, Munro & Zipser, 1987; Cottrell & Fleming 1990; Cottrell & Metcalfe 1990; O’Toole et al. 1991)

A self-organizing network that learns whole-object representations
(Features, Principal Components, Holons, eigenfaces)

Holons
(Gestalt layer)

Input from
Perceptual Layer
How to do PCA with a neural network
(Cottrell, Munro & Zipser, 1987; Cottrell & Fleming 1990; Cottrell & Metcalfe 1990; O’Toole et al. 1991)

A self-organizing network that learns whole-object representations
(features, Principal Components, Holons, eigenfaces)

Holons
(Gestalt layer)

Input from Perceptual Layer
How to do PCA with a neural network
(Cottrell, Munro & Zipser, 1987; Cottrell & Fleming 1990; Cottrell & Metcalfe 1990; O’Toole et al. 1991)

A self-organizing network that learns whole-object representations
(features, Principal Components, Holons, eigenfaces)

Holons
(Gestalt layer)

Input from Perceptual Layer

...
How to do PCA with a neural network
(Cottrell, Munro & Zipser, 1987; Cottrell & Fleming 1990; Cottrell & Metcalfe 1990; O’Toole et al. 1991)

A self-organizing network that learns whole-object representations
(features, Principal Components, Holons, eigenfaces)

Holons
(Gestalt layer)

Input from Perceptual Layer
How to do PCA with a neural network
(Cottrell, Munro & Zipser, 1987; Cottrell & Fleming 1990; Cottrell & Metcalfe 1990; O’Toole et al. 1991)

A self-organizing network that learns whole-object representations

Holons
(Gestalt layer)

Input from Perceptual Layer
Holons

- They act like face cells (Desimone, 1991):
  - Response of single units is strong despite occluding eyes, e.g.
  - Response drops off with rotation
  - Some fire to my dog’s face
- A novel representation: Distributed templates --
  - each unit’s optimal stimulus is a ghostly looking face (template-like),
  - but many units participate in the representation of a single face (distributed).
The Final Layer: Classification


The holistic representation is then used as input to a categorization network trained by supervised learning.

*Output: Cup, Can, Book, Greeble, Face, Bob, Carol, Ted, Happy, Sad, Afraid, etc.*

- **Categories**
- **Holons**
- **Input from Perceptual Layer**

- Excellent generalization performance demonstrates the sufficiency of the holistic representation for recognition
The Final Layer: Classification

- Categories can be at different levels: basic, subordinate.
- Simple learning rule (~delta rule). It says (mild lie here):
  - *add* inputs to your weights (synaptic strengths) when you are supposed to be *on*,
  - *subtract* them when you are supposed to be *off*.
- This makes your weights “look like” your favorite patterns – the ones that turn you on.
- When no hidden units => No back propagation of error.
- When hidden units: we get task-specific features (most interesting when we use the basic/subordinate distinction)
Outline for the next two parts

- What is perceptual expertise?
  - Behavior, fMRI, and ERPs
- A model of perceptual expertise
Are you a perceptual expert?

Take the expertise test!!!**

“Identify this object with the first name that comes to mind.”

**These slides courtesy of Jim Tanaka, University of Victoria
“Car” - Not an expert

“2002 BMW Series 7” - Expert!
“Bird” or “Blue Bird” - Not an expert

“Indigo Bunting” - Expert!
“Face” or “Man” - Not an expert
“Trump!” - Expert!
“Jerk” or “Megalomaniac” - Democrat!
How is an object to be named?

Animal

Superordinate Level

Bird

Basic Level
(Rosch et al., 1971)

Indigo Bunting

Subordinate Species Level
Entry Point Recognition

Animal
Semantic analysis

Bird
Visual analysis

Downward Shift Hypothesis

Indigo Bunting
Fine grain visual analysis
Dog and Bird Expert Study

- Each expert had at least 10 years experience in their respective domain of expertise.

- None of the participants were experts in both dogs and birds.

- Participants provided their own controls

Tanaka & Taylor, 1991
Object Verification Task

Superordinate
- Animal
- Plant

Basic
- Bird
- Dog

Subordinate
- Robin
- Sparrow
Dog and bird experts recognize objects in their domain of expertise at subordinate levels.

The graph shows the mean reaction time (msec) for novices and experts across three levels: superordinate (animal), basic (bird/dog), and subordinate (robin/beagle). The graph indicates a downward shift in reaction time as the level of expertise increases, with experts showing faster reaction times than novices at all levels.
Is face recognition a general form of perceptual expertise?

Donald J Trump

Indigo Bunting

2002 Series 7 BMW
Face experts recognize faces at the individual level of unique identity.

Diagram: Mean Reaction Time (msec) vs. Category (Superordinate, Basic, Subordinate).
- Objects: Orange line with square markers.
- Faces: Blue line with circle markers.

Downward Shift.

Event-related Potentials and Expertise

Face Experts

- Human Faces
- Animal Faces
- Hands
- Furniture

Bentin, Allison, Puce, Perez & McCarthy, 1996

Object Experts

- Bird Experts
- Dog Experts

Bird Pictures — Dog Pictures


Novice Domain

Expert Domain
Neuroimaging of face, bird and car experts

Fusiform Gyrus

"Face Experts"

Gauthier et al., 2000

Fusiform Gyrus

Car Experts

Bird Experts

Cars-Objects

Birds-Objects
How to identify an expert?

**Behavioral benchmarks of expertise**
- Downward shift in entry point recognition
- Improved discrimination of novel exemplars from learned and related categories

**Neurological benchmarks of expertise**
- Enhancement of N170 ERP brain component
- Increased activation of fusiform gyrus
End of Tanaka Slides
Kanwisher et al., 1997: Took BOLD signal activation of faces and subtracted the BOLD activation of:
- random objects
- scrambled faces
- houses

Every time she got the same spot – the “Fusiform Face Area”

Hence Kanwisher claimed that the FFA is a “module” specialized for faces

But she didn't control for what?
Greeble Experts (Gauthier et al. 1999)

- Subjects trained over many hours to recognize individual Greebles.
- Activation of the FFA increased for Greebles as the training proceeded.
The “visual expertise mystery”

- If the so-called “Fusiform Face Area” (FFA) is specialized for face processing, then why would it also be used for cars, birds, dogs, or Greebles?
- Our view: the FFA is an area associated with a *process*: fine level discrimination of homogeneous categories.
- But the question remains: why would an area that presumably starts as a face area get recruited for these other visual tasks? Surely, they don’t share features, do they?

Solving the mystery with models

- Main idea:
  - There are multiple visual areas that could compete to be the Greeble expert - "basic" level areas and the "expert" (FFA) area.
  - The expert area must use features that distinguish similar looking inputs -- that’s what makes it an expert.
  - Perhaps these features will be useful for other fine-level discrimination tasks.

Solving the mystery with models

- We will create
  - Basic level models - trained to identify an object’s class
  - Expert level models - trained to identify individual objects.
- Then we will put them in a race to become Greeble experts.
- Then we can deconstruct the winner to see why they won.

- A network that can differentiate faces, books, cups and cans is a “basic level network.”
- A network that can also differentiate individuals within ONE class (faces, cups, cans OR books) is an “expert.”
Model

- Pretrain two neural networks on different tasks — Expertise, and Basic-level classification.
- The hidden layer in the expert network corresponds to the FFA.
- The hidden layer in the basic-level network corresponds to the Lateral Occipital Complex.
- Compare their ability to learn a new individual Greeble classification task.
Expertise begets expertise

Amount Of Training Required To be a Greeble Expert

Training Time on first task
- Learning to individuate cups, cans, books, or faces first, leads to faster learning of Greebles (can’t try this with kids!!!).
- The more expertise, the faster the learning of the new task!
- Hence in a competition with the object area, FFA would win.
- If our parents were cans, the FCA (Fusiform Can Area) would win.
Entry Level Shift:
Subordinate RT decreases with training
(Reaction Time = uncertainty of response = 1.0 - max(output))

Human data  Network data

--- Subordinate
--- Basic

RT

# Training Sessions
How do experts learn the task?

- Expert networks must be *sensitive* to within-class variation:
  - Representations must *amplify* small differences
  - Basic networks must *ignore* small differences
  - Representations should *reduce* differences
Observing hidden layer representations

- Principal Components Analysis (PCA) on hidden unit activation:
  - PCA of hidden unit activations allows us to reduce the dimensionality (to 2) and plot representations.
  - We can then observe how tightly clustered stimuli are in a low-dimensional subspace.

- We expect basic level networks to separate classes, but not individuals.

- We expect expert networks to separate classes and individuals.
Subordinate level training magnifies small differences *within* object representations.
The clumping transform also clumps Greebles

The spreading transform generalizes to Greeble representations
Examining the Net’s Representations

- We want to visualize “receptive fields” in the network.
- But the Gabor magnitude representation is noninvertible.
- We can learn an approximate inverse mapping, however.
- We used linear regression to find the best linear combination of Gabor magnitude principal components for each image pixel.
- Then projecting each hidden unit’s weight vector into image space with the same mapping visualizes its “receptive field.”
Two hidden unit receptive fields

AFTER TRAINING AS A FACE EXPERT

AFTER FURTHER TRAINING ON GREEBLES

NOTE: Practically no difference! Very little learning was required.
Controlling for the number of classes

- We obtained 13 classes from hemera.com:
  - 10 of these are learned at the basic level.
  - 10 faces, each with 8 expressions, make the expert task
  - 3 (lamps, ships, swords) are used for the novel expertise task.
Controlling for the number of classes

- We obtained 13 classes from hemera.com – these are the 10 training categories:

- 10 of these are learned at the *basic* level.
- 10 faces, each with 8 expressions, make the *expert* task
- 3 (lamps, ships, swords) are used for the novel expertise task.
Results: Pre-training

- New *initial* tasks of similar difficulty: In previous work, the basic level task was much easier.

![Training Error and Training Time Graph](image)

- These are the learning curves for the 10 object classes and the 10 faces.
Results

- As before, expert networks still learned new expert level tasks faster

Number of epochs to learn swords after learning faces or objects

Number of training epochs on faces or objects
Conclusions
(of the talk, or of this part, depending on time!)

- There is nothing *special* about faces!
- *Any* kind of expertise is sufficient to create a fine-level discrimination area!
- It is the kind of discrimination (fine-level, i.e., individual or species-level) that matters, not the domain of expertise.
- Again, if our parents were cans instead of people, the Fusiform Can Area would be recruited for Greebles.

- We predict that if activations of neurons in FFA could be measured at a fine grain, we should see high variance in response to different faces.