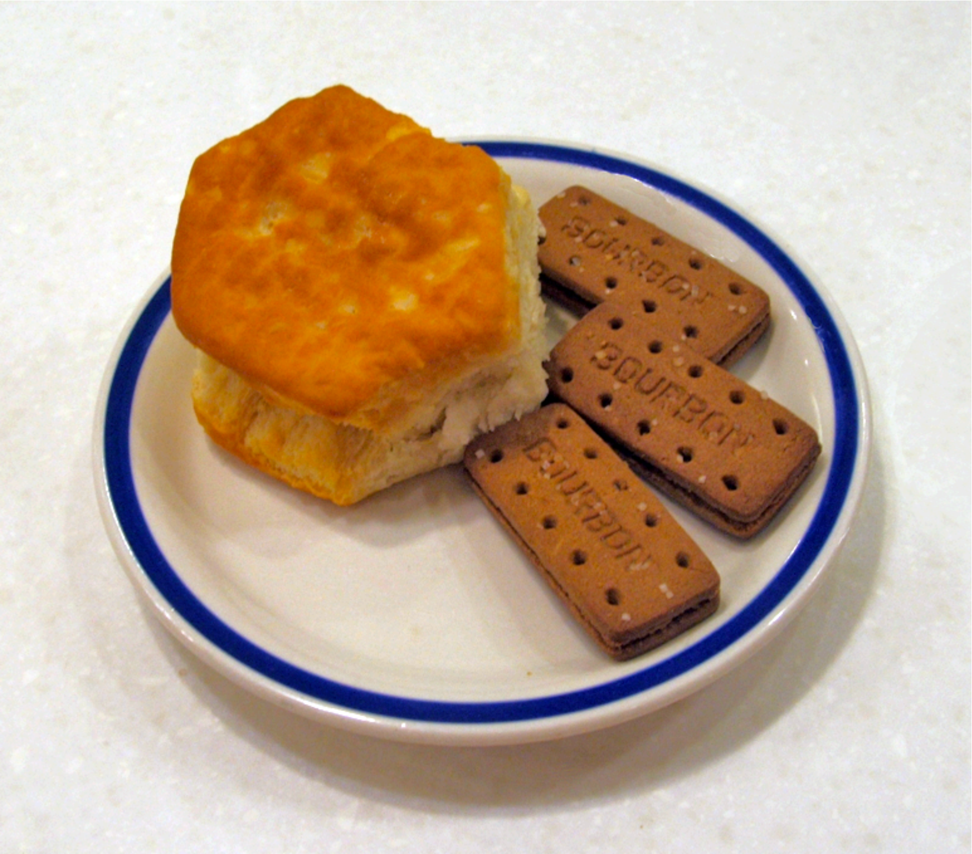
University of Illinois at Urbana-Champaign
Dept. of Electrical and Computer Engineering

ECE 101: Exploring Digital Information Technologies for Non-Engineers

Speech and Natural Language

Ambiguity in Human Language

Biscuits



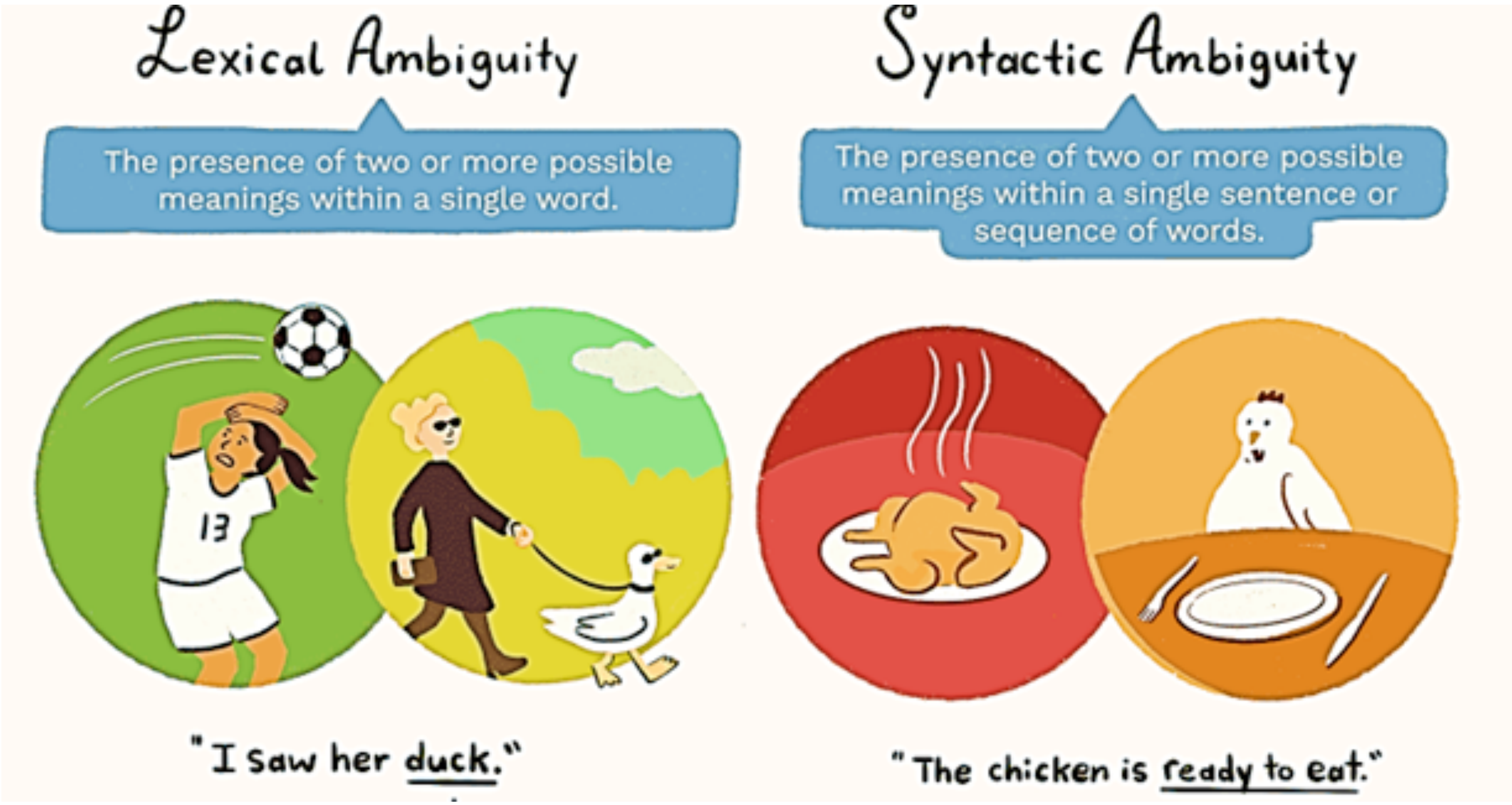
Ambiguity in Human Language

Torch **Flashlight**

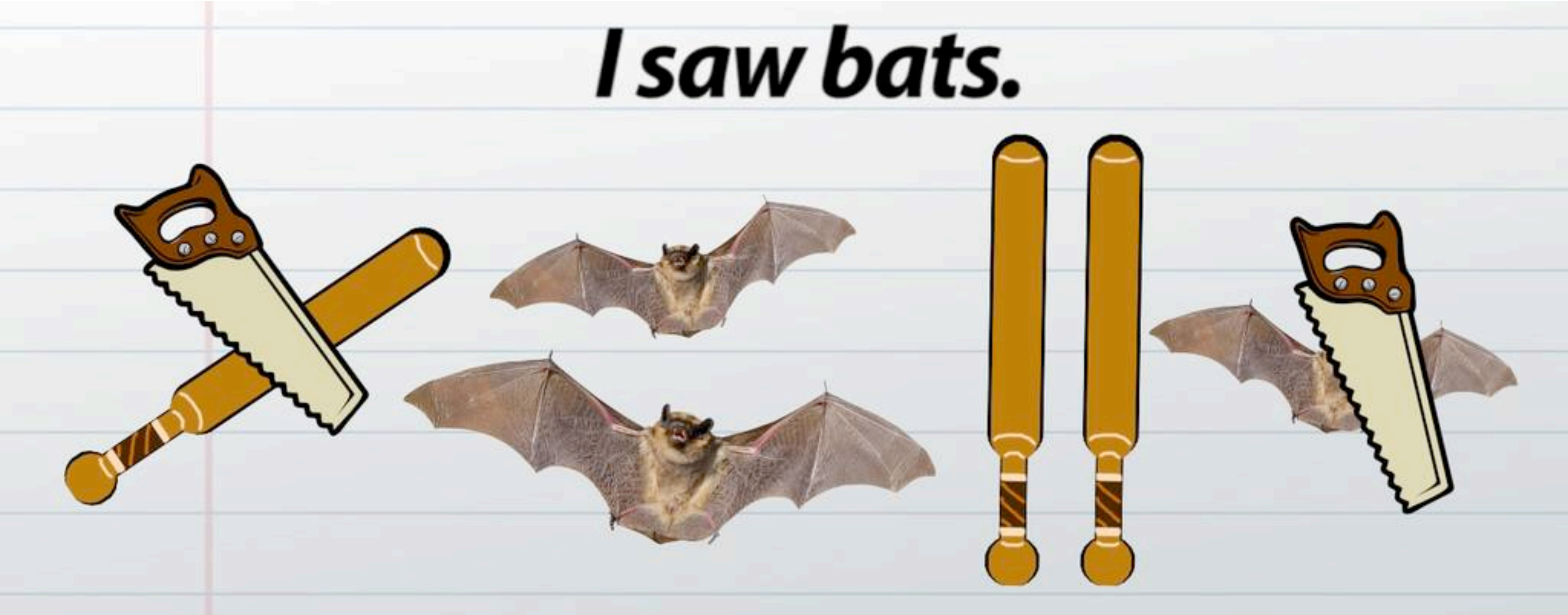
Ambiguity in Human Language

Chill

Context Understanding

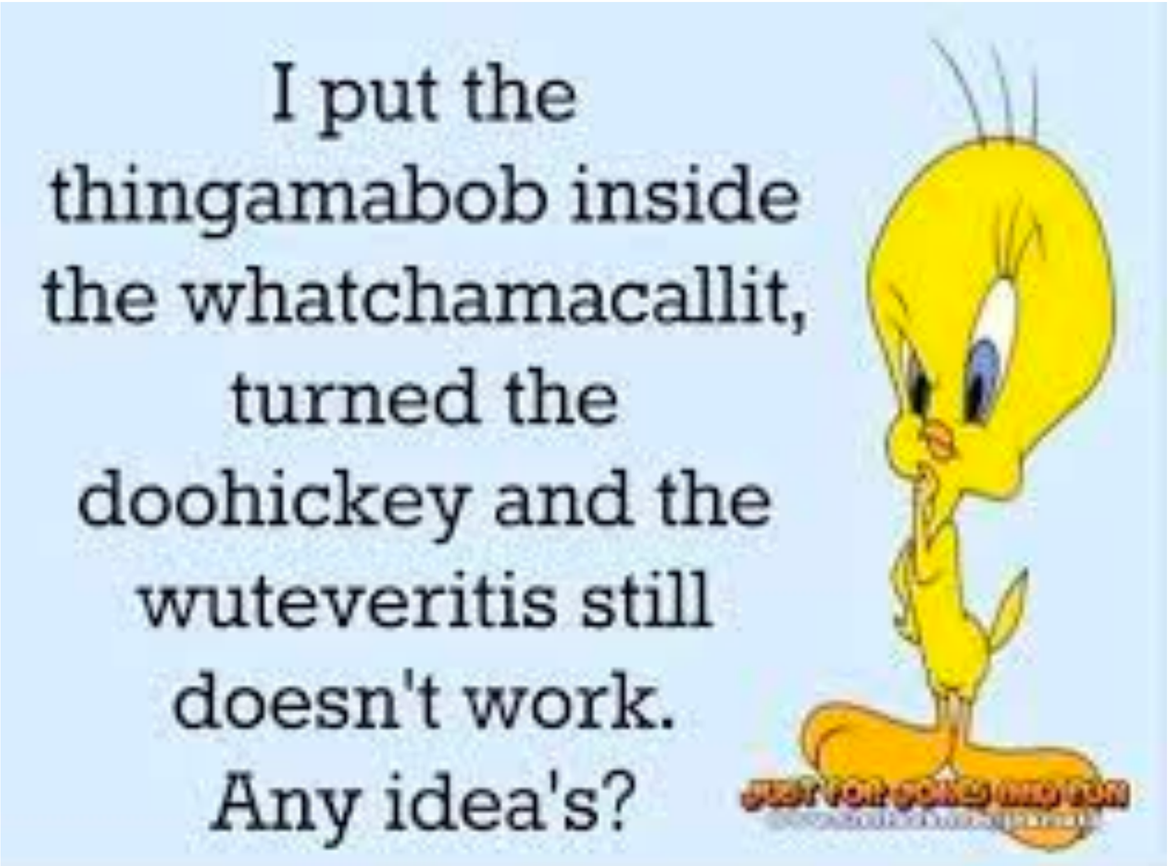


Context Understanding



Sarcasm and Irony

Understanding Human Speech and Language is Hard



Devices that Need to Understand Humans

Consider a voice-controlled smart home
(or an online assistant).

Remember our theme:
sense, **compute**, **communicate** and **actuate**.

What is being sensed when
interacting with the homeowner?

A human voice, captured by a mic.

Devices that Need to Understand Humans

Consider a voice-controlled smart home (or an online assistant).

Sense: human voice

What is computed?

Here, we may **need several steps**.

Computing Steps

Consider a voice-controlled smart home
(or an online assistant).

Computing:

- **get rid of noise**: other voices, music, television, video games, pets, and so forth.
- **perform "voice recognition"**: translate an audio signal into a sequence of words.
- **understand** what the human is trying to communicate: process their **natural language** (English, for example).

Noise will Always Impair the Process

Step 1: get rid of noise.

This task is easy now, but controlled environments always make it easier, and results better.

In 2006, IBM transcribed news from Al Jazeera: formal tone, little/no background noise, intended
for clarity, prosaic content
(not poems!!)

THE IBM 2006 GALE ARABIC ASR SYSTEM

Hagen Soltau, George Saon,
Daniel Povey, Lidia Mangu, Brian Kingsbury, Jeff Kuo, Mohamed Omar and Geoffrey Zweig

IBM T. J. Watson Research Center, Yorktown Heights, NY, 10598
e-mail: {hsoltau,gsaon}@us.ibm.com

شبكـــة الجزيـــرة الإعلاميـــة
ALJAZEERA MEDIA NETWORK

using a super-computer

Unauthorized Voices Can be Treated as Noise

Today "**noise**" may also include unauthorized voices.

By parametrizing the **range of frequencies**, **speeds**, and **accents** for human speech in a given language,

- modern systems are able to record a **voiceprint** (a set of parameter values) and
- **verify that the speaker is authorized** to make use of the system.

Voice Recognition Success Depends Strongly on Context

Step 2: voice recognition

The **context** matters here.

Recognizing "zero" to "nine" in clear, crisp, and unaccented speech has been **possible for decades**.

Understanding a **non-native speaker** who may **mispronounce** words and use unexpected **grammar**, on an **unknown topic** is still **years away** on edge devices.

Variations of Speech Affect Success

Success depends on several aspects…

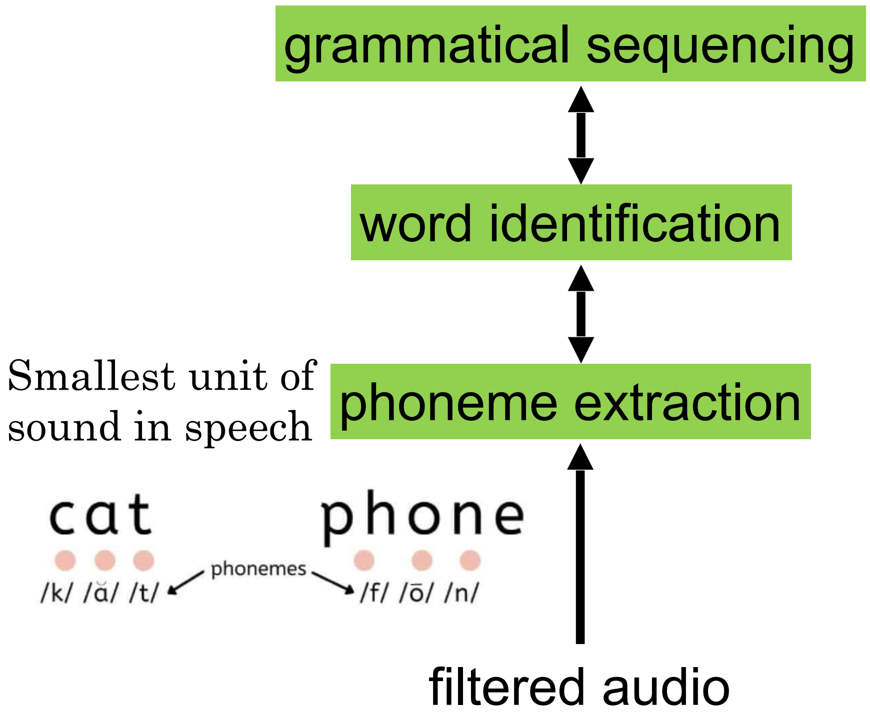
- **How many words** in the vocabulary?
- Do speakers **need to enunciate clearly?** (Didja getdat?)
- Are **euphemisms and idioms** allowed ("collateral damage" instead of "innocent people killed in the war")?
- How **precise**ly must speakers use **grammar**?
- Are **different accents** handled?

It's **easier to provide support for multiple languages** than to understand the vast number of pidgin languages that humans develop spontaneously as they learn new languages.

Hierarchical Models Share Information

Modern voice recognition uses a **hierarchy of interacting, probabilistic models**.

Machine learning is now **used to solve** specific **sub-problems**.



Processing from the Nouns Up

Step 3: natural language processing (NLP)—understanding what the human meant

The most basic form is keyword search.

E.g. What does a human want to see when they type "Pizza" into Google?

Interrogative Adverbs Add Clarity … Sometimes

What if we start to add grammatical elements?

- “Pizza”
  - millions of results: recipes, restaurants, history, memes.
- “How … pizza?”
  - instructions, recipes or tutorials
- “Where … pizza?”
  - location-based intent, so you’ll get local restaurants

Imagine Playing Jeopardy

You don't have to "think" if you can extract keywords and crawl the web

- **Step 1:** Extract keywords from the clue.
  - "This Italian dish is traditionally topped with tomato sauce and cheese."
  - **Keywords**: **Italian dish**, **tomato sauce**, **cheese**
- **Step 2:** Add grammatical cues for context.
  - Jeopardy clues often imply a question like *"What is ___?"*.
  - Add interrogative structure (like **what**, **where**, **how**)
  - "What Italian dish has tomato sauce and cheese?".
- **Step 3:** Search and rank answers
  - crawled the web and use keyword matching plus statistical models to rank likely answers.
  - **Pizza** would score highest because it appears frequently near those keywords

IBM Watson: Jeopardy Champion through Web Crawling

In 2011, **IBM Watson**

- became the **world champion of Jeopardy**,
- a game in which a host gives an answer to a question of the form, "**What is X?**"

Example: "To marry Elizabeth, Prince Philip had to renounce claims to this southern European country's crown."

The question? **"What is Greece?"**

To compete, **Watson crawled the web and built a knowledge base** from which it could draw answers.

using a super-computer

Natural Language Models are Complex and Expensive

Natural language processing today uses a combination of probabilistic inference and machine learning.

One study*, as early as 2019, estimated that training a modern NLP model releases as much carbon as manufacturing and using five cars for their entire lifetimes.

*E. Strubell, A. Ganesh, A. McCallum, "Energy and Policy Considerations for Deep Learning in NLP," *57th Annual Meeting of the Association for Computational Linguistics*, 2019.

Time Check

Let's Learn Some Probability Tools

To better understand the ideas in voice recognition and NLP, let's talk about using probabilities to make good guesses.

Probabilities are always more fun as games…

Rules for 20 Questions

Have you played 20 Questions?

- One person picks a "thing."
- Others ask twenty yes/no questions (the first is allowed to have three answers: animal, vegetable, or mineral?).
- First person to guess the "thing" wins ("Is it a 'thing?' " Yes!).
- Picker wins if no one guesses within 20 questions.

Here’s a Sample Game

Question 1: Animal, vegetable, or mineral?

Answer: Animal.

Question 2: Is it bigger than a dog?

????? Which dog ?????

Need Intuition about Dogs to Answer the Question!

To answer, we have to think …

- What's the typical size of a **dog**?
- What's the typical size of a **thing**?
- **How likely** is a **thing** to be bigger than a **dog**?

A Similar Question of Imagination

Similarly, if I tell you …

"The dog knocked over the child."

In your imagination, how big is the dog?

(Perhaps you want to know the child's age?)

Our Brains Use Probability … Minds … Maybe Not

How do we come up with probabilities based on observed facts?

Humans are generally pretty **bad**
- **at** reasoning consciously about **probability**
- AND at using probability subconsciously.

But **our brains are** reasonably **good at using probability** unconsciously **for language**.

Y0UR M1ND 15 R34D1NG 7H15 4U70M471C4LLY W17H0U7 3V3N 7H1NK1NG 4B0U7 17

Estimate Highly Biased by Experience

What if we ask two people:
how big is an average dog?

- Pat, who grew up in an apartment in downtown Singapore, and
- Jan, who grew up on a farm in US Midwest?

Pat will probably give a
smaller size than Jan.

Why?

Their experience with dogs is likely to differ.

Using Probabilities in Reverse Makes No Sense

In many problems, however, we must **estimate values based on observations**.

Probabilities are **not invertible**:

- if I tell you that I flipped a coin,
- and the result was "tails,"*
- **what can you say about my coin?**

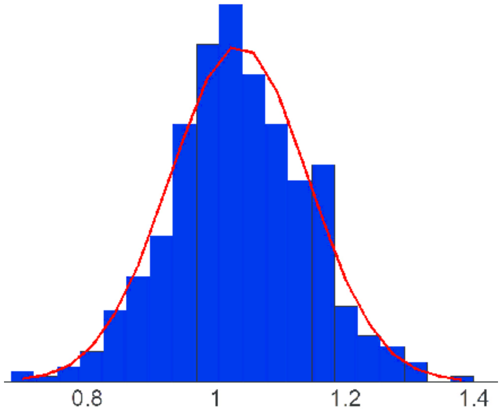
Only that one side is marked as "tails"—not two "heads".

*"Heads" means the side with a person's head, and "tails" means the other side. Most coins in most countries allow this distinction.

MLE: Explanation Most Likely to Lead to Observation

To address this issue,
- people often use a technique called
- **maximum likelihood estimation** (MLE).

Given an observation,
- **choose** the **explanation** that is
- **most likely to produce** the **observation.**



Applying MLE in Casino: Watch and Learn

In casinos, for example…
- **people think** that slot machines pay at different rates.
- **One machine may pay more** money more **often** than another.
- So they **stand and watch** other people play.

If one machine pays **10 times out of 100** plays, and a second machine pays **5 times out of 100** plays, the person then sits down at the first machine.

Likelihood Used to Estimate Win Probabilities

Why?

They are applying primitive MLE!

Assume each play is random, but
- first machine pays with probability $\mathbf{P_1}$, and
- second machine pays with probability $\mathbf{P_2}$.

If one sees 100 plays on a machine,
- and the machine pays **N** times,
- probability **N/100** is most likely for that machine.

Compare the Frequency of Payouts to Pick a Machine

If first machine pays **X** times, $P_1 = X/100$.

If second machine pays **Y** times, $P_2 = Y/100$.

So **X > Y implies P_1 is probably > than P_2**!

Most gamblers
- couldn't explain why at this level of detail
- let alone prove the MLE claims.

Here's a Easy Game to Play

Let's think about another game.

Pat will roll either
- one (six-sided) die or
- two dice and add up the numbers.

Then Pat tells us the amount rolled.

Can we guess whether Pat rolled one or two dice?

Some Cases are Easy, but Others are Hard

Some cases are easy.

For example, Pat rolled an 11. **One or two dice?**

Pat rolled a 1. **One or two dice?**

Other cases are harder…

Pat rolled a 4. **One or two dice?**

Calculate the Chance of a 4 for Each Choice

Let's imagine that Pat rolled one die.

What is the chance that Pat rolled a 4?

1 in 6

Now imagine that Pat rolled two dice.

What is the chance that Pat rolled a 4 (total)?

1+3, 2+2, or 3+1

3 in 36 (same as 1 in 12)

Choice Most Likely to Report 4 is the Best!

With maximum likelihood estimation,

- **we choose "one die"** because
- **probability (if Pat rolls one die, Pat gets a 4)**

  >

  probability (if Pat rolls two dice, Pat gets a 4).

But there's a tricky point.

What does "if Pat rolls one die" mean?

Conditional Probabilities: Chances in Specific Conditions

"If Pat rolls one die" is a **condition**.

In math and engineering,

- we call such probabilities
- **conditional probabilities**
- and we write them this way:

probability (get a 4 | Pat rolls one die)

The meaning is the same:

if Pat rolls one die, Pat gets a 4.

Did We Compute the Wrong Values?

But that's NOT what we wanted to know!

We **wanted to compare**

probability (Pat rolled one die AND got a 4)

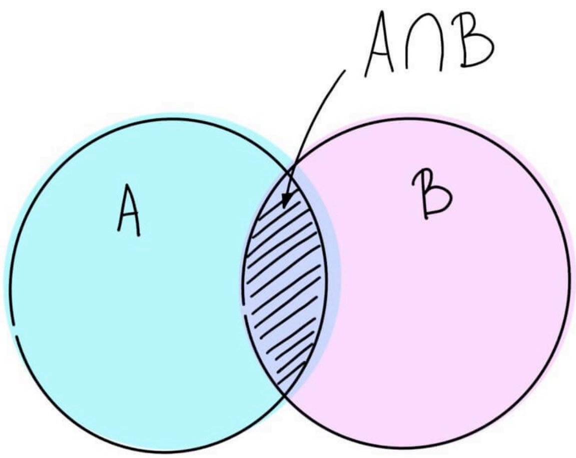
with

probability (Pat rolled two dice AND got a 4)

What can we do?

Bayes' Theorem to the Rescue

Fortunately, we can make use of a famous fact about probability called **Bayes' Theorem**:



Probability (A AND B) =
Probability (A) x Probability (B | A)

The chance of **A and B both happening** is equal to the **product** of the chance of **A happening** and the chance of **B happening if A has happened**.

Apply Bayes' Theorem to Find Our Answer

So to find

probability (Pat rolled one die AND got a 4),

we compute

probability (Pat rolled one die) X
probability (got a 4 | Pat rolled one die)

We know the second number,
i.e. probability (got a 4 | Pat rolled one die): 1/6

But how can we know how Pat makes decisions?

We can't. Pat is a fictional character!

Assume Equal Chance of Both Options

In such cases, we often **assume** that **all** such **events** are **equally likely**.

It's a dumb assumption.

But what else can we do?

In that case, our earlier comparison makes sense

½ · probability (got a 4 | Pat rolled one die)

= ½ · 1/6 = 1/12

>

½ · probability (got a 4 | Pat rolled two dice)

= ½ · 1/12 = 1/24

Initial Probabilities are Important to Correct Choices

What **if Pat tells us** that

probability (Pat rolls one die) = ¼ and

probability (Pat rolls two dice) = ¾ ?

In that case, **our guess changes**, as

¼ · probability (got a 4 | Pat rolled one die)

= ¼ · 1/6 = 1/24

<

¾ · probability (got a 4 | Pat rolled two dice)

= ¾ · 1/12 = 1/16

Recognizing Digits Also Uses MLE

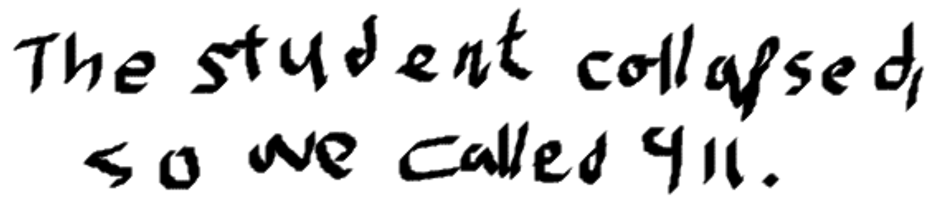
One can also interpret systems that we've already seen as examples of MLE …

Given a picture of a digit, which digit most likely produced the picture?

And context (initial probabilities) DOES matter.

What is this number?

And when it's in context?



MLE Solves the Voice Recognition Problem

How is MLE useful in speech recognition?

Voice recognition answers the question, **"Given an audio input, what sequence of words was spoken?"**

A solution is generated by finding the sequence of words that is most likely to have generated the audio input.

(Our brains are good with this question.)

MLE Solves the Natural Language Problem

How is MLE useful in NLP?

Natural language processing answers the question, "**Given a sequence of words, what did the speaker want to communicate?**"

A solution is generated by finding the meaning that is most likely to have generated the sequence of words.

(Our brains are also good with this question.)

Terminology You Should Know from These Slides

- voice/speech recognition
- Natural Language Processing (NLP)
- Maximum Likelihood Estimation (MLE)
- conditional probability
- Bayes' Theorem

Concepts You Should Know from These Slides

- steps computation: audio → noise removal → word sequence → meaning
- sources of noise
- challenging aspects of speech recognition
- hierarchy of models for speech: phonemes, words, and grammar
- impact of human experience on probabilistic "reasoning"
- how MLE can be used to solve problems