Cognitive Models

CS 347
Michael Bernstein

Last time

Visualizations can be represented as **encodings** that map from **data to marks & visual attributes** based on **data types**

Our **cognitive and perceptual systems determine which encodings are effective**: we (mis)read data if encoded poorly

Active research at frontiers investigating **how users can create effective visualizations** and **how readers take information away from them**

Today

The model human processor

- GOMS, KLM

Generative AI simulation models

Building a better mouse(trap)

[Card and Moran 1988]

Doug Engelbart and Bill English felt that their **mouse was an interim device**, and wanted to make something better

But none of their inventions were actually improving target acquisition speeds

So, Stu Card and Tom Moran tested the mouse in the lab on a variety of pointing tasks

Building a better mouse(trap)

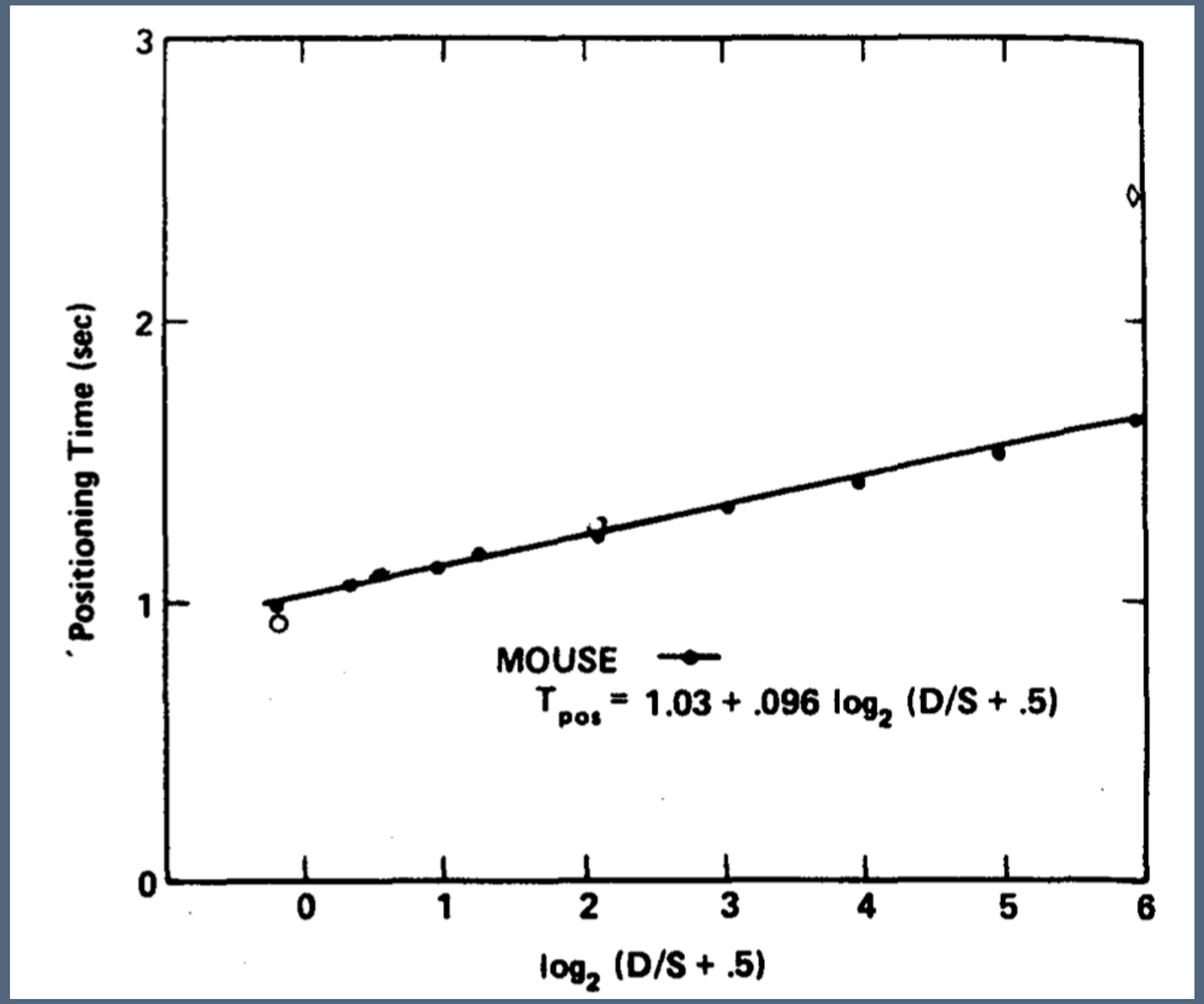
[Card and Moran 1988]

Performance was very well modeled by **Fitts's Law**.
(Fitts's Law is about human pointing, not mouses.)

$T = a + b \log(D/S + 0.5)$, D = distance, S = target size

Moreover, the mouse's constant of proportionality (b = 0.96 sec/bit = 10.4 bits/sec) is approximately the same with the mouse as with the hand alone — **so the mouse is near optimal, you actually can't do better!**

Here, modeling solved a problem that engineering couldn't.



Line = Fitts's Law prediction
Dots = measured mouse time

"User technology includes hardware and software techniques […] but it must include a **technical understanding of the user** and of the nature of human-computer interaction. This latter part, **the scientific base of user technology**, is necessary in order to understand **why** interaction techniques are (or are not) successful, to help us **invent** new techniques, and to pave the way for machines that aid humans in performing significant intellectual tasks."

[Card and Moran 1988]

Model Human Processor

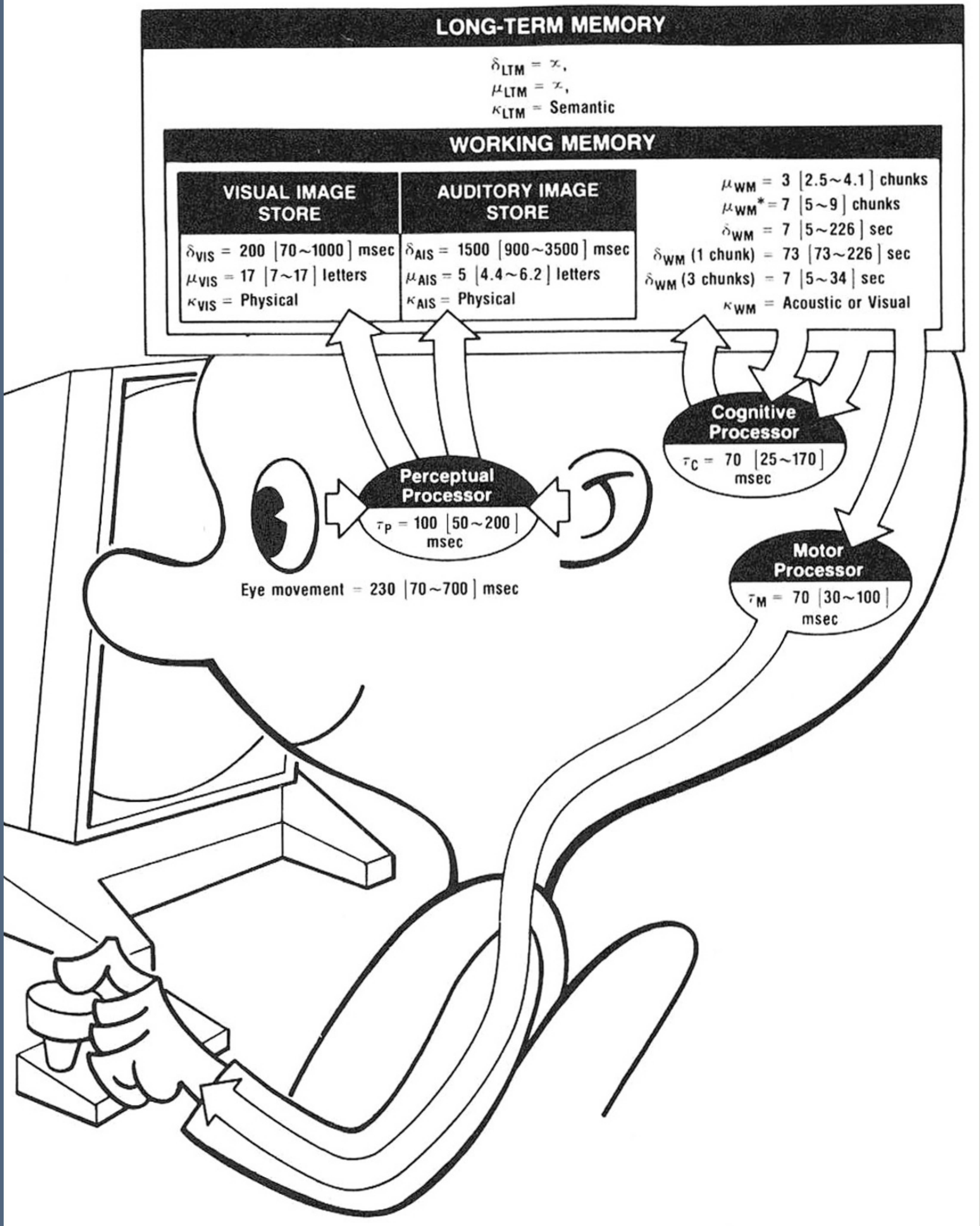
The Model Human Processor

[Card, Moran and Newell 1983]

A unified, low-level engineering model of user task completion

Based on empirical research, contains internal processors for cognition and motor movement

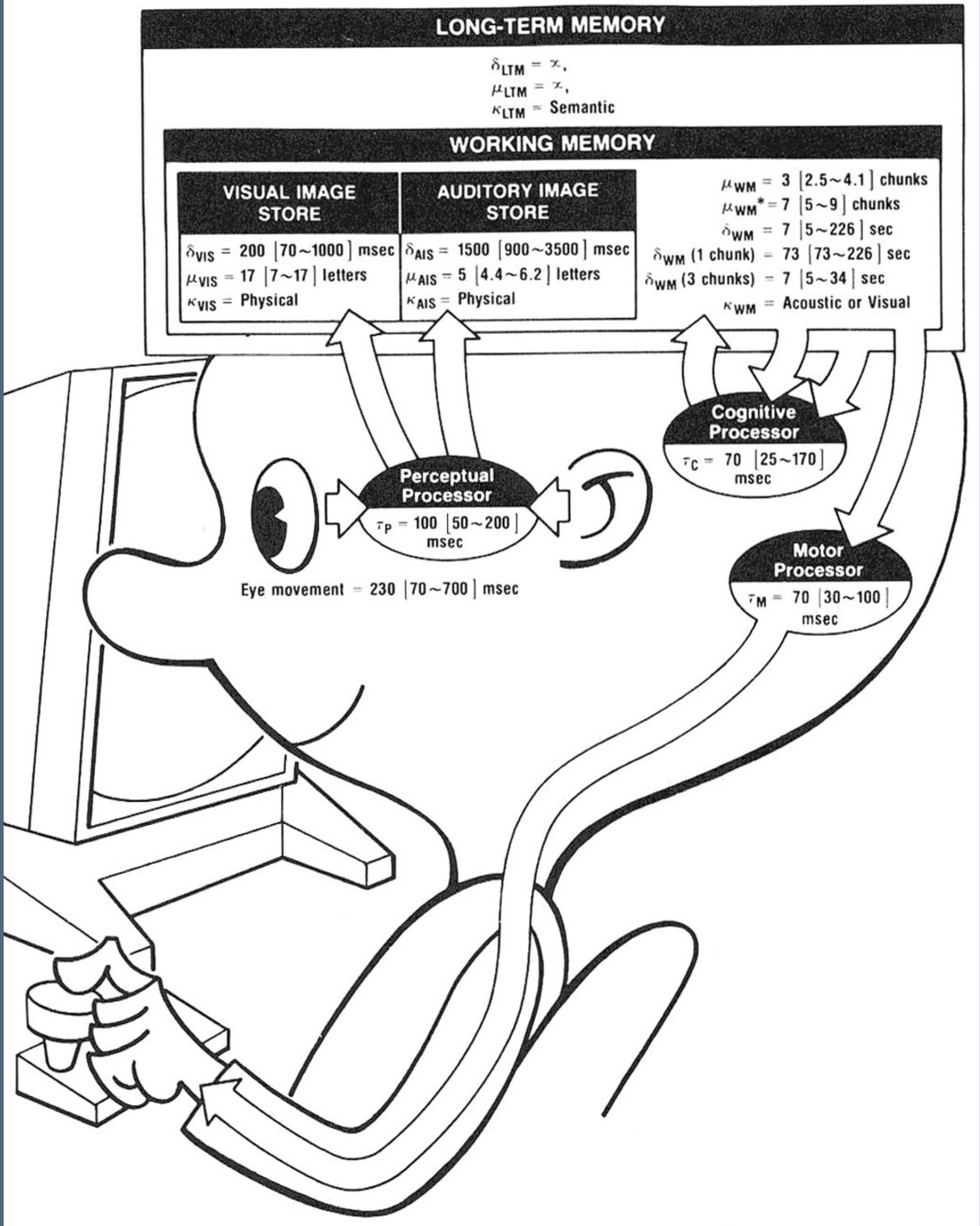
Similarly, empirically-estimated working and long term memory



Why Model Humans?

So we can **better understand why** what works works, and **why** what doesn't work is broken

- Apply MHP to **predict time and accuracy** of using interface
- Apply MHP as a **simulation of human user** (with constraints) **to evaluate** interface designs



What to pay attention to in this section

Track the **translation** from **empirical psychology** into **engineering decisions** in how the Model Human Processor works

The point is less to memorize specific results, but instead to understand how these results get translated, and why

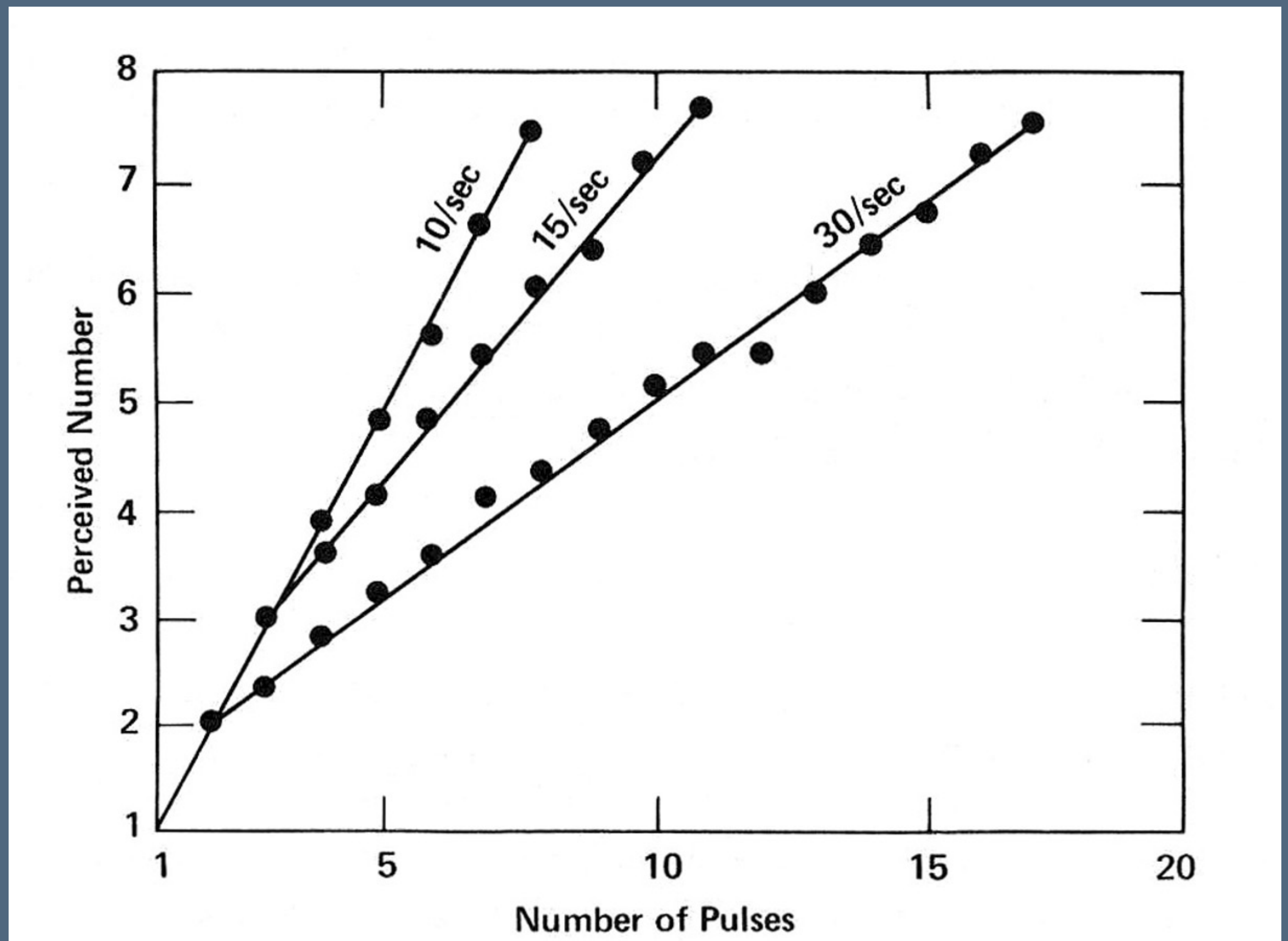
The result is, effectively, **a human "algorithm" that mimics the basics of our perceptual, motor, and cognitive behaviors**.

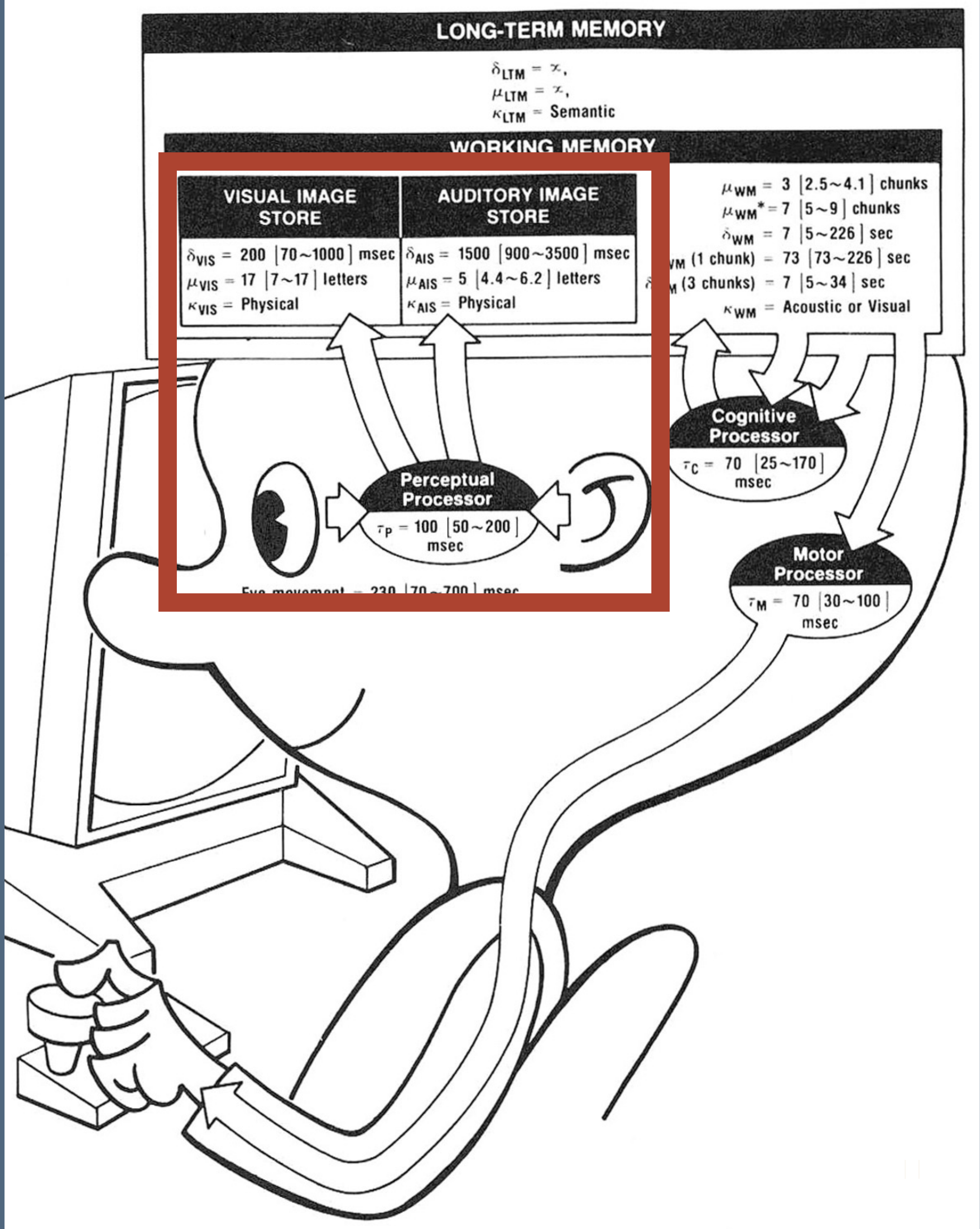
Perception

Time needed to integrate/fuse perceptual experience of the world

T_P = 100 msec ("quantum of experience")

= 10 fps: Rate needed for film to be perceived as continuous





Perception of Causality

[Michotte 1946]

What do you see?

Perception of Causality

[Michotte 1946]

What do you see?

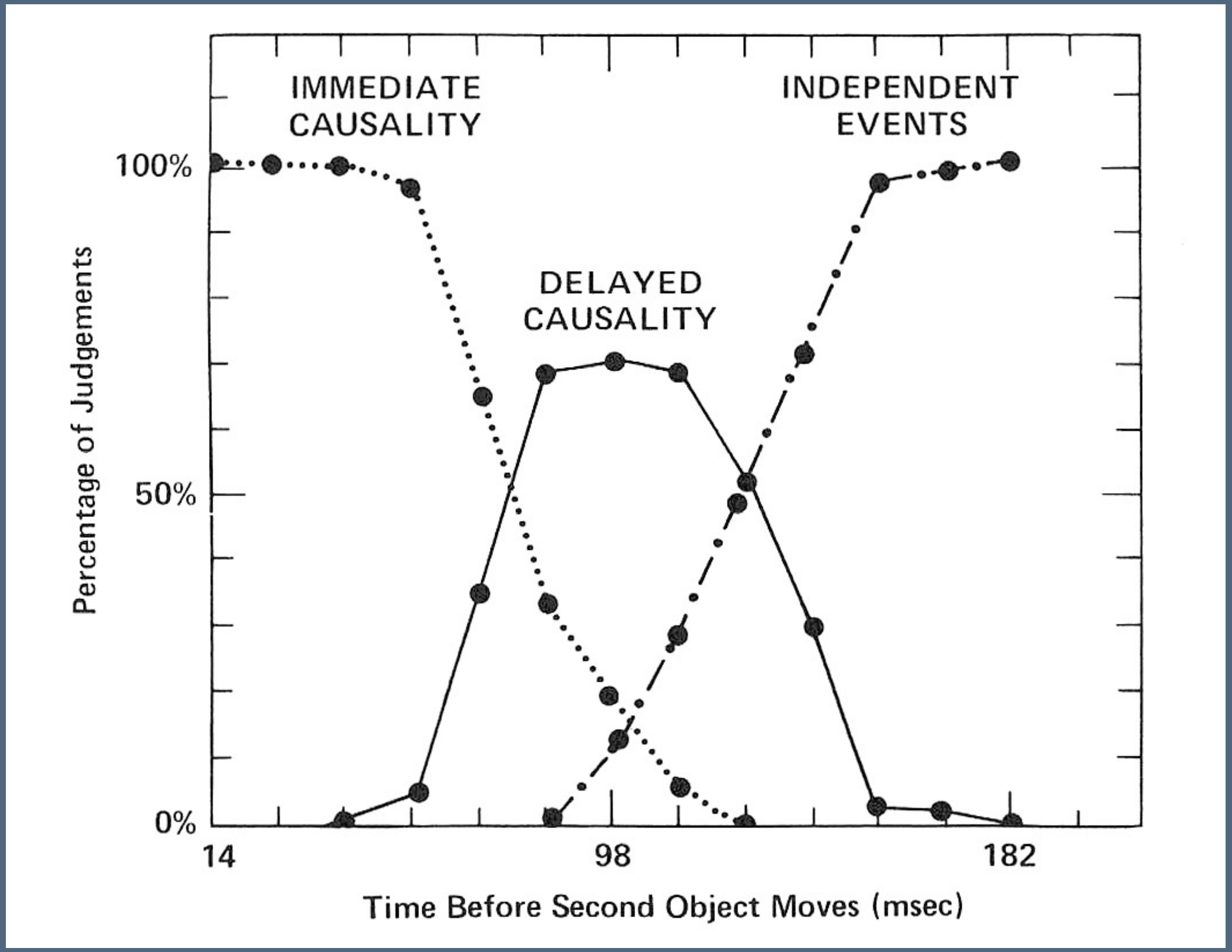
Perceptual Processor

Time needed to integrate/fuse perceptual experience of the world

T_P = 100 msec (“quantum of experience”)

= 10 fps: Rate needed for film to be perceived as continuous

= **rate needed to imply causality**



Memory

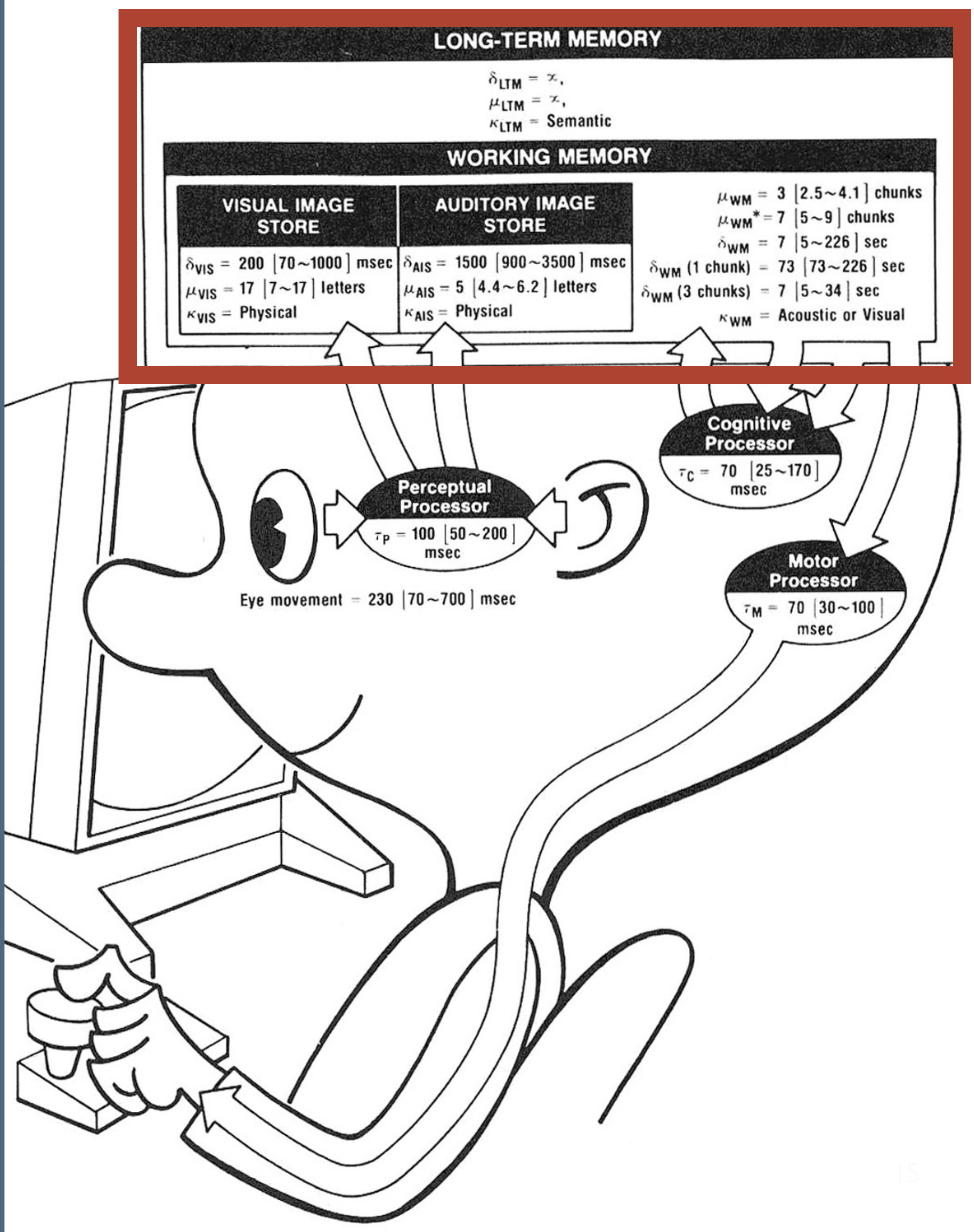
Perceptual processor puts information into (vis/aud/…) **sensory store**

- very fast decay 200-1500 msec
- small units of information (e.g. letters)

Some info then **chunked** and put into longer decay **working memory**

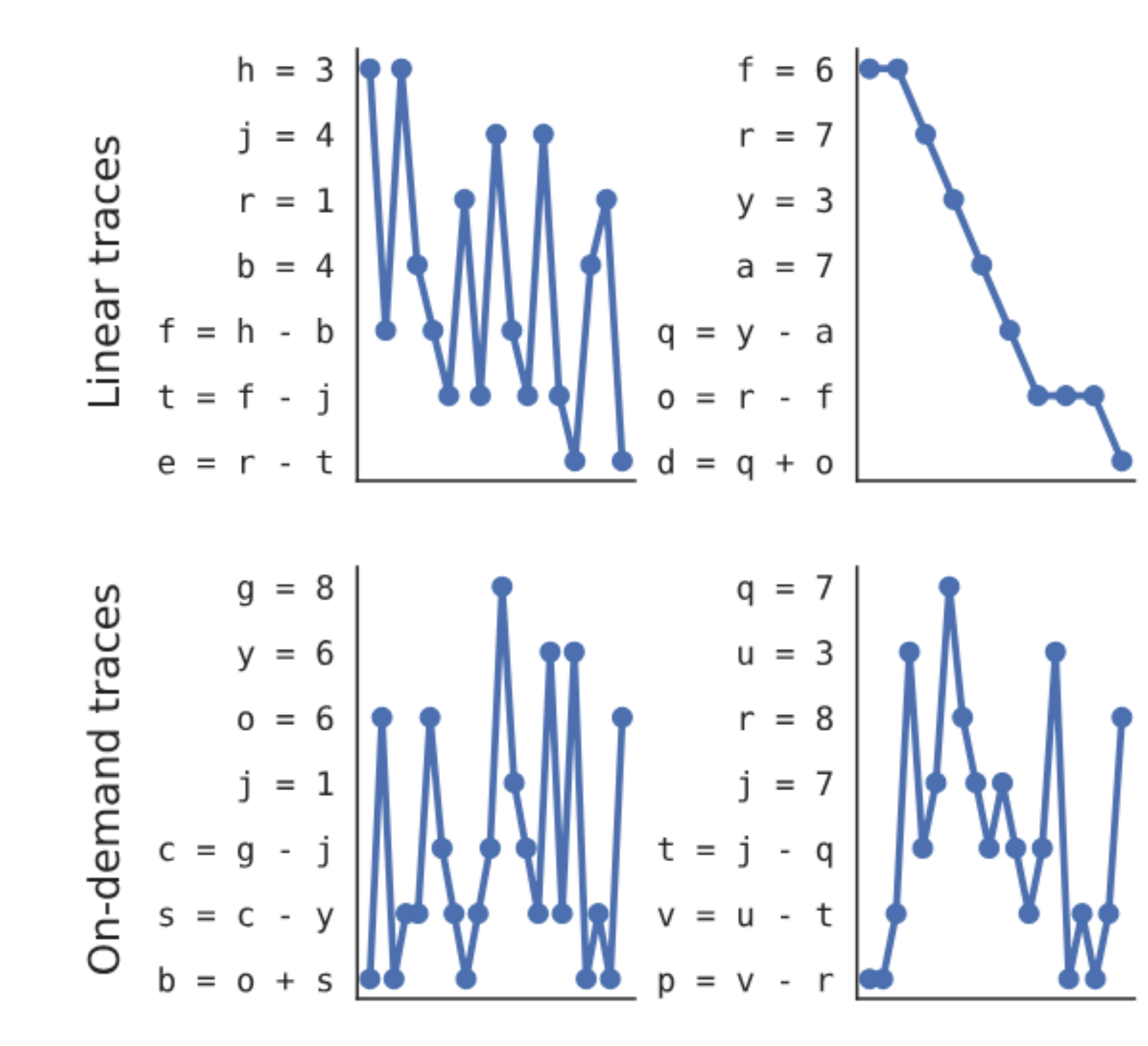
- decay 5-225 sec (content dependent)
- 7 +/- 2 chunks (e.g. words)

Some info then recoded (semantically) and put into **non-decaying long-term memory**



WM and Program Tracing

[Crichton, Agrawala, Hanrahan 2021]



Examines how people trace simple programs

- Order in which lines are exposed (linear vs. on-demand)
- How often need to re-visit a line already seen

WM holds ~7 (variable, value) pairs

Both linear and on-demand orderings frequently used

People make different WM errors depending on ordering strategy with more errors using on-demand

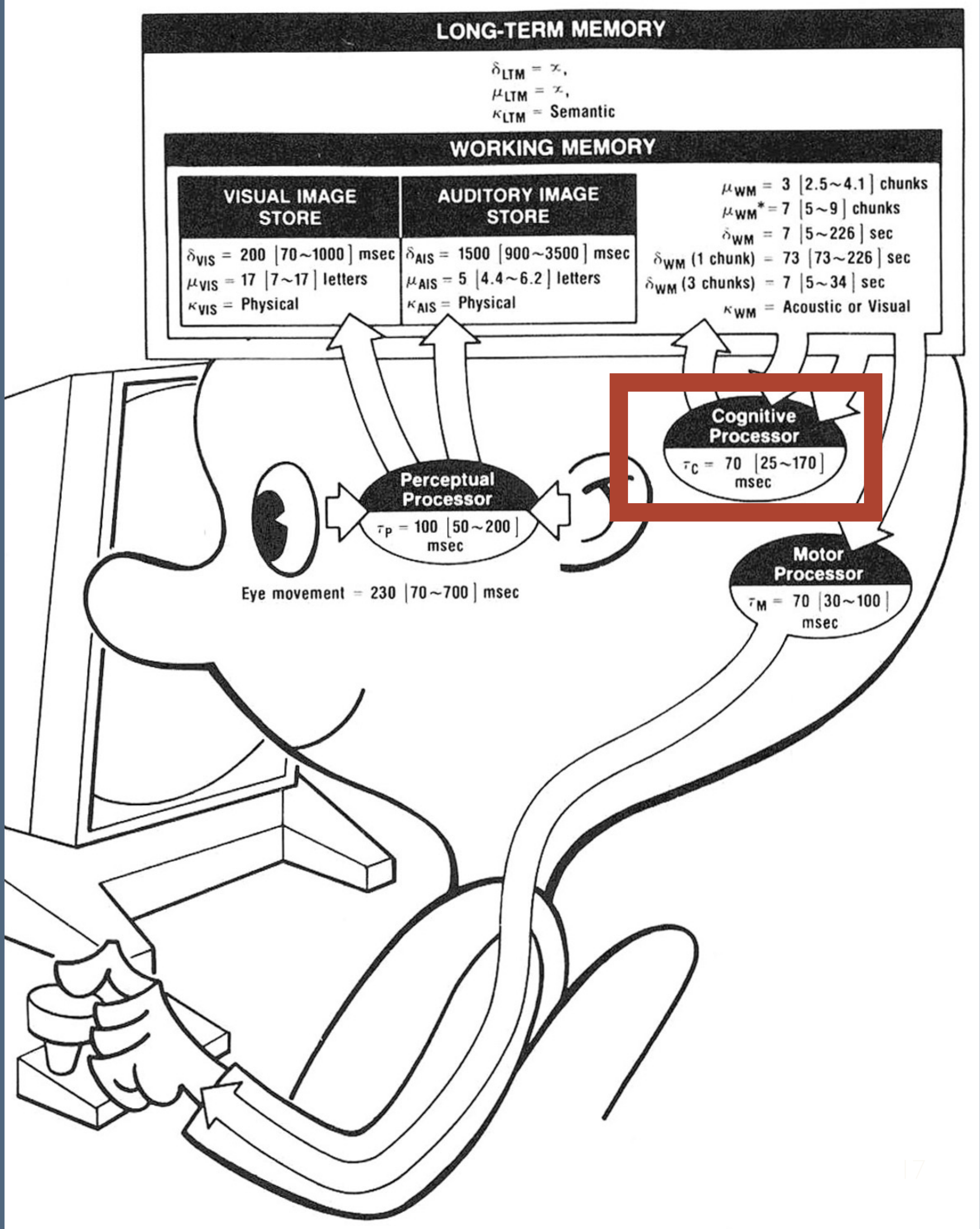
Cognition

Time needed to observe WM and operate on it (e.g. check if 2 chunks match)

T_C = 70 msec

Fundamentally serial

One locus of attention at a time

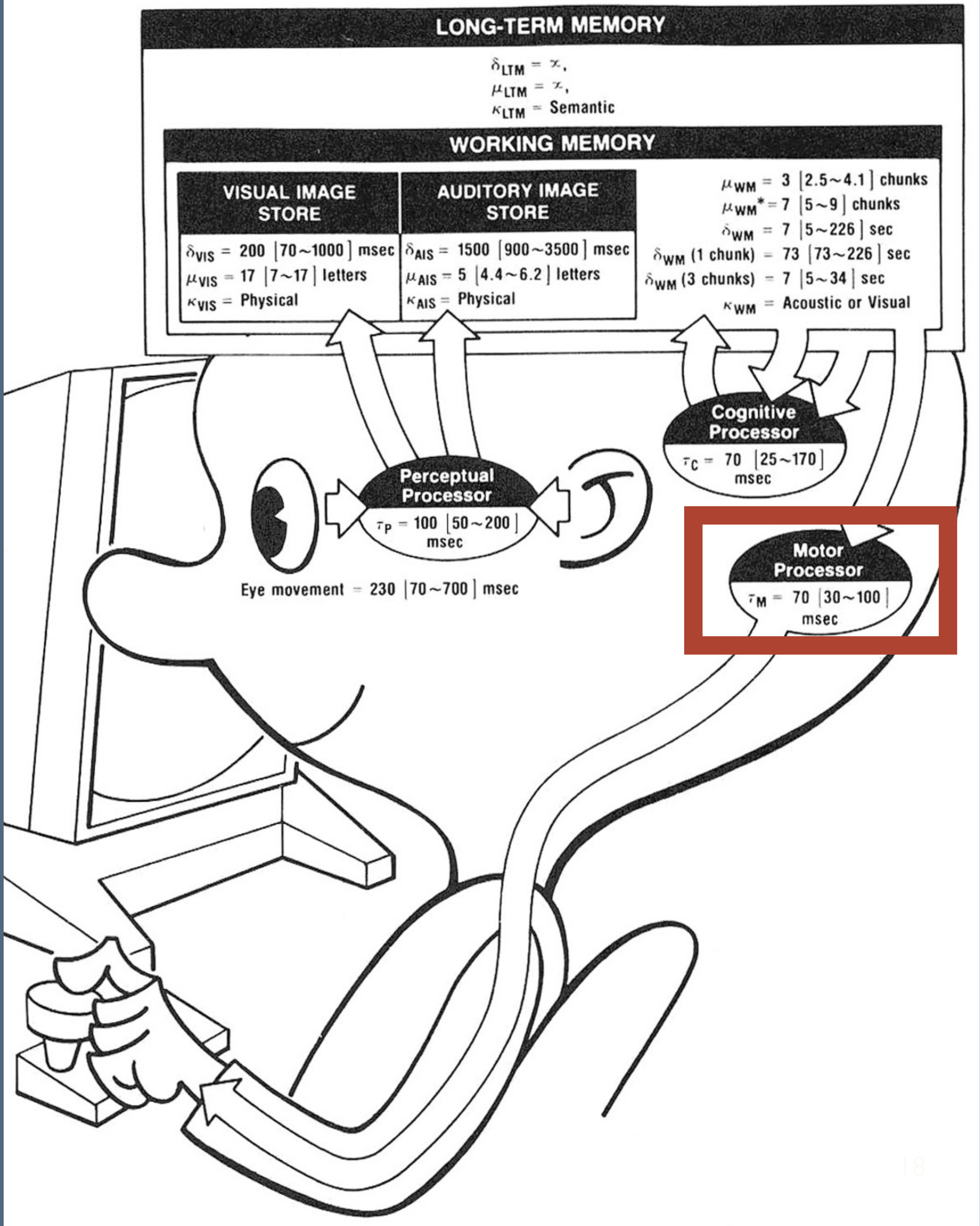


Motor Proc.

Time needed to take input command from cognitive processor and execute it with body

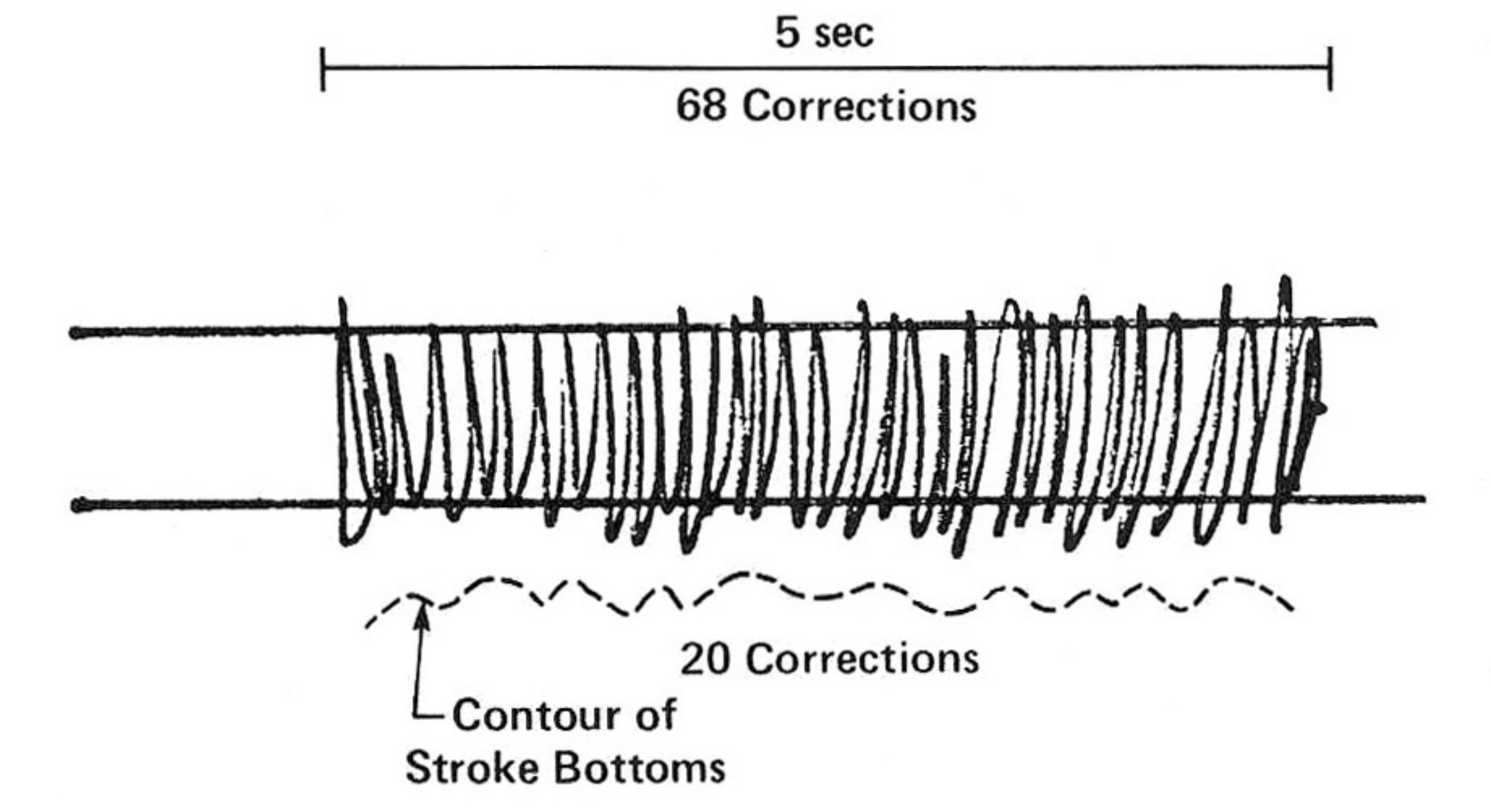
T_M = 70 msec

Example: pianist (up to 16 finger movements/sec)



Motor Experiment

Ask person move pen back and forth as quickly as possible:



Open loop: 68 reversals/5sec = 74 msec/reversal

Closed loop: Participant perceives if stroke is staying within lines, sends info to the cognitive processor, which advises the motor processor to adjust.

Predicted time = $\mathbf{T_P} + \mathbf{T_C} + \mathbf{T_M}$ = ~240msec

Empirical time = 20 corrections/5 = 250msec

Using the Model Human Processor

Low level task: I will flash 2 symbols ***x*** and **y** on screen serially, press a key if they are both numbers

Clocks starts when 2nd symbol **y** is flashed Move symbol **y** into visual store WM	Tp
Recognize both symbols ***x*** and **y** as codes	+ Tc
Classify the both codes as numbers	+ Tc
Match the fact that they are both numbers	+ Tc
Initiate motor response	+ Tc
Process motor command	+Tm

Approx 450 (180-980) msec

GOMS

Goals: what the user seeks to achieve

Operators: low-level operations

Methods: compositions of operations together

Selection rules: how to decide between multiple available methods

Given this specification, a system can trace a path that a user would take through a system to achieve their goal and report how long it would take

KLM [Card, Moran and Newell 1980] [Raskin 2000]

Keystroke Level Model: a specific model in the GOMS family. Designed to be quick and easy to use, no need to build a prototype.

Provides a bunch of operators and methods: not GOMS from scratch

Six operators: push a key, point to a target on the display, moving hands between keyboard/mouse/etc., drawing a line (seems extraneous to me), making a decision about the next step, waiting for system response

Operator	Time
Key/Click	0.20
Point	1.1
Homing	0.4
Draw	$.9n_D + .16\, l_D$
Mental	1.35
Sys. **R**esp.	Depends

KLM [Raskin 2000]

Example [via Wael Aboelsaadat]: minimize a window by clicking on the button, or using the keyboard shortcut (hands initially on mouse)?

Clicking on the button:

Mentally prepare, Point, Click:
MPK = 1.35 + 1.1 + 0.2 = 2.65sec

Keyboard shortcut:

Home to keyboard, Mentally prepare, Key (type) cmd-M:
HMKK = 0.4 + 1.35 + 0.4 = 2.15sec

Operator	**Time**
Key/Click	0.20
Point	1.1
Homing	0.4
Draw	$.9n_D + .16\, l_D$
Mental	1.35
Sys. **R**esp.	Depends

Raskin's KLM Rules

First break task into H,P, K,D, R (then use rules)

R0: Insert M

In front of all K

In front of all P's selecting a command (not setting args)

R1: Remove M btw fully anticipated operators

PMK to PK

R2,R3: if MKs form cog. unit delete all Ms except first

typing "4564.23": MKMKMKMKMKMKMK to MKKKKKKK

typing "enter" "enter": MKMK to MKK (redundant terminator)

R4: if K terminates freq. used fixed length string (e.g. cmd) delete M in front of it

typing "cd" "enter": MKKMK to MKKK

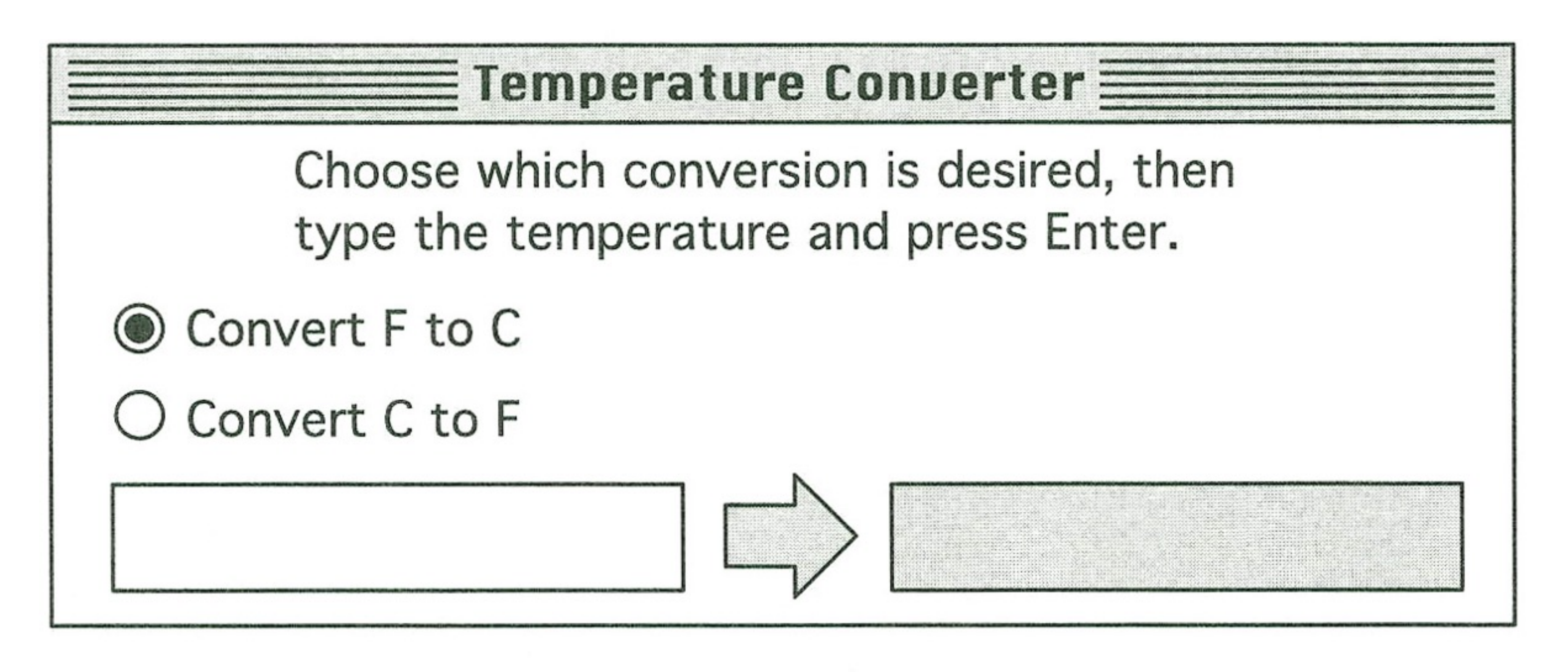
typing "cd" "class" "enter": MKKKMKKKKKMK (do not remove last M)

Operator	Time
Key/Click	0.20
Point	1.1
Homing	0.4
Draw	$.9n_D + .16\ l_D$
Mental	1.35
Sys. **R**esp.	Depends

Converting Temperature

Convert 92.5

Assume focus on dialog, hands at keyboard, typing enters text into text field



Assuming goal C to F

H PK H KKKK K to H MPMK H MKMKMKMK MK to H MPK H MKKKK MK (7.15sec)

Assuming goal F to C

KKKK K to MKMKMKMK MK to MKKKK MK (3.7sec)

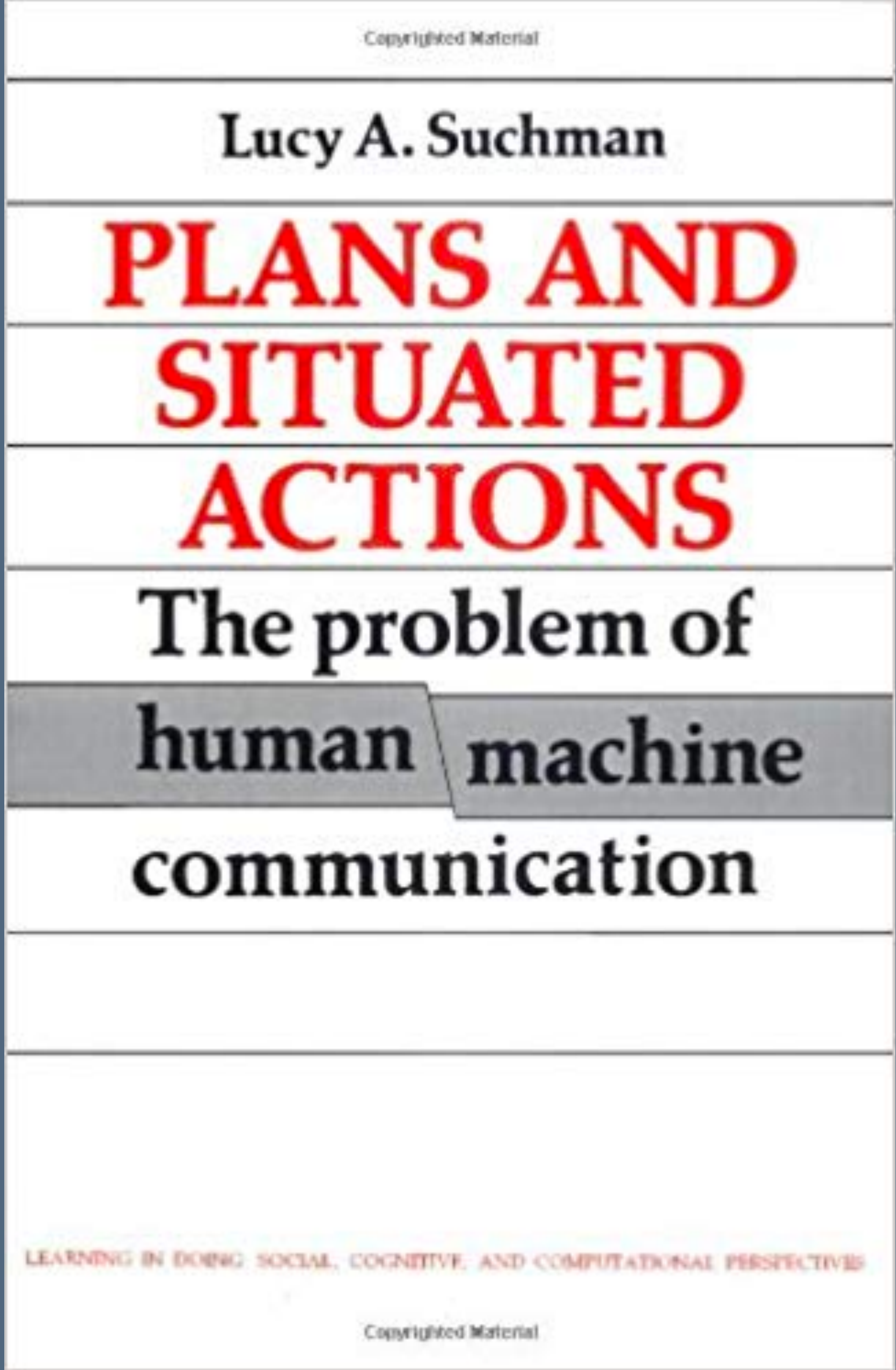
Avg time: 5.4sec

Where are they now?

Models as inhuman models of how we act

Plans cannot succeed in complex environments, which instead require constant reflection and reorientation [Suchman 1988]

Anthropological comparison: how people perform wayfinding



GOMS is Sensitive to Methods and Operators

In GOMS, the **developer defines operators** and **methods**. Need to be **careful** to make sure they are **appropriate to task** and **context**

> "There's no accounting for taste" — GOMS will not object to a baroque set of operations that a user might never use in practice

> Outcomes will depend strongly on exactly which operators and methods you define and make available to the model

GOMS Is (Relatively) Quantitative

GOMS explicitly **capture low-level cognitive behaviors** of interest **quantitatively**

- The Model Human Processor estimates were based on careful lab studies
- But **absolute numbers are less reliable than relative values**

Can be less work than a user study

Today's state

Low-level cognitive models (e.g. GOMS and KLM) have fallen out of favor, largely because they require **substantial effort** to create, vs. directly prototyping

However, for low-level optimizations and interface decisions, cognitive models can be very useful

And, they remain important to HCI as an example of how **grounding our designs in psychological methods** and results can lead to more effective approaches and insights

Generative AI Simulation Models

A Complementary Strategy

There are two strategies to creating models [Bruch & Atwell 2015]:

- Low-dimensional realism, with a **specific but interpretable** model like the Model Human Processor
- High-dimensional realism, with a **broad but uninterpretable model**

Modern large language models have learned much about human cognition, attitudes, and behavior through their training data: can they power this latter class of cognitive model?

AIs have trained on a range of human behaviors

Large language models can be prompted to take on a variety of backgrounds, experiences and traits [Park et al. 2022]

Generative agents [Park et al. 2023]

Generative agents [Park et al. 2023]

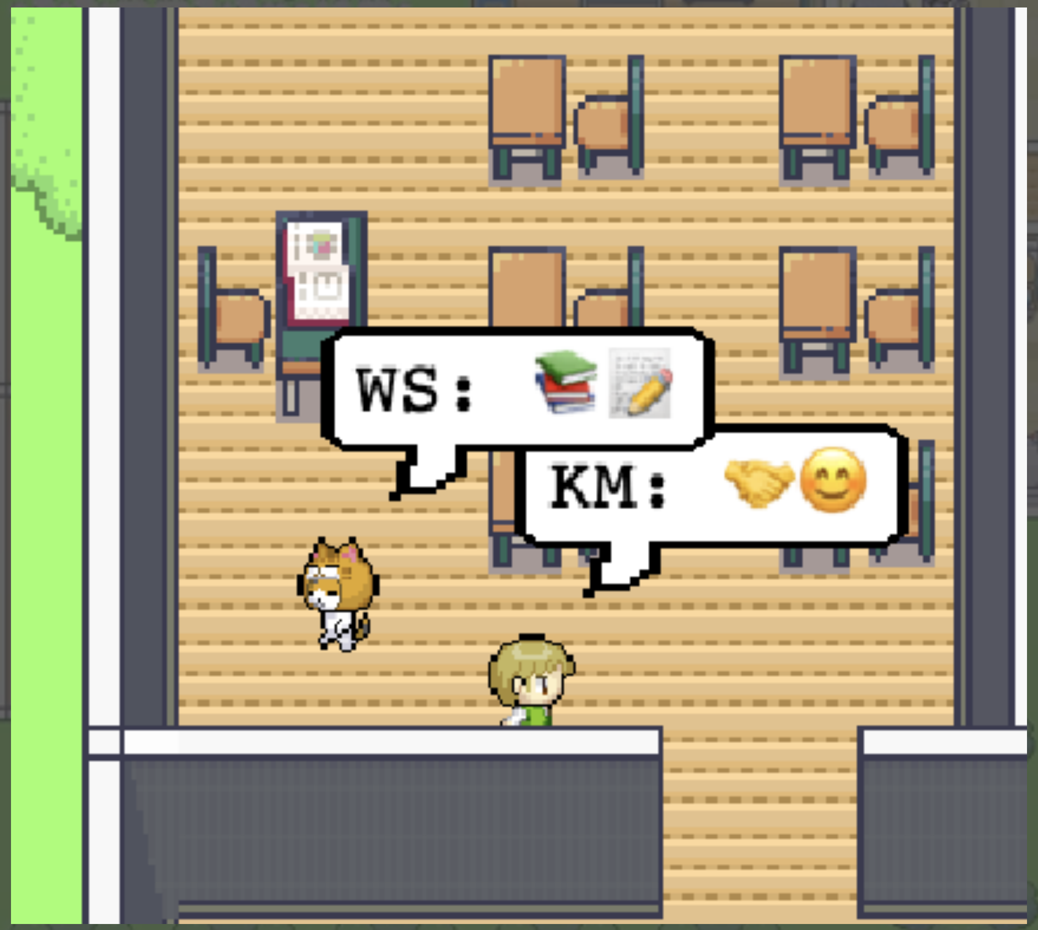
Agents that draw on generative models to simulate believable human behavior

A student athlete agent in the morning wakes up and:

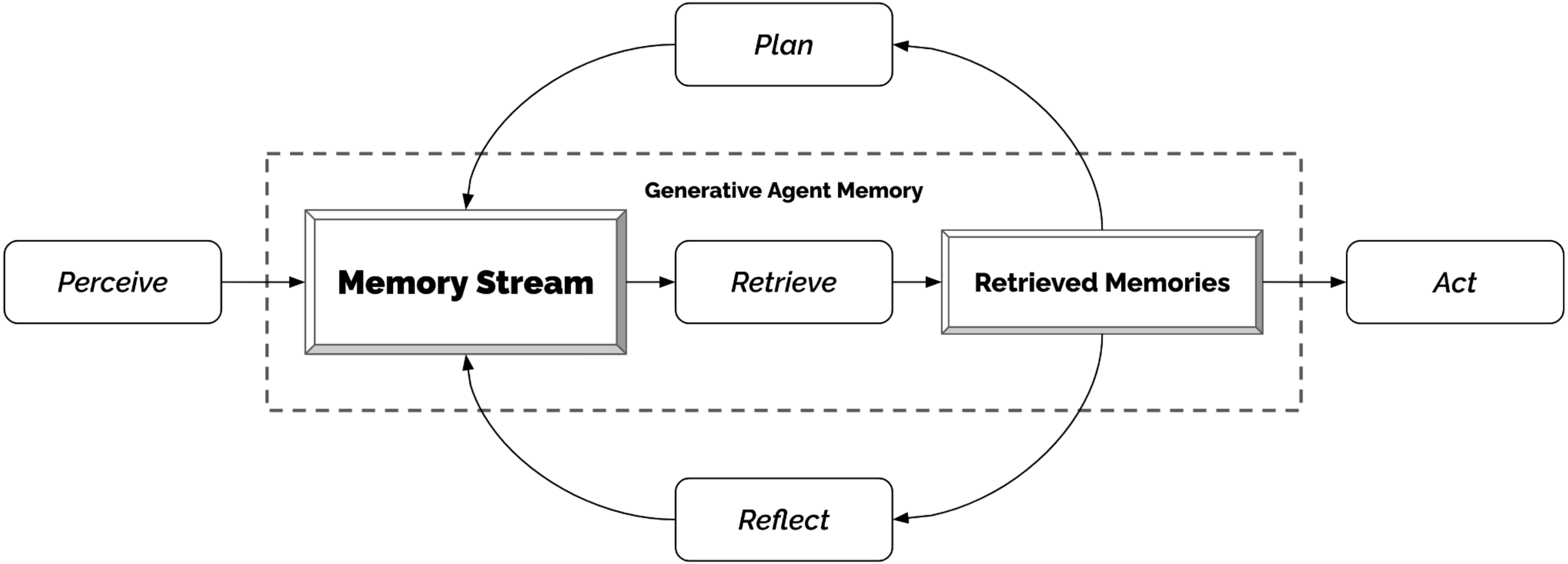
Brushes teeth

Goes for a run

Cooks breakfast



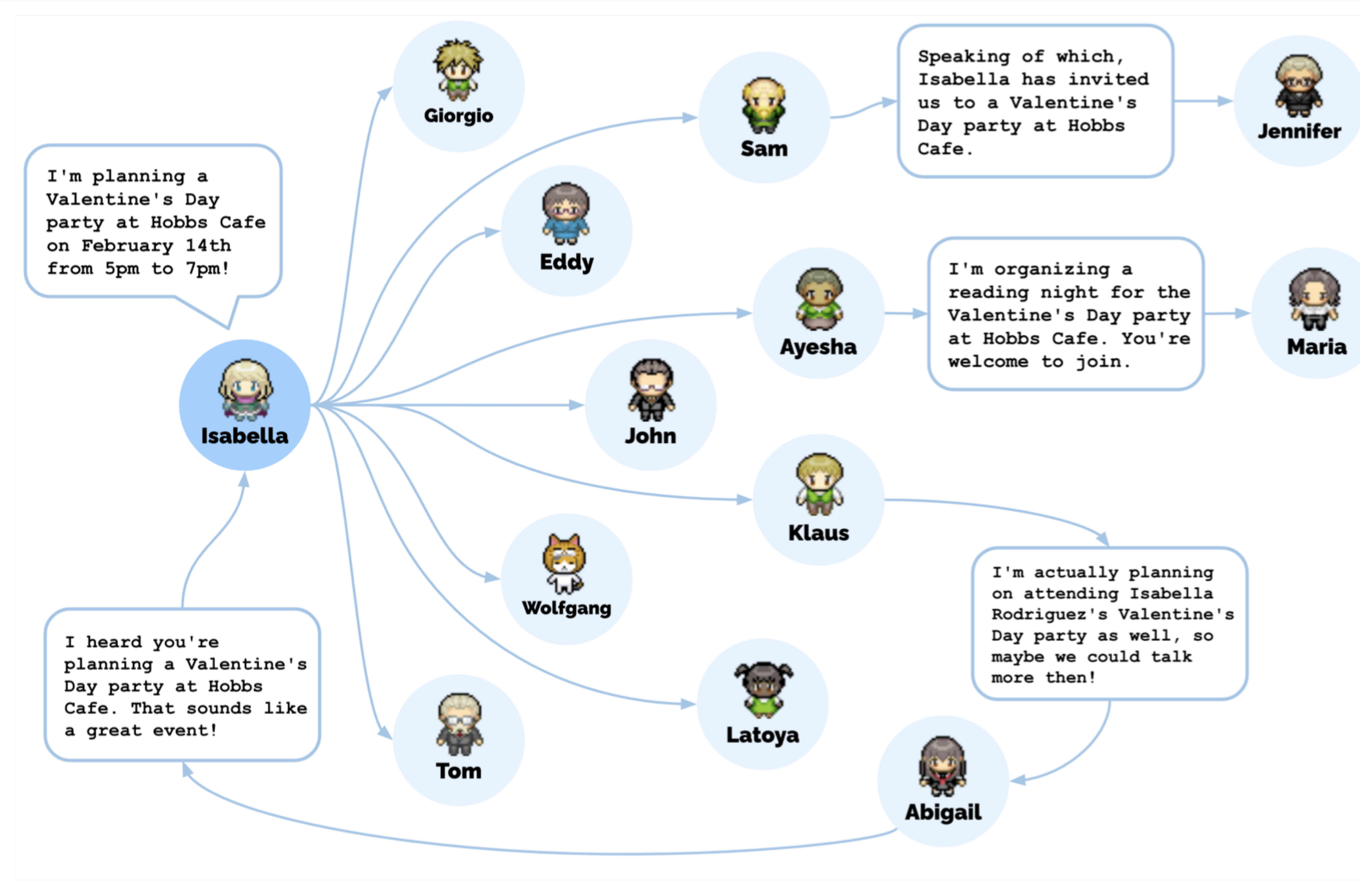
Heads to class


Plan
Generative Agent Memory
Perceive
Memory Stream
Retrieve
Retrieved Memories
Act
Reflect

Isabella is initialized with an intent to plan a Valentine's Day Party

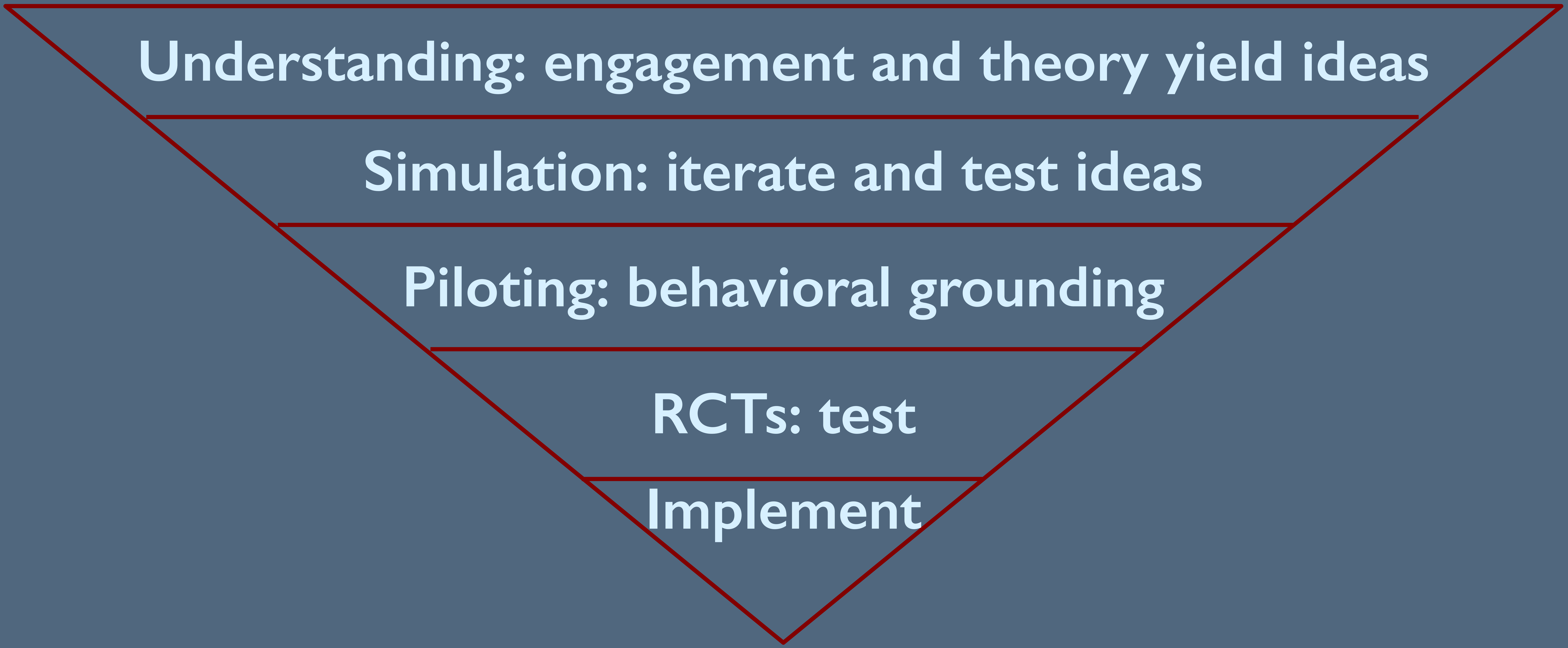
She tells customers at her cafe

- Twelve hear about it
- Five came
- Three cited conflicts: e.g., Rajiv, a painter: "I'm focusing on my upcoming show, and I don't really have time."
- Four showed interest but didn't come
- Maria asks her crush, Klaus, to the party



Simulation: iterate and test ideas

Implement


Understanding: engagement and theory yield ideas
Simulation: iterate and test ideas
Piloting: behavioral grounding
RCTs: test
Implement

Stepping back, what other models of cognition might inform HCI?

Thinking in the world

Cognition for ubiquitous computing environments

Recall: "Pictures Under Glass"

[Victor 2011]

Embodied cognition

[Dourish 2004; Klemmer, Hartmann, Takayama 2006]

Our cognition leverages **embodiment**—our bodies:

- We learn through interaction with the world
- We leverage the environments around us to make us smarter
- We communicate our intent through much broader mechanisms than just our fingertips: consider musicians, dancers, construction workers, professors on stage trying to get your attention

Epistemic action

[Kirsh and Maglio 1994]

Tetris as an example task to study cognition

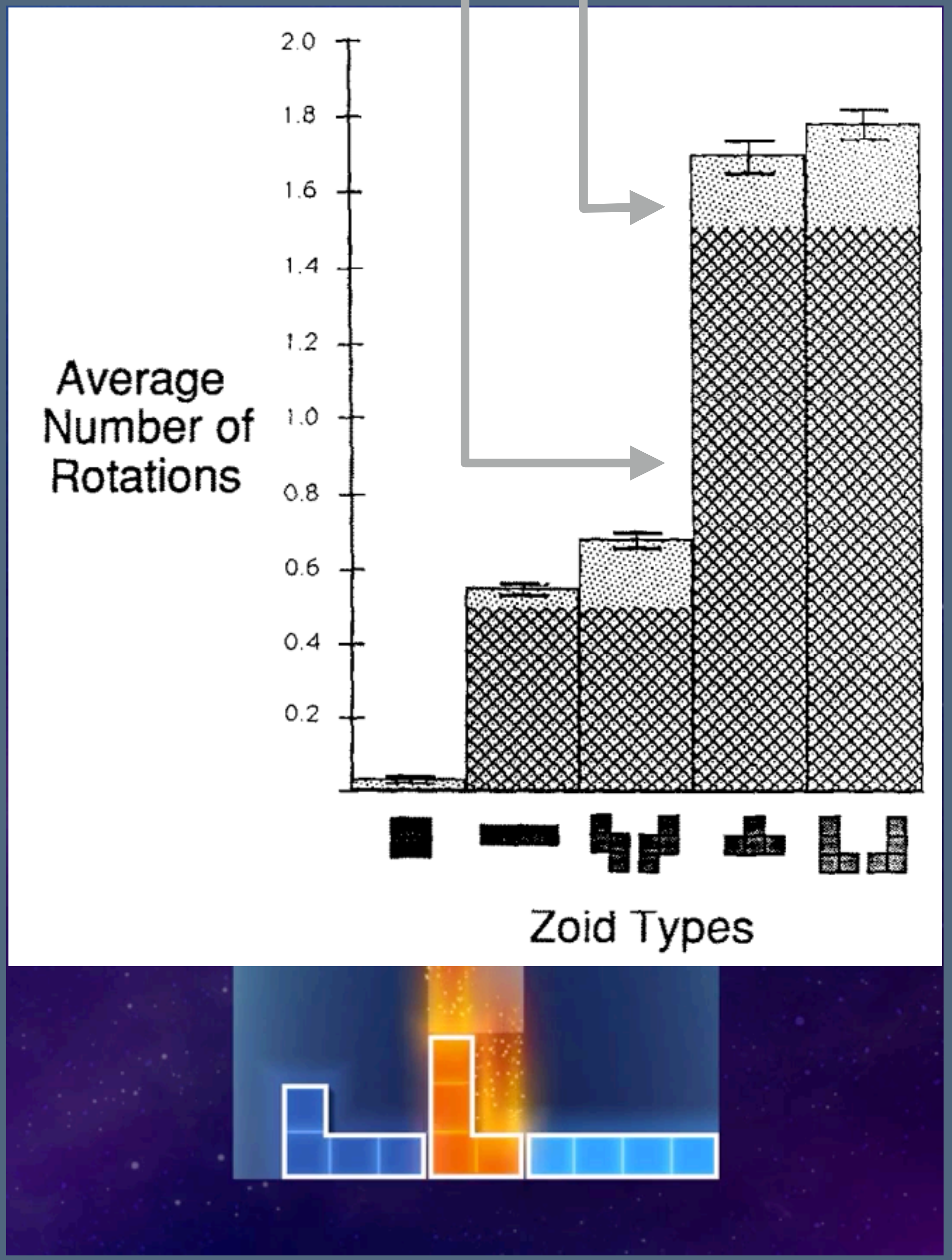
Players see a piece, rotate it, and drop it into position

However, experts perform more rotations than strictly needed to position the piece. Why?

We perform actions in the world to uncover information that is hard for us to compute mentally

Hatched area: required to position the piece

Gray area: extraneous rotations



Distributed cognition

[Hutchins 1995]

Theory: social and physical environments, not just people, can exhibit intelligence

Source: ethnography on the navigation bridge of Navy ships

> Intelligent navigation is **emergent** — from people who coordinate via structured codes, and from their tools
>
> **Intelligent navigation does not reside within any single individual**

Implication: when analyzing a system, **look for cognition that arises between people or between people and artifacts**

Cognitive limitations

"In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes.

What information consumes is rather obvious: it consumes the attention of its recipients.

Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."

- Herb Simon [1971]

Information overload

As we get more and more information in our environments, we cease being able to make effective use of it — our decision making stops improving or even gets worse

Yerkes-Dodson Law: as arousal (not volume of information) increases, performance increases, but only to a point [Yerkes and Dodson 1908]

Multitasking has costs

People have ~10 different working spheres per day, and spend 11.5 min per working sphere before switching [González and Mark 2004]

- When someone gets interrupted, they take 25 minutes on average before resuming [Mark, González, and Harris 2005]

People who self-report as high multitaskers are actually worse at multitasking [Ophir et al. 2009]

- Proposed mechanism: worse at filtering out irrelevant stimuli

Summary

Cognitive models create **computational proxies of human behavior**, to help us characterize and understand how we will engage with a piece of technology

- Model human processor: GOMS, KLM
- These models are powerful but can err when they make predictions out of domain

Generative AI models can also offer us broad simulation capabilities, at the risk of introducing other forms of error due to their black-box nature

References

Card, Stuart K., Allen Newell, and Thomas P. Moran. "The Psychology of Human-Computer Interaction." (1983)

Card, Stuart K., and Thomas P. Moran. "User technology: From pointing to pondering." A history of personal workstations. 1988. 489-526.

Card, Stuart K., Thomas P. Moran, and Allen Newell. "The keystroke-level model for user performance time with interactive systems." Communications of the ACM 23.7 (1980): 396-410.

Dourish, Paul. Where the action is: the foundations of embodied interaction. MIT press, 2004.

González, Victor M., and Gloria Mark. "" Constant, constant, multi-tasking craziness" managing multiple working spheres." Proceedings of the SIGCHI conference on Human factors in computing systems. 2004.

Hutchins, Edwin. Cognition in the Wild. MIT press, 1995.

Kirsh, David, and Paul Maglio. "On distinguishing epistemic from pragmatic action." Cognitive science 18.4 (1994): 513-549.

Klemmer, Scott R., Björn Hartmann, and Leila Takayama. "How bodies matter: five themes for interaction design." Proceedings of the 6th conference on Designing Interactive systems. 2006.

References

Lu, Yao, et al. "Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity." arXiv preprint arXiv:2104.08786 (2021).

Mark, Gloria, Victor M. Gonzalez, and Justin Harris. "No task left behind? Examining the nature of fragmented work." Proceedings of the SIGCHI conference on Human factors in computing systems. 2005.

Ophir, Eyal, Clifford Nass, and Anthony D. Wagner. "Cognitive control in media multitaskers." Proceedings of the National Academy of Sciences 106.37 (2009): 15583-15587.

Park, Joon Sung, et al. "Social Simulacra: Creating Populated Prototypes for Social Computing Systems." Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. 2022.

Simon, Herbert A. "Designing organizations for an information-rich world." Computers, communications, and the public interest 72 (1971): 37.

Suchman, Lucy A. Plans and situated actions: The problem of human-machine communication. Cambridge university press, 1987.

Tan, Desney, and Anton Nijholt. Brain-computer interfaces and human-computer interaction. Springer London, 2010.

Victor, Bret. A Brief Rant on the Future of Interaction Design. 2011. http://worrydream.com/ABriefRantOnTheFutureOfInteractionDesign/

Yerkes, Robert Mearns, and John D. Dodson. "The relation of strength of stimulus to rapidity of habit-formation." (1908): 459-482.