The neuroscience of probability

Wei Ji Ma
Baylor College of Medicine
10^{10-10^{11}} neurons

Each neuron has 10^{3-10^{4}} synaptic connections
Neuron diversity

Optic tectum of the sparrow
Ramón y Cajal, 1905

R. Stufflebeam, The Mind Project
Based on drawings by Cajal
Circuit diagram of macaque visual areas

Felleman and Van Essen, 1991
Levels of organization

Figure 2.1 Levels of organization in the nervous system, as characterized by Gordon Shepherd (1986a).
Difficult question:

How to model the brain?
Simulating the brain bottom-up

Model neocortical microcircuit
What?

Why?
Why do we have a brain?

Essential function: using sensory information to guide behavior
Top-down view: Start with behavior

Mental states

Neural states

mapping?
Goal: to understand the relationship between neural and mental states in quantitative terms

Approach: normative/optimality: what should the brain be doing? → is it really doing that?

System: perception in humans
Gareth Oliver – Britain’s Got Talent 2009
How is our brain fooled into thinking that the puppet is talking?

Hypothesis 1: The puppet is talking.
Support: - We see that the puppet’s movements match the speech.
    - We see that the human’s face isn’t moving.

Hypothesis 2: The human is talking.
Support: - We know that most puppets don’t come with sound.
    - We (kind of) hear the sound coming from the human.
How does our brain decide between these two hypotheses?

• Optimal: compute the *probability* of each from the observations and prior knowledge.

  probability that the puppet is talking given observations
  probability that the human is talking given observations

• What is perceived is the hypothesis with the highest probability.

• **Claim**: *all perception* consists of computing probabilities
Computing probabilities means we are uncertain.

Why can’t we be certain?
Low-quality input

It’s dark
Visibility is low
Stuff is far away

Stuff happens in the periphery
Noise in the brain
Ambiguity

a trapezoid

a rectangle on a road

a weird wire frame
Perception as inference

• The brain, forced to interpret *low-quality* and *ambiguous* observations, computes probabilities: *(probabilistic) inference.*

\[
p \left( \text{hypothesis} \mid \text{observations} \right)
\]

• The brain is not a recording device!
Interpretation

↓

Inference
Al Hazen (Ibn al-Haytham), 965-1040
“Perception requires unnoticed judgments.”

Pierre-Simon Laplace, 1749-1827
“One may even say, strictly speaking, that almost all our knowledge is only probable.”

Hermann von Helmholtz, 1821-1894
“Perception is unconscious inference.”
Is almost all our knowledge only probable?

Do we compute $p$ (hypothesis $|$ observations) ?

Let’s look at daily life!
Perception...

\[ \rho(\text{that is my friend} \mid \text{visual information}) \]
$p(\text{memo is present on desk | messy visual information})$
Prediction...

\[ p(\text{it will rain | atmospheric data}) \]
$p(\text{I will get sick if I eat this apple} \mid \text{look, smell})$
\( p(\text{my teammate will catch the ball} \mid \text{peripheral visual information}) \)
\( p(\text{I can jump over the stream} \mid \text{visual information, jumping ability}) \)
Complex decision-making...

How did you decide to come here?

$p(\text{this is an interesting REU} \mid \text{announcement})$
$p(\text{this is a nice place to work} \mid \text{first impression})$

$p(\text{this is the guy we need} \mid \text{first impression})$
\[ p(\text{he's the one} \mid \text{his behavior}) \]
\[ p(\text{it will rain} \mid \text{atmospheric data}) \]

\[ p(\text{I will get sick if I eat this apple} \mid \text{look, smell}) \]

\[ p(\text{my teammate will catch the ball} \mid \text{peripheral visual information}) \]

\[ p(\text{I can jump over the stream} \mid \text{visual information, jumping ability}) \]

\[ p(\text{that is my friend} \mid \text{visual information}) \]

\[ p(\text{this is a nice place to work} \mid \text{first impression}) \]

\[ p(\text{this is the guy we need} \mid \text{first impression}) \]

\[ p(\text{he’s the one} \mid \text{his behavior}) \]

\[ \text{\(p(\text{hypothesis} \mid \text{observations})\)} \]

The brain interprets.
Probabilities are everywhere.

All yes/no variables!
... but easily extends to variables with >2 possible values

\[ p(\text{stress} \mid \text{headache}) \]
\[ p(\text{hangover} \mid \text{headache}) \]
\[ p(\text{brain tumor} \mid \text{headache}) \]
\[ p(\text{other} \mid \text{headache}) \]
... and to *continuous* variables

Less uncertainty! More uncertainty!
“We (kind of) hear the sound coming from the human.”

A small experiment...

Tons of uncertainty!
How does the brain decide who’s talking?

Prior knowledge:

\[ p(\text{puppet talking}) = 0.20 \]
\[ p(\text{human talking}) = 0.80 \]

Probabilities given observations:

\[ p(\text{puppet talking} \mid \text{visual observations}) = 0.95 \]
\[ p(\text{human talking} \mid \text{visual observations}) = 0.05 \]

\[ p(\text{puppet talking} \mid \text{auditory observations}) = 0.40 \]
\[ p(\text{human talking} \mid \text{auditory observations}) = 0.60 \]

Optimally combined probabilities:

\[ \frac{p(\text{puppet talking} \mid \text{observations})}{p(\text{human talking} \mid \text{observations})} = \frac{0.20 \times 0.95 \times 0.40}{0.80 \times 0.05 \times 0.60} = 3.2 \]

\[ \Rightarrow \text{It is more probable that the puppet is talking!} \]
Bayes’ rule

posterior probability

\[ p(\text{hypothesis} \mid \text{observations}) \propto \]

\[ p(\text{observations} \mid \text{hypothesis}) \cdot p(\text{hypothesis}) \]

likelihood of hypothesis prior probability
Example: object recognition

\[
s: \text{object identity}
\]

\[
\text{image data } I
\]

Kersten and Yuille, 2003
Prior over objects
$p(s)$

Likelihood over objects given 2D image
$L(s) = p(l | s)$

Kersten and Yuille, 2003
Posterior over objects 
\( p(s \mid l) \)

Perceive the object with the highest posterior probability

\[
\begin{align*}
p(s_1 \mid l) &= p_1 \\
p(s_2 \mid l) &= p_2 \\
p(s_3 \mid l) &= p_3
\end{align*}
\]

\( p(s_3 \mid l) = p_3 \text{ is biggest} \)

Kersten and Yuille, 2003
Many perceptual effects can be explained as consequences of Bayesian inference.
Ponzo illusion
Checkerboard shadow illusion
Uninformative likelihood: Necker cube

Prior probability comes to the rescue:

\[
\frac{p(s = \text{Cube 1} | I)}{p(s = \text{Cube 2} | I)} = \frac{p(I | s = \text{Cube 1}) p(s = \text{Cube 1})}{p(I | s = \text{Cube 2}) p(s = \text{Cube 2})} \approx \frac{p(s = \text{Cube 1})}{p(s = \text{Cube 2})}
\]
Uninformative likelihoods in nature

Peppered moth caterpillar
*Biston betularia*
Noor et al., PLOS ONE 2008
Uninformative likelihood: hidden messages?

Led Zeppelin, *Stairway to heaven* (1971)

Played backwards

Satanism?
Aidan and the Quicksand: In a faraway land, a little boy named Aidan wanders lost in the wilderness. Desperate, his father consults a shaman who can hear the “broken music” of the spirit world. The shaman reports that Aidan made a campfire for warmth, but that the fire burned out as the wood turned to ash. Ash (“quicksand”) is an evil power that, unchecked by the heat of fire, escapes to engulf its victims. Aidan’s father sings this lament: ...
How to do science on this?

• **Theory**: the human brain optimally computes probabilities in making perceptual judgments

• **Experimental test**: measure human responses in a controlled behavioral task (psychophysics)

• Compare with alternative theories
Speech recognition gone wild

Demo from
http://www.media.uio.no/personer/arntm/McGurk_english.html
McGurk effect

McGurk and MacDonald, Nature 1976
Why does this happen?

- Hypotheses: “ba”, “ga”, ”da”, other syllables
- Noisy auditory (A) evidence for “ba”
- Noisy visual (V) evidence for “ga”
- The brain computes $p(\text{syllable} \mid A,V)$: cue combination
Generative model

Bayes’ rule →

\[
p(s \mid x_A, x_V) \propto p(x_A \mid s) \cdot p(x_V \mid s)
\]
Computing the posterior

degree of belief (probability)

auditory

auditory-visual

visual

hypothesized stimulus

$X_A$ $X_V$
What was $s$?

Hypothesized stimulus

Weighted average:

$$\hat{S} = \frac{W_A x_A + W_V x_V}{W_A + W_V}$$

where $w_A = \frac{1}{\sigma_A^2}$ and $w_V = \frac{1}{\sigma_V^2}$
The Bayesian observer weighs cues by their reliabilities, on a single trial.

Do humans do this?
Example: ventriloquist effect
Humans integrate visual and haptic information in a statistically optimal fashion

Robert J. van Beers • Anne C. Sittig
Jan J. Denier van der Gon

Marc O. Ernst* & Martin S. Banks

Vision Science Program/School of Optometry, University of Calif
94720-2020, USA

The Ventriloquist Effect Results from Near-Optimal Bimodal Integration

Optimal integration of texture and motion cues to depth

Robert A. Jacobs *

Center for Visual Science, University of Rochester, Rochester, NY 14627, USA

Motion illusions as optimal percepts

Yair Weiss1, Eero P. Simoncelli2 and Edward H. Adelson3

1 School of Computer Science and Engineering, Hebrew University of Jerusalem, Givat Ram Campus, Jerusalem 91904, Israel
2 Howard Hughes Medical Institute, Center for Neural Science and Courant Institute of Mathematical Sciences, New York University,
4 Washington Place, New York, New York, 10003, USA
3 Brain and Cognitive Sciences Department, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, Massachusetts 02139, USA

Correspondence should be addressed to...

Lip-Reading Aids Word Recognition: Noise: A Bayesian Explanation Using Feature Space

Wei Ji Ma1*, Xiang Zhou2*, Lars A. Ross3, John J. Foxe4,5, Lucas

1 Department of Neuroscience, Baylor College of Medicine, Houston, Texas, United States of America, 2 Dept of Linguistics, University of Washington, Seattle, Washington, United States of America, 3 Program in Cognitive Neuroscience, Department of Psychology, New York University, New York, New York, United States of America, 4 Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, 5 Institute for Learning, New York University, New York, New York, United States of America

Do humans optimally integrate stereo and texture information for judgments of surface slant?

David C. Knill *, Jeffrey A. Saunders

Center for Visual Science, University of Rochester, 274 Main Hall, Rochester, NY 14627, USA

Received 2 December 2002; revised in revised form 22 April 2003
Is something wrong with the brain?

• We perceive a puppet to be talking...
  We hear speech that is not there...
  Are we delusional?!

• Nothing wrong! The brain uses probabilities that are correct based on *normally occurring stimuli*.

• In illusions, the stimuli/task are artificially created to make those probabilities misleading.
Mental states mapping?

Neural states
Two questions about neural states

• How do neurons encode probability distributions? (*representation*)

• How do neurons perform Bayesian inference? (*computation*)
Hubel and Wiesel
Tuning curve of a single neuron

Response (spk/sec)

A: CV=0.2, BW=20deg

B: CV=0.77, BW=20deg

C: CV=0.33, BW=28deg

D: CV=0.85, BW=30deg

Orientation (deg)

Macaque V1
Shapley et al., 2003
Tuning curve of a single neuron

Macaque S2
Pruett et al., 2000
Idealized tuning curve

Mean response as a function of the stimulus

Activity (spikes/s) vs. Stimulus $s$:
- Preferred stimulus of this neuron

Mathematical function:
$$f(s)$$
Models of tuning curves

- Gaussian
- Von Mises
- Rectified cosine
- Piecewise linear
- Sigmoid

These have a preferred stimulus.

These don’t.
Variability around the mean response

Response distribution: \( p(r | s) \)

What functional form?

---

Trial 1: 7 spikes
Trial 2: 5 spikes
Trial 3: 3 spikes
Trial 4: 6 spikes
Poisson variability

• Discrete distribution (spike counts)

\[ p(r \mid s) = \frac{e^{-f(s)} f(s)^r}{r!} \]

• \( r \) is an integer, \( f(s) \) not necessarily

• Mean of \( r \):

\[ \langle r \rangle = \sum_{r=0}^{\infty} rp(r \mid s) = f(s) \]
Histograms of a Poisson random variable

Mean rate 3.2 spikes

Mean rate 9.7 spikes
Fano factor

\[
\text{Fano factor} = \frac{\text{variance of spike count}}{\text{mean spike count}}
\]

Poisson process: Fano factor = 1

Physiology: Fano factor in range 0.3 to 1.8
Single neuron – response variability

$p(r | s)$

Variability depends on stimulus
Population of neurons

$f_1(s), f_2(s), \ldots, f_N(s)$

$N$: number of neurons

**Activity (spikes/s)**

Stimulus

Different preferred stimuli

What is unrealistic about this picture?
Population activity on a single trial

$S$ given stimulus

\[ r = (r_1, r_2, \ldots, r_N) \]

Activity

Preferred stimulus

\( N \) neurons

These are now different neurons, not different stimuli!
Population activity – variability

\[ \mathbf{r} = (r_1, r_2, ..., r_N) \]

Response distribution (noise distribution): \[ p(\mathbf{r} | s) \]
Independent Poisson variability

One neuron: \[ p(r \mid s) = \frac{e^{-f(s)} f(s)^r}{r!} \]

Population: \[ p(r \mid s) = \prod_{i=1}^{N} \frac{e^{-f_i(s)} f_i(s)^{r_i}}{r_i!} \]

If not independent, then *noise correlations*
Population codes in the brain

- Primary visual cortex (orientation, spatial frequency)
- MT (motion direction, velocity)
- IT (human faces, objects)
- SC (saccade direction)
- Primary motor cortex (arm movement direction)
- Hippocampus in rat (self location)
- Cercal interneurons in cricket (wind direction)
- Prefrontal cortex (numerosity)

Why population coding and not single-neuron coding?
stimulus $s$ encoding population activity $r$ decoding stimulus estimate $\hat{s}$ probability distribution
Decoding a probability distribution

Activity

Preferred stimulus

Bayes’ rule

Probability

$p(s | r)$

Hypothesized stimulus

$p(s | r) \propto p(r | s)p(s)$

population variability
Computation with population codes

Ma et al., Nature Neuroscience 2006
Integrate-and-fire network

Auditory-visual layer: conductance-based integrate-and-fire neurons

Visual input

Auditory input

Synaptic weights

Decoder
Network performs near-optimally

Mean estimates

Same tuning curves, same covariance matrices

Estimate variance

Same tuning curves, different covariance matrices

Different tuning curves, different covariance matrices
Testing predictions...

Neural correlates of multisensory cue integration in macaque MSTd

Yong Gu¹, Dora E Angelaki¹,³ & Gregory C DeAngelis¹−³

Multisensory Integration in Macaque Visual Cortex Depends on Cue Reliability

Michael L. Morgan,¹ Gregory C. DeAngelis,²,³ and Dora E. Angelaki¹,³,*
¹Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, MO 63110, USA
²Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, NY 14627, USA
³These authors contributed equally to this work
*Correspondence: angelaki@pcg.wustl.edu
DOI 10.1016/j.neuron.2008.06.021
More complex forms of inference

Cue integration
- speech
  - sound
  - lips

Invariant perception (discounting)
- object
  - viewpoint
  - image

Causal inference
- one source
  - s
    - x_A
    - x_V
- two sources
  - s_A
  - s_V

Visual search
- one or two sources?
  - T_1, ..., T_N
    - where’s my memo?
      - memos on desk
      - retinal images

- s_1
- s_2
- ...
Your brain is an inference machine

• The brain *interprets*, not just transmits, sensory input.
• The brain *computes posterior probabilities* to interpret the world: Bayesian inference
• Many *illusions* can be explained as consequences of probabilistic inference.
• *Humans are Bayesian observers* in many psychophysical tasks. They weigh observations by reliability.
• Explaining human *behavior* using a Bayesian model can elucidate underlying *neural processes*. 
"It's not an optical illusion, Madame, it just looks that way!"