CS 380: Artificial Intelligence

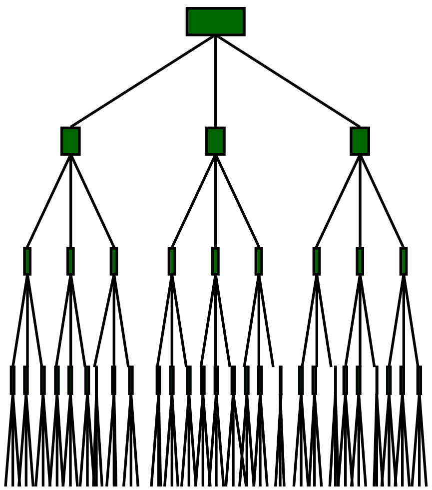
Lecture 17: Cognitive Models I

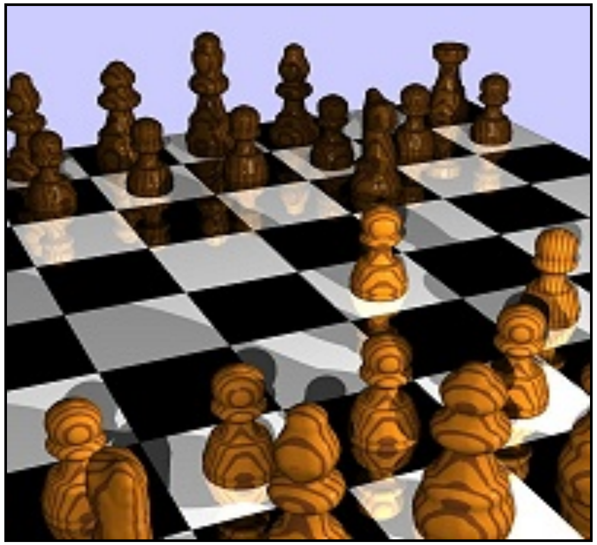
Machine vs. Human Intelligence

- Machines do lots of “intelligent” things
  - add really big numbers really quickly
  - diagnose patient symptoms
  - predict earthquakes, volcanic eruptions
  - etc., etc., etc. …
- Typically, they don’t do things like humans
  - sometimes, methods inspired by humans
    - e.g., neural networks
  - but usually, the goal is to solve a problem, not imitate humans

Example: Playing chess

- Computers search lots and lots of possible moves and their implications





- This can beat the best human players.
- But how do humans play chess?

Example: Sorting numbers

- Consider these numbers:

23	78	32	44	52	17	89	41

- Now sort them in increasing order…

Example: Sorting numbers

- How did you do it?

17	**23**	**32**	**41**	**44**	**52**	**78**	**89**

- Computers: many efficient algorithms
- Humans (typically): find smallest, write it, find 2nd smallest, write it, …
  - What's the time complexity of this algorithm?

Example: Sorting numbers

- How did you do it?

17	**23**	**32**	**41**	**44**	**52**	**78**	**89**

- Computers: many efficient algorithms
- Humans (typically): find smallest, write it, find 2nd smallest, write it, …
  - What's the time complexity of this algorithm?
    - $O(n^2)$
  - Why don't people use a more efficient algorithm?

Cognitive Models

- We want a machine (computer) that thinks and acts like a person, at some convenient level of detail / abstraction.
- In other words, a *model* of human behavior
- The two branches of this endeavor:
  - Studying how people think / act
    - generally comes from experimentation and observation
  - Modeling these thoughts and actions
    - develop models that do the thinking and acting
- Both branches go hand-in-hand!

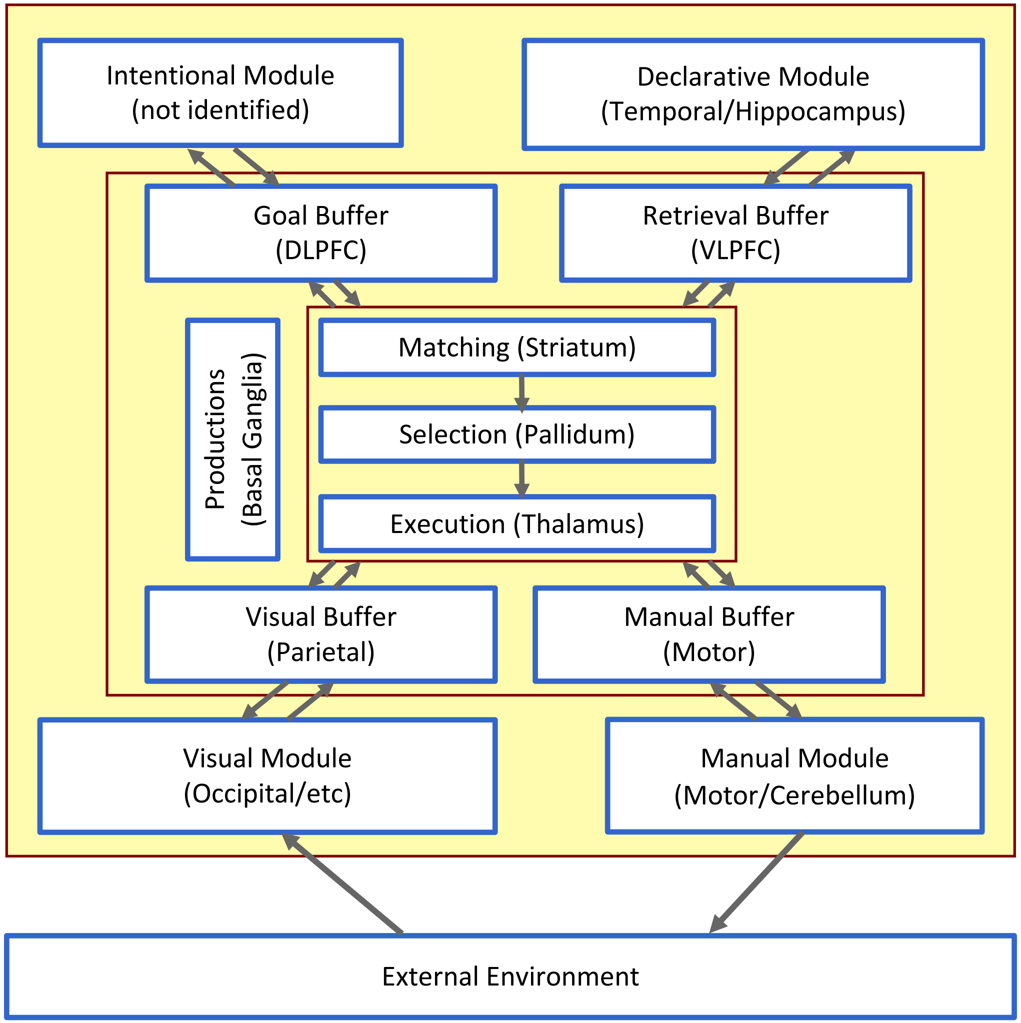
Empirical study

- The Psychology side of our endeavor
- How can we study human thought?
  - general observation
    - "as time passes, I forget things"
  - rigorous experimentation
    - data collection and analysis
  - anatomical study
    - dissect the brain and figure it out (!)

Experiment Demos

A Cognitive Architecture

ACT-R



Perceptual/Motor Overview

- Perception and motor modules act in parallel
  - can speak, type, listen simultaneously (in theory)
- But there's a central bottleneck: Cognition
  - for perception, cognition must process information waiting at perceptual modules
  - for motor actions, cognition must initiate movement by issuing motor command
- So... The model can do many things at once, but can only think about one thing at a time
- Let’s first look at the different components in turn…

Vision

- Key #1: “Spotlight” theory of attention

what the world looks like

how we see the world

  - limited “fovea” of high resolution
  - large “periphery” of degrading resolution

Vision

Find a location in my visual field
that satisfies a set of constraints

"Where"

Given that I have the location, move
my attention to that location

"What"

Attend to that object

Vision

- Example written in the ACT-R modeling language:

```
(p read*find-word
    =goal>
        isa read
    ?visual-location>
        state free
        buffer empty
    ?visual>
        state free
        buffer empty
==>
    +visual-location>
        isa visual-location
        kind text
)
```

```
(p read*encode
    =goal>
        isa read
    =visual-location>
        isa visual-location
        kind text
    ?visual>
        state free
        buffer empty
==>
    +visual>
        isa move-attention
        screen-pos
          =visual-location
)
```

```
(p read*done
    =goal>
        isa read
    =visual>
        isa text
        value =value
==>
    << use =value >>
)
```

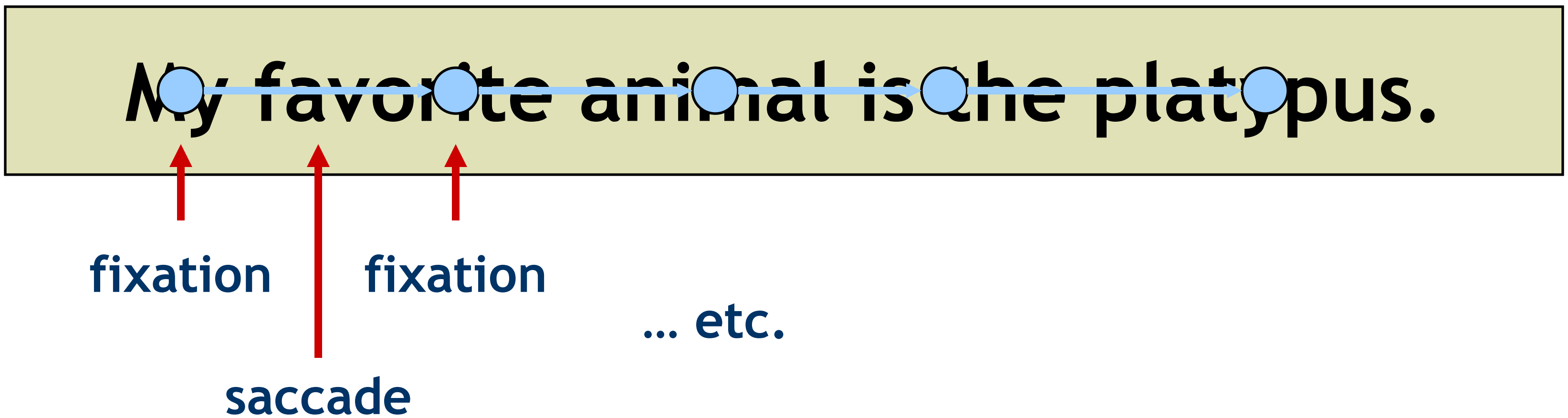
Visual + Eye movements

- How do our eyes move when we read?

My favorite animal is the platypus.

Visual + Eye movements

- How do our eyes move when we read?



  - attention moves from word to word; fixation does not (necessarily)
  - tend to skip short or high-frequency words (e.g., “a”, “the”)

Visual + Eye movements

- Key #2: Fixation ≠ visual attention

Where you're looking isn't always the same as where you're attending!

  - fixation = pause while looking (~200-500ms)
  - saccade = rapid, ballistic movement (~20-30ms)
  - eye movements follow visual attention
    - like a rubber band!
  - note: other types of eye movements as well (smooth pursuit, vergence)

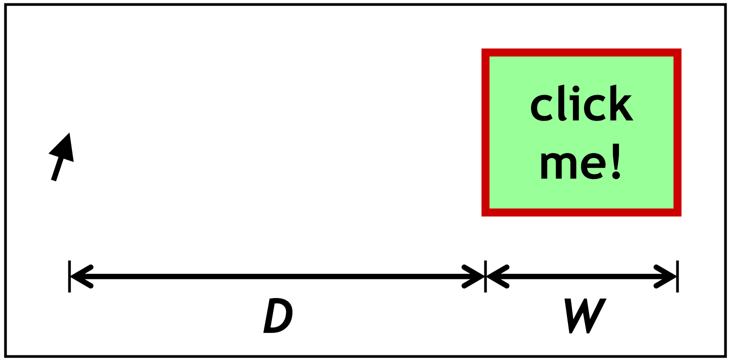
Motor Movements

- Motor movements often specific to HCI actions
  - keystrokes, mouse movements, mouse clicks
- Cognition issue motor commands,
  motor resource responds
- Movement is divided into two phases: preparation and execution
- Can only prepare one movement at a time, and can only execute one movement at a time
  - but, can prepare one while executing another

Execution Time

- One of the earliest and most common models of motor movement… **Fitts' law**
  - D = distance to move
  - W = width of target
  - b = scaling constant

$$T = b \cdot \log_2\left(\frac{D}{W} + 0.5\right)$$



Audition (a.k.a. Hearing)

- Works a lot like vision
- *Where* vs. *what* processing
  - Find sound (or have it “stuffed” into the system)
  - Shift auditory attention with aural request
  - Harvest result from the aural buffer
- The timing is different from vision
  - Events take time to enter the “audicon”
  - Attentional latency is longer than vision
  - Detection and encoding depend on type of sound
    - Tones, words, numeric digits, user defined

Speech

- Rough approximation for short utterances
  - Execution time is a function of syllables
- Prepare, then execute
- Sample commands:
  - Speak
    - Output speech which the model and others may hear
  - Subvocalize
    - Output speech which only the model can hear

21

Other P/M processes?

- What’s missing? Plenty.
  - Basically, everything is tuned for HCI-flavored tasks centered on a desktop computer
  - Mobile devices / touch screens?
    - Some new work on this recently
  - Full cockpit, e.g. car or plane?
    - We're not there yet, although there are related “body models” out there (e.g., HUMOSIM)
      - e.g., imagine pressing a button in a bumpy vehicle!
  - Full locomotion, and so on
    - Would need a lot of integration with other research areas — but the benefit for cognitive models is unclear

Declarative Memory

- Declarative memory = memory of facts
  - 3+2=5
  - The past tense of “go” is “went”
  - Last week I had blueberry soup (episodic)
- These are sometimes referred to as *chunks*

Activation

- Activation is a “subsymbolic” feature of a chunk that determines how easily it’s recalled
- Activation determines…
  - *Whether* a chunk can be retrieved
  - *Which* chunk is retrieved
  - *How long* it takes for a chunk to be retrieved

Activation

$$A_i = \boxed{B_i} + \boxed{C_i} + \boxed{\epsilon}$$

base-level activation

spreading activation

noise

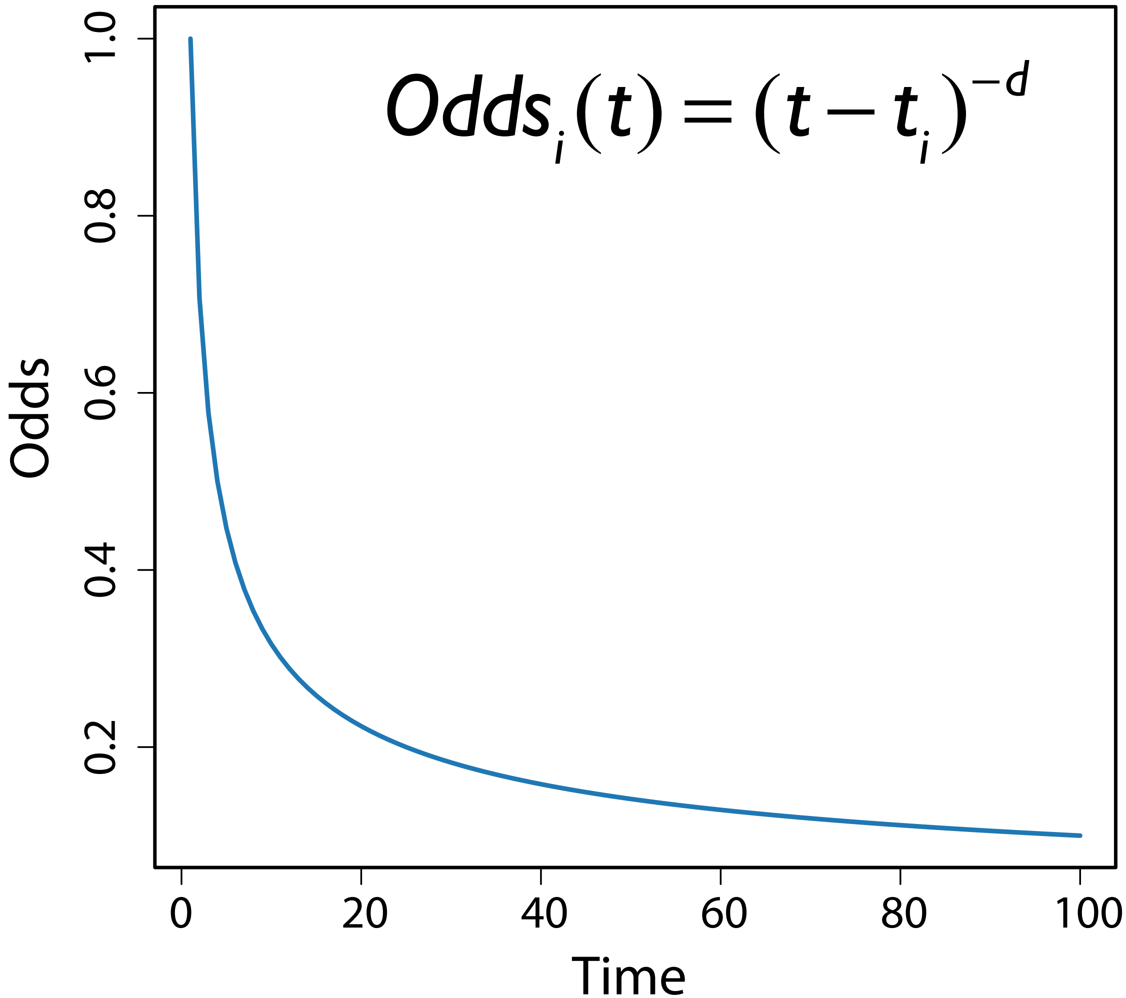
Base-Level Activation

- Reflects the odds that you need a chunk based on frequency and recency
- For a single time the chunk is used…

decay parameter

$$Odds_i(t) = \underbrace{(t - t_i)}_{\text{time since last use}}{}^{\boxed{-d}}$$

Base-Level Activation



Base-Level Activation

- For n uses of the chunk (practice):

$$Odds_i(t) = \sum_{k=1}^{n} (t - t_k)^{-d}$$

- What counts as a "use"?
  - Chunk creation: first "knowing" a chunk
  - Memory retrieval
  - Chunk merging: re-creating the chunk

Base-Level Activation

n presentations

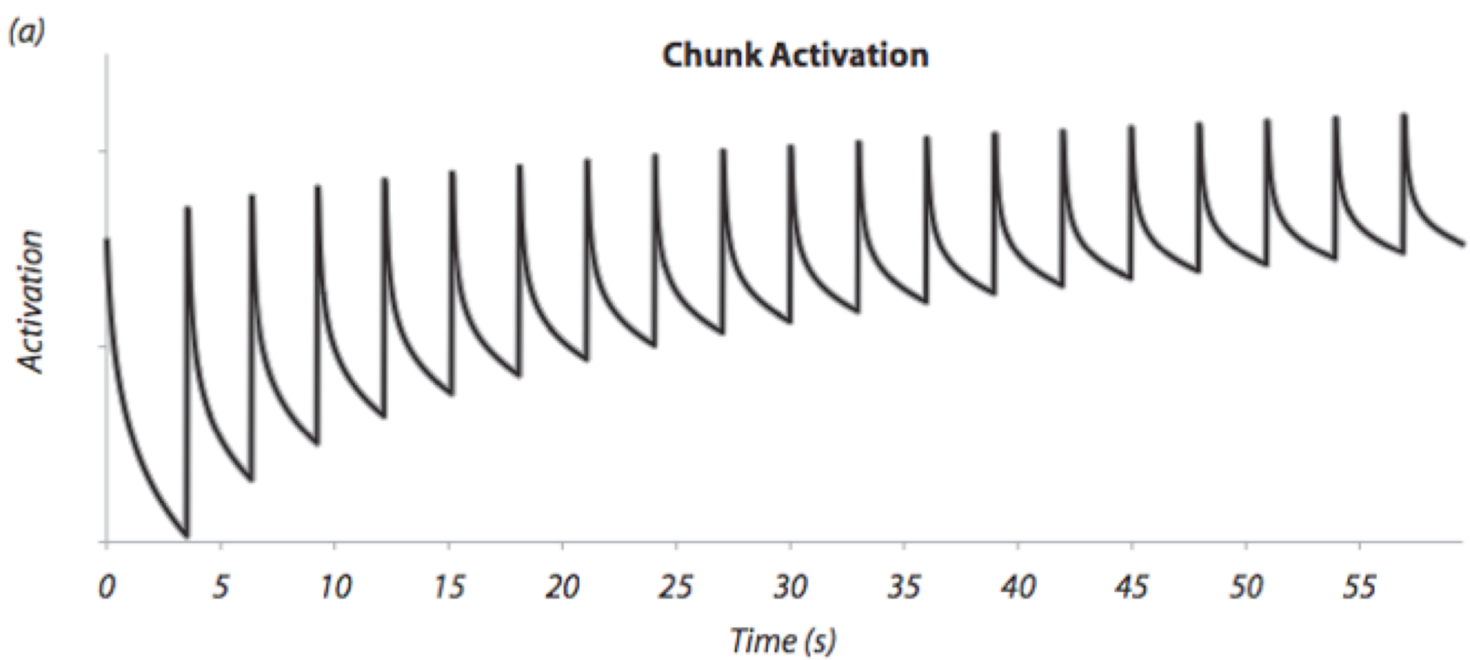
decay parameter

$$B_i(t) = \ln\left(\sum_{k=1}^{\boxed{n}} \boxed{(t - t_k)}^{\boxed{-d}}\right)$$

time since *n*th presentation

Base-Level Activation

$$B_i(t) = \ln\left(\sum_{k=1}^{n}(t - t_k)^{-d}\right)$$



Retrieval Time

- F is a parameter giving the time scale (called the latency factor)
  - defaults to 1.0
- A is the activation of the retrieved chunk
- Retrieval Threshold τ
  - A ≥ τ , or retrieval fails (causes "state error")
  - defaults to 0.0

$$Time = Fe^{-A_i}$$

Retrieval Probability

- Given a chunk's activation, what's the probability that it can be recalled?
  - Based on threshold and noise

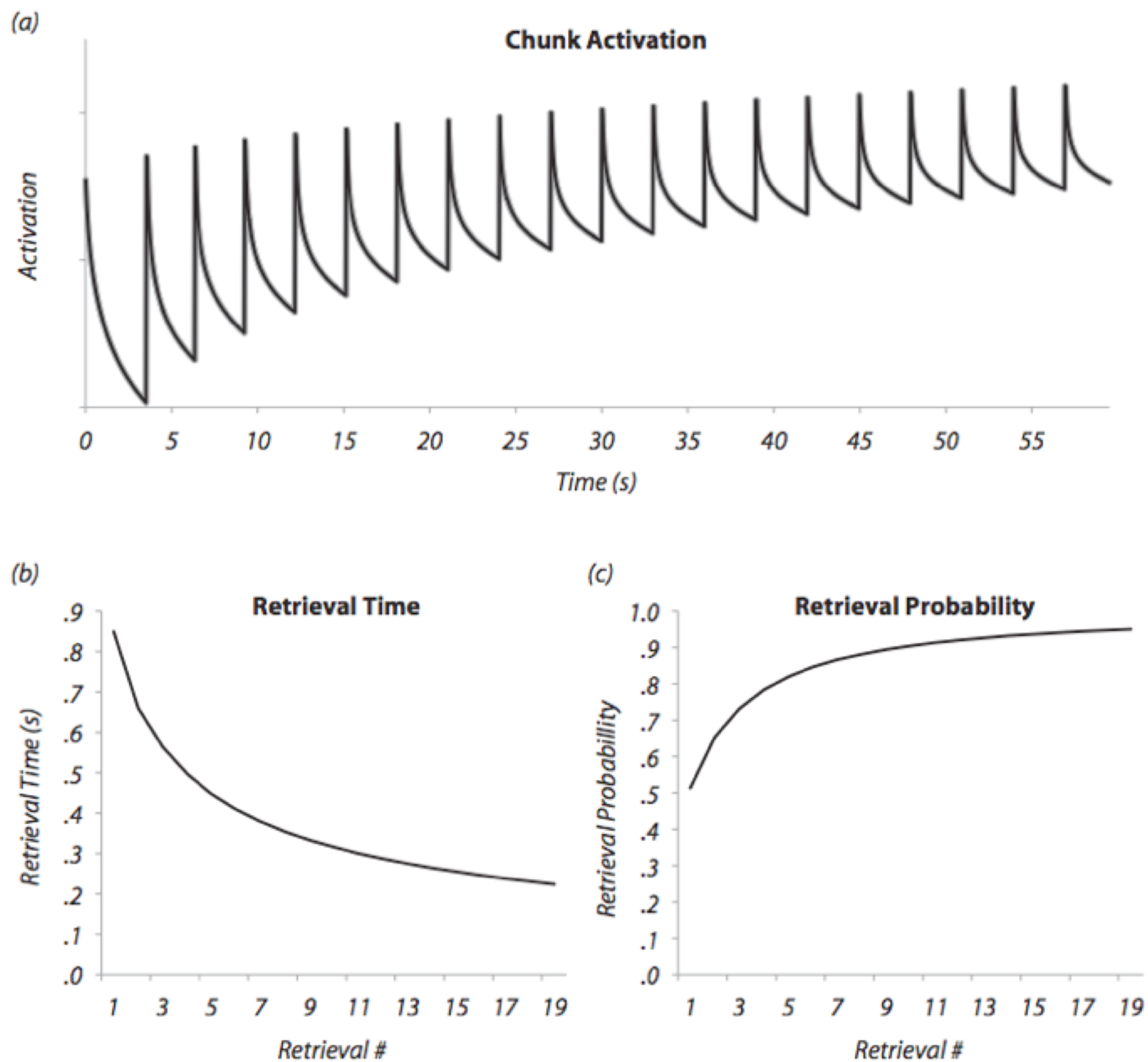
retrieval threshold

$$recall\ probability_i = \frac{1}{1 + e^{\frac{\boxed{\tau} - \boxed{A_i}}{\boxed{s}}}}$$

activation

noise parameter

Activation


(a)
Chunk Activation
Activation
0
5
10
15
20
25
30
35
40
45
50
55
Time (s)
(b)
Retrieval Time
Retrieval Time (s)
.9
.8
.7
.6
.5
.4
.3
.2
.1
.0
1
3
5
7
9
11
13
15
17
19
Retrieval #
(c)
Retrieval Probability
Retrieval Probability
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
.0
1
3
5
7
9
11
13
15
17
19
Retrieval #

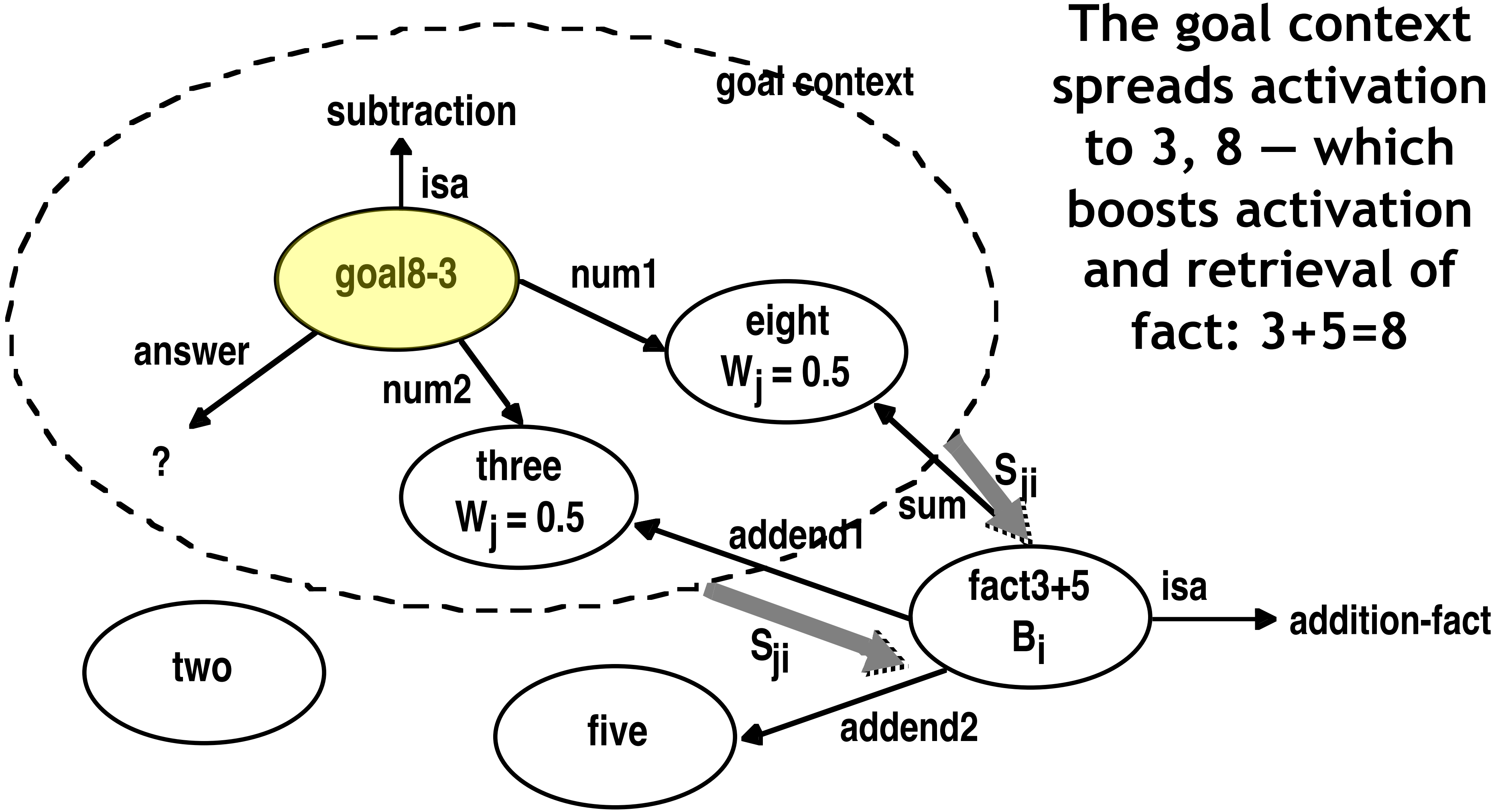
Spreading Activation

- Spreading activation reflects increased probability of needing a chunk if it is associated with the current "context"

$$A_i = B_i + \boxed{C_i} + M_i + \varepsilon$$

spreading
activation

Spreading Activation



The goal context spreads activation to 3, 8 — which boosts activation and retrieval of fact: 3+5=8

World Knowledge

- Past efforts have generally relied on hand-constructed knowledge (e.g., Cyc, WordNet)
- Instead, what if we derive knowledge from the largest knowledge source in the world: Wikipedia
  - More specifically, from DBpedia — a structured representation of information in Wikipedia
  - Includes (1) relations, (2) types, and (3) names (described next)

Knowledge Content

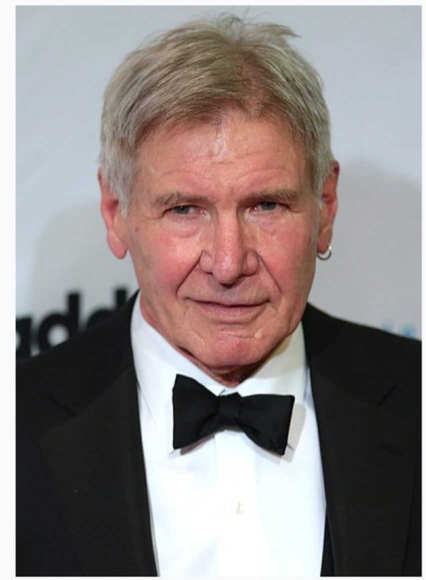
- (1) Relations
  - Wikipedia includes “infoboxes” with basic information about an item
  - DBpedia contains a cleaned-up version of this information
  - This information is translated into relation chunks as triplets <object, attribute, value>

Harrison Ford	birth date	1942-07-13
Harrison Ford	birth place	Chicago
Harrison Ford	spouse	Calista Flockhart

Raiders of the Lost Ark	starring	Harrison Ford

Star Wars Episode IV	starring	Harrison Ford

Harrison Ford



Harrison Ford in 2017

Born	July 13, 1942 (age 76) Chicago, Illinois, U.S.
Occupation	Actor · producer
Years active	1966–present
Spouse(s)	Mary Marquardt (m. 1964; div. 1979) Melissa Mathison (m. 1983; div. 2004) Calista Flockhart (m. 2010)
Children	5

Knowledge Content

- (1) Relations

Alabama	demonym	Alabamian
Alabama	capital	Montgomery, Alabama
Alabama	largest city	Birmingham, Alabama
Alabama	area total	1.35765E11
Abraham Lincoln	birth date	1809-02-12
Abraham Lincoln	birth place	Hardin County, Kentucky
Abraham Lincoln	death date	1865-04-15
Abraham Lincoln	resting place	Oak Ridge Cemetery
Abraham Lincoln	spouse	Mary Todd Lincoln
Algeria	anthem	Kassaman
Algeria	currency	Algerian dinar
Algeria	capital	Algiers
Algeria	official language	Arabic
Algeria	official language	French language
Amphibian	kingdom	Animal
Amphibian	phylum	Chordate

Knowledge Content

- (2) Types: Wikipedia "types" translate easily to an *isa* property…

Aristotle	isa	person
Aristotle	isa	philosopher
Abraham Lincoln	isa	person
Abraham Lincoln	isa	politician
Abraham Lincoln	isa	president
Algeria	isa	populated place
Algeria	isa	country
Amphibian	isa	species
Amphibian	isa	eukaryote

Knowledge Content

- (3) Names: Wikipedia "redirects" can be used as a name property linked to the canonical symbol…

Mary (mother of Jesus)	name	"Madonna"
Madonna (entertainer)	name	"Madonna"
Madonna (entertainer)	name	"Lourdes Leon Ciccone"
Madonna (entertainer)	name	"Madonna Louise Veronica Ciccone"
Madonna (entertainer)	name	"Madonna Ciccone"
Muammar al-Gaddafi	name	"Muammar Qaddafi"
Muammar al-Gaddafi	name	"Mohammar Qaddafi"
Muammar al-Gaddafi	name	"Gadaffi"
Muammar al-Gaddafi	name	"Gadhafi"
Muammar al-Gaddafi	name	"Gaddafi"
Muammar al-Gaddafi	name	"Muammar Gaddafi"
Muammar al-Gaddafi	name	"Mu'ammar Al Qathafi"
Muammar al-Gaddafi	name	"Moammar Ghadafi"

Knowledge Accessibility

- The core knowledge treats all facts equally
  - e.g., facts about Bob Dylan vs. <local artist>
  - e.g., facts about New York City vs. <small town>
- How can we estimate the accessibility of facts with respect to the "average" person?
  - (ignoring individual variability as a first cut)
- Accessibility as activation
  - Base-level activation → general accessibility
  - Associative activation → accessibility in context

Knowledge Accessibility

- Base-level activation
  - Assume that a person's exposure to each fact is proportional to N = # times the object appears in the core knowledge base
  - Derivation yields:
    Base-level activation = log(2N)
  - Example: retrieve
    "isa musician"
    (out of >37,000)

    David Bowie
    Prince
    Bob Dylan
    Kanye West
    James Brown
  - [For now, all chunks about an object have the same base-level activation; we need more data to do better.]

Knowledge Accessibility

- Associative activation
  - Given a "context" of symbols, this defines how the context boosts activation of related facts
  - Context thus guides retrieval
    - e.g., What is the capital of New York?

$$A_i = B_i + \sum_k \sum_j W_{kj} S_{ji} + \varepsilon$$

$$S_{ji} = S - \ln(fan_j)$$

New York	name	"New York"
New York	capital	Albany, New York
…		
New York City	name	"New York"
New York City	leader title	Mayor
…		

Examples

- We can build a cognitive model to demonstrate usage of the declarative knowledge base
  - E.g., a simple question-answering system that parses questions and retrieves answers
  - when a phrase is heard, retrieve "name" chunk to map phrase → symbol (with context)
    - e.g., "New York" → *New_York_City* or *New_York*
  - then, retrieve the answer to the question (guided by both base-level and associative activation)

Examples

- Q1: What is the capital of Pennsylvania?
  Q2: What is the population of Quebec City?
  - straightforward retrieval of a single fact
- A1: Harrisburg, Pennsylvania
  A2: 516,622

- Q: What is Philadelphia?
  - the city? the film? the magazine?
  - base-level activation guides the most likely response
- A: city, place

Examples

- Q1: Who is the author of No Country for Old Men?
  Q2: Who is a star of No Country for Old Men?
  - “author” and “star” guide correct retrieval, book vs. film
- A1: Cormac McCarthy
  A2: Tommy Lee Jones, Javier Bardem

- Q1: What actor is a star of Airplane?
  Q2: What athlete is a star of Airplane?
  - “star” guides retrieval to film interpretation of “Airplane”
  - “actor” and “athlete” guide retrieval of the responses
- A1: Robert Hays, Leslie Nielsen
  A2: Kareem Abdul-Jabbar

Examples

- Q1: Where is the Baseball Hall of Fame?
  Q2: Who is Theodore Geisel?
  – name attribute guides retrieval to the canonical forms
- A1: Cooperstown, New York [National_Baseball_Hall_of_Fame_and_Museum]
  A2: cartoonist [Dr._Seuss]

- Q1: Name an actor born in Philadelphia.
  Q2: Name a musician born in New Jersey.
  – “actor” and “musician” guide retrieval to these types
  – base-level activation guides to well-known people
- A1: Bill Cosby
  A2: Bruce Springsteen

Summary

- Cognitive models are a powerful tool for exploring how humans think
  - Remember this 2x2 matrix to understand AI?

Thinking Humanly	Thinking Rationally
Acting Humanly	**Acting Rationally**

  - We've focused almost entirely on **Acting Rationally**
  - Cognitive models focus on **Thinking Humanly**
- Next time, we'll look at some practical applications of cognitive modeling…