

multimodal HCI

CS 347
Yujie Tao

announcements

Quiz #4 on Wednesday! Cover up to today's lecture.

today's lecture

- What are multimodal interfaces?
- How does interfacing with different sensory modalities differ?
- What are the strengths and limitations of multimodal interfaces?

recall the last time you
were in a coffee shop...





but with our devices...



how can we enable computing devices
to equip with rich multimodal output?

to understand multimodal interfaces

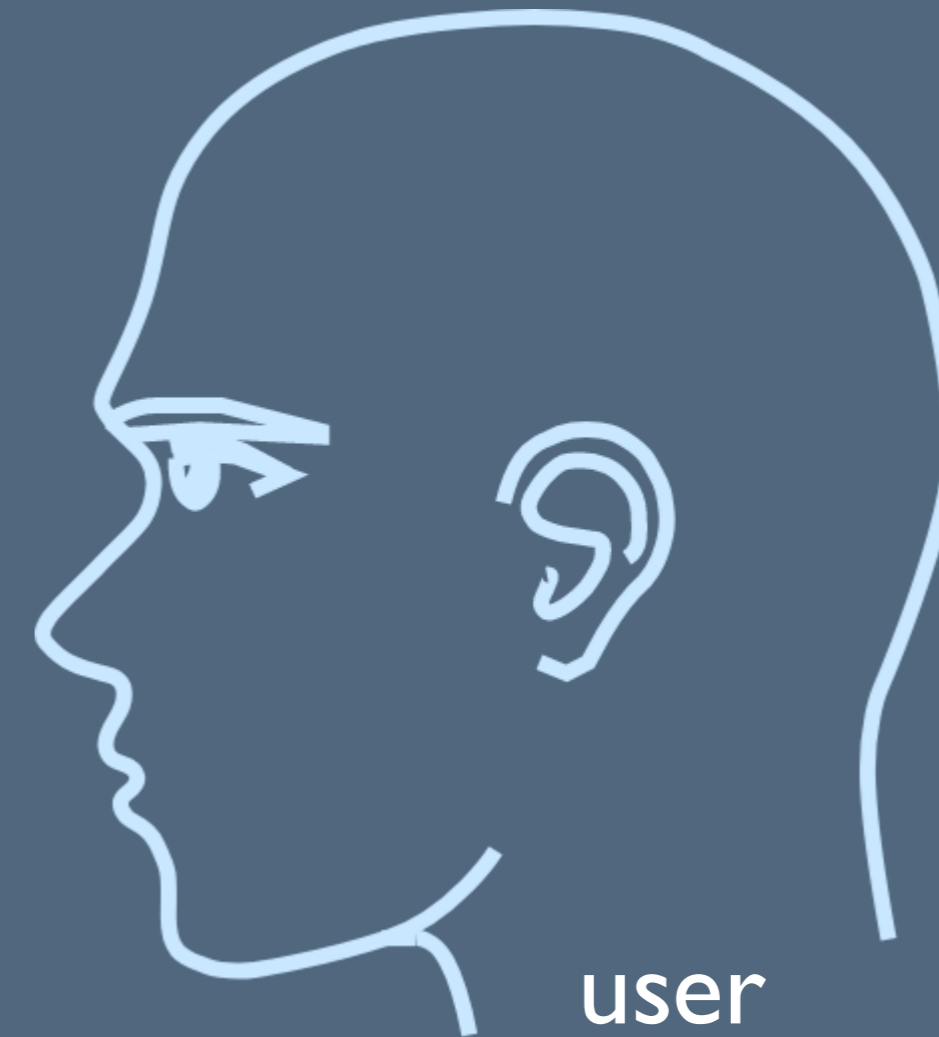
we need to first understand the physiological process of perception

to understand multimodal interfaces

we need to first understand the physiological process of perception



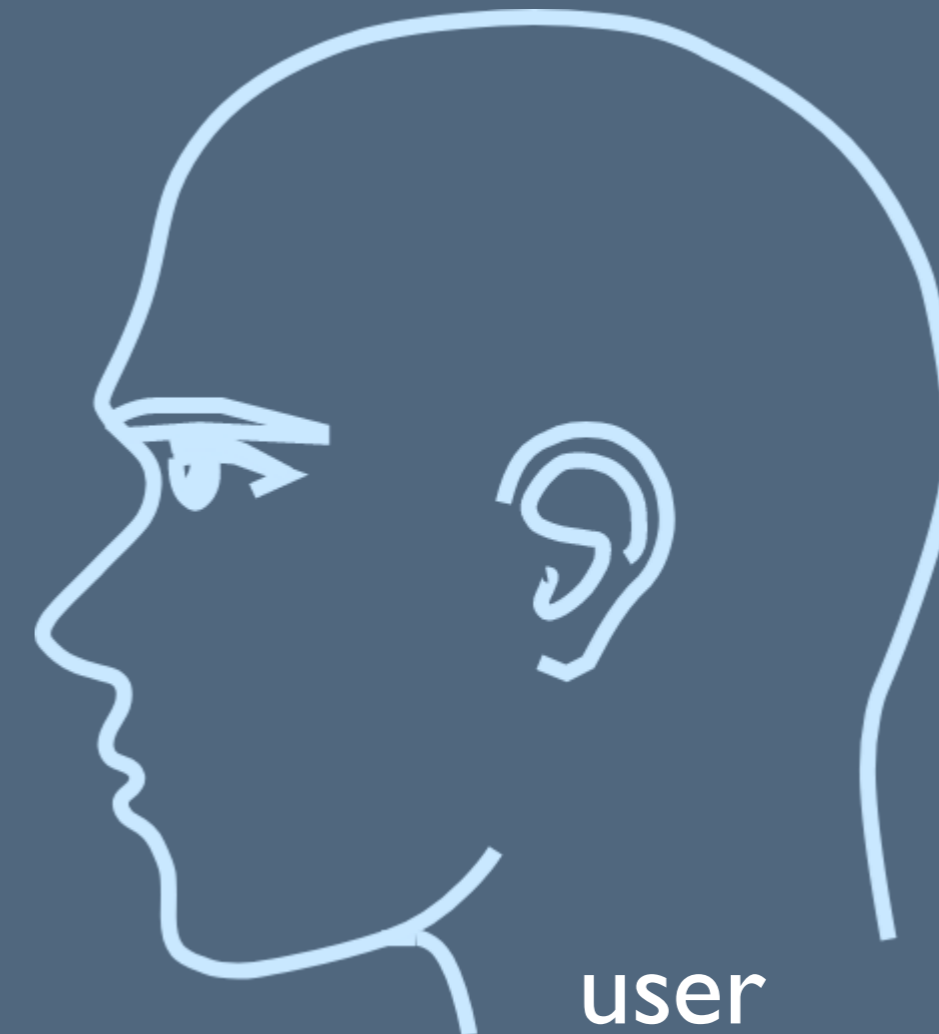
physical
stimuli



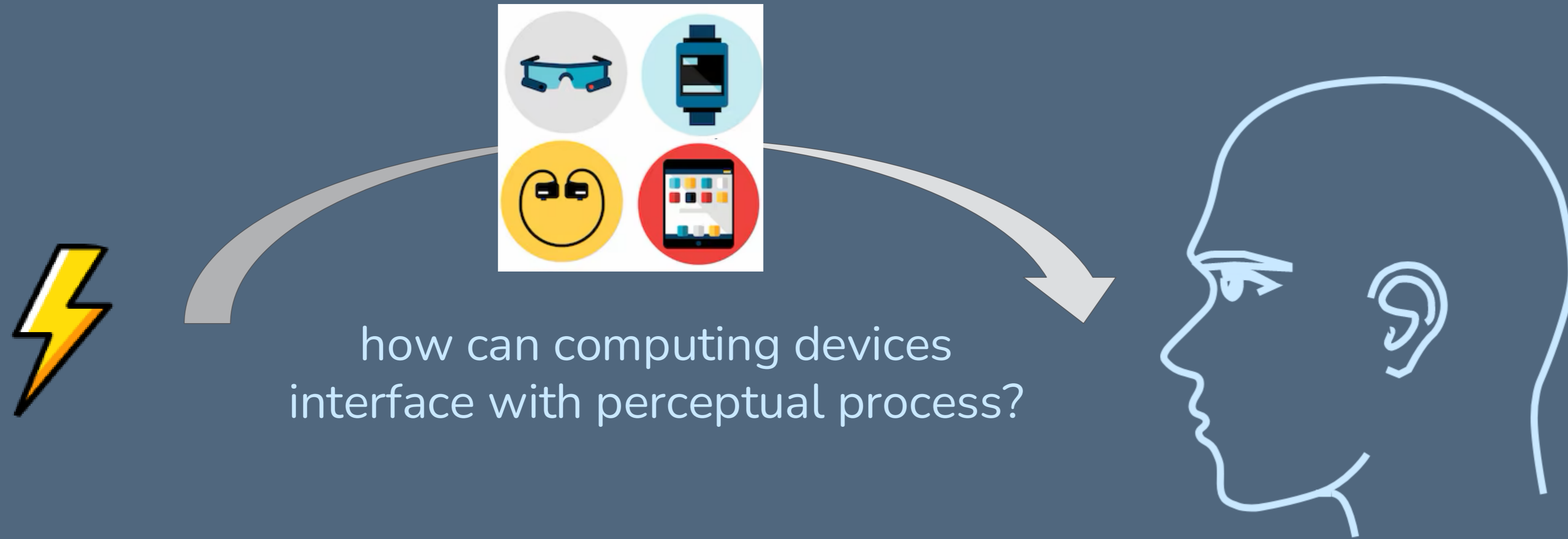
to understand multimodal interfaces

we need to first understand the physiological process of perception

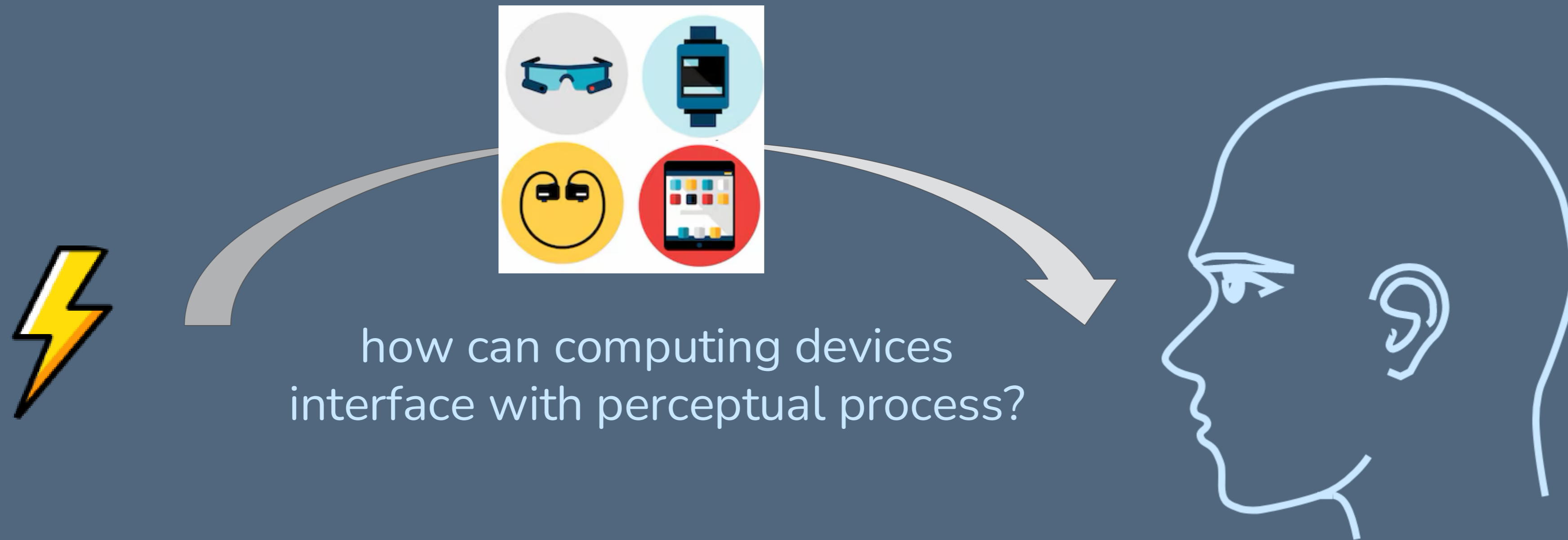
transduction: translating arriving stimulus to electrical signals by sensory receptors



to understand multimodal interfaces



to understand multimodal interfaces



1. simulation: replicate natural sensory experiences.

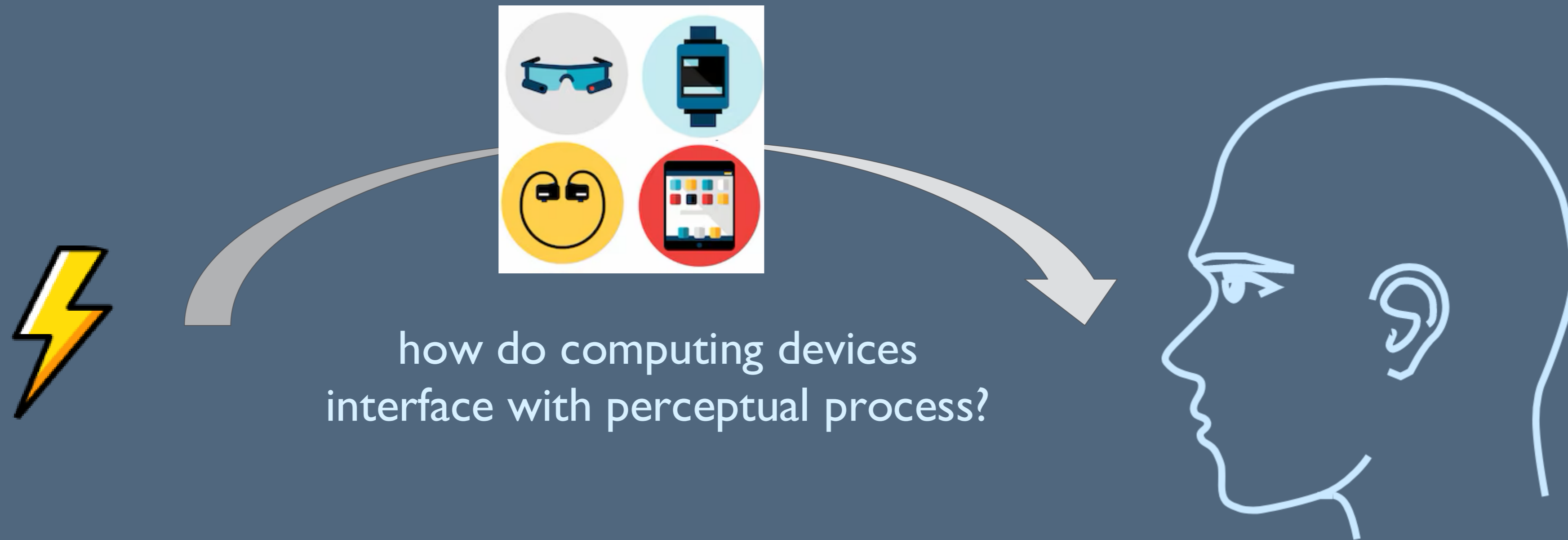
to understand multimodal interfaces

how do computing devices interface with perceptual process?



1. simulation: replicate natural sensory experiences.

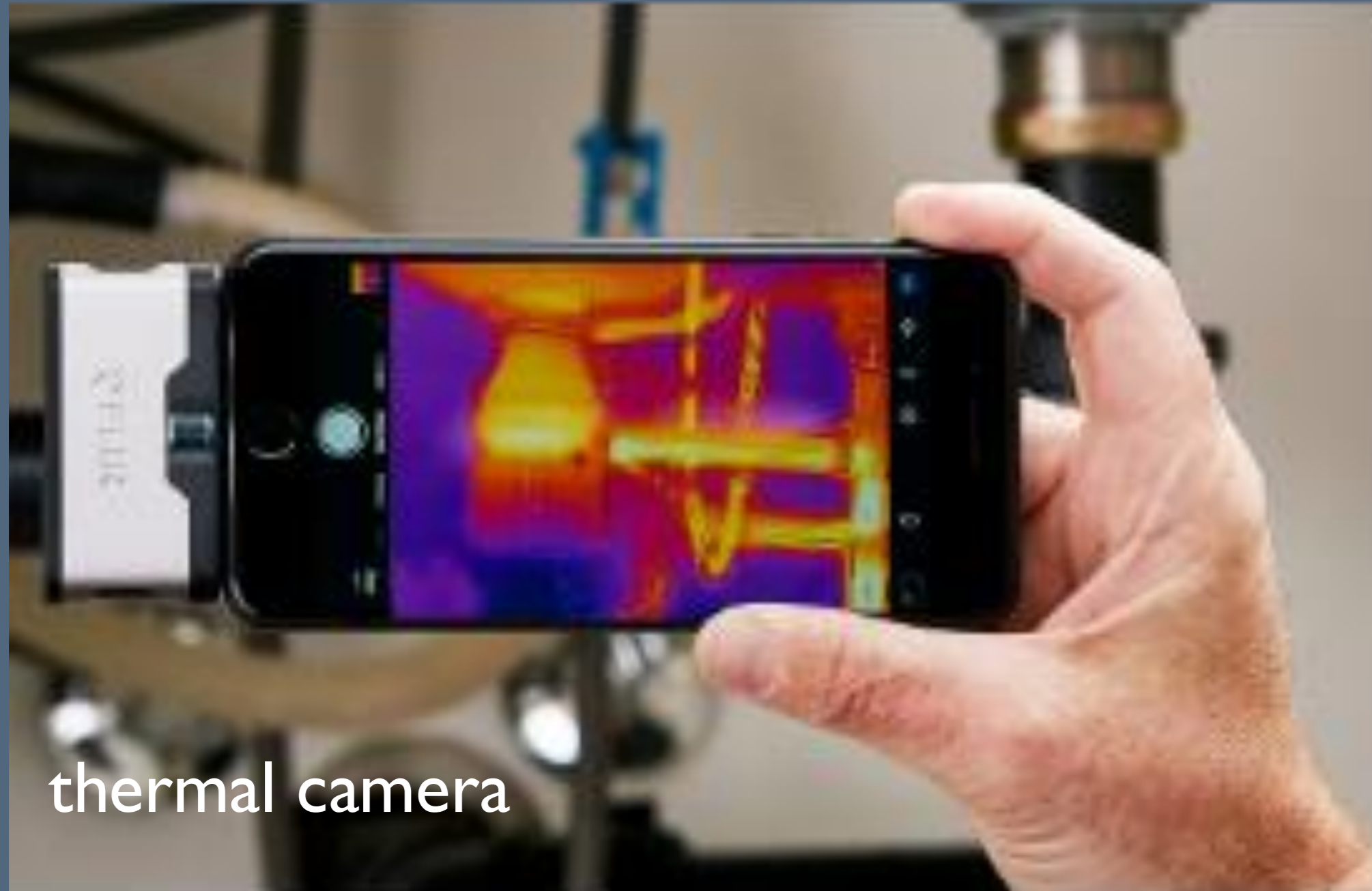
to understand multimodal interfaces



1. **simulation:** replicate natural sensory experiences.
2. **mediation:** alter natural sensory experiences.

to understand multimodal interfaces

how do computing devices play a role in the process?



thermal camera



noise-cancellation
headphone

1. **simulation:** replicate natural sensory experiences.
2. **mediation:** alter natural sensory experiences.

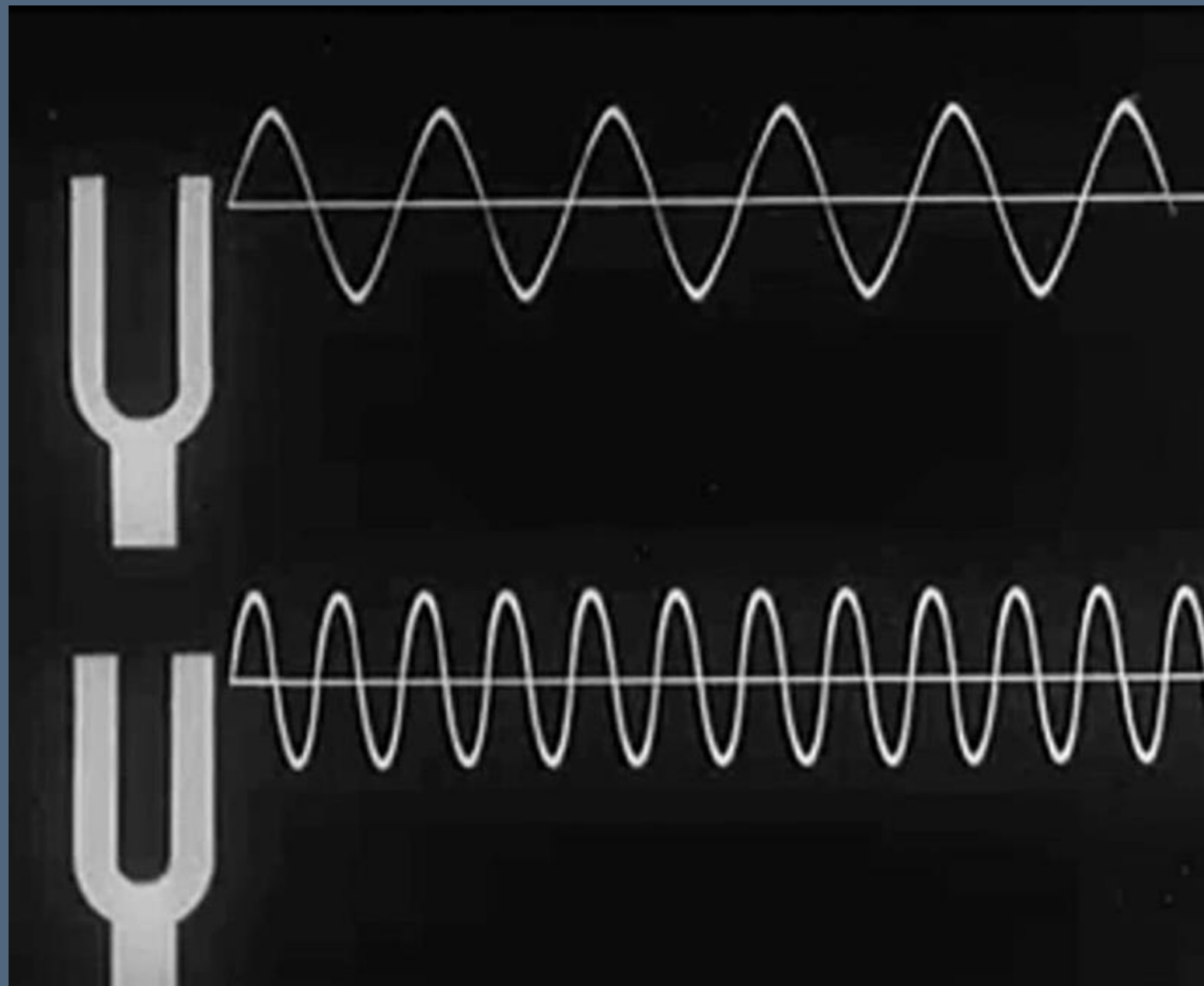


audio-based interfaces

how do we perceive sounds?

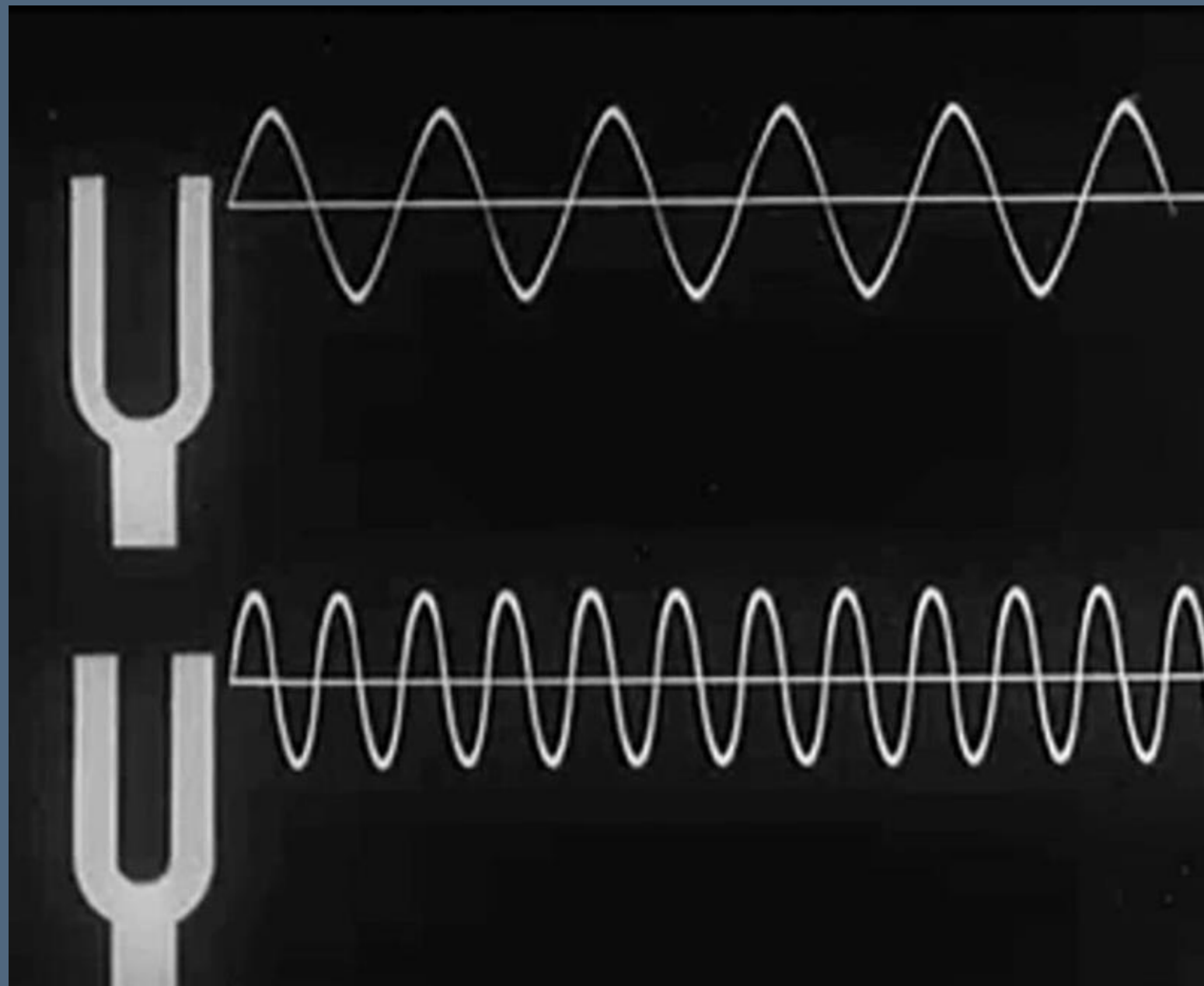
sound source

an object that vibrates

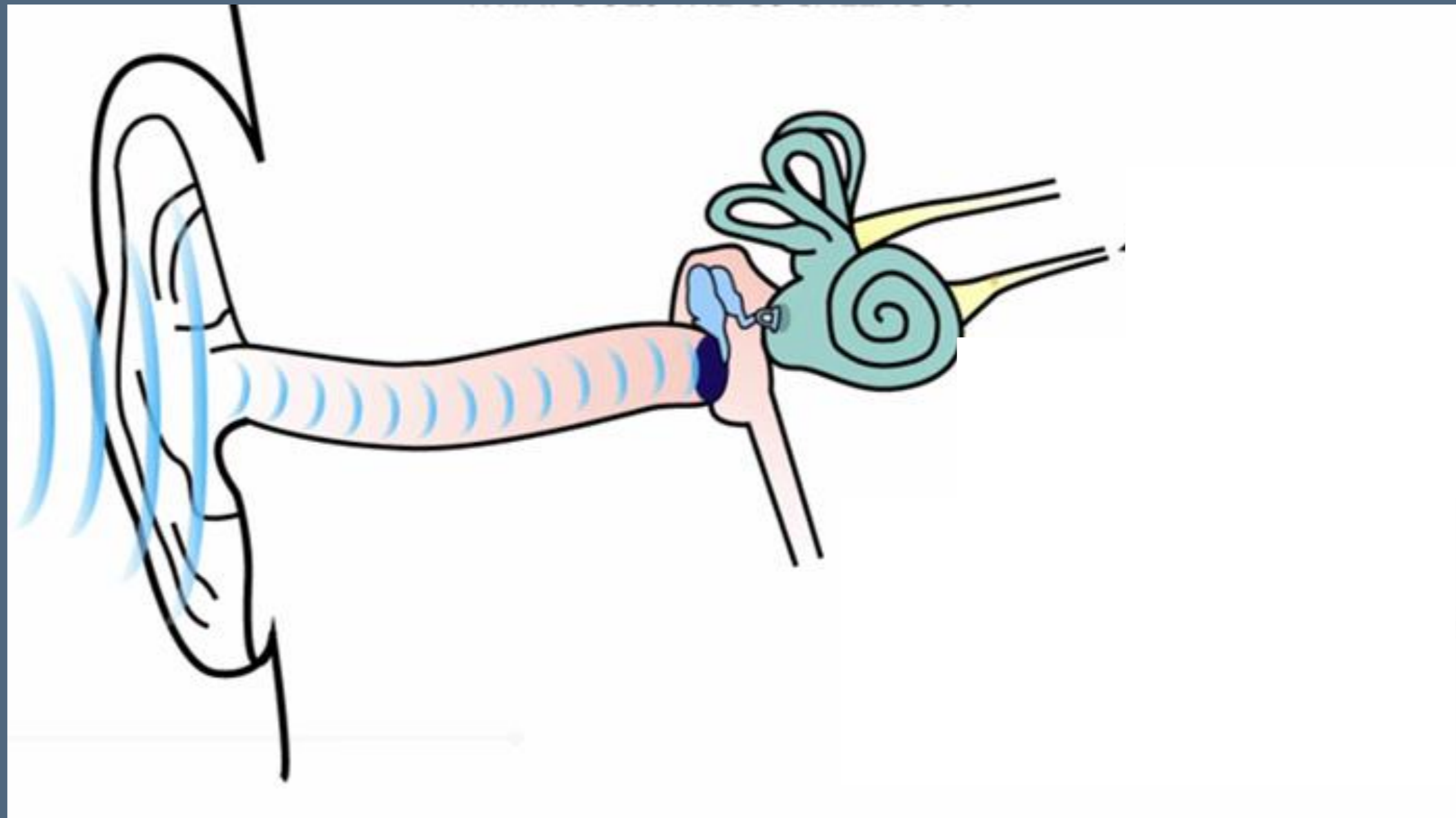


how do we perceive sounds?

sound source
an object that vibrates

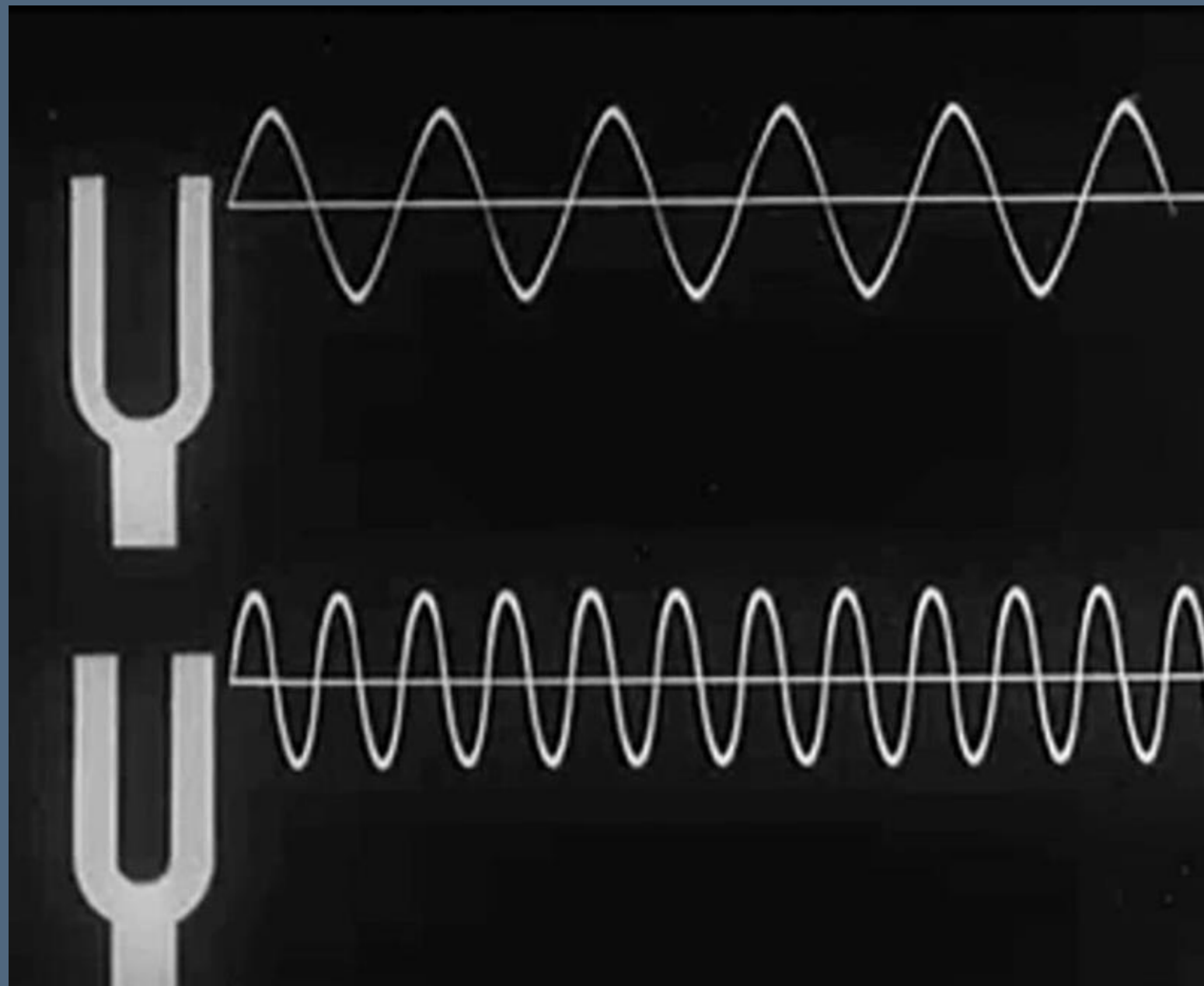


sound waves enter ear canal and strike the eardrum, causing it to vibrate back and forth.

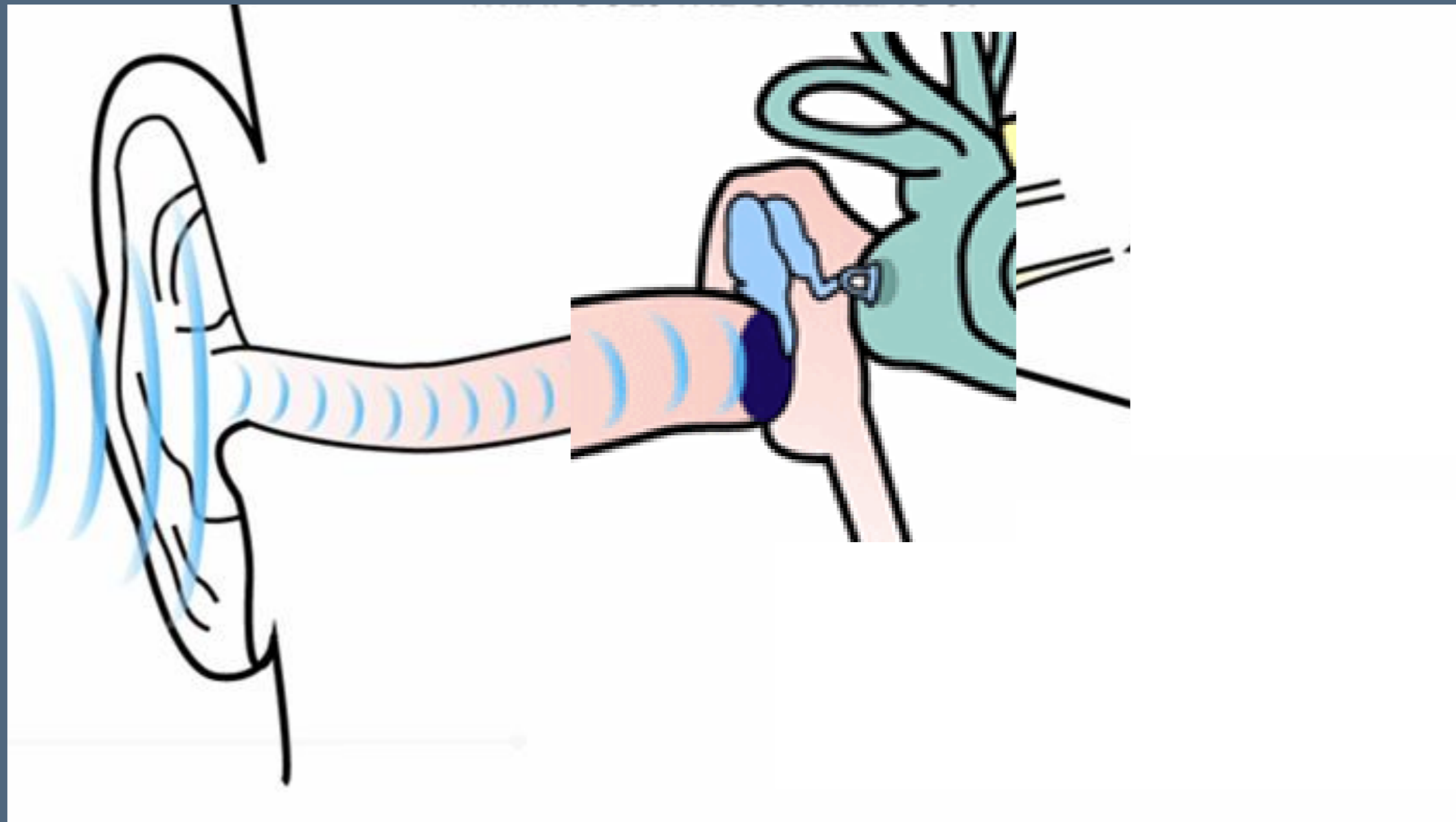


how do we perceive sounds?

sound source
an object that vibrates

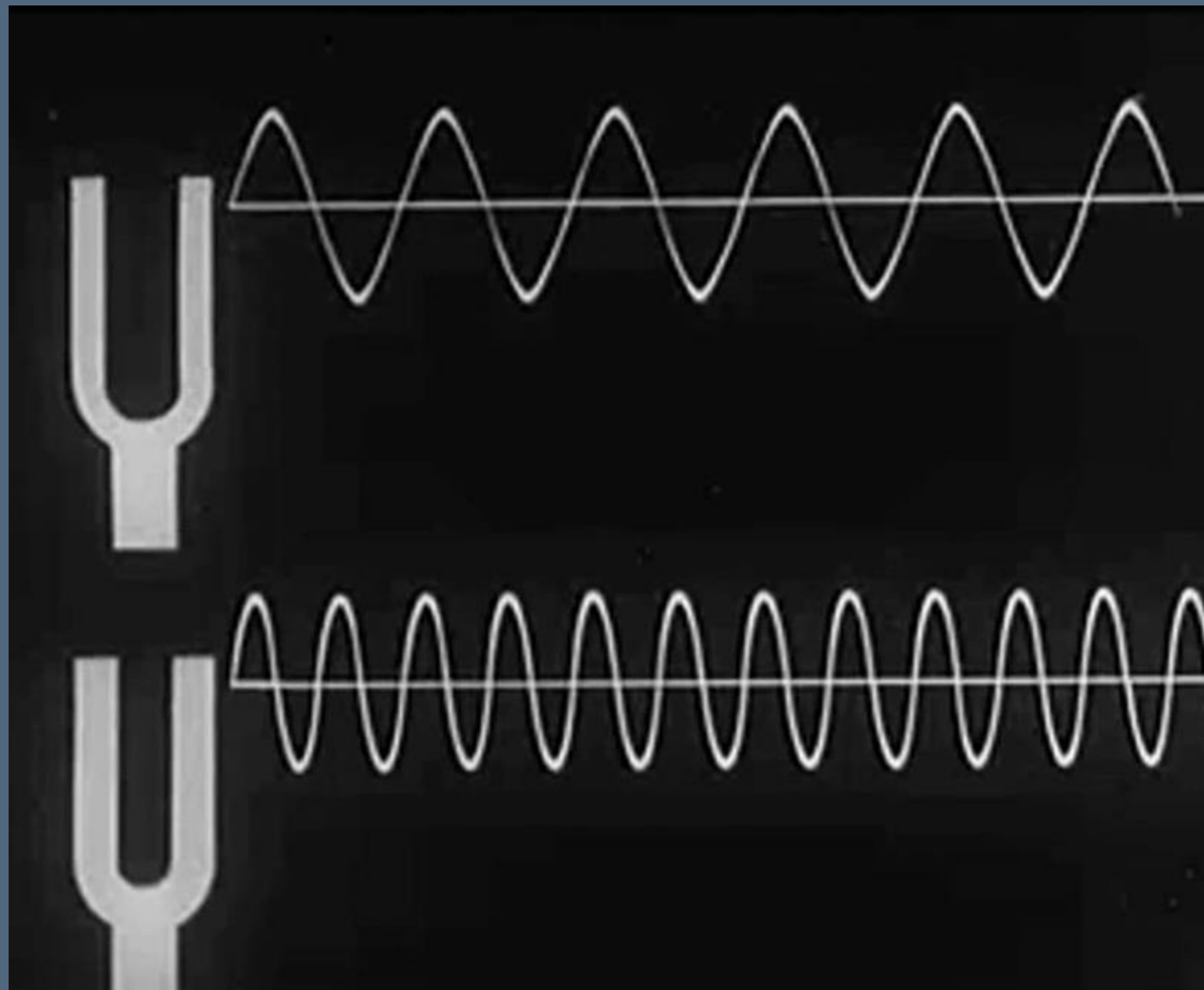


sound waves enter ear canal and strike the eardrum, causing it to vibrate back and forth.

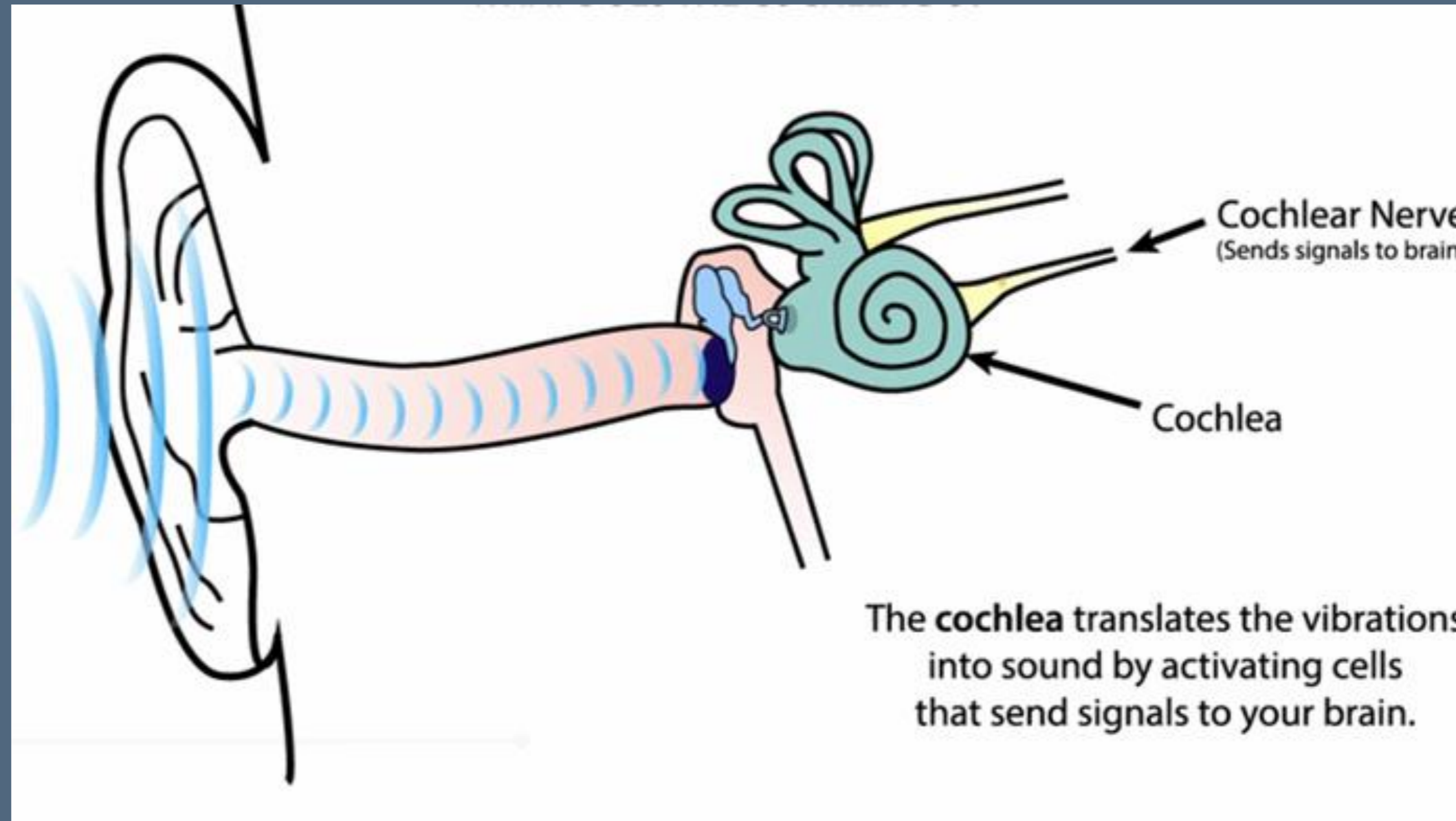


how do we perceive sounds?

sound source
an object that vibrates



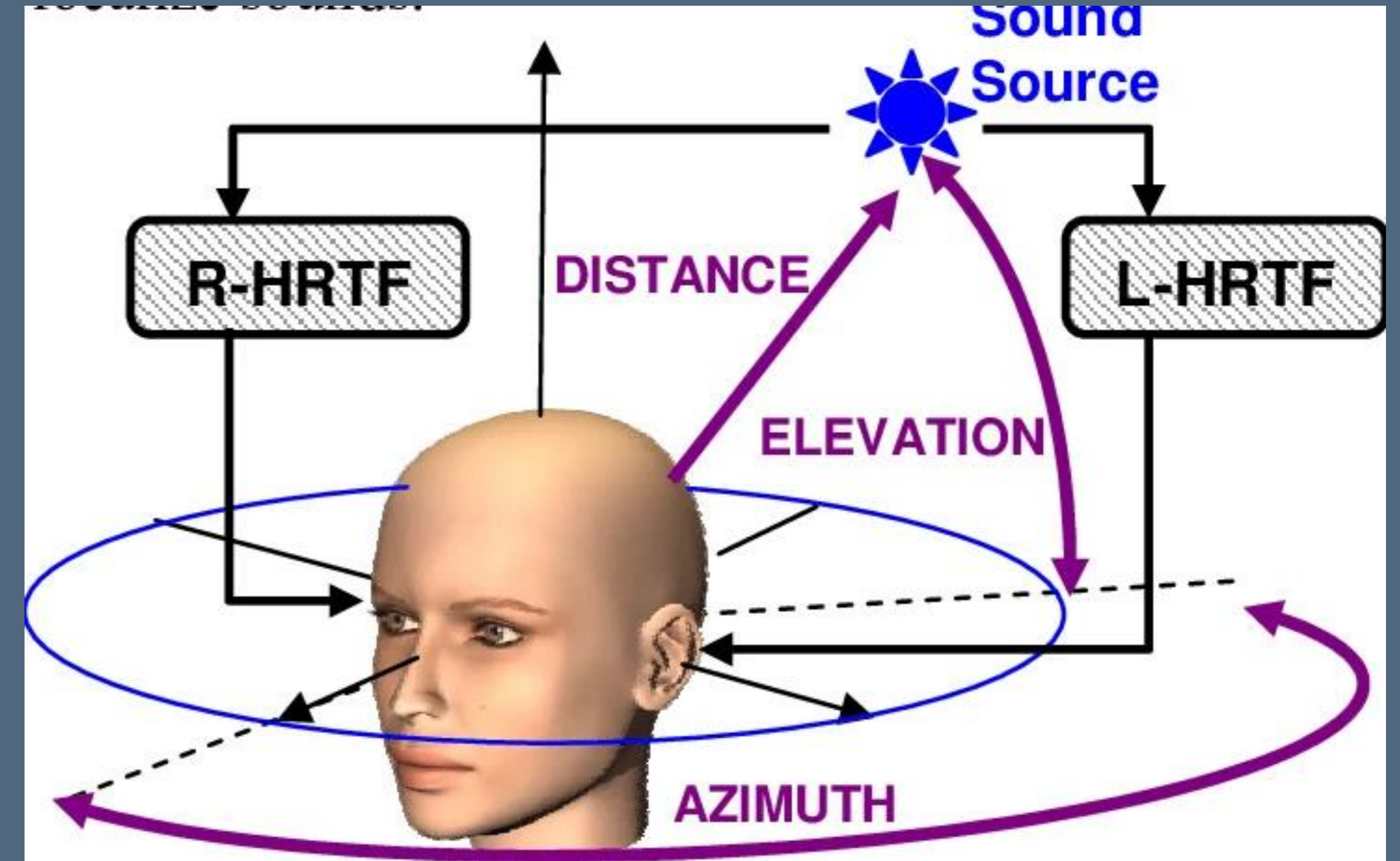
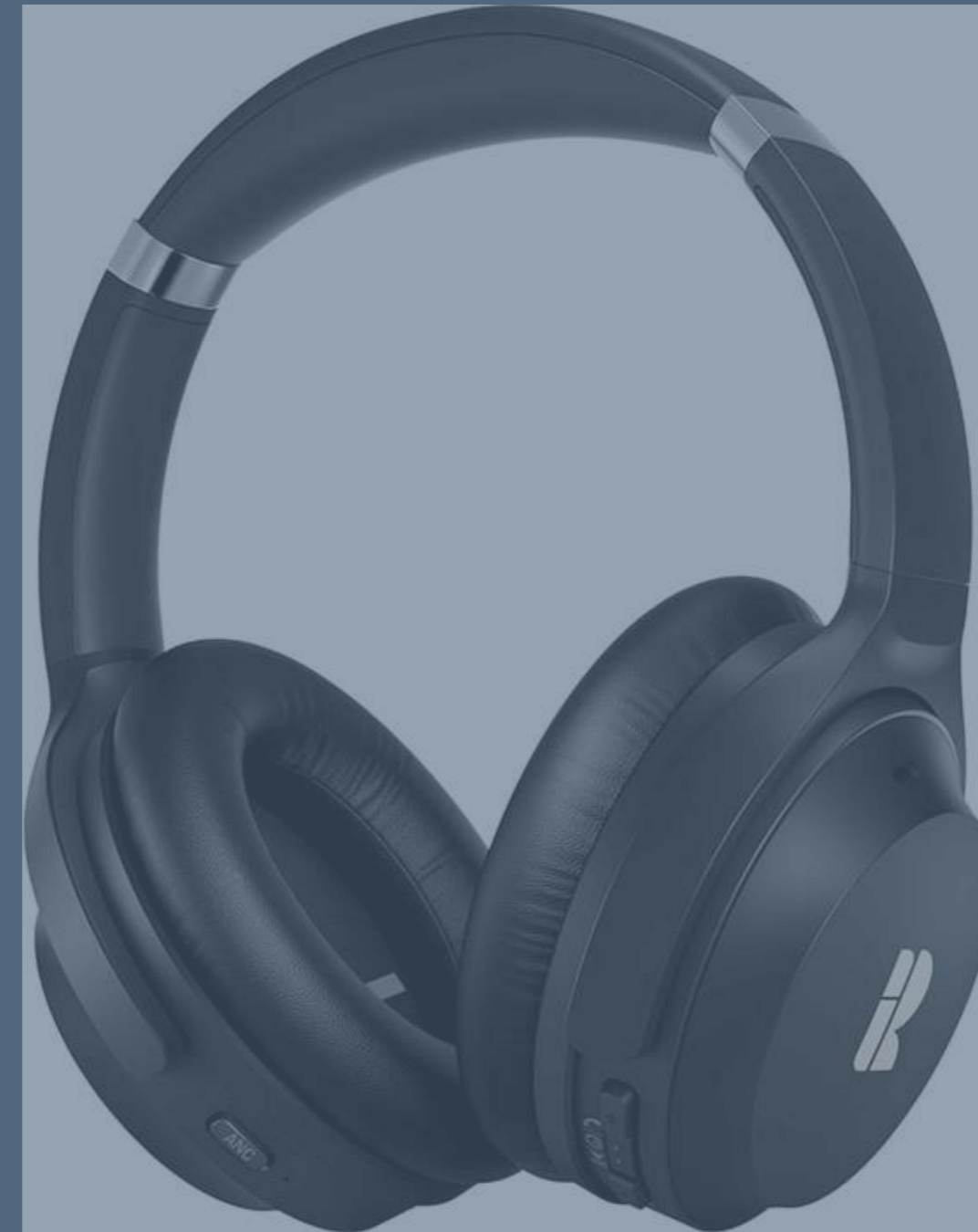
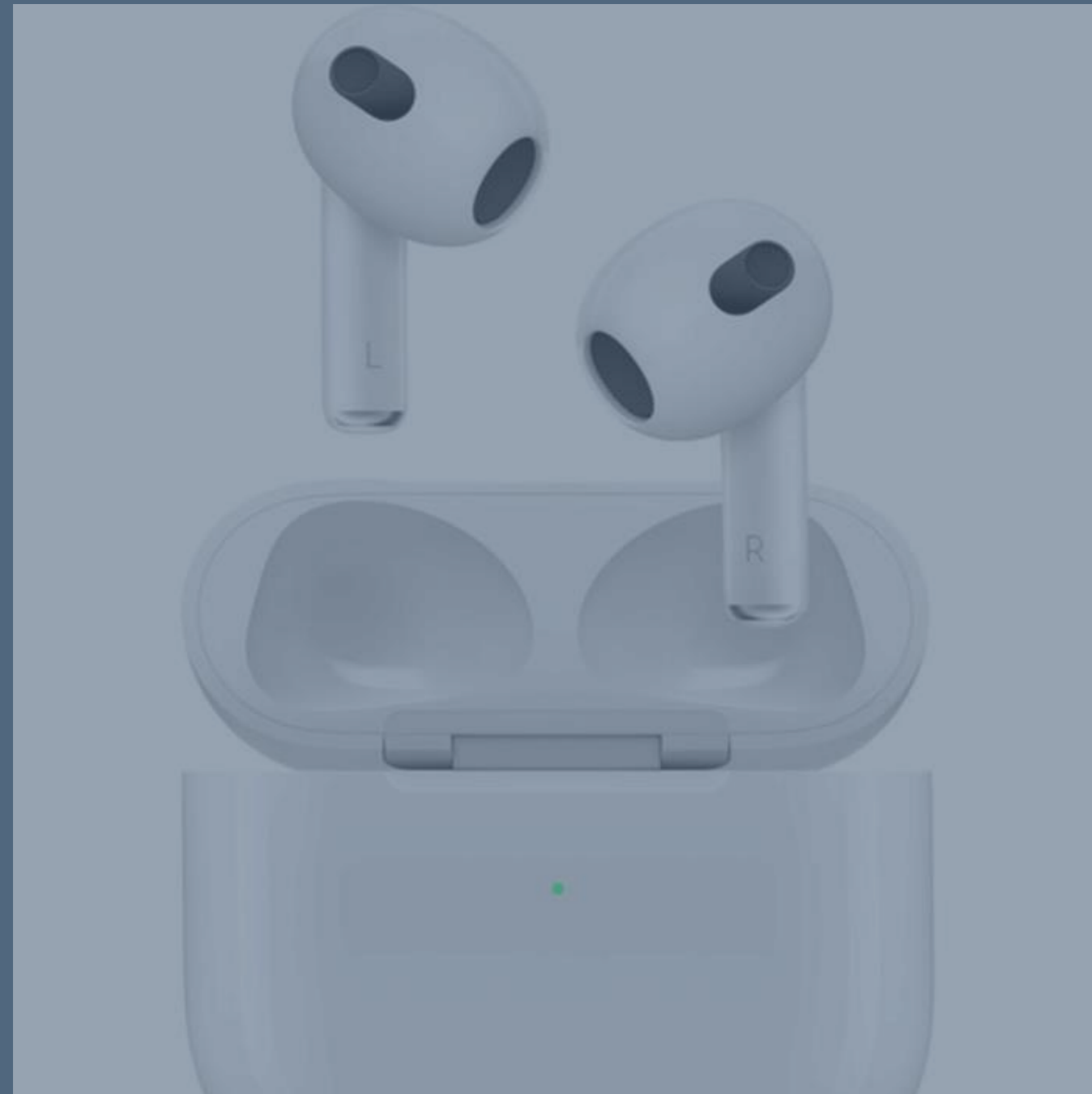
sound waves enter ear canal and strike the eardrum, causing it to vibrate back and forth.



simulate auditory experience



simulate auditory experience



A head-related transfer function (HRTF) characterizes how an ear receives a sound from a point in space

CCRMA listening room



simulate acoustic experiences
room-scale with multiple
speakers (7 from ceiling, 8 at ear
level and 7 below grid floor)

mediate auditory experience

mediate auditory experience



selectively hearing
environmental sounds
[Veluri & Itani et al., 2023]

mediate auditory experience

Auditory Seasoning Filters



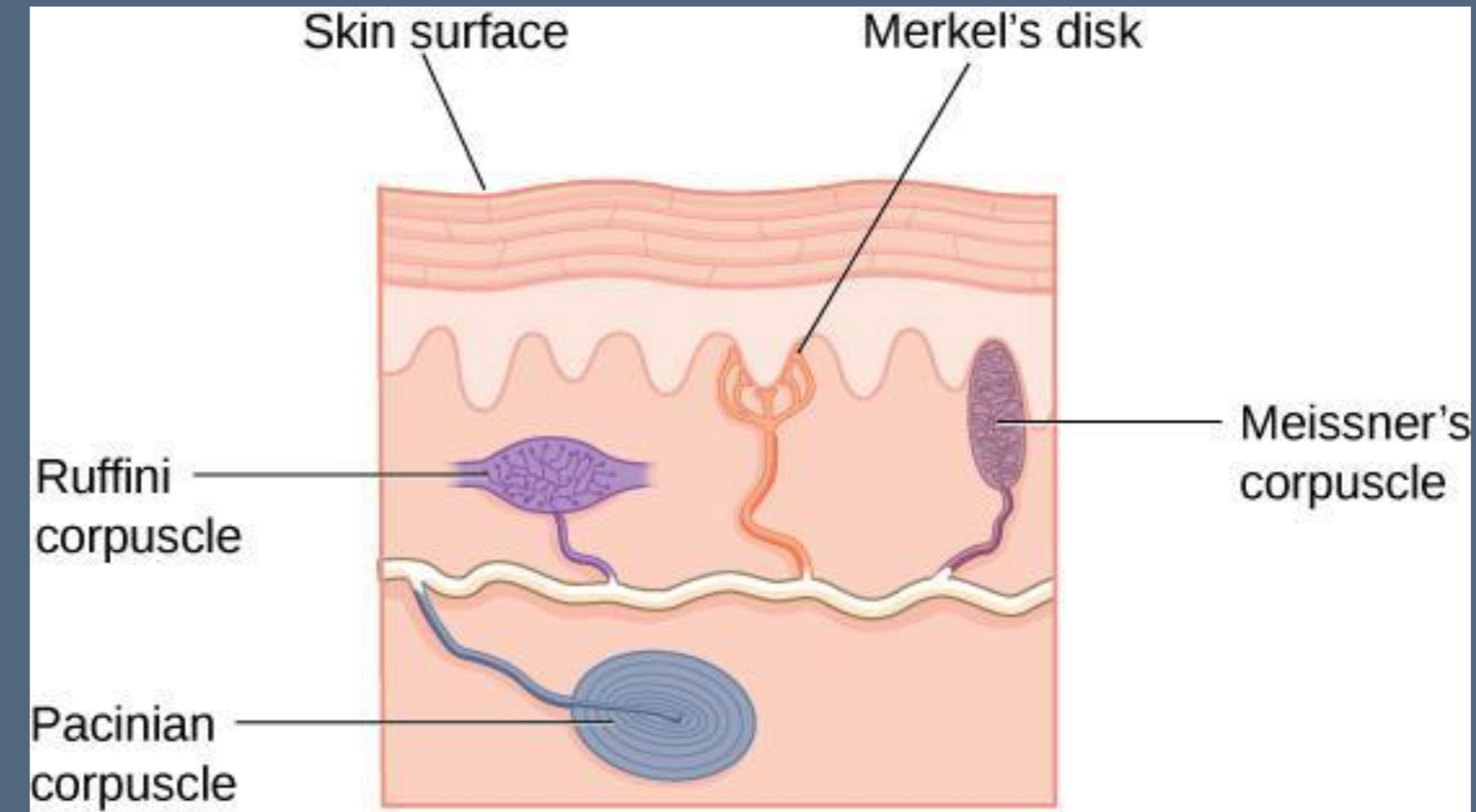
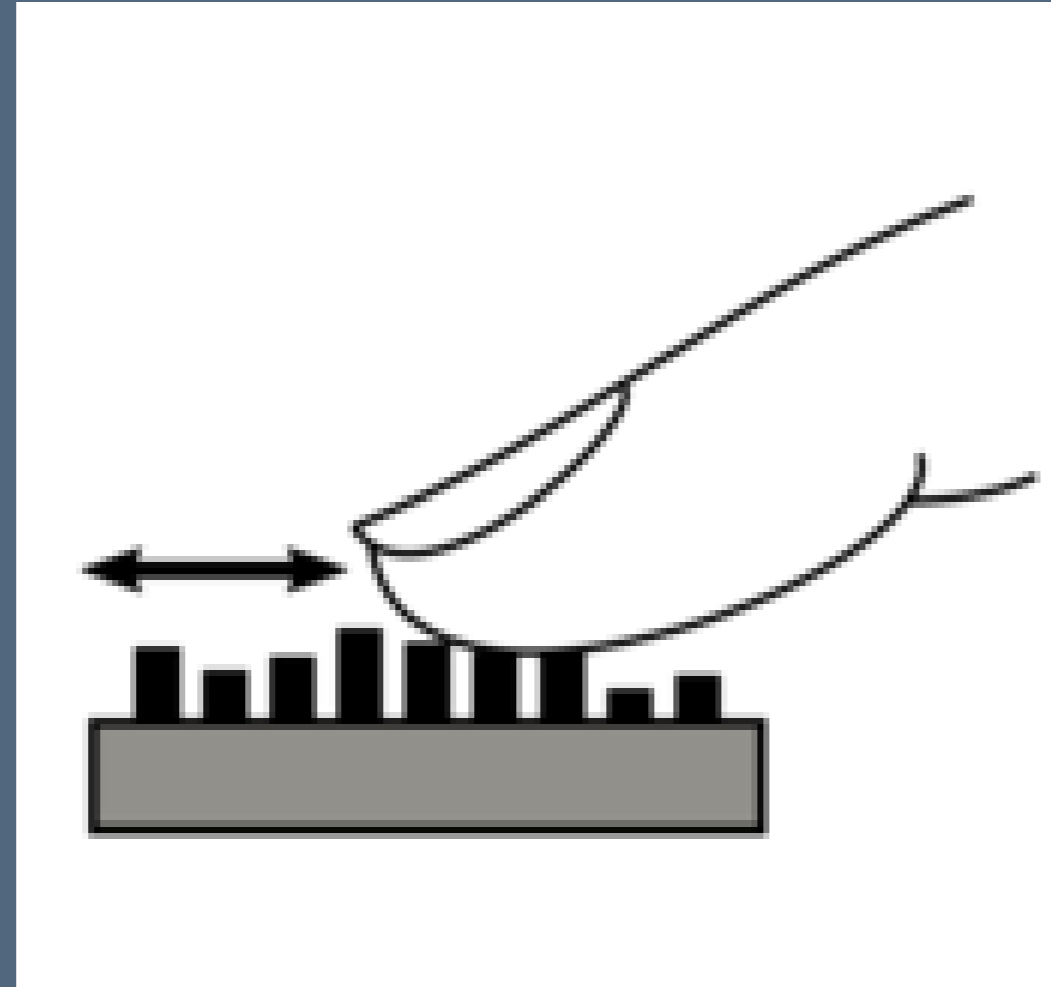
mediating chewing sounds
[Kleinberger, et al., 2023]

interface with the sense of touch

- tactile devices
- kinesthetic devices

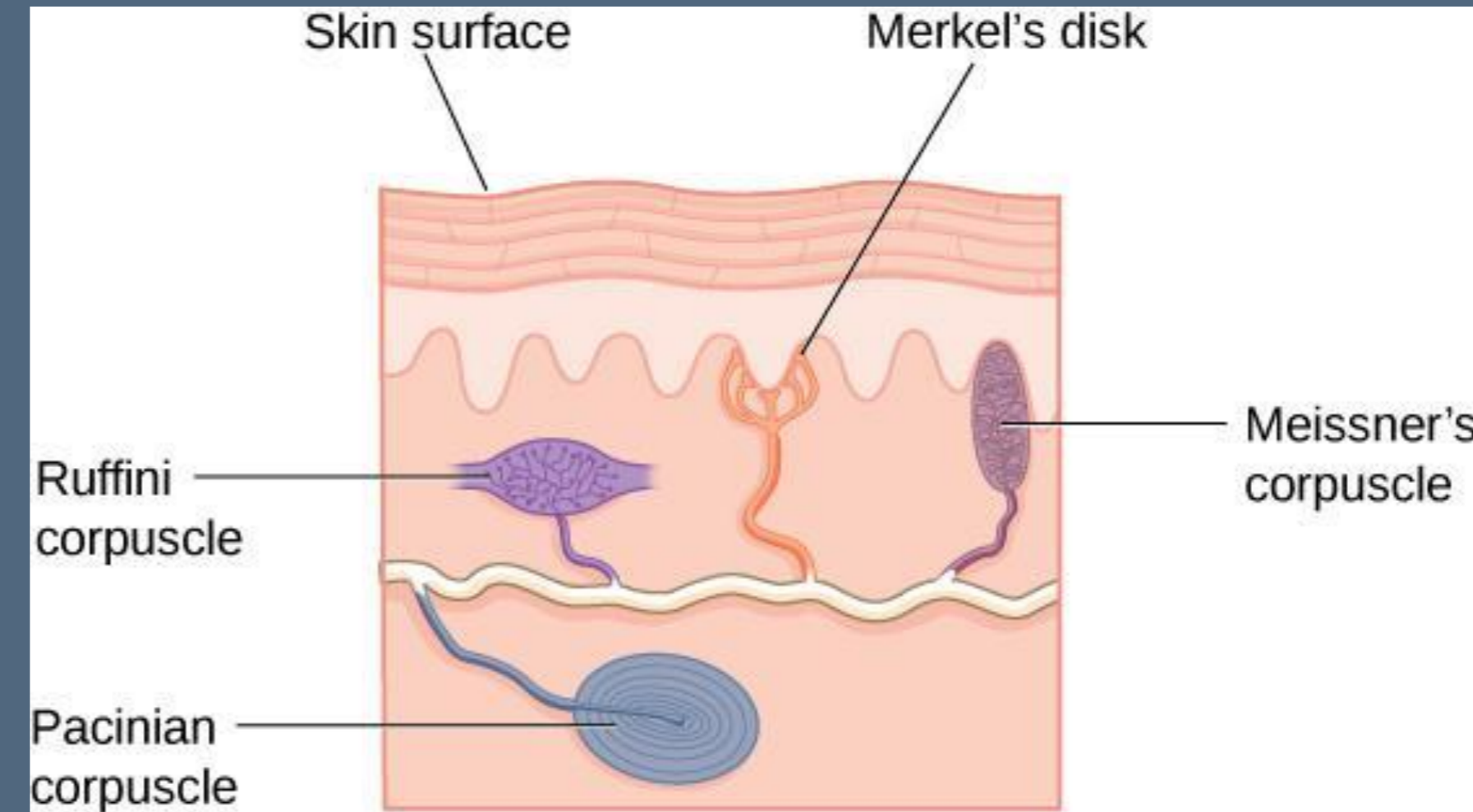
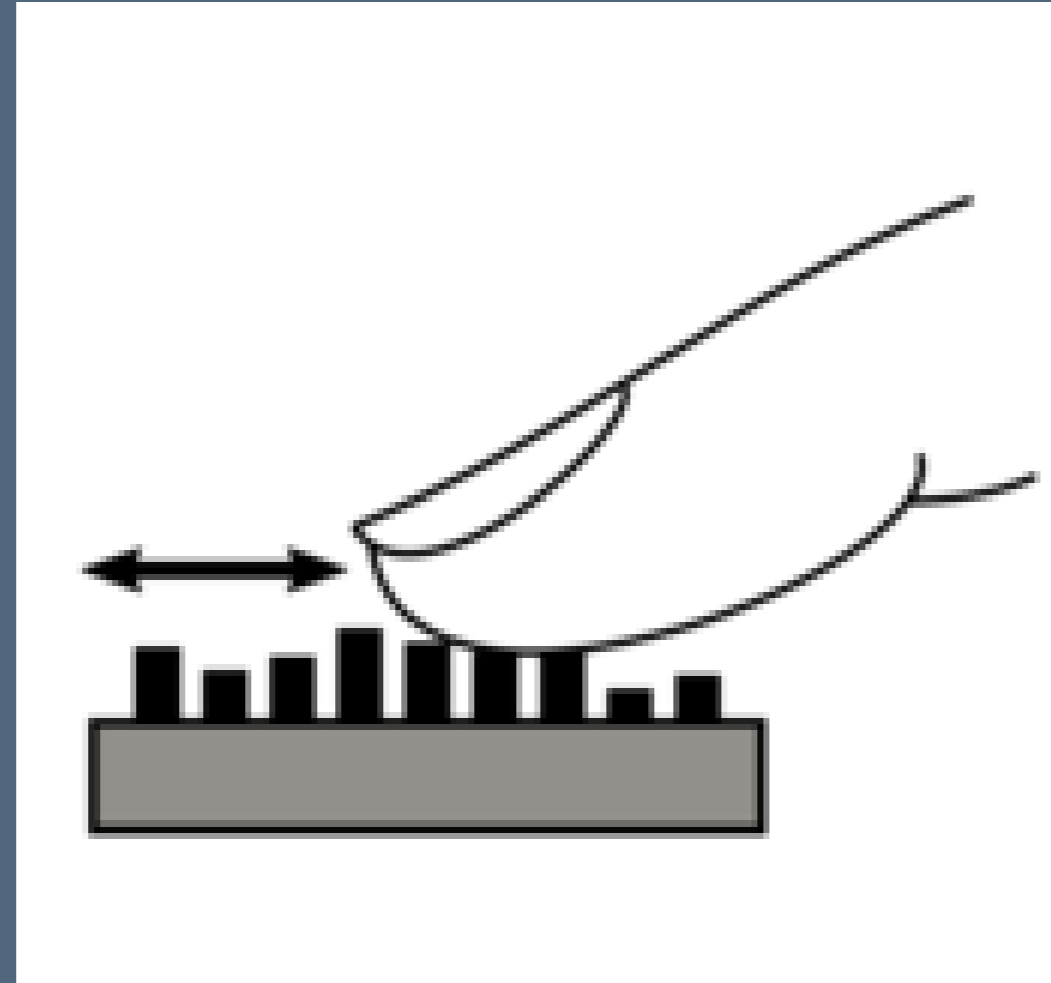
tactile devices

tactile
stimulate the skin



tactile devices

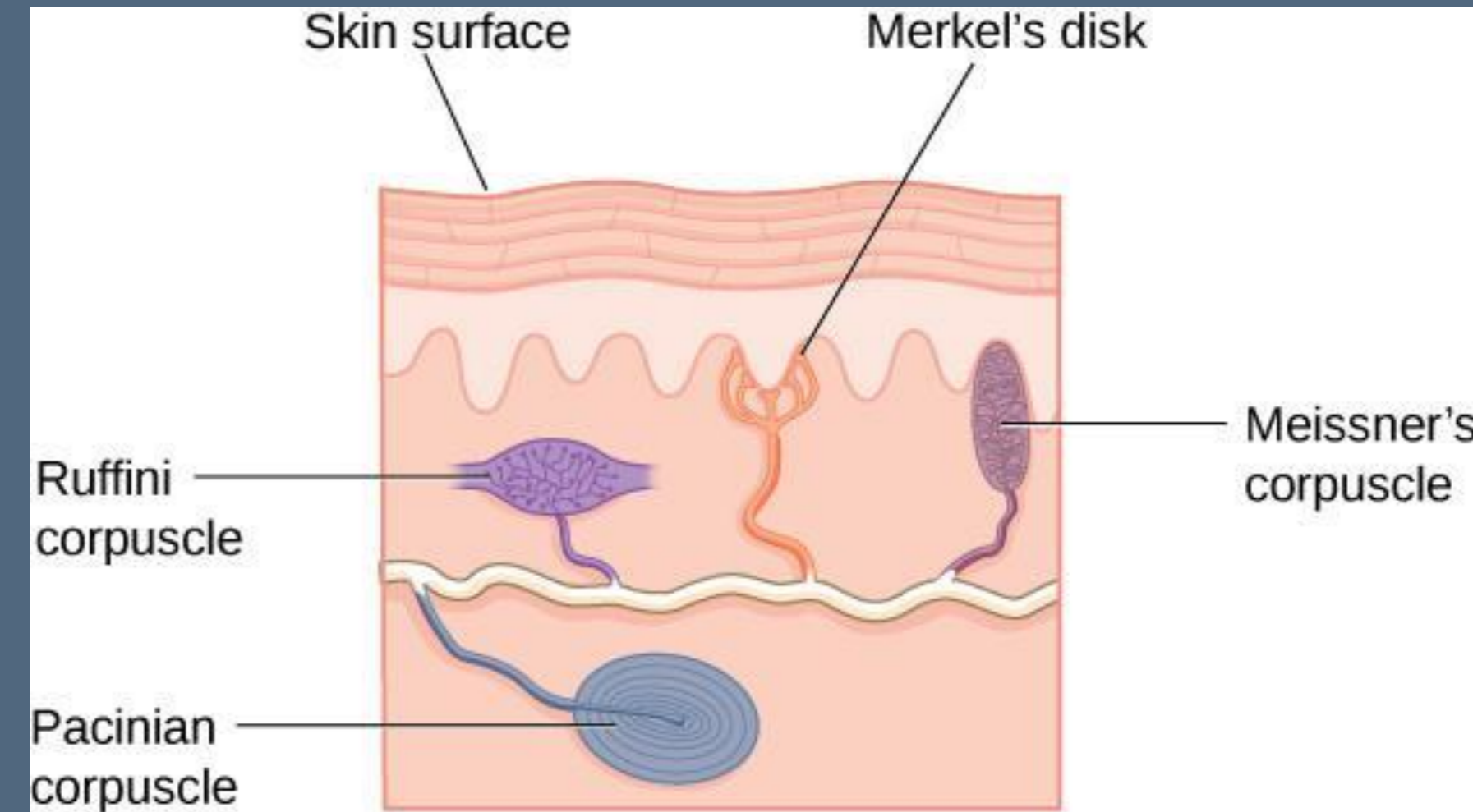
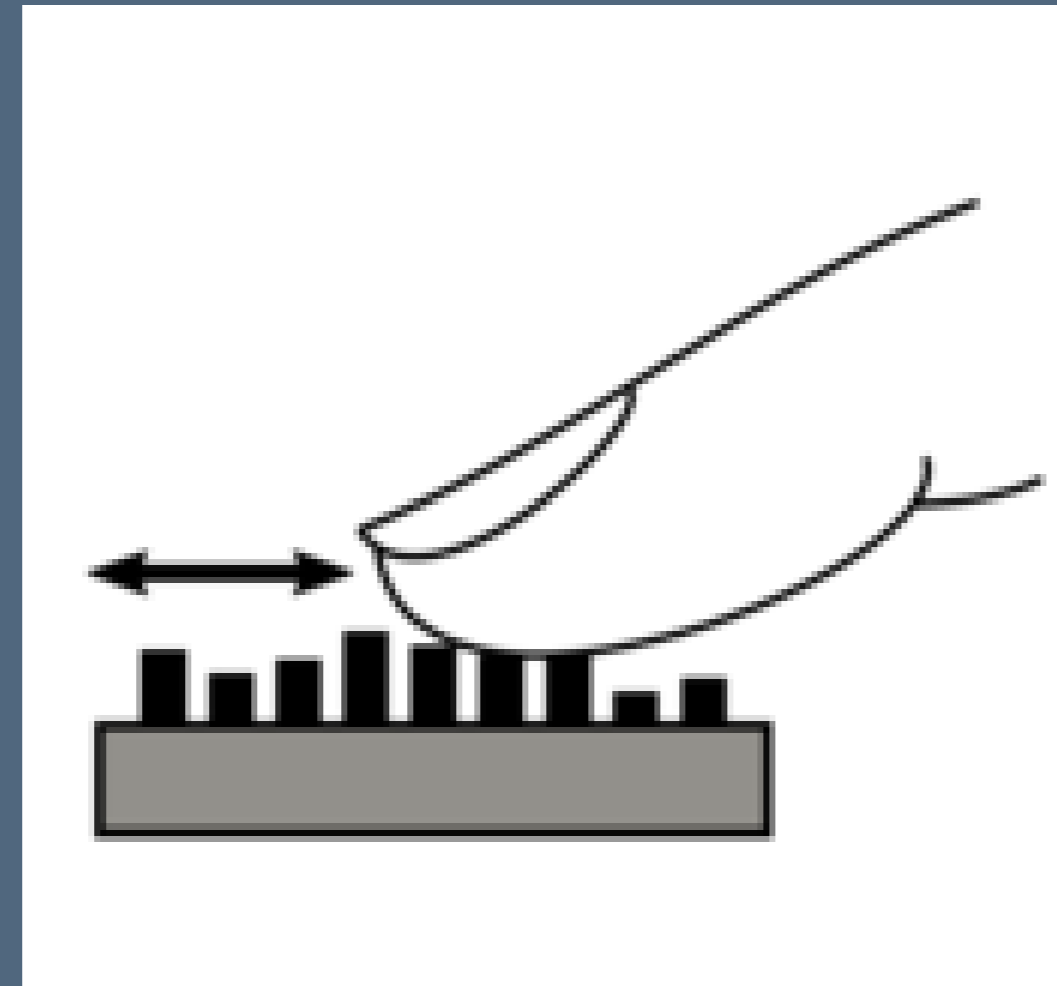
tactile
stimulate the skin



- Merkel cells detect fine details and texture,
- Meissner's corpuscles perceive light touch,
- Pacinian corpuscles sense deep pressure & vibrations
- Ruffini endings respond to continuous skin stretch & prolonged pressure

tactile devices

tactile
stimulate the skin



Q: what is the most common tactile device we have in the room right now?

tactile devices

stimulate the skin



tactile devices

stimulate the skin



tactile devices

stimulate the skin



pneumatics-based actuator
embedded in knit sleeve
[du Pasquier et al, 2023]

tactile devices

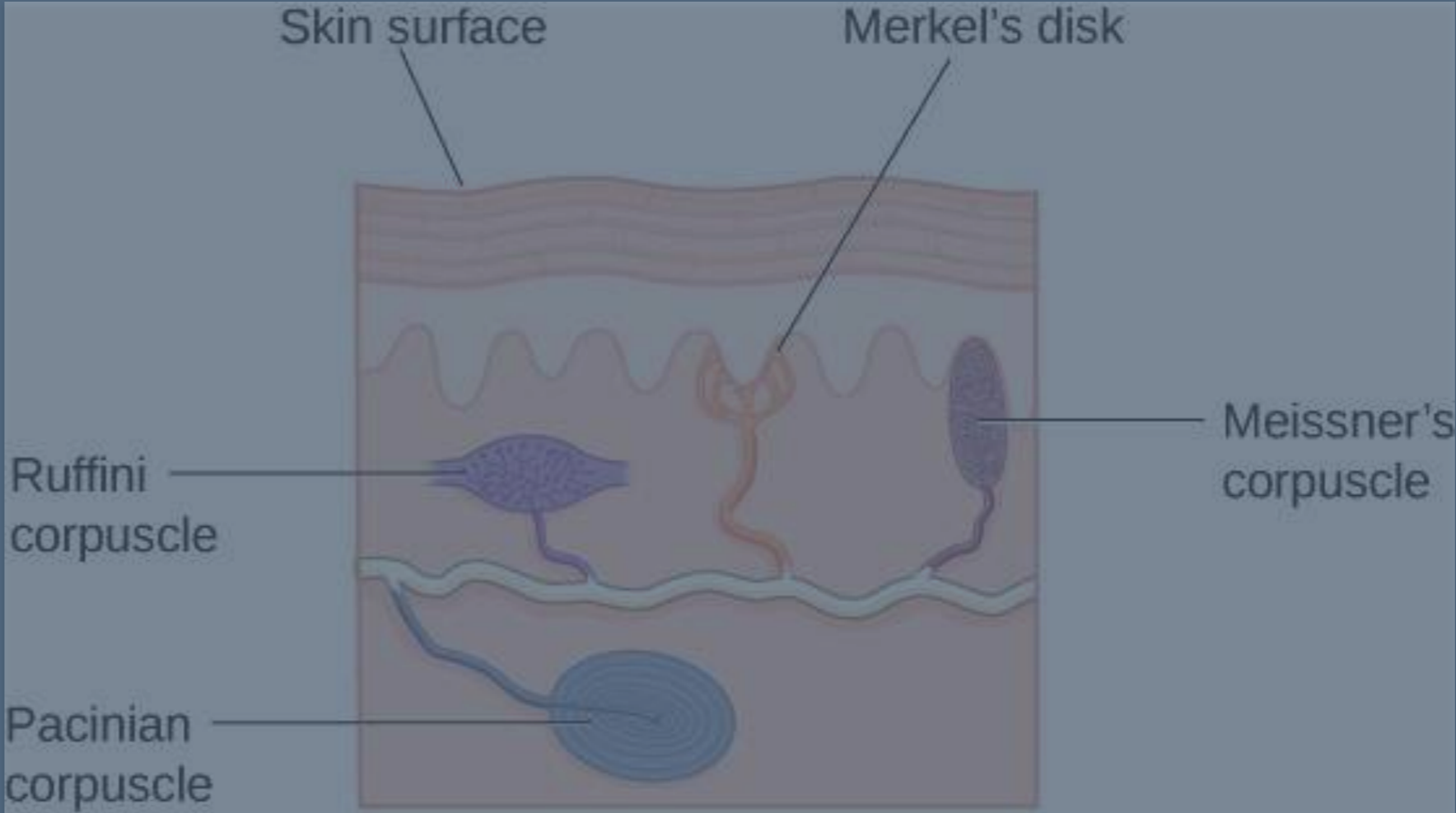
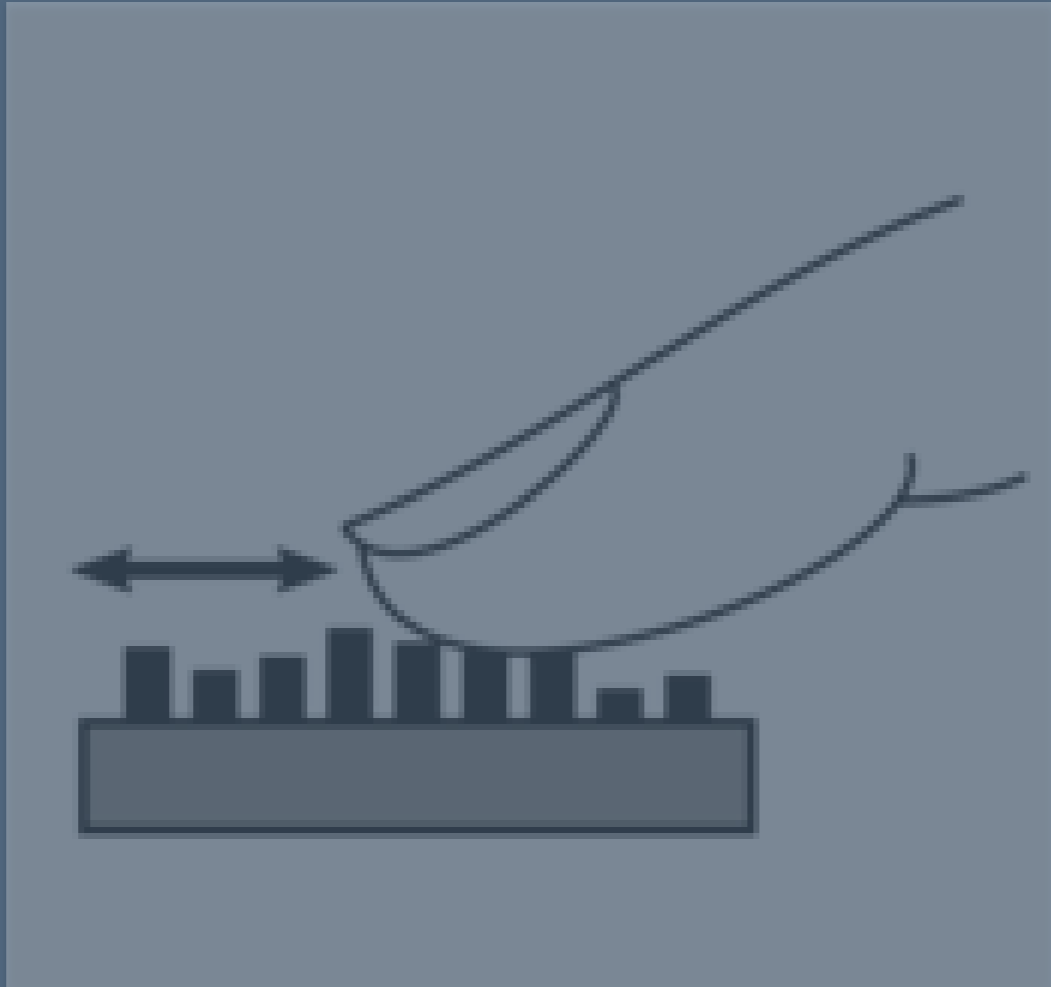
stimulate the skin (our largest organ!)



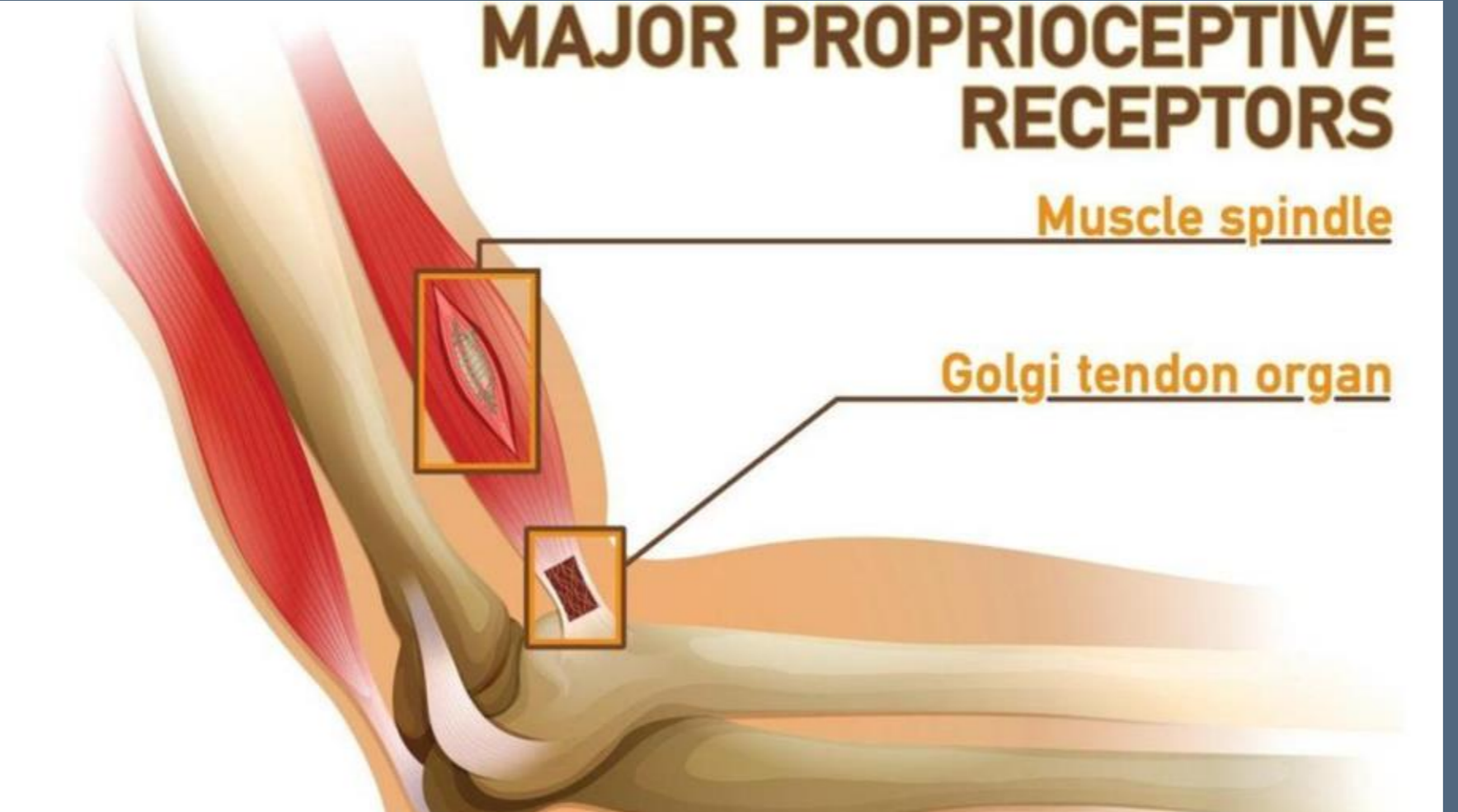
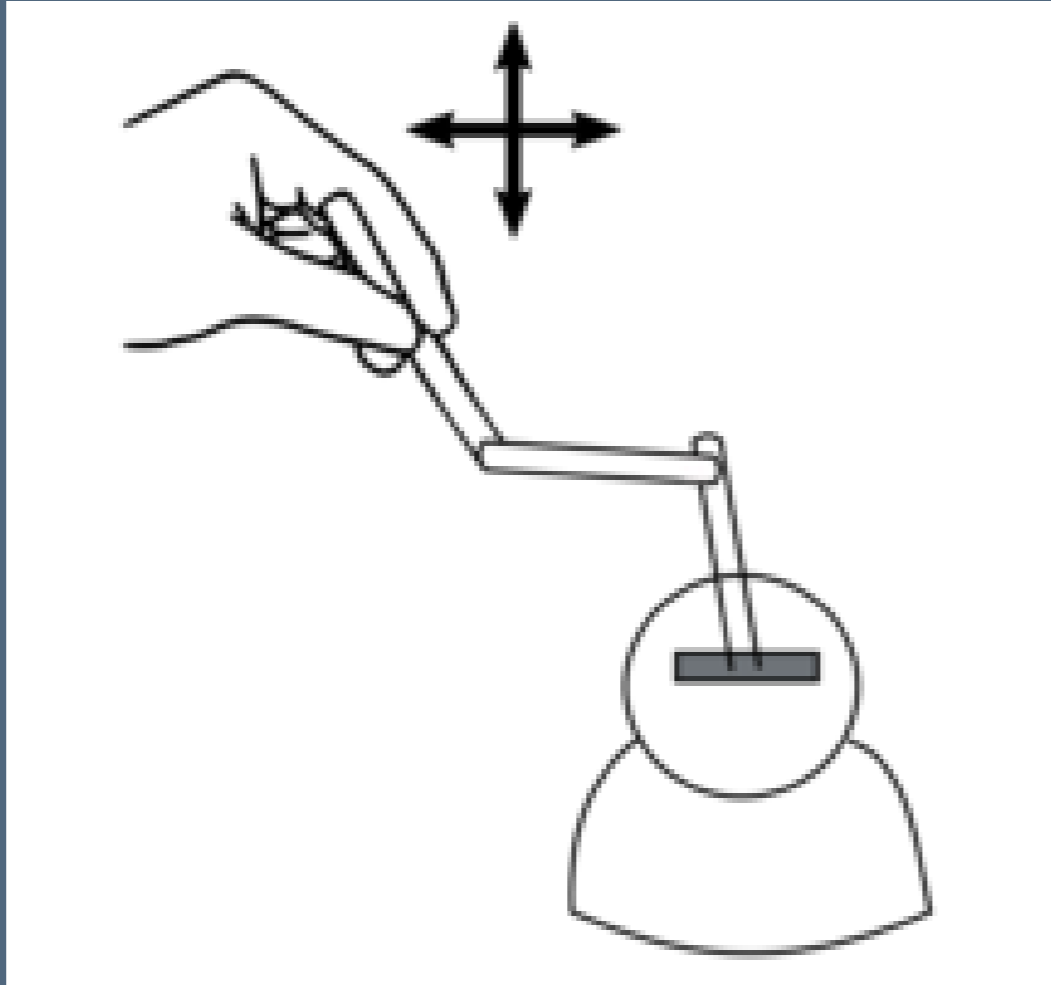
ultrasound to render haptic effects onto the mouth
[Shen et al., 2023]

kinesthetic devices

tactile
stimulate the skin



kinesthetic
display force or motions



kinesthetic devices

display forces or motions

phantom haptic device



[Molinaro et al., 2024]

kinesthetic devices

display forces or motions



electrical muscle stimulation
to render force feedback
[Lopes et al., 2019]

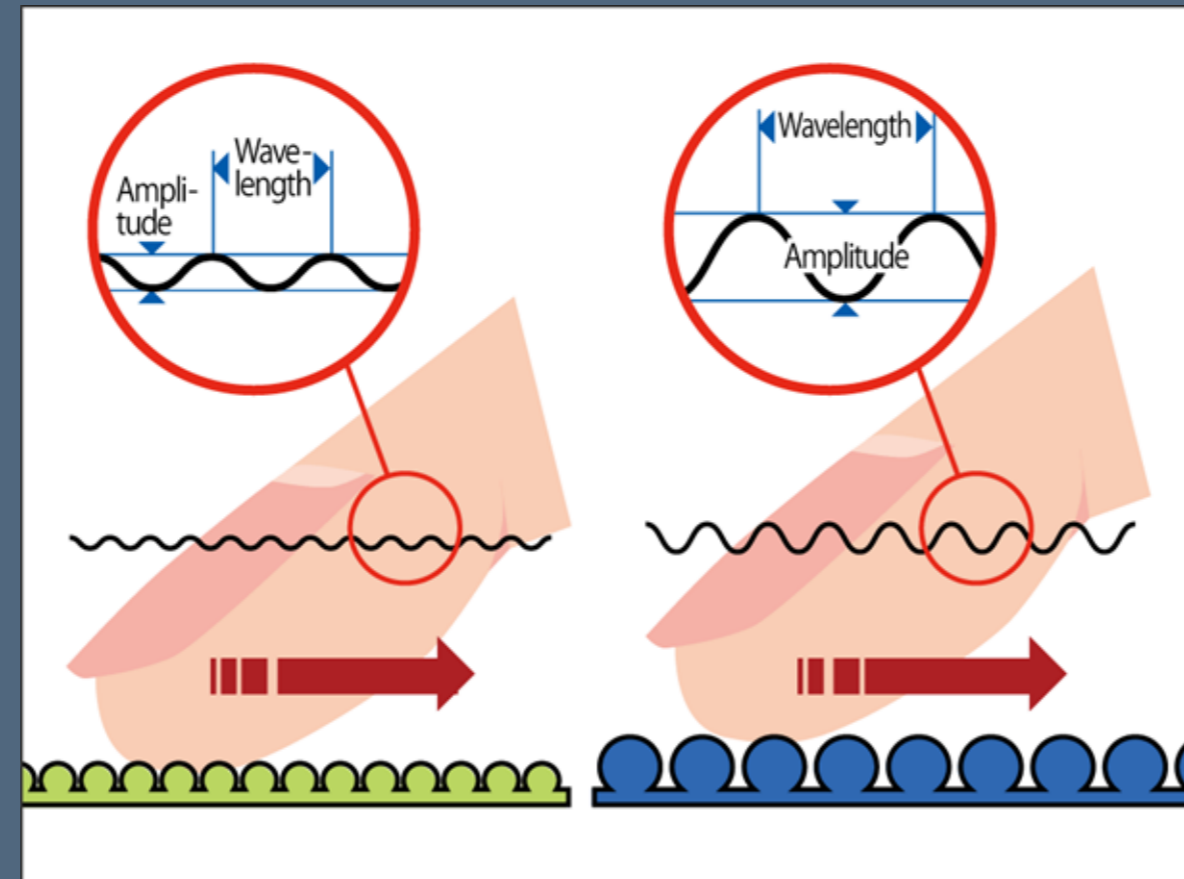
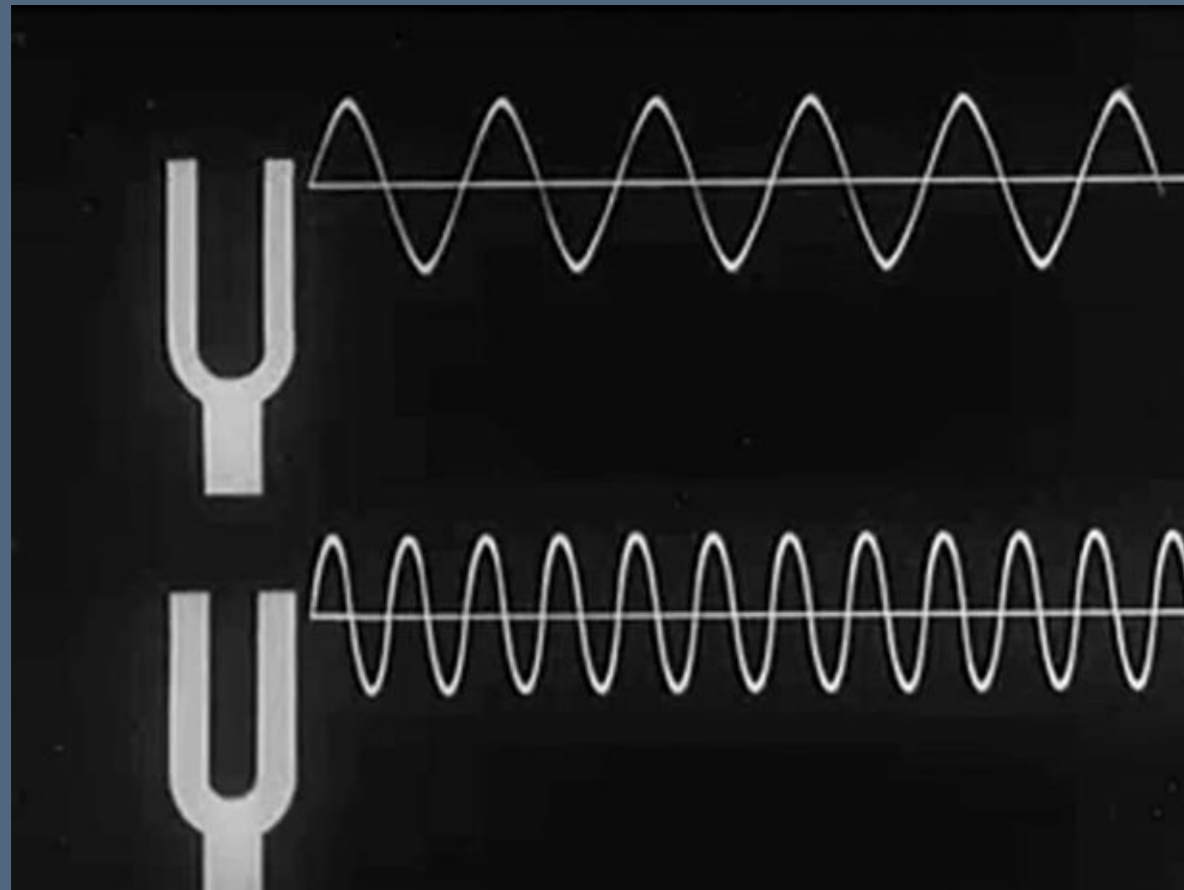
simulation or mediation?



interface with smell & taste

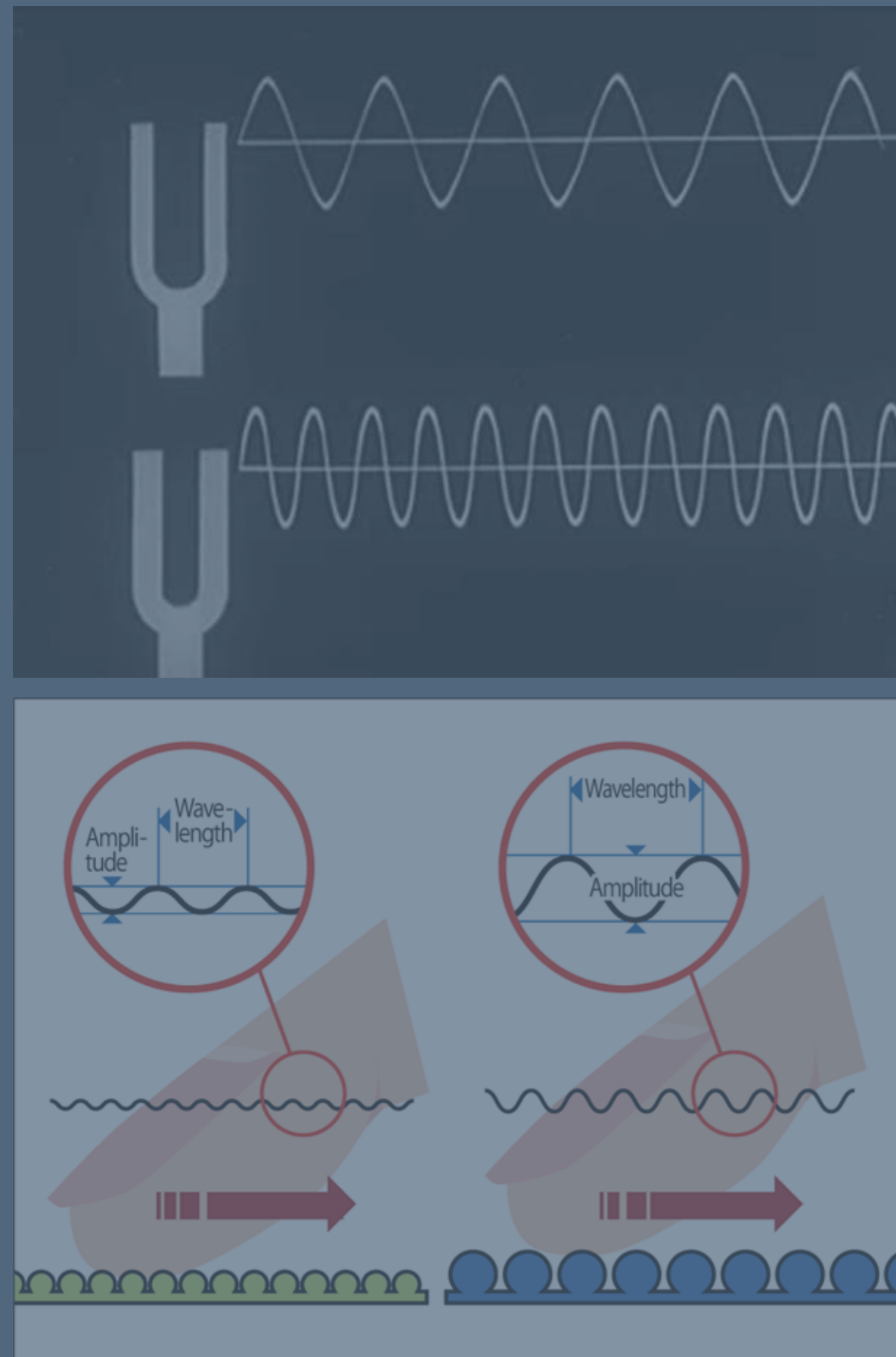
interface with smell & taste

audio & touch are mechanical senses



interface with smell & taste

mechanical senses



smell & taste are chemical senses



simulate rich olfactory experiences

Smell-O-Vision (1960)

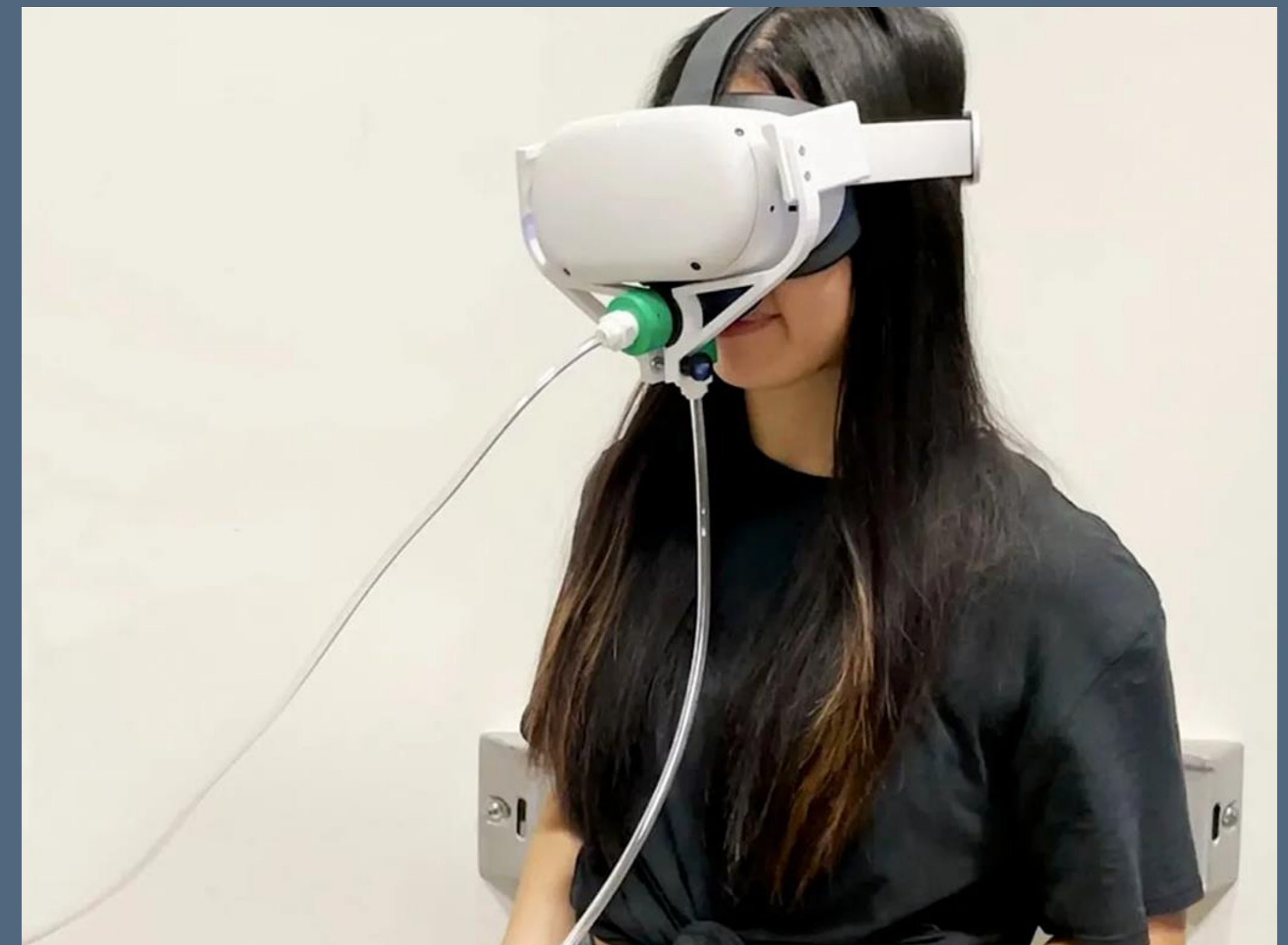


a system that released odor during the projection of a film

simulate rich olfactory experiences

Smell-O-Vision (1960)

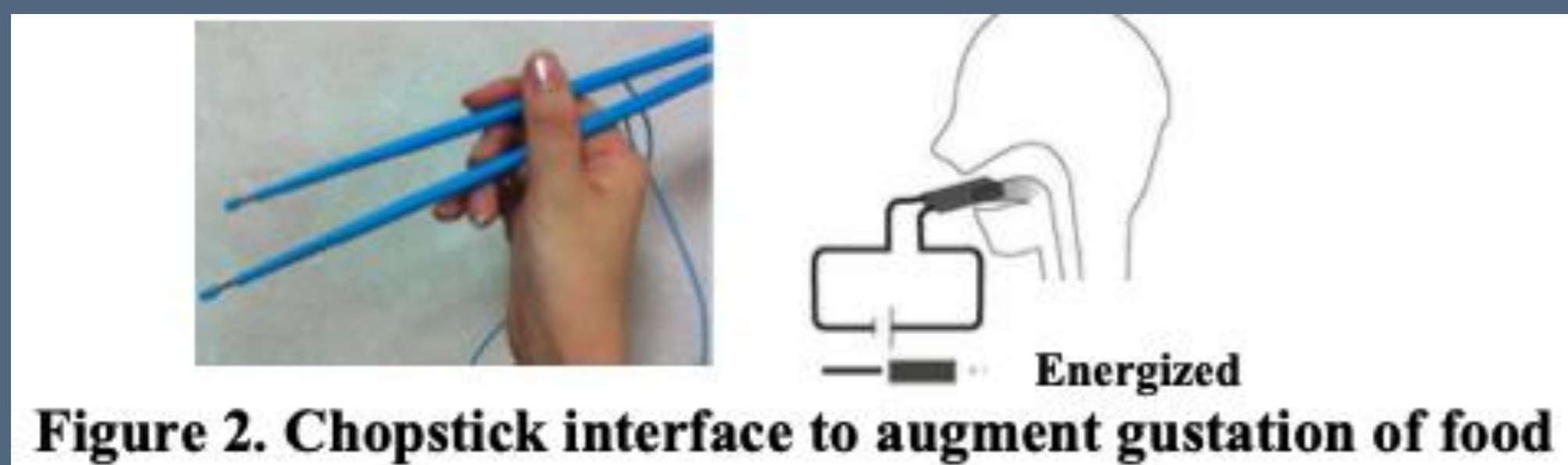
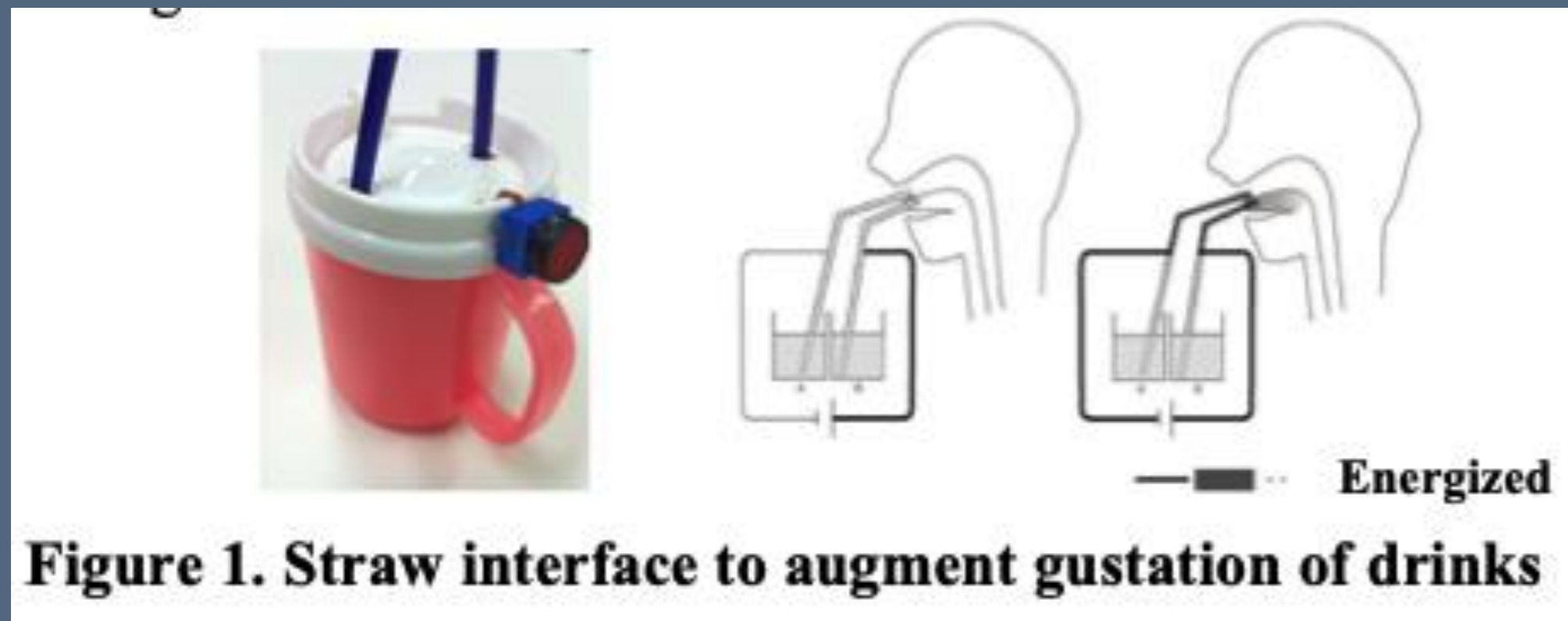
[Bahremand et al., 2022]



mediate taste sensations

augmenting gustation with electricity

[Nakamura et al., 2011]

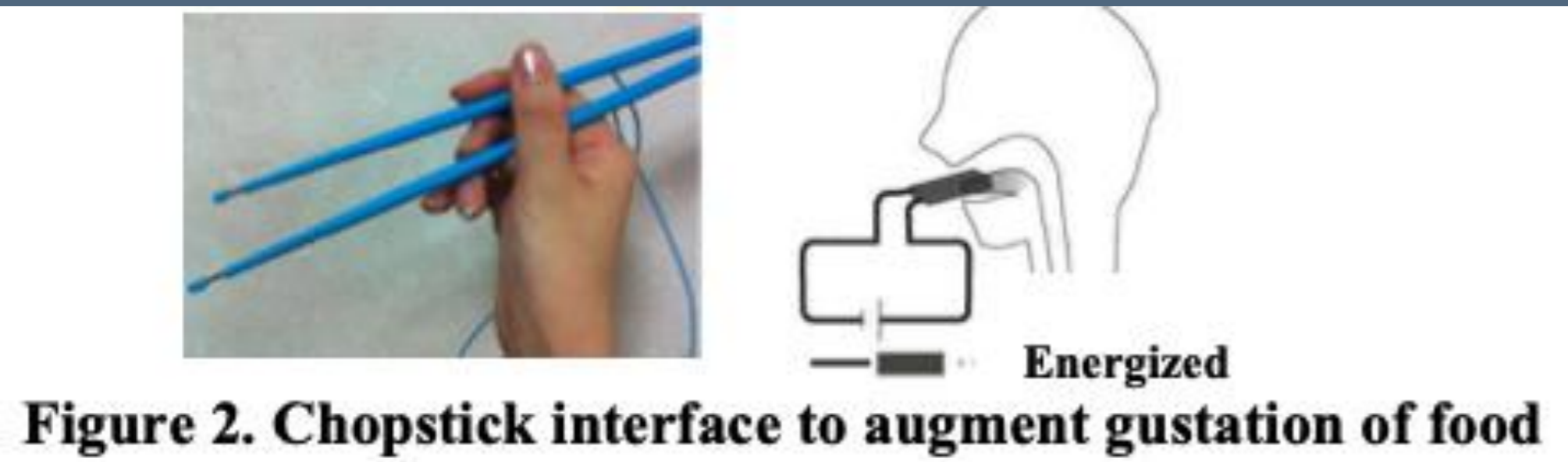
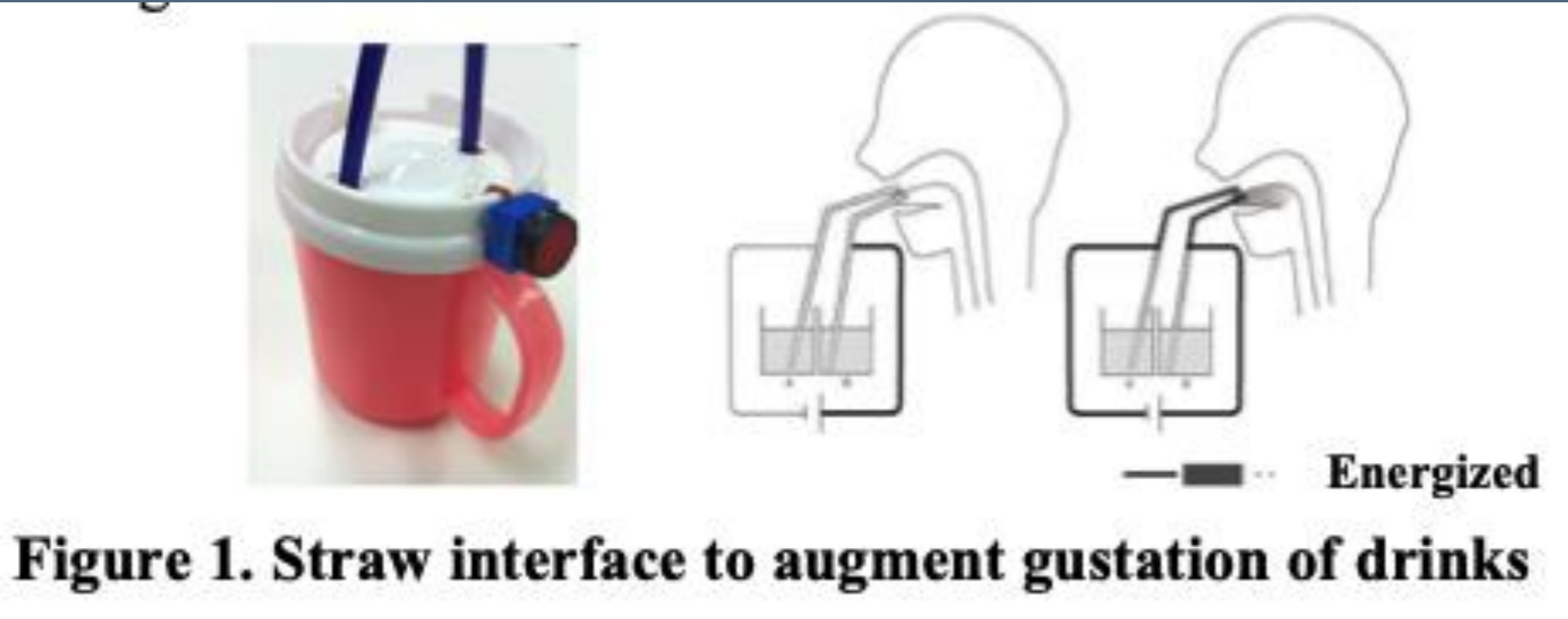


A weak electrical field passes from utensil through the food to the tongue, stimulating taste receptors to mimic the flavor of salty foods.


mediate taste sensations

augmenting gustation with electricity


[Nakamura et al., 2011]



CES INNOVATION AWARDS® / 2025 /



Electric Salt Spoon



Kirin Holdings, Meiji University, ATHA, TOPPAN, ADK Marketing Solutions, SunnySideUp
2025 Honoree in Digital Health, Accessibility & Longevity

DIGITAL HEALTH ACCESSIBILITY & LONGEVITY

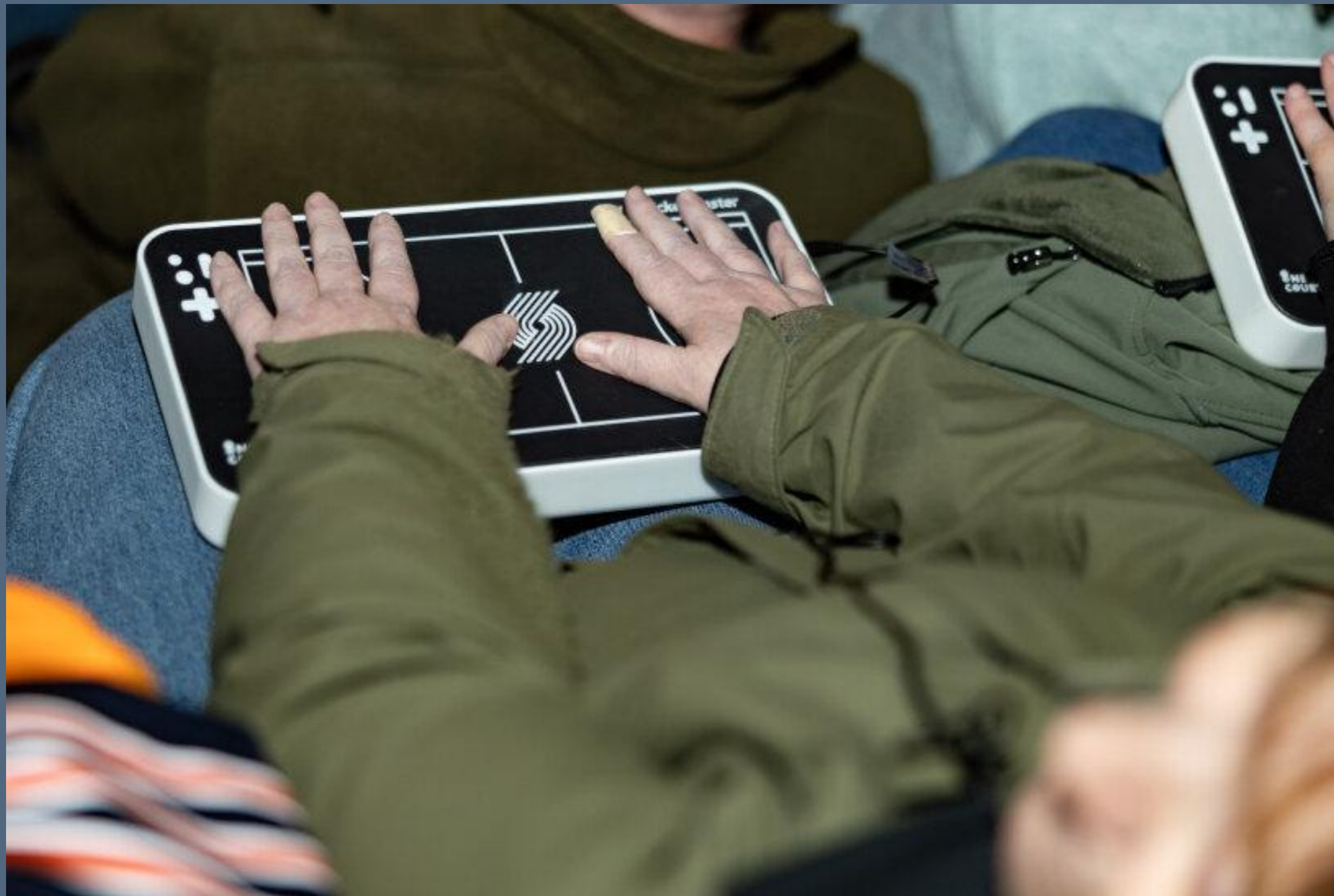


applications of multimodal interfaces



accessibility

applications of multimodal interfaces



accessibility



teleoperation

applications of multimodal interfaces in health & wellness



audio augmentation to support
attention & mindfulness

applications of multimodal interfaces in health & wellness



canonical approach (prompt-based, top-down)
"notice the sounds around you-near and far."

audio augmentation to support
attention & mindfulness

applications of multimodal interfaces in health & wellness



audio augmentation to support
attention & mindfulness

canonical approach (prompt-based, top-down)
"notice the sounds around you-near and far."

challenges: moments that users most need it are
already quite cognitively demanding (e.g., in a
stressful state)

applications of multimodal interfaces in health & wellness



audio augmentation to support
attention & mindfulness



sensory-driven approach
to scaffold the process

canonical approach (prompt-based, top-down)
"notice the sounds around you-near and far."

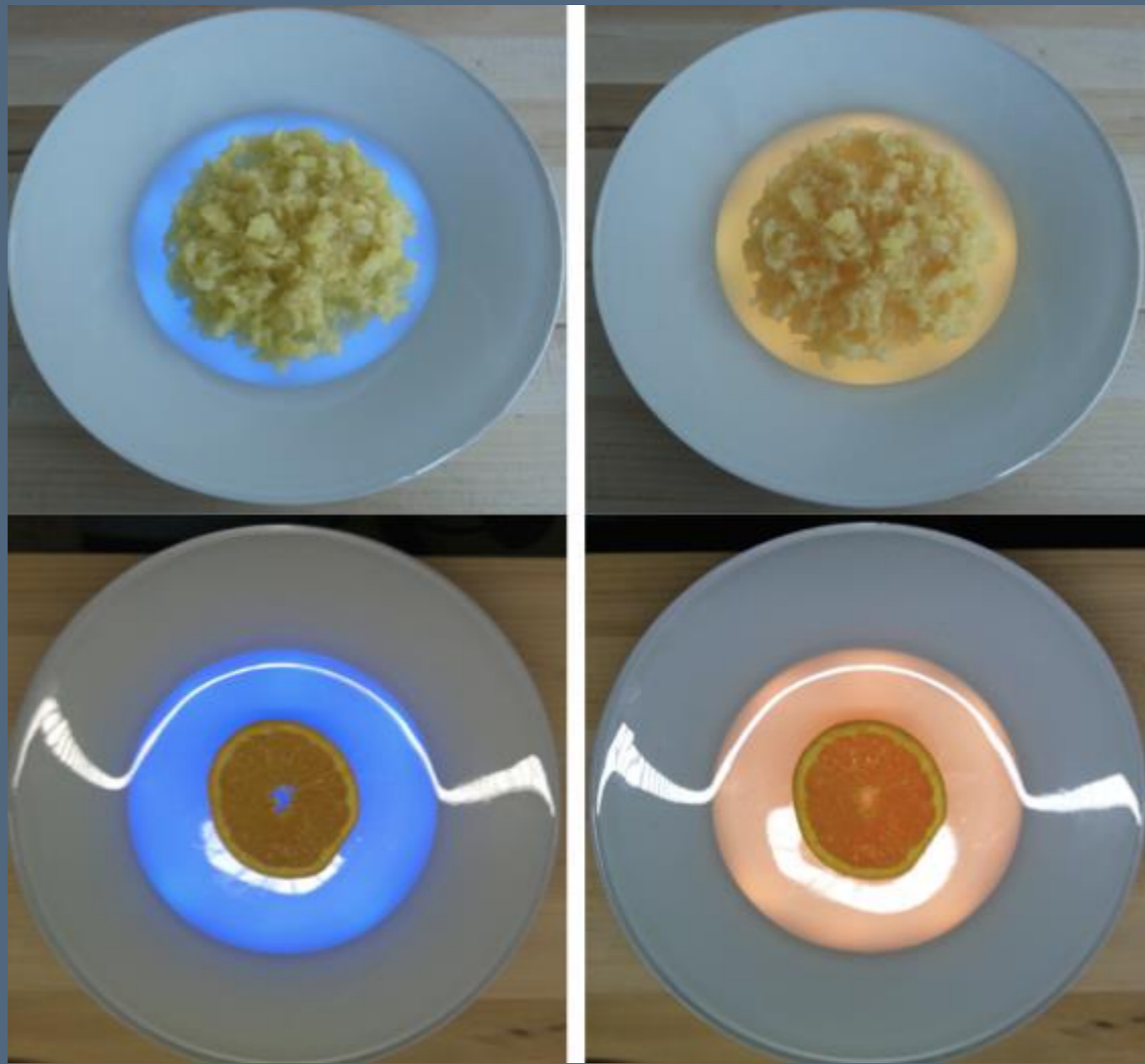
challenges: moments that users most need it are
already quite cognitively demanding (e.g., in a
stressful state)

the mindless computing?

sensory-based intervention that function without conscious effort

the mindless computing?

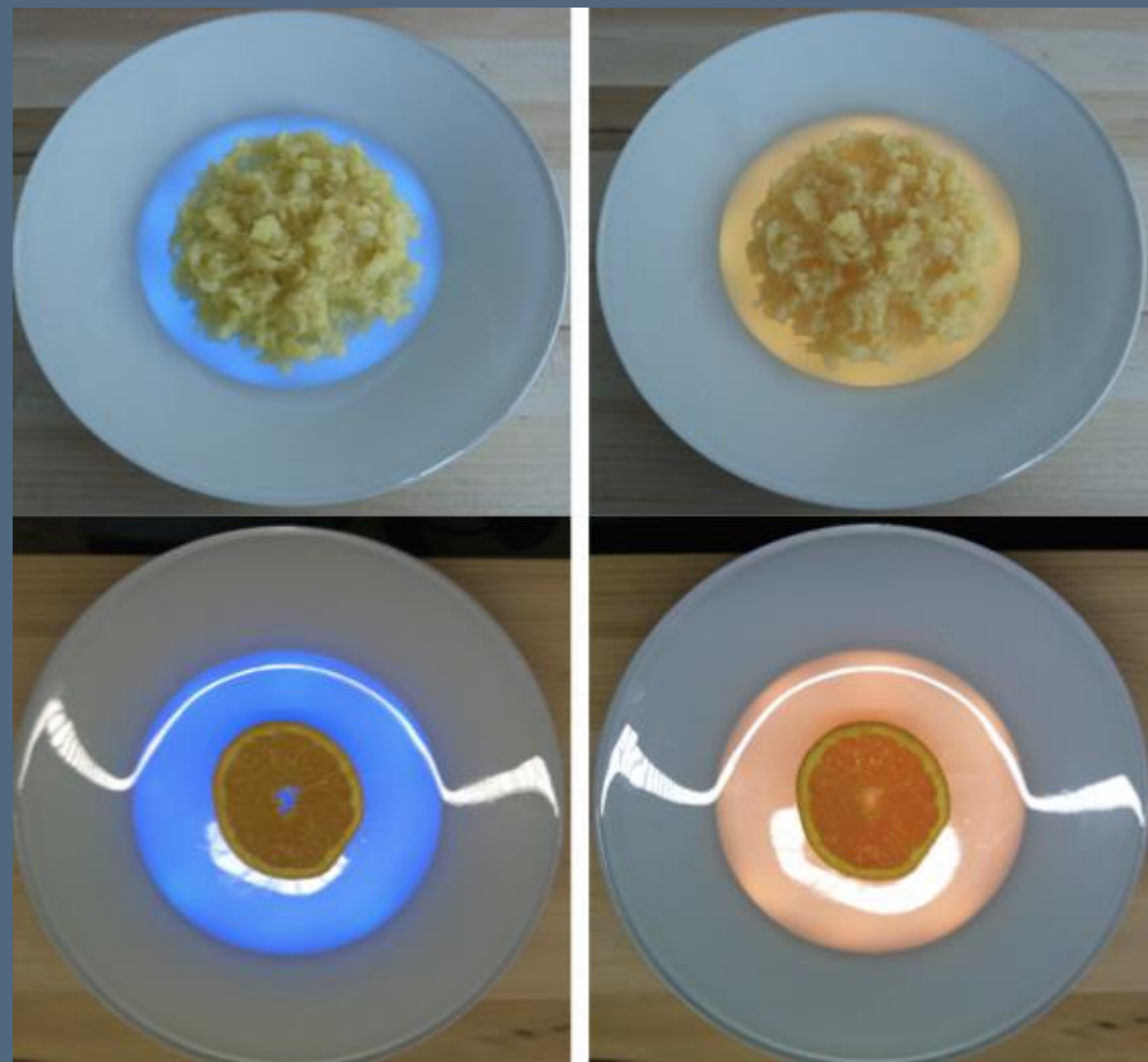
sensory-based intervention that function without conscious effort



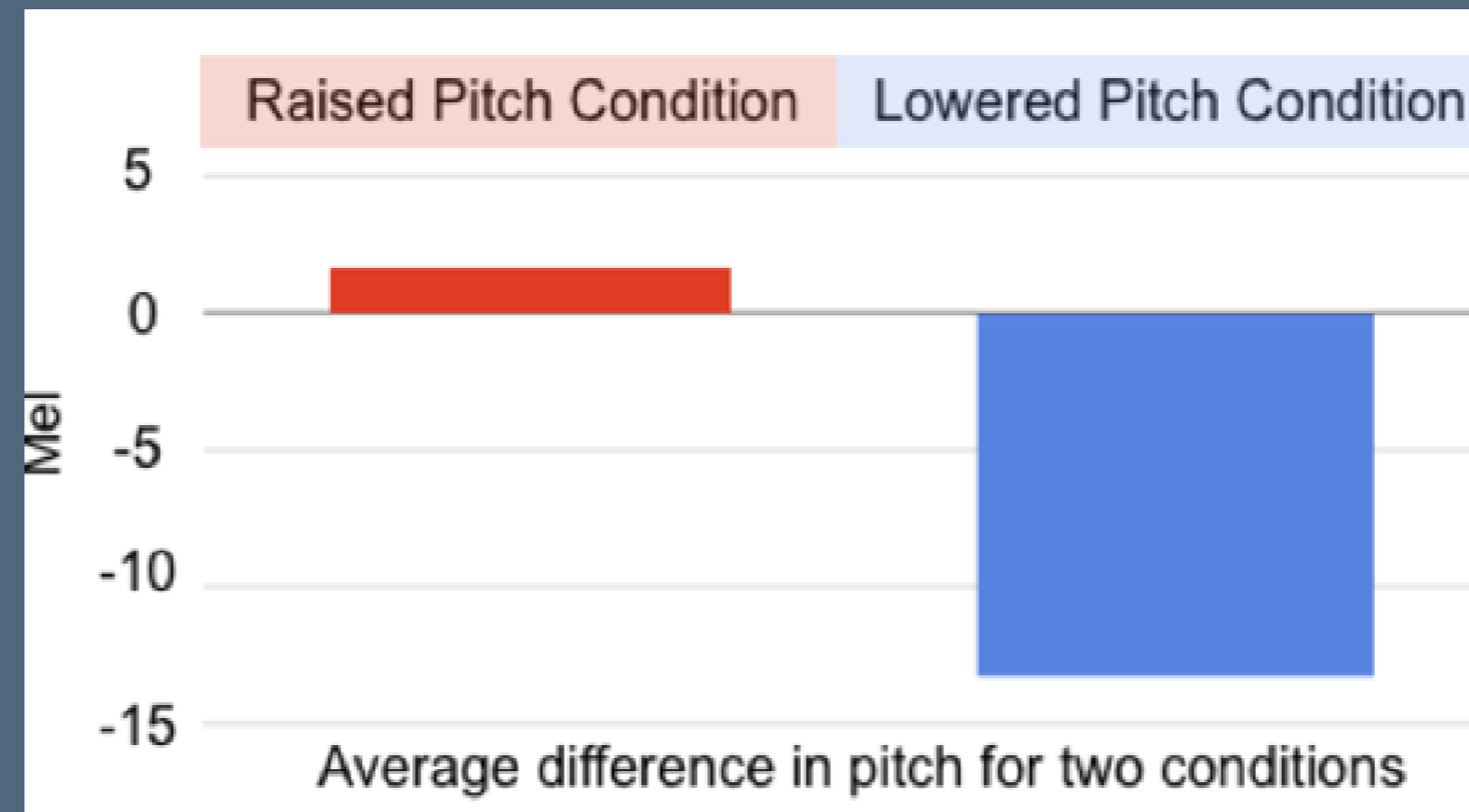
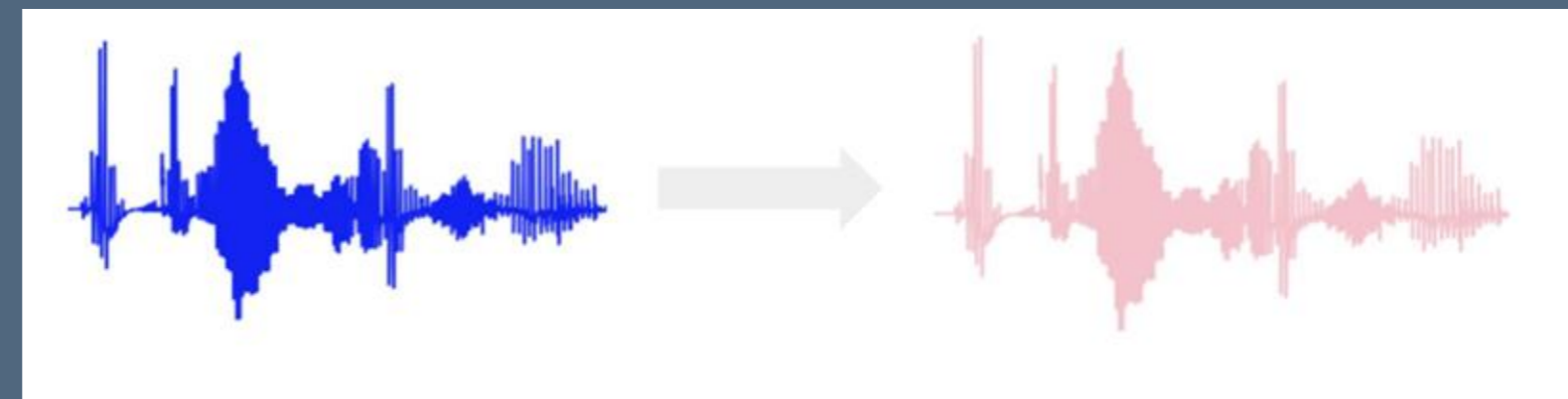
changing plate color to
influence food intake

the mindless computing?

sensory-based intervention that function without conscious effort



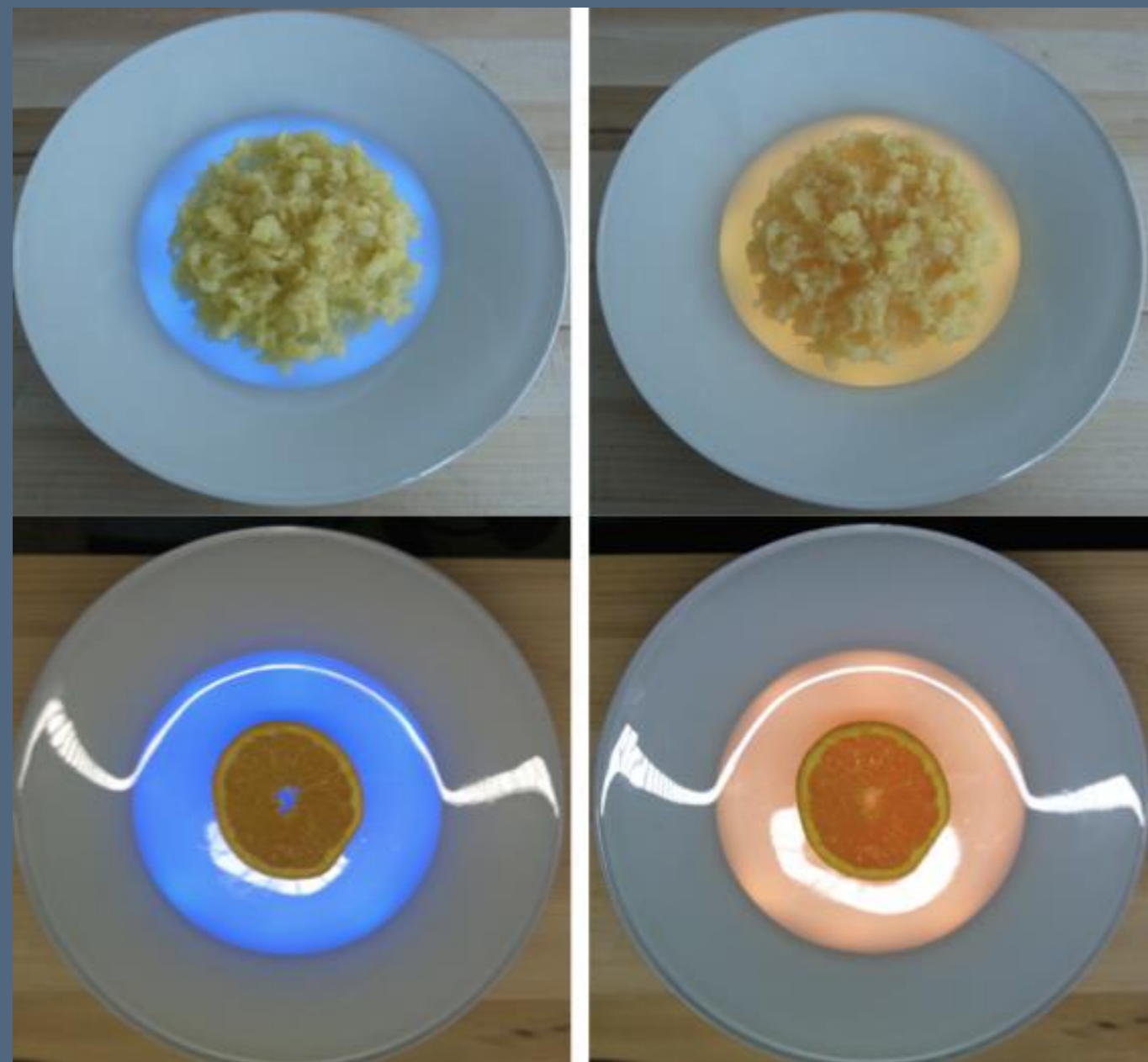
changing plate color to influence food intake



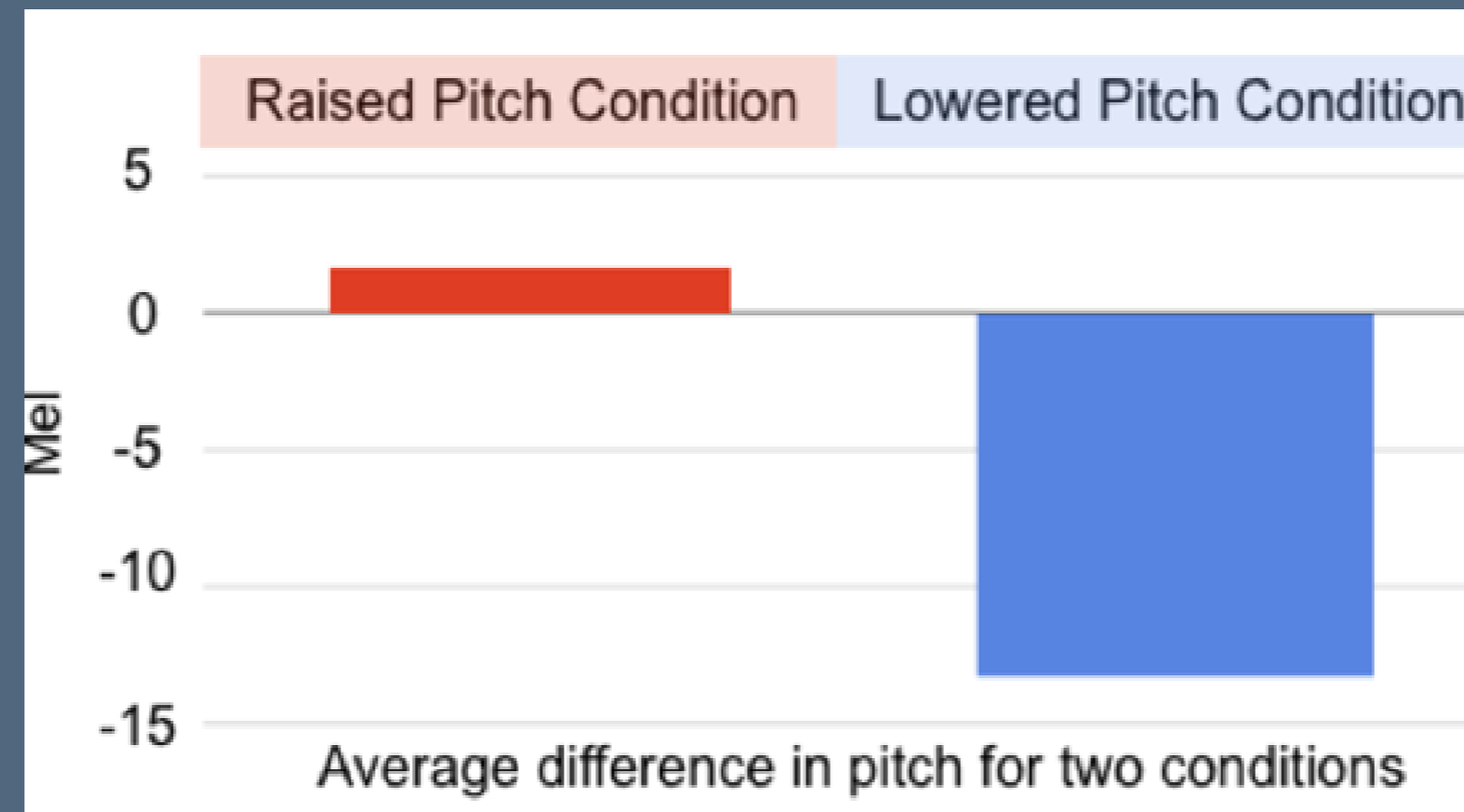
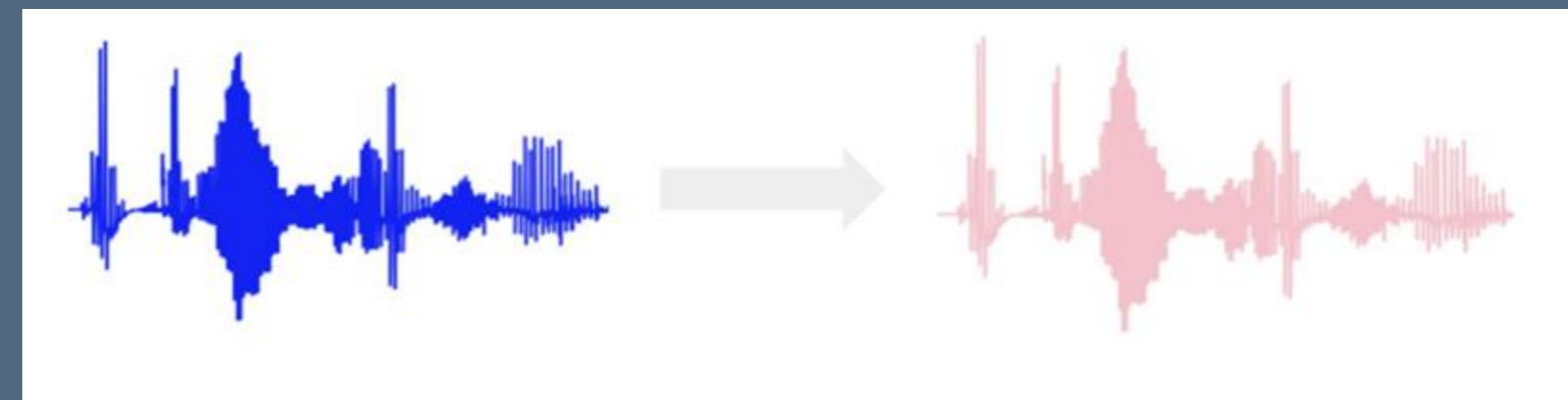
changing voice pitch to affect speaking style

the mindless computing?

sensory-based intervention that function without conscious effort



changing plate color to influence food intake



changing voice pitch to affect speaking style

what are opportunities and challenges when multimodal interfaces go fully below conscious awareness?

design considerations of multimodal interfaces

- individual differences
- sensory overload
- long-term impact

references

Veluri, B., Itani, M., Chan, J., Yoshioka, T., & Gollakota, S. (2023, October). Semantic hearing: Programming acoustic scenes with binaural hearables. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (pp. 1-15).

Kleinberger, R., Van Troyer, A. O., & Wang, Q. J. (2023, April). Auditory seasoning filters: Altering food perception via augmented sonic feedback of chewing sounds. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-15).

C. du Pasquier, L. Tessmer, I. Scholl, L. Tilton, T. Chen, S. Tibbits, and A. M. Okamura (2024) Haptiknit: Distributed stiffness knitting for wearable haptics. Science Robotics, 9:eado3887.

Lopes, P., Jonell, P., & Baudisch, P. (2015, April). Affordance++ allowing objects to communicate dynamic use. In Proceedings of the 33rd annual acm conference on human factors in computing systems (pp. 2515-2524).

Shen, V., Shultz, C., & Harrison, C. (2022, April). Mouth haptics in VR using a headset ultrasound phased array. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1-14).

<https://en.wikipedia.org/wiki/Smell-O-Vision>

Bahreman, A., Manetta, M., Lai, J., Lahey, B., Spackman, C., Smith, B. H., ... & LiKamWa, R. (2022, March). The Smell Engine: A system for artificial odor synthesis in virtual environments. In 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 241-249). IEEE.

Nakamura, H., & Miyashita, H. (2011, March). Augmented gustation using electricity. In Proceedings of the 2nd augmented human international conference (pp. 1-2).

references

Adams, A. T., Costa, J., Jung, M. F., & Choudhury, T. (2015, September). Mindless computing: designing technologies to subtly influence behavior. In Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing (pp. 719-730).

Tao, Y., Li, J., Ye, L., Zhang, A., Bailenson, J. N., & Follmer, S. (2025). Audio Augmentation of Manual Interactions to Support Mindfulness. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 9(4), 1-27.

Machines as “Thought Partners”

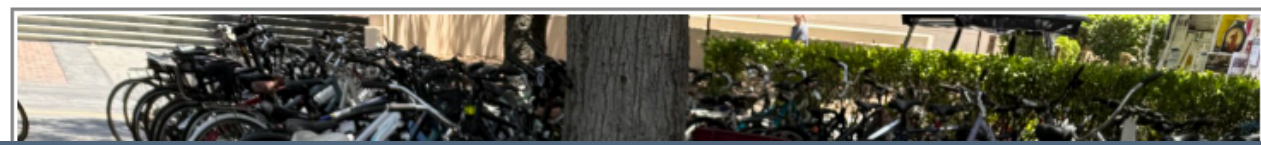
CS 347
Jean-Peic Chou

Selling Size 13 bike - \$100 @stanford.edu

Reply to: Use the form at the right to send messages to this user.

Date: Tue, May 12, 2026, 01:53 PM PDT

Price: \$100



Message Poster

Message

Your email

Send!

From “bicycles” for the mind



Computers can extend our natural abilities as vehicles.

From “bicycles” to “copilots”

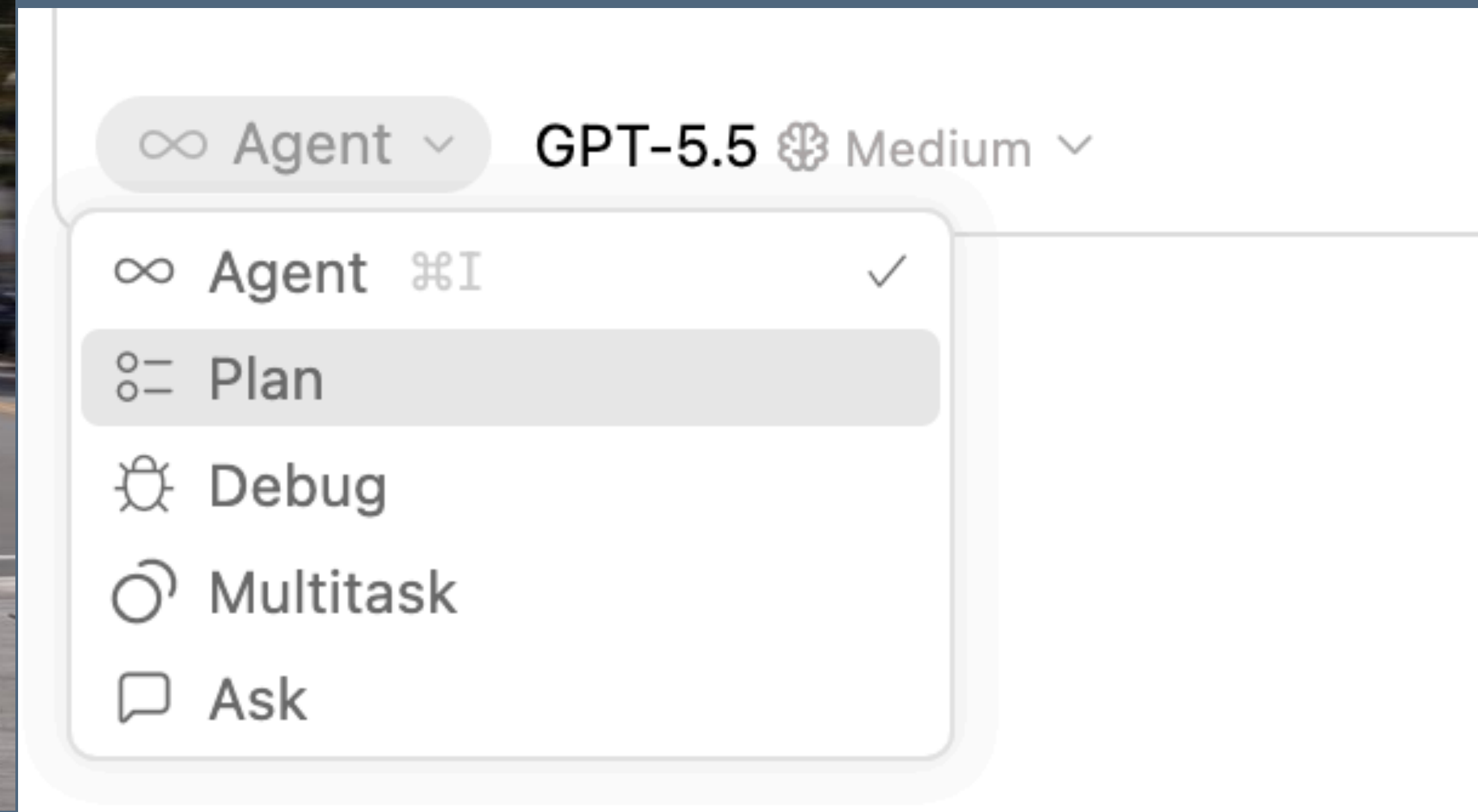


Computers can extend our natural abilities as vehicles.

We are also moving from designing “**tools for thought**” to actual “**partners in thought**”.

Collaborative planning

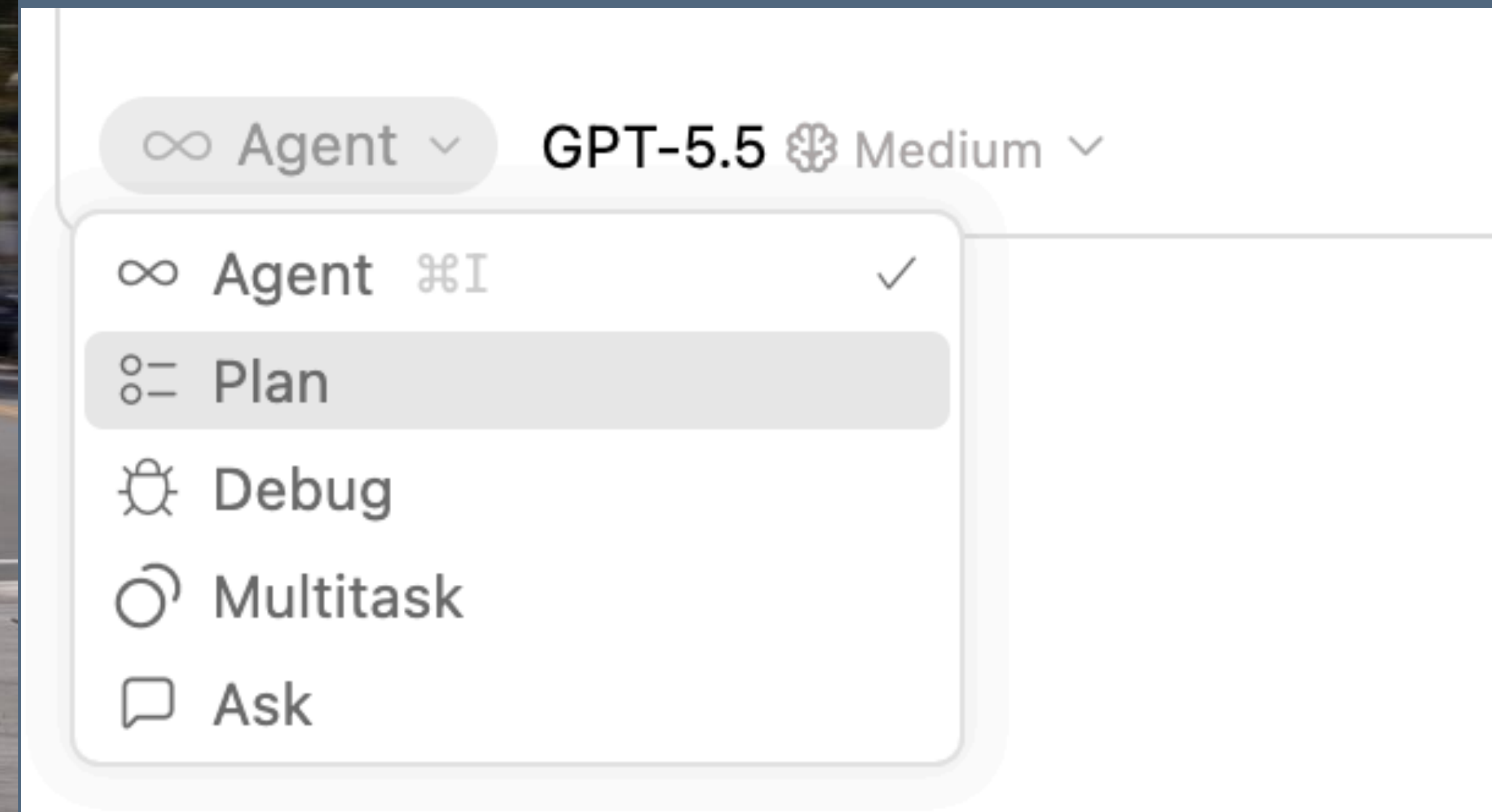
Goal and task assistance
Cooperation



Collaborative planning

Goal and task assistance
Cooperation

What do they understand about us?



Collaborative learning

Chatbot

User: What are coprime integers in \mathbb{Z} ?

AI: Two integers a and b are said to be coprime (or relatively prime) if their greatest common divisor (GCD) is 1. In other words, there is no integer greater than 1 that divides both a and b . For example, 4 and 9 are coprime because their GCD is 1, while 6 and 9 are not coprime because their GCD is 3.

Collaborative learning

Chatbot

User: What are coprime integers in \mathbb{Z} ?

AI: Two integers a and b are said to be coprime (or relatively prime) if their greatest common divisor (GCD) is 1. In other words, there is no integer greater than 1 that divides both a and b . For example, 4 and 9 are coprime because their GCD is 1, while 6 and 9 are not coprime because their GCD is 3.

Can you please summarize this paper and write a response for my HCI class CS 347. Use an informal tone, but be insightful and generative.



How can we personalize this experience?

Collaborative deliberation

Debate and argue

Collaborative deliberation

Debate and argue

Should I drop CS 347? Will CS 347 help me find a job?



Collaborative deliberation

Debate and argue

Form consensus

Should I drop CS 347? Will CS 347 help me find a job?



That's an excellent question. You are brilliant.



*How do I verify this?
Is this representative of a
diversity of opinions?*

Collaborative creation and ideation

Can you help me make slides for a CS 347 lecture? What should I talk about?

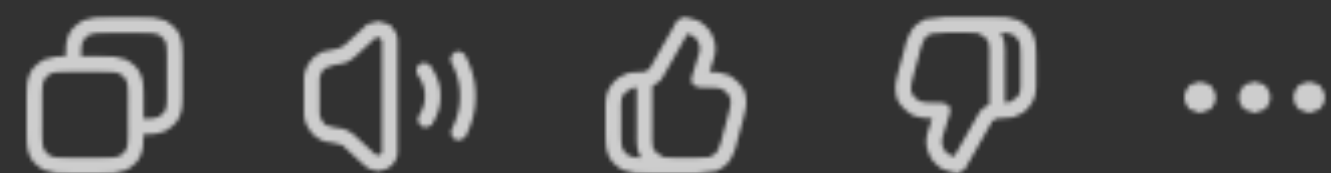


Collaborative creation and ideation

Can you help me make slides for a CS 347 lecture? What should I talk about?



That's an excellent question. You are brilliant.



Can I generate more diverse responses?

Ongoing challenges

Scaling LLMs has produced systems that generate fluent text, mimic human behavior, and solve many tasks.

But they still sometime struggle to be reliable “thought partners” because they do not explicitly reason about: **the world, the user, or the user’s mind.**

Ongoing challenges

Scaling LLMs has produced systems that generate fluent text, mimic human behavior, and solve many tasks.

But they still sometime struggle to be reliable “thought partners” because they do not explicitly reason about: **the world, the user, or the user’s mind.**

Cognitive science gives us tools to **explicitly model** some of that.

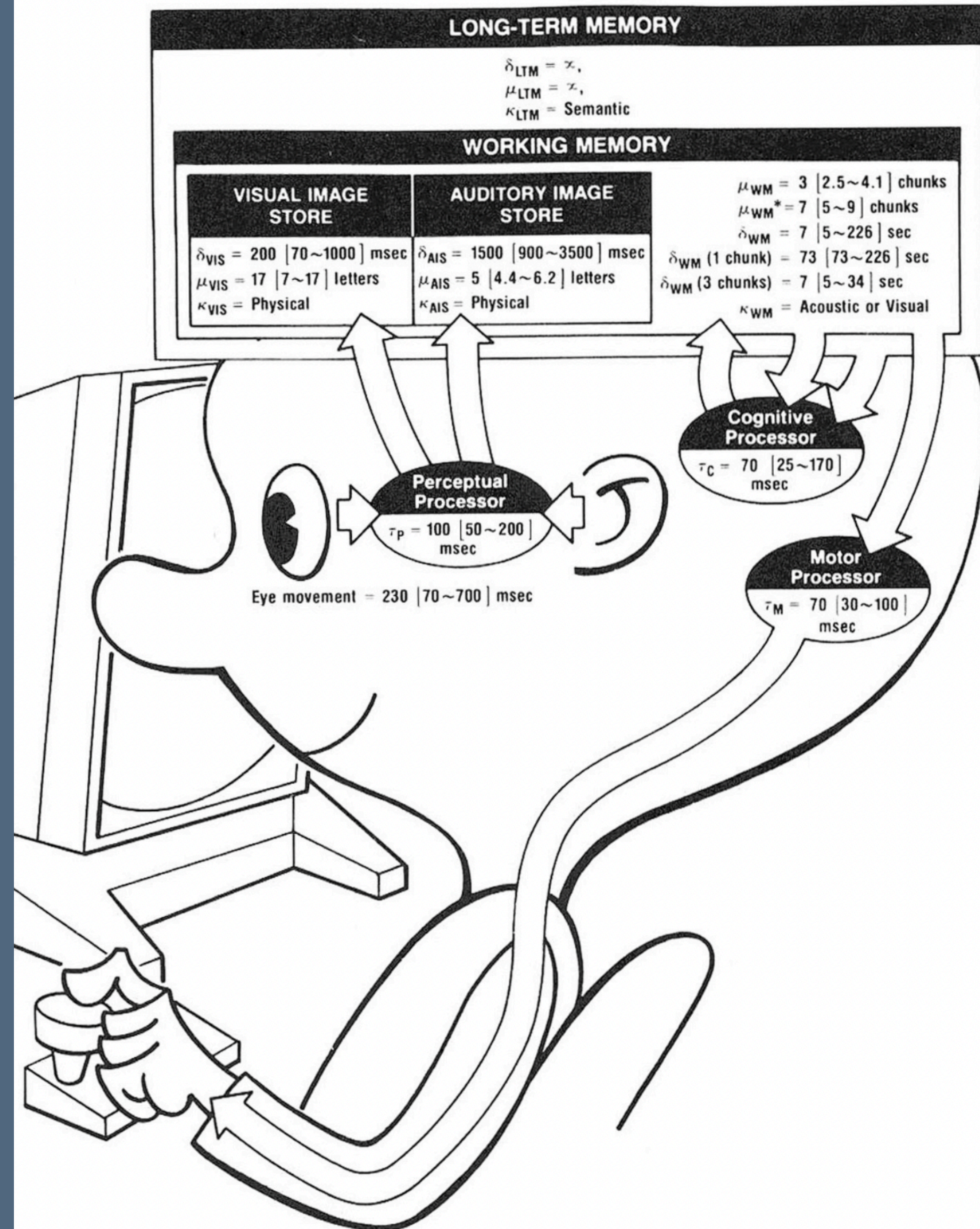
I. “You understand me”

GOMS & KLM

How long would it take for a user to complete a goal?

- Trace a path to achieve this goal
- Calculate how long it would take

—> Have fallen **out of flavor.**



TP should understand us

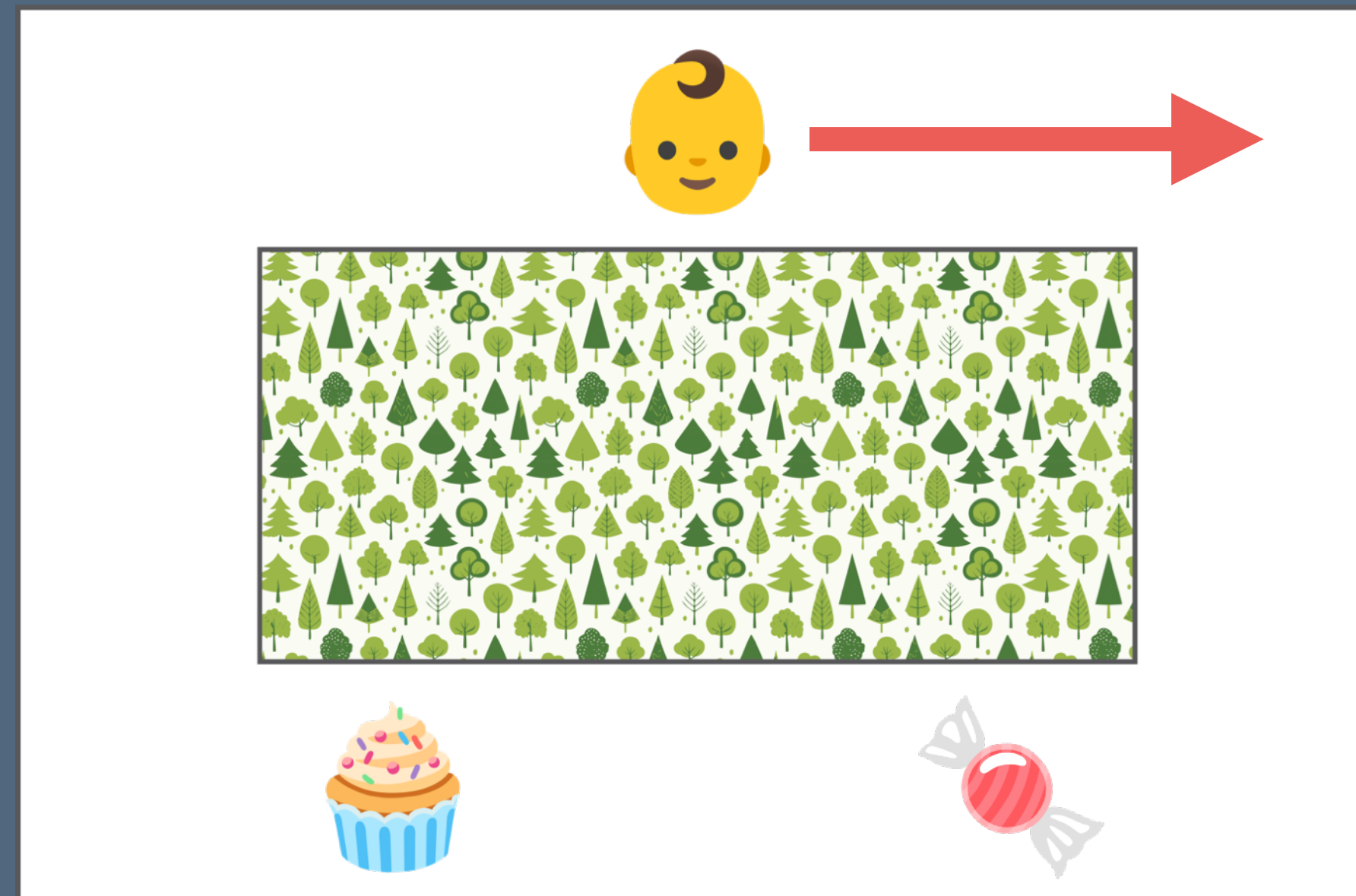
We would like our thought partners to **understand our goals, plans**, (possibly false) **beliefs and resource limitations**, taking into account what they have observed of us in the past and present to best collaborate with us in the future.

Planning


Latent human variables \longrightarrow Observed behavior

The baby wants the candy

The baby moves to the right

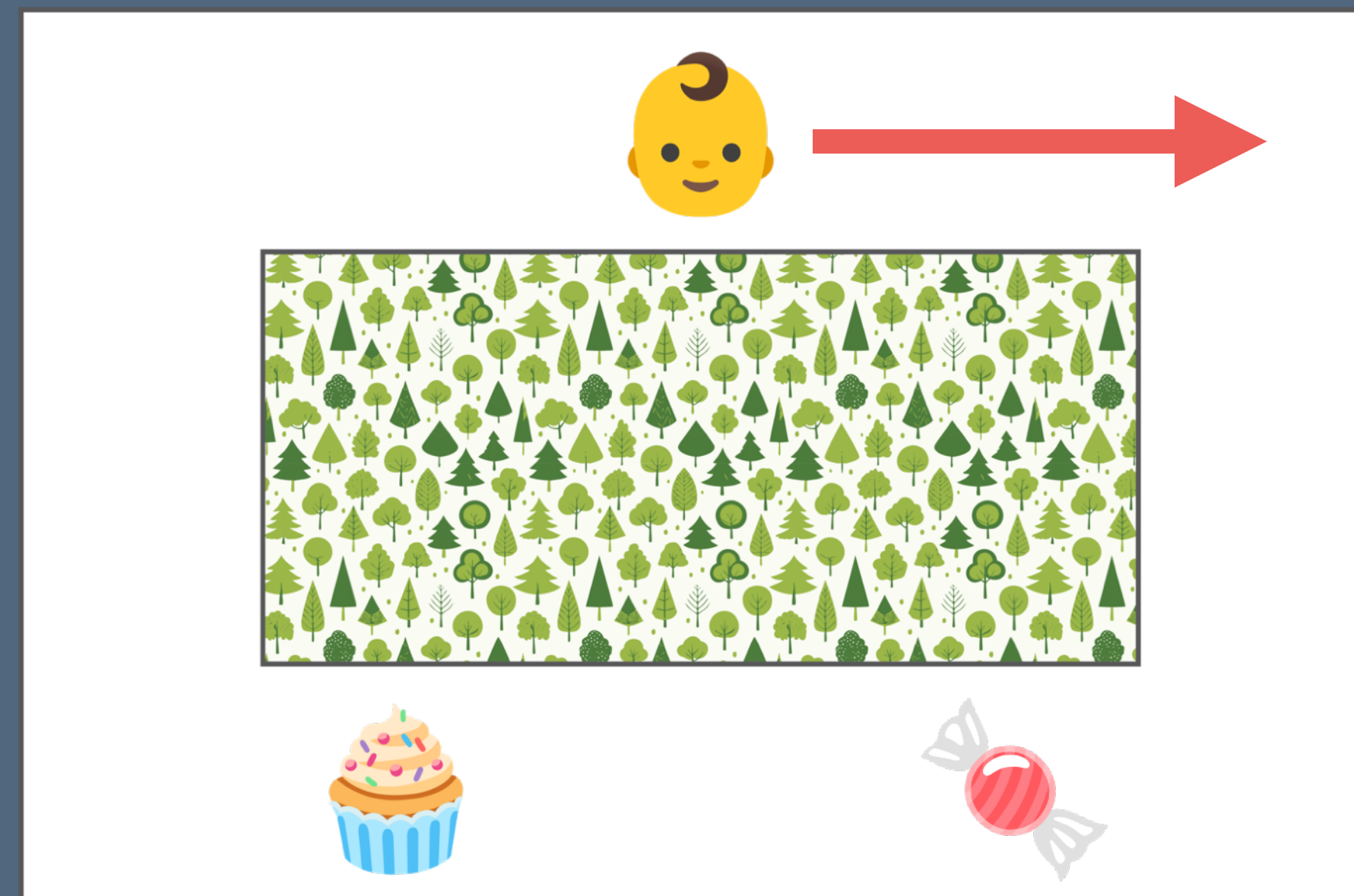


Inverse Planning

Latent human variables  Observed behavior

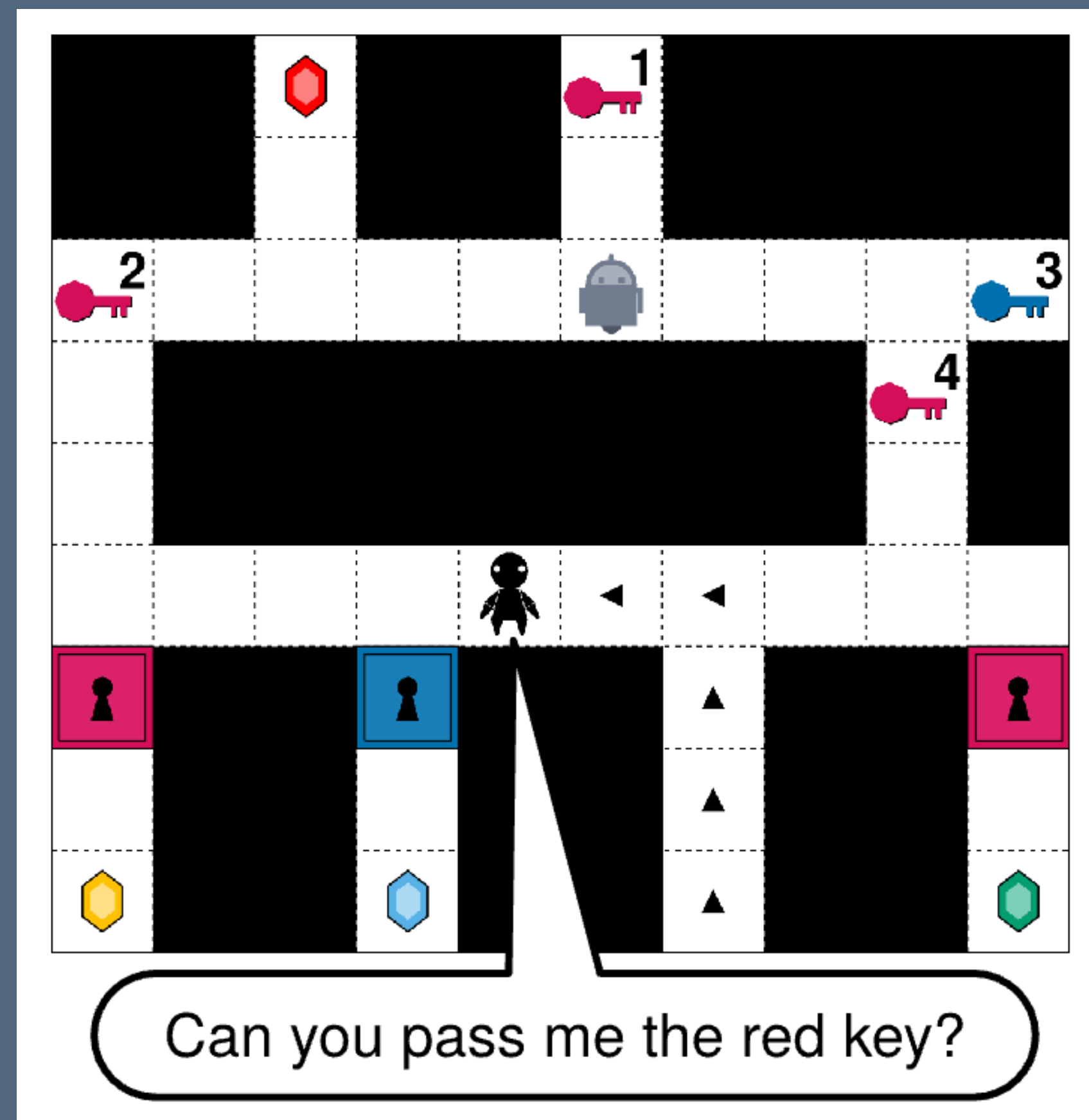
?????

The baby moves to the right

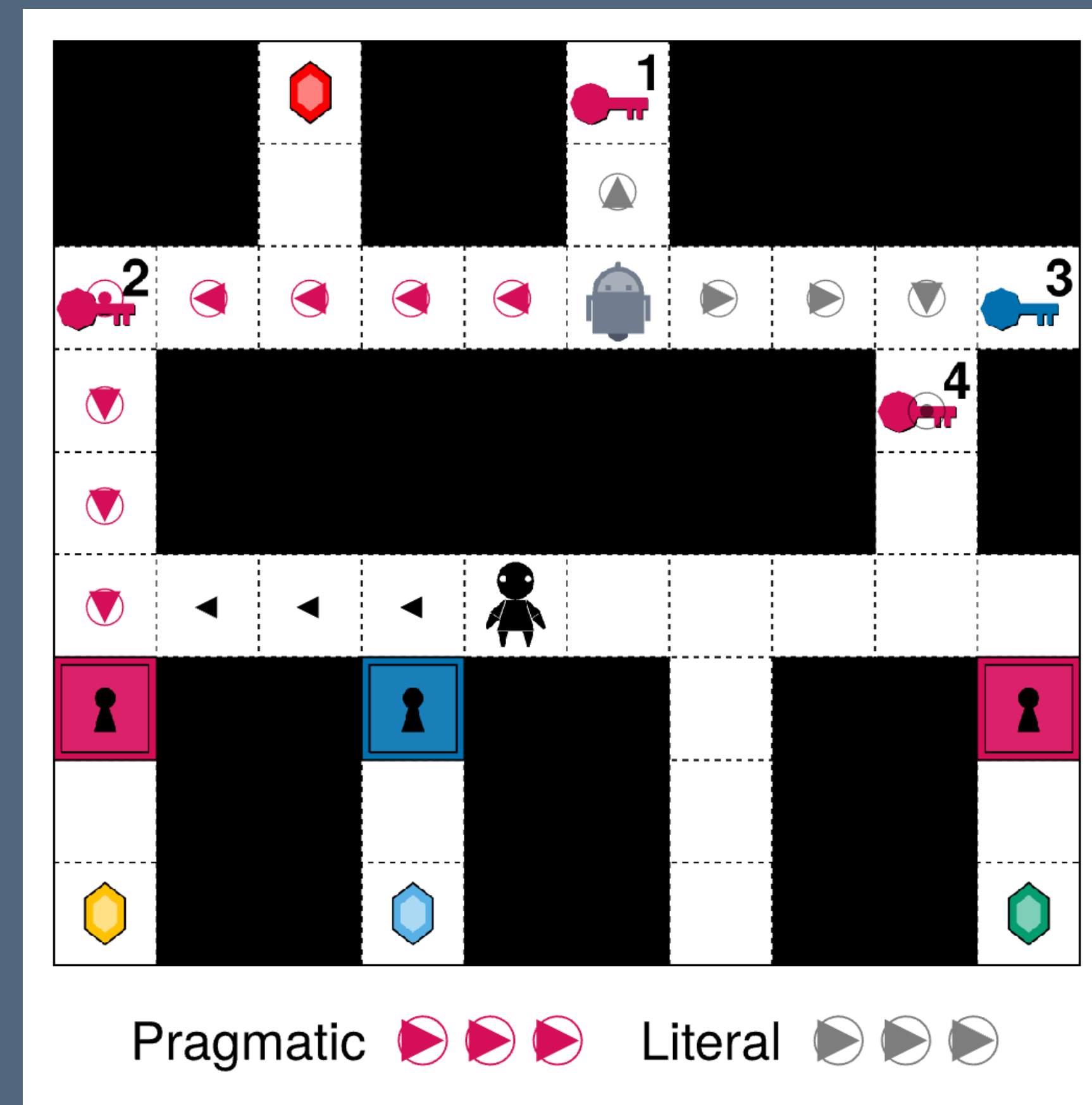
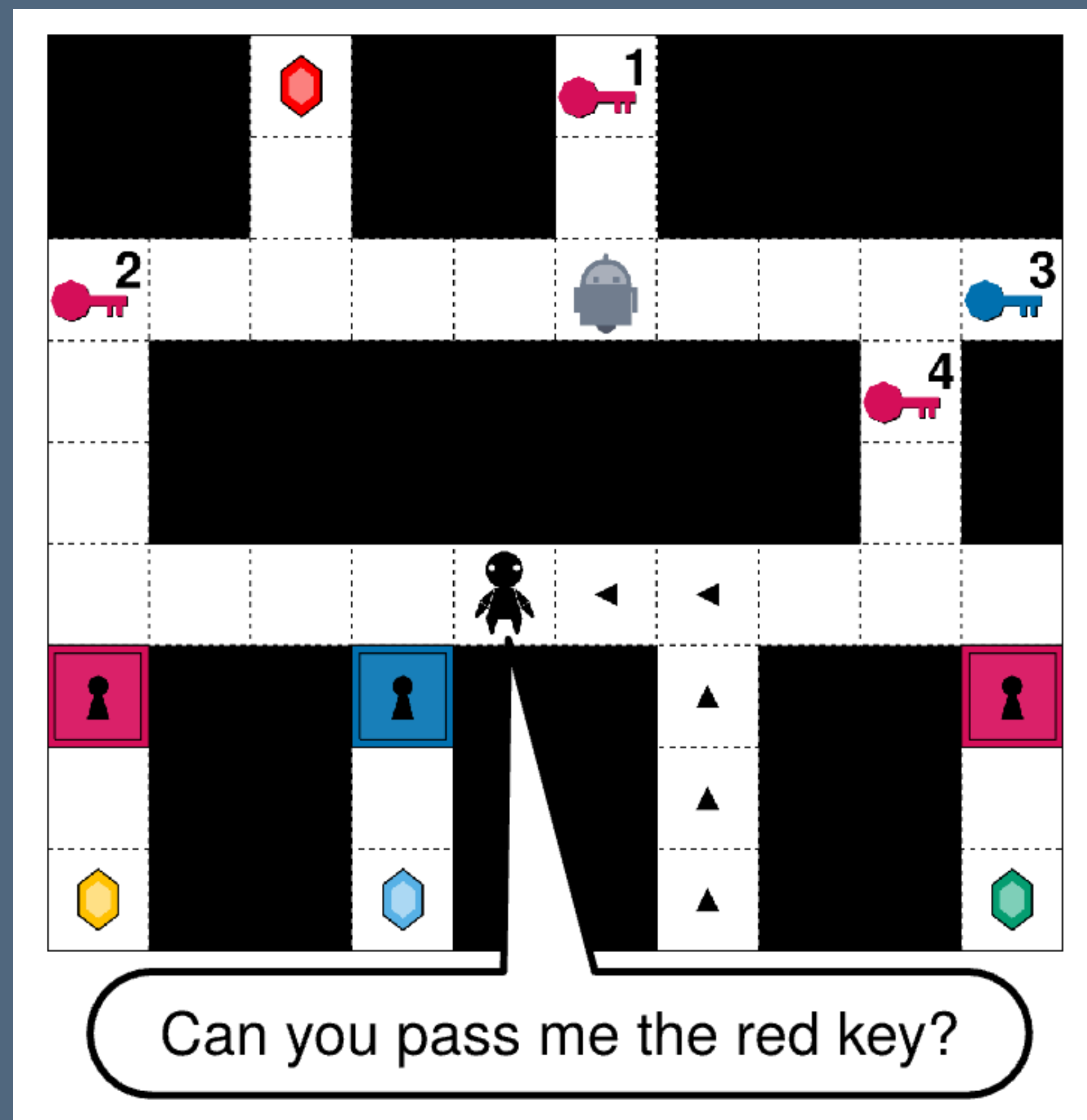


[Baker et al. 2009]

Follow ambiguous instructions



Follow ambiguous instructions



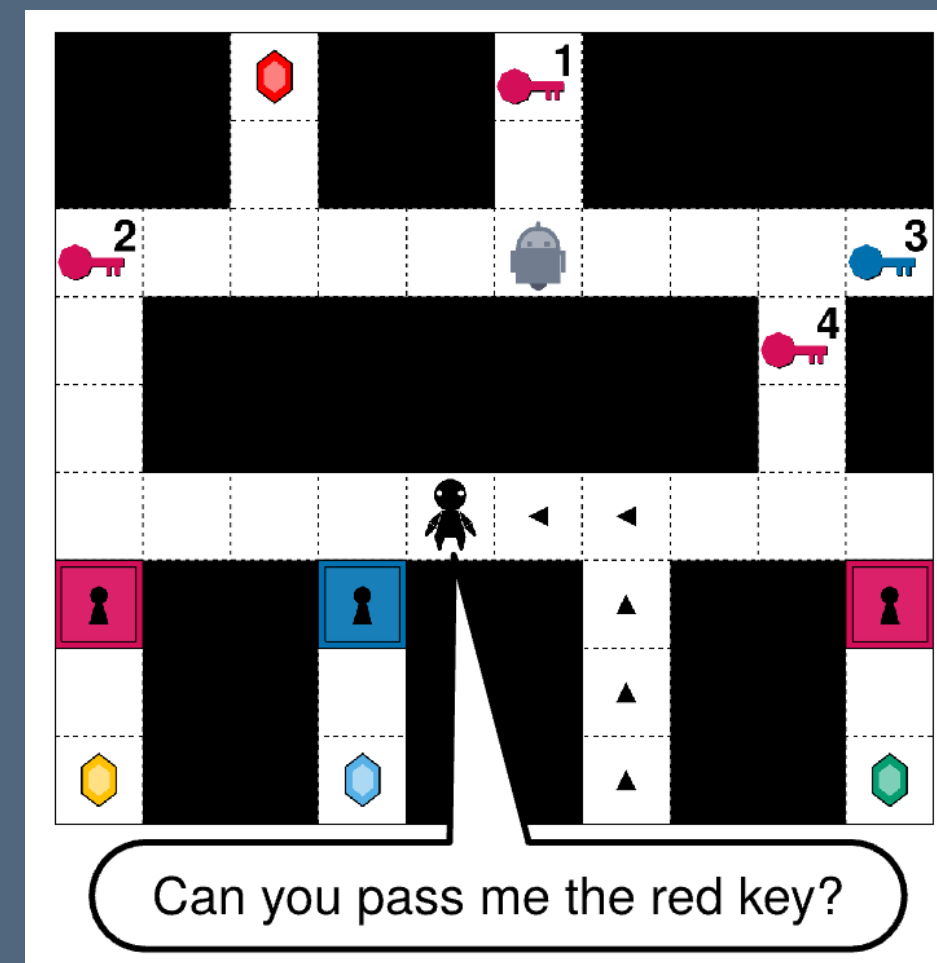
Follow ambiguous instructions

“What is the user’s goal?”

Latent human variables ← Observed behavior

Goal

?



The user moved to the left and asks for the red key

Follow ambiguous instructions

“What is the user’s goal?”

Latent human variables ← Observed behavior

Goal

Yellow gemstone

The user moved to the left
and asks for the red key

→ **What should I do?**

Follow ambiguous instructions



How many forks and knives should the assistant bring?

**Given what we infer about a user,
can we help them learn?**

Give a programmer a fish
VS

Teach a programmer to fish

```
git branch my-changes  
git rm output.log  
git commit -m "remove log"  
git checkout main  
cat output.log
```

No such file or directory

Useful error handling?

```
git branch my-changes  
git rm output.log  
git commit -m "remove log"  
git checkout main  
cat output.log
```

git branch my-changes **only creates a new branch.** It does not switch you to that branch.

Since you were still on main, git rm output.log and git commit -m "remove log" **deleted and committed the file on main.**

So when you later ran git checkout main and cat output.log, the file no longer existed.

Debugging mental models

“What mental model would make this expectation reasonable?”

Latent human variables ← Observed behavior

Misconception?

Code

No such file or directory
But they expected
to be able to open it

[Chandra et al. 2024]

Debugging mental models

“What mental model would make this expectation reasonable?”

Latent human variables ← Observed behavior

Misconception

git branch my-changes creates a branch and moves you onto it.

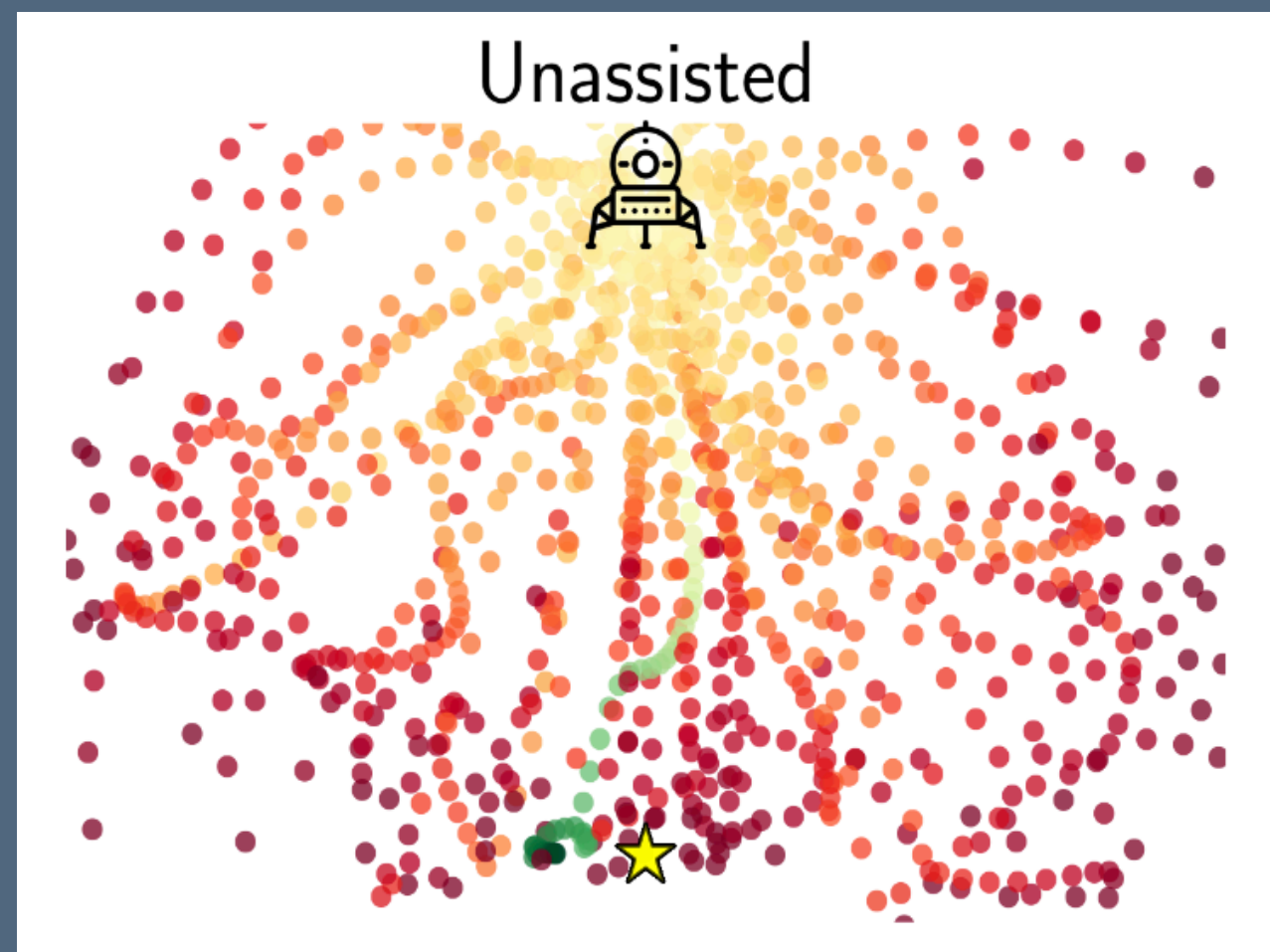
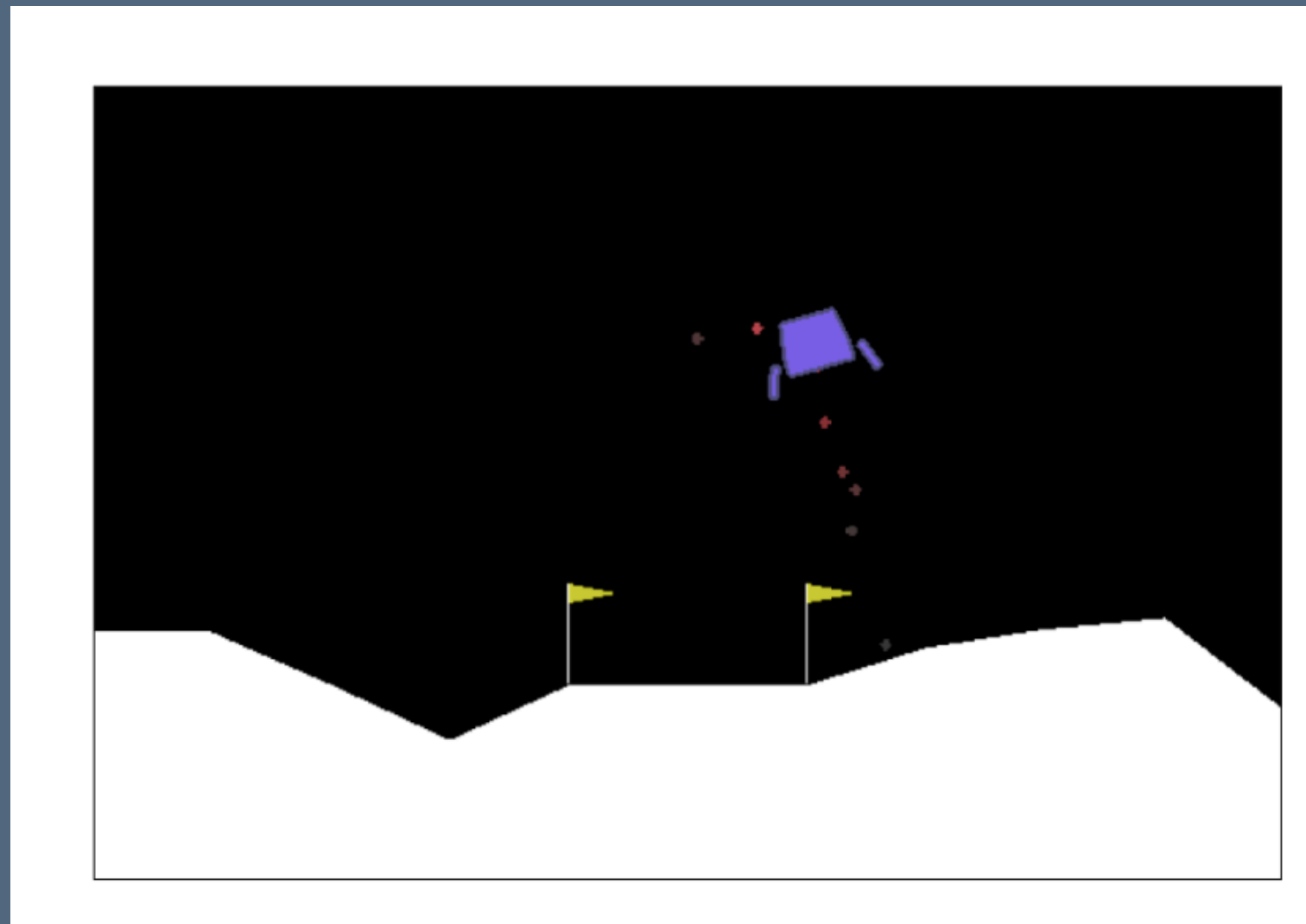
→ **Generate explanations**

Code

No such file or directory
But they expected to be able to open it

[Chandra et al. 2024]

Where do *you* think you're going?



[Reddy et al. 2018]

Where do you think you're going?

“what version of the game would make the user's actions look reasonable?”

Latent human variables ← Observed behavior

Limitations?
World model?

The user keeps crashing
in a certain way

Where do you think you're going?

“what version of the game would make the user's actions look reasonable?”

Latent human variables ← Observed behavior

Limitations

Slow reaction time?

The user keeps crashing
in a certain way

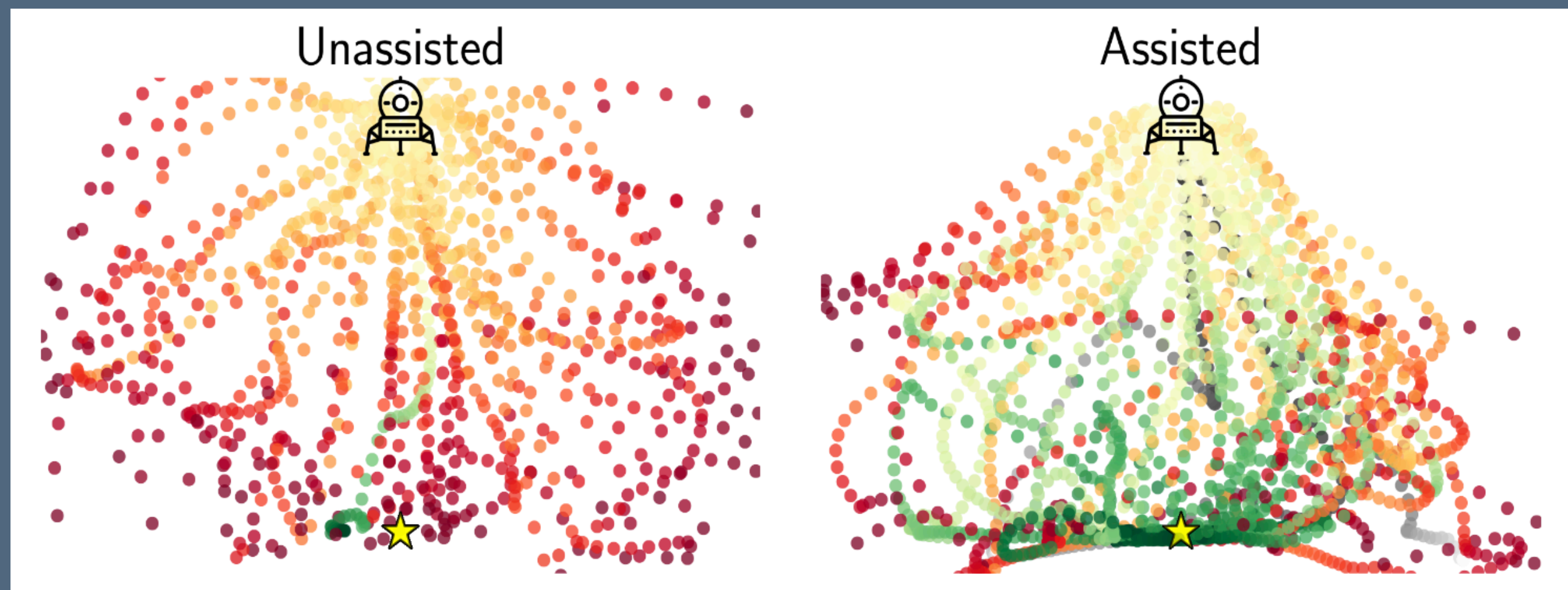
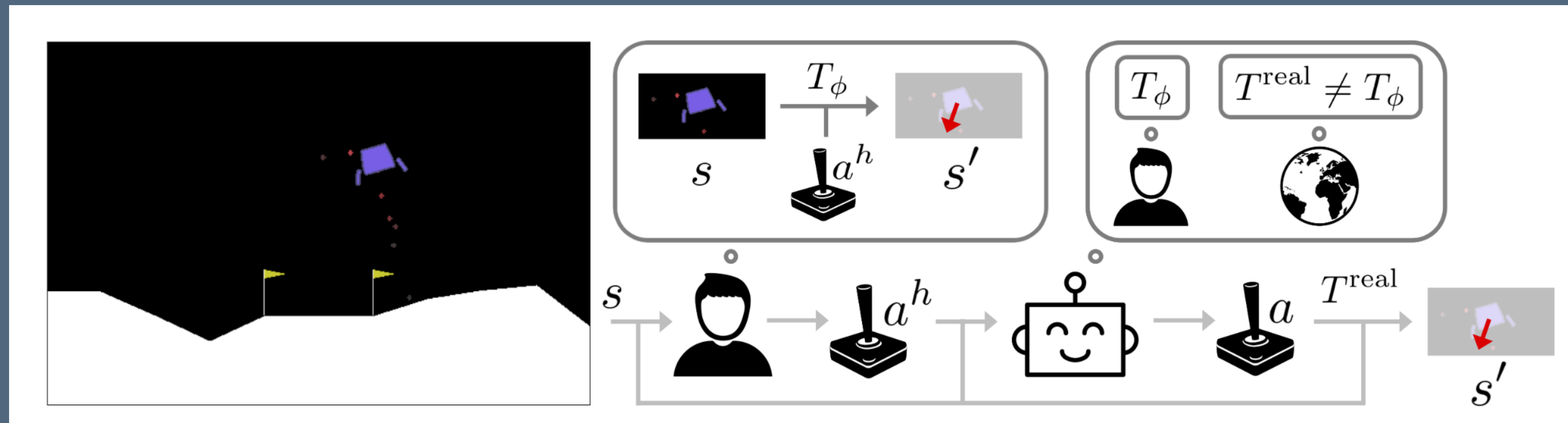
World model

User thinks gravity is stronger

→ **How can we assist?**

[Reddy et al. 2018]

Where do you think you're going?

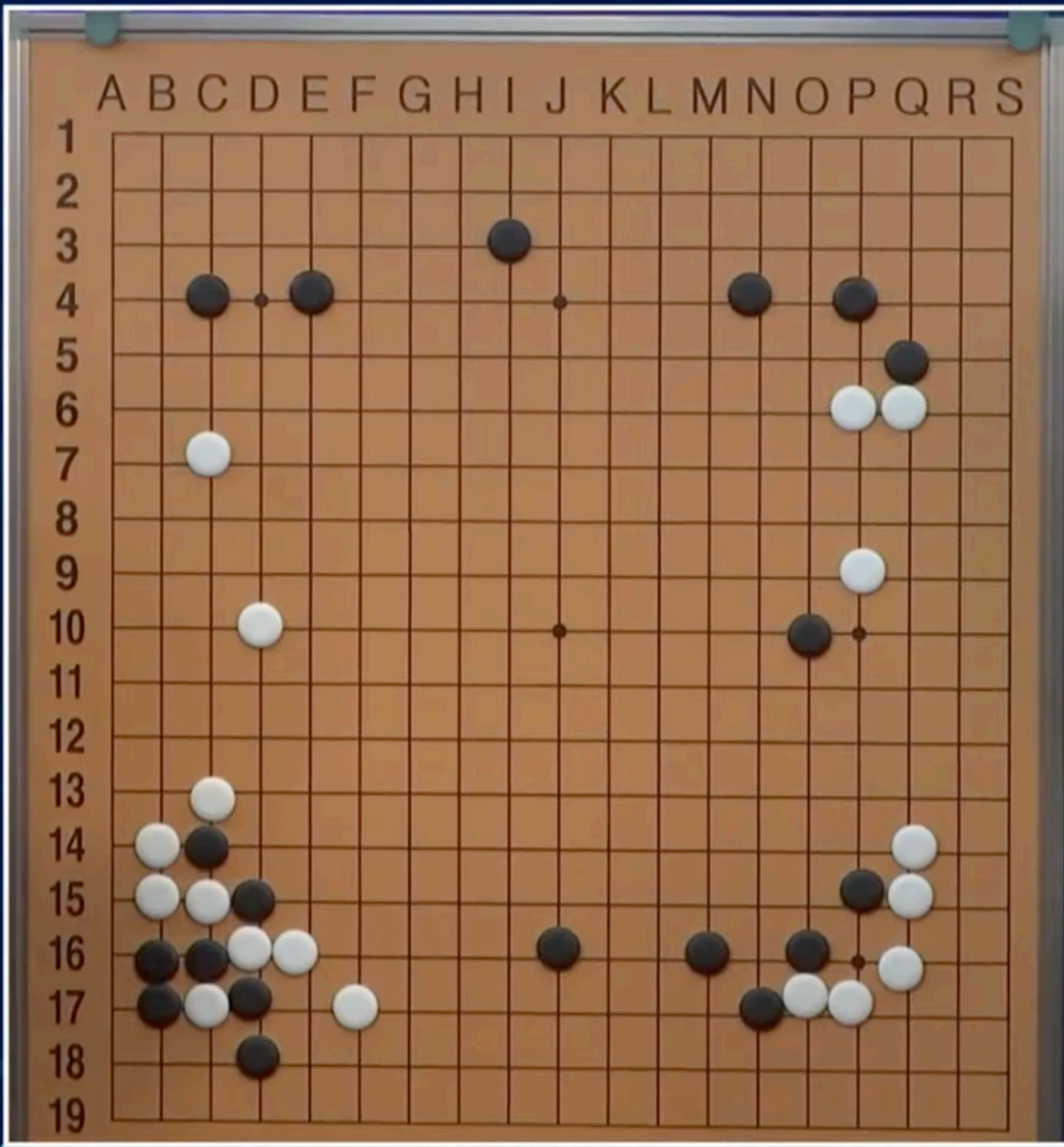


[Reddy et al. 2018]

2. “I understand you”

Understanding machines

Move 37



● LEE SEDOL
01:34:22

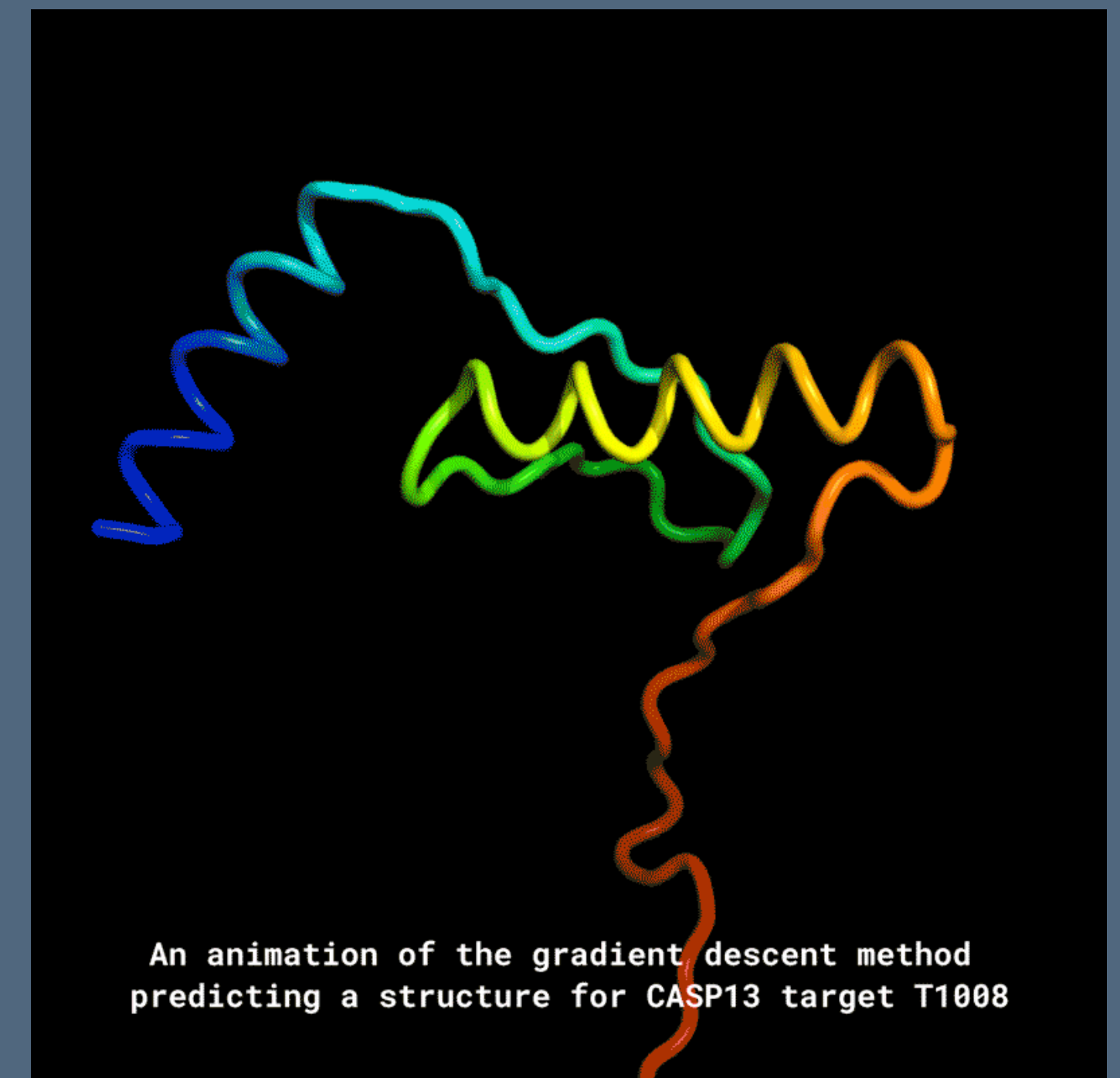
● ALPHAGO
01:38:39



AI can generate solutions not understandable

Lack of mental model and potentially
control

Oftentimes **does not teach us** much



Users should understand TP

We would like our thought partners to **act in a way that is legible** to us and communicate with us in the way **we “intuitively” understand.**

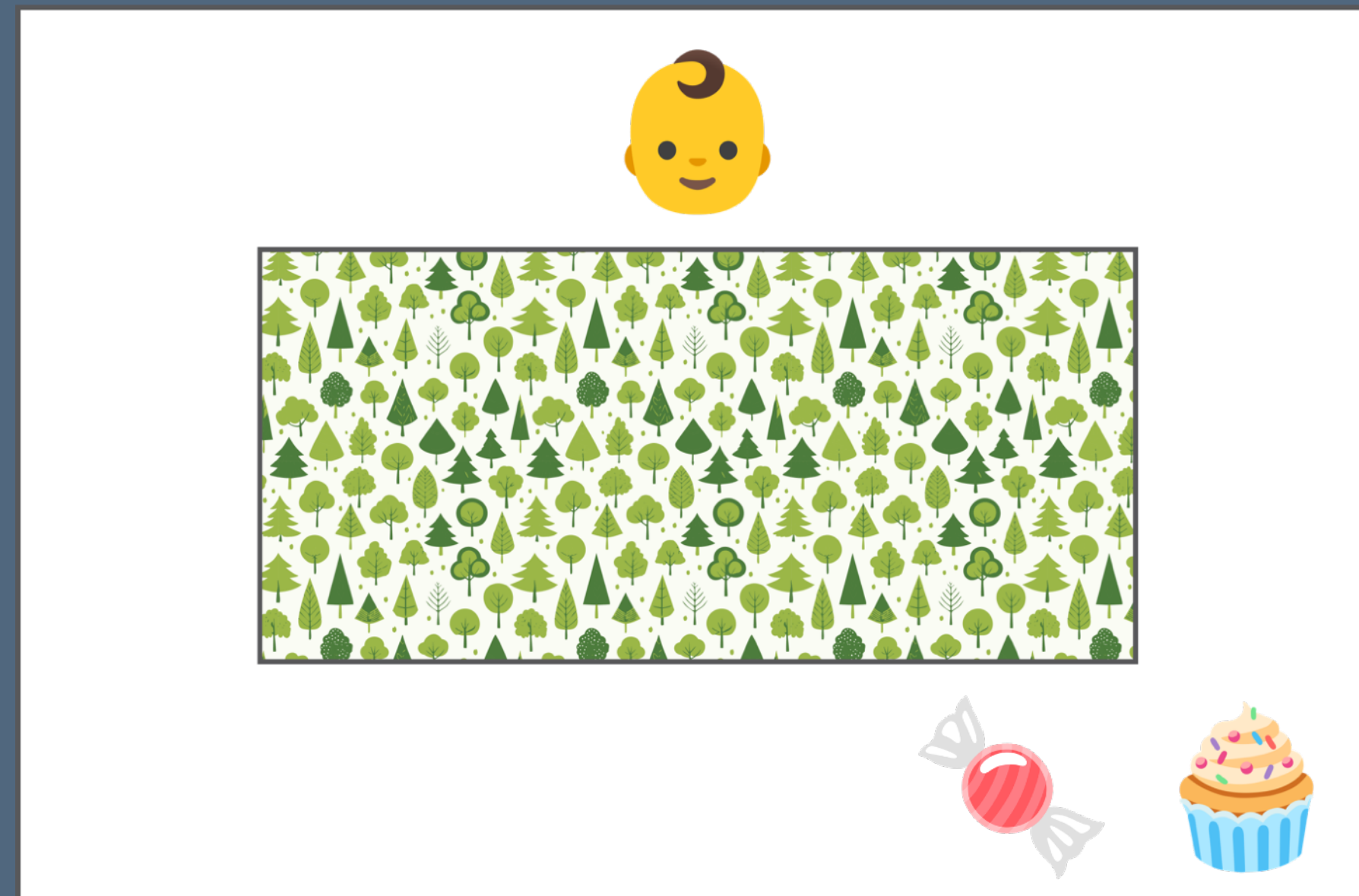
Planning

Latent human variables

The baby wants the candy



Actions to take



Planning

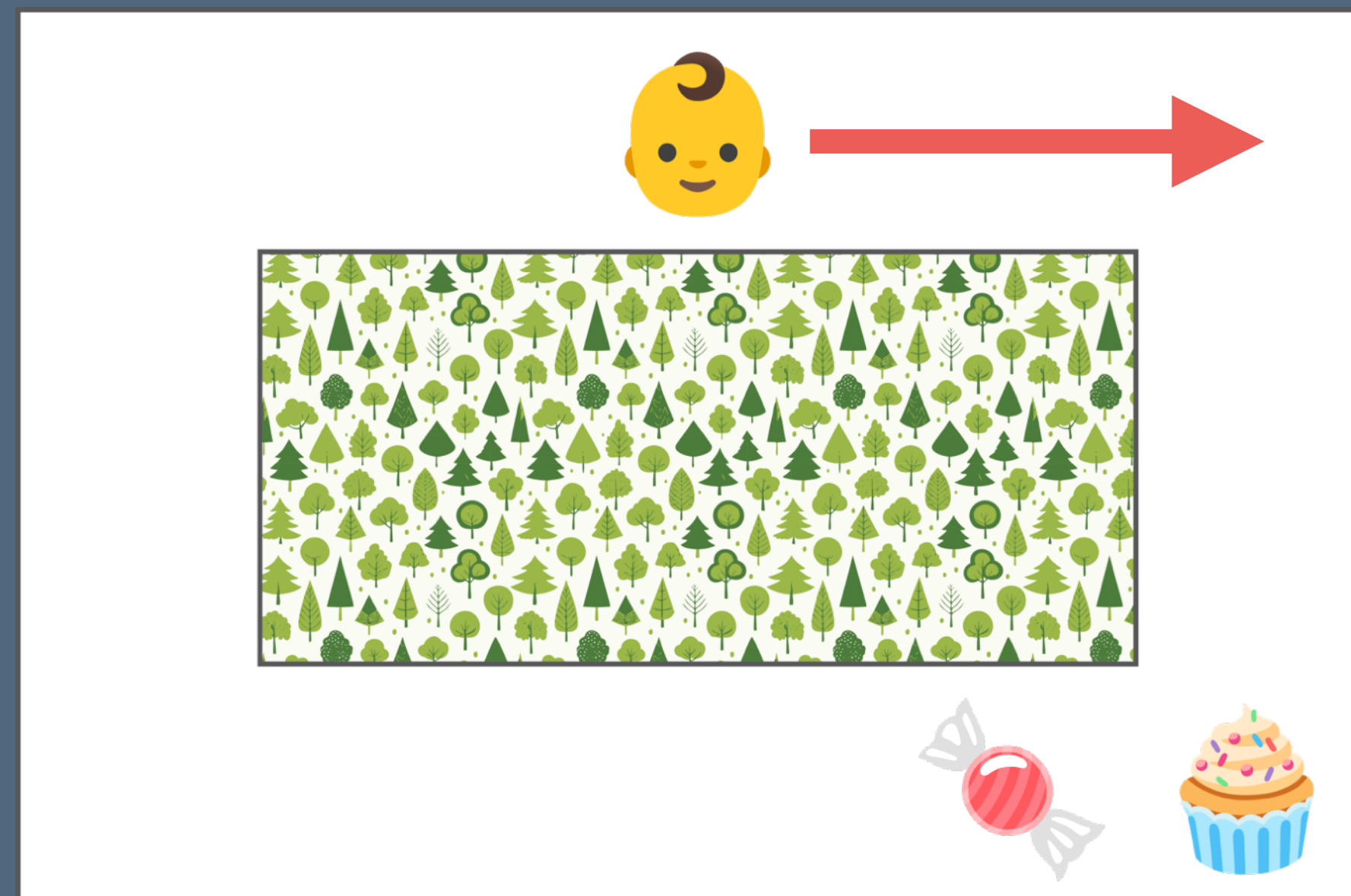
Latent human variables

The baby wants the candy



Actions to take

The baby moves to the right



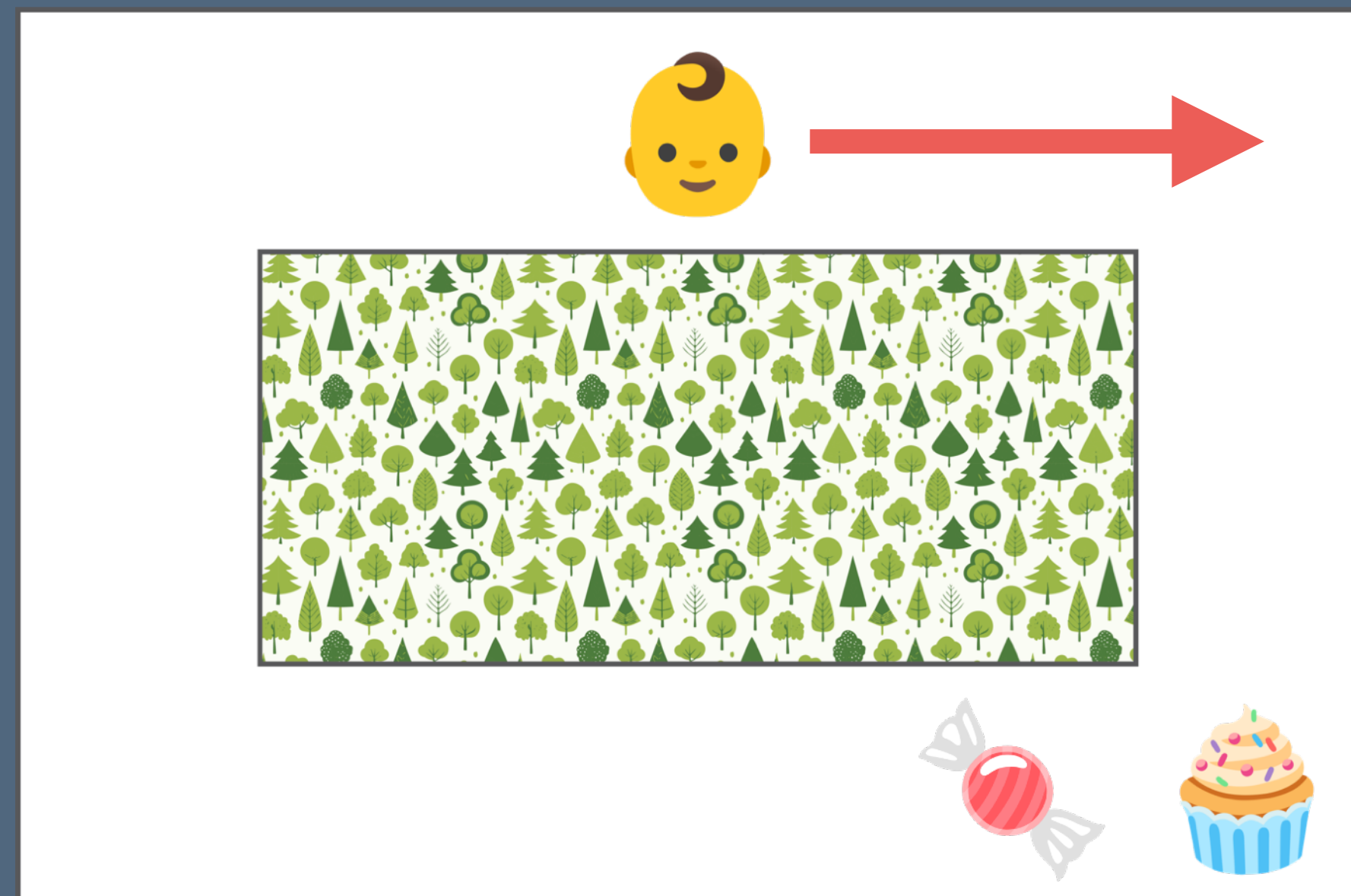
Inverse Planning (audience)

Latent human variables



Observed behavior

The baby moves to the right



Inverse Planning (audience)

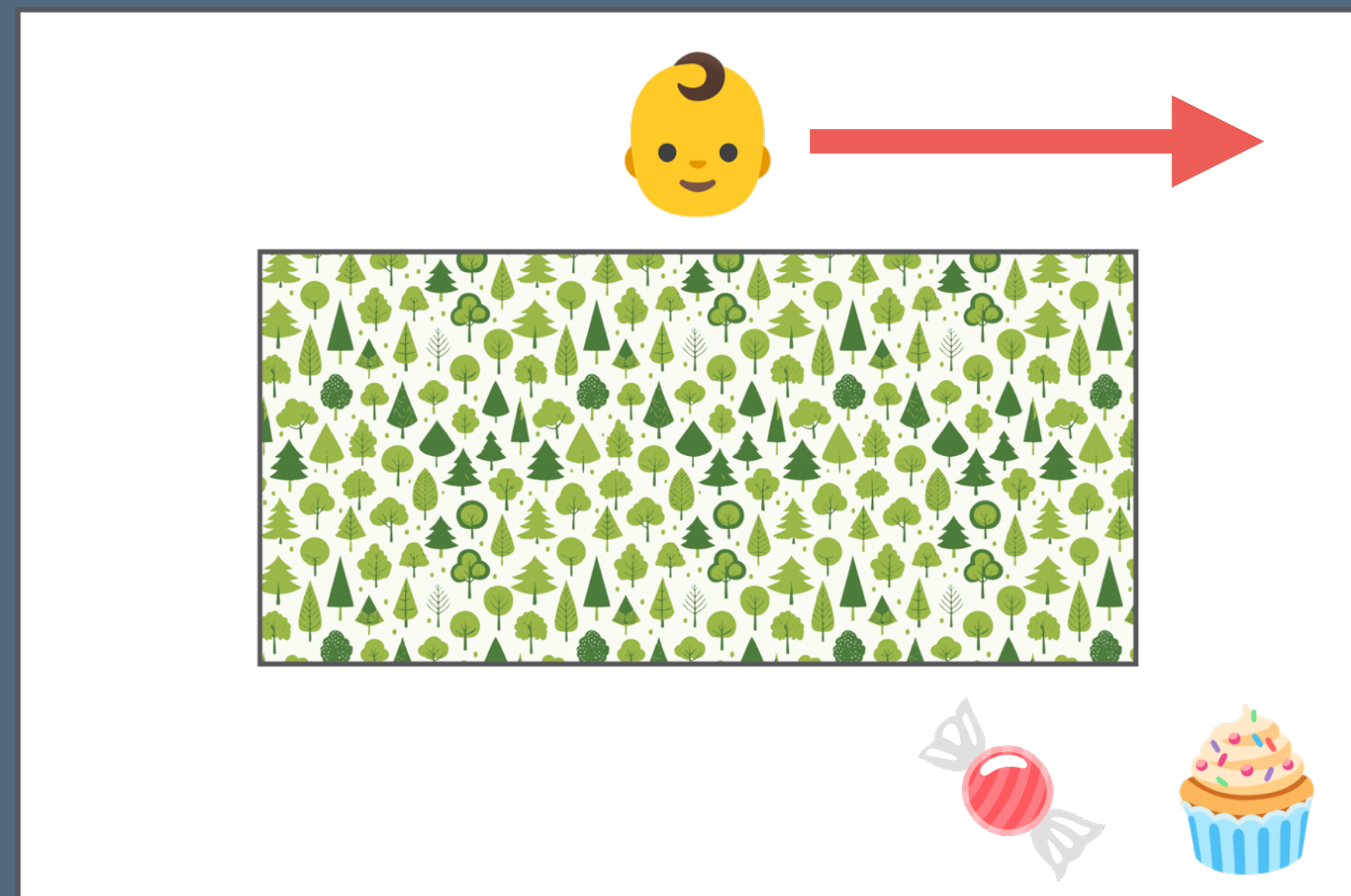
Latent human variables

????



Observed behavior

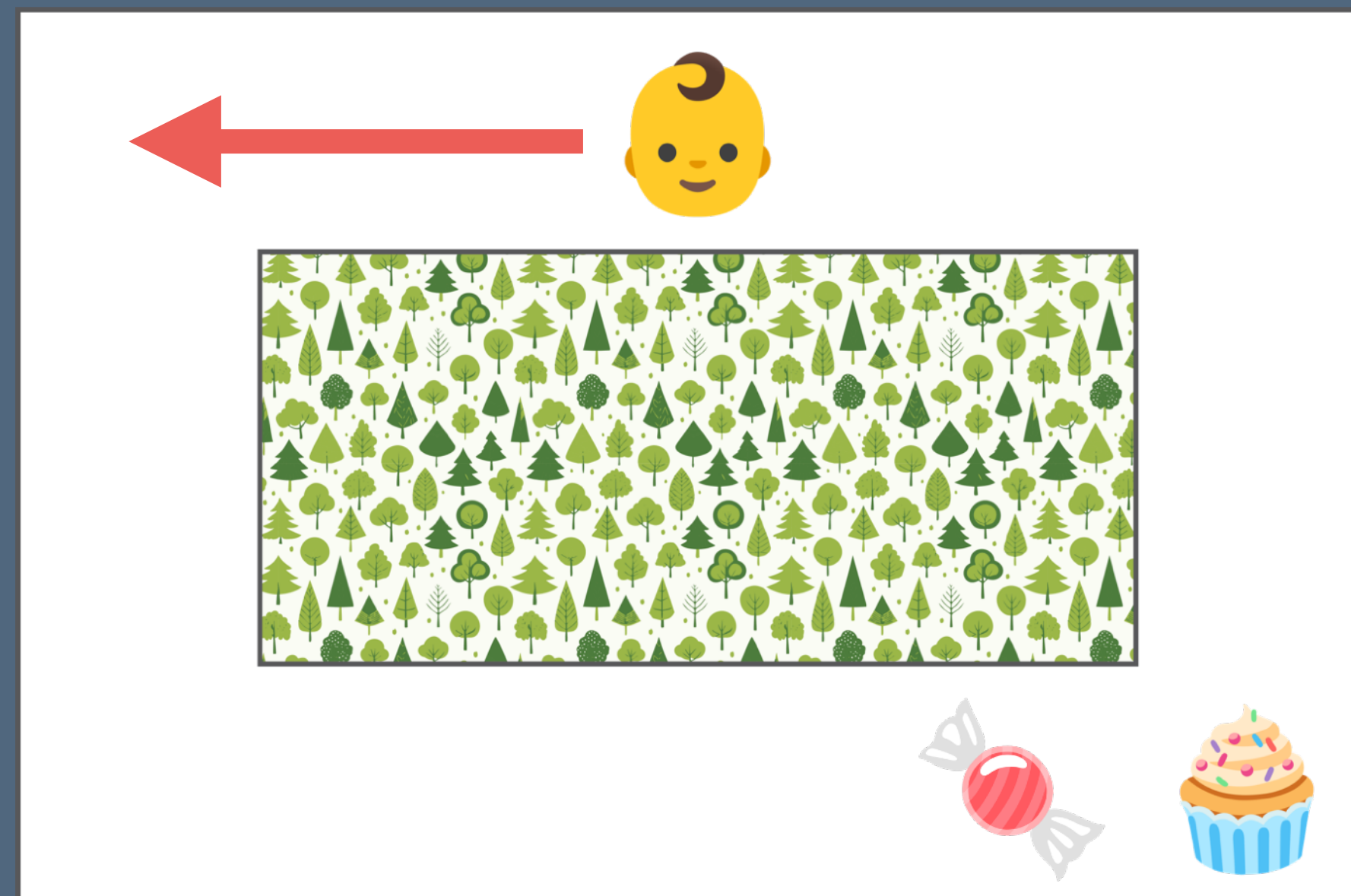
The baby moves to the right



Inverse Inverse Planning

Latent human variables ← Observed behavior

How should I act so that an observer believes I want the candy?



Inverse Inverse Planning

Planning to best expresses a goal to a viewer

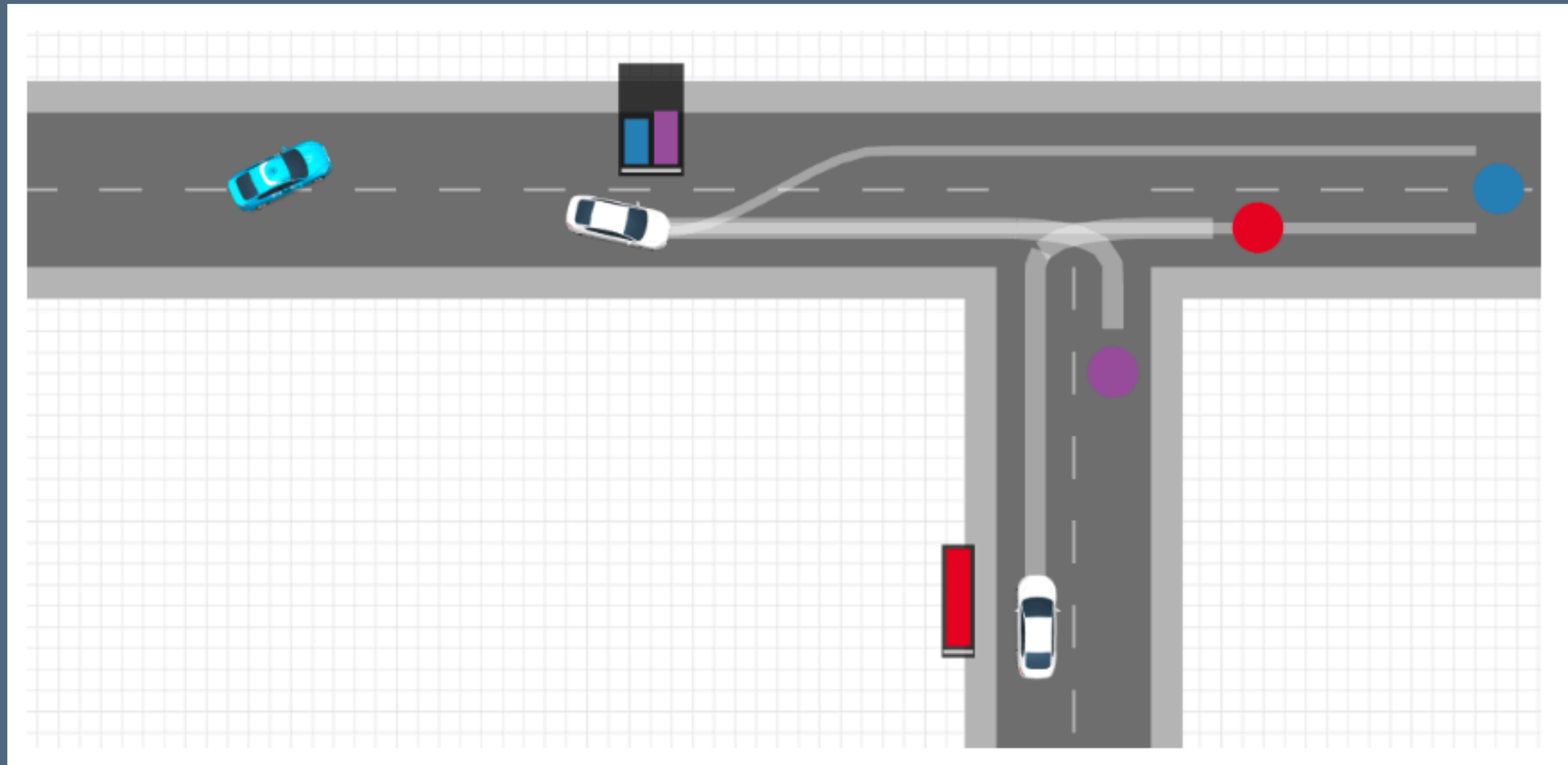


Robot legibility

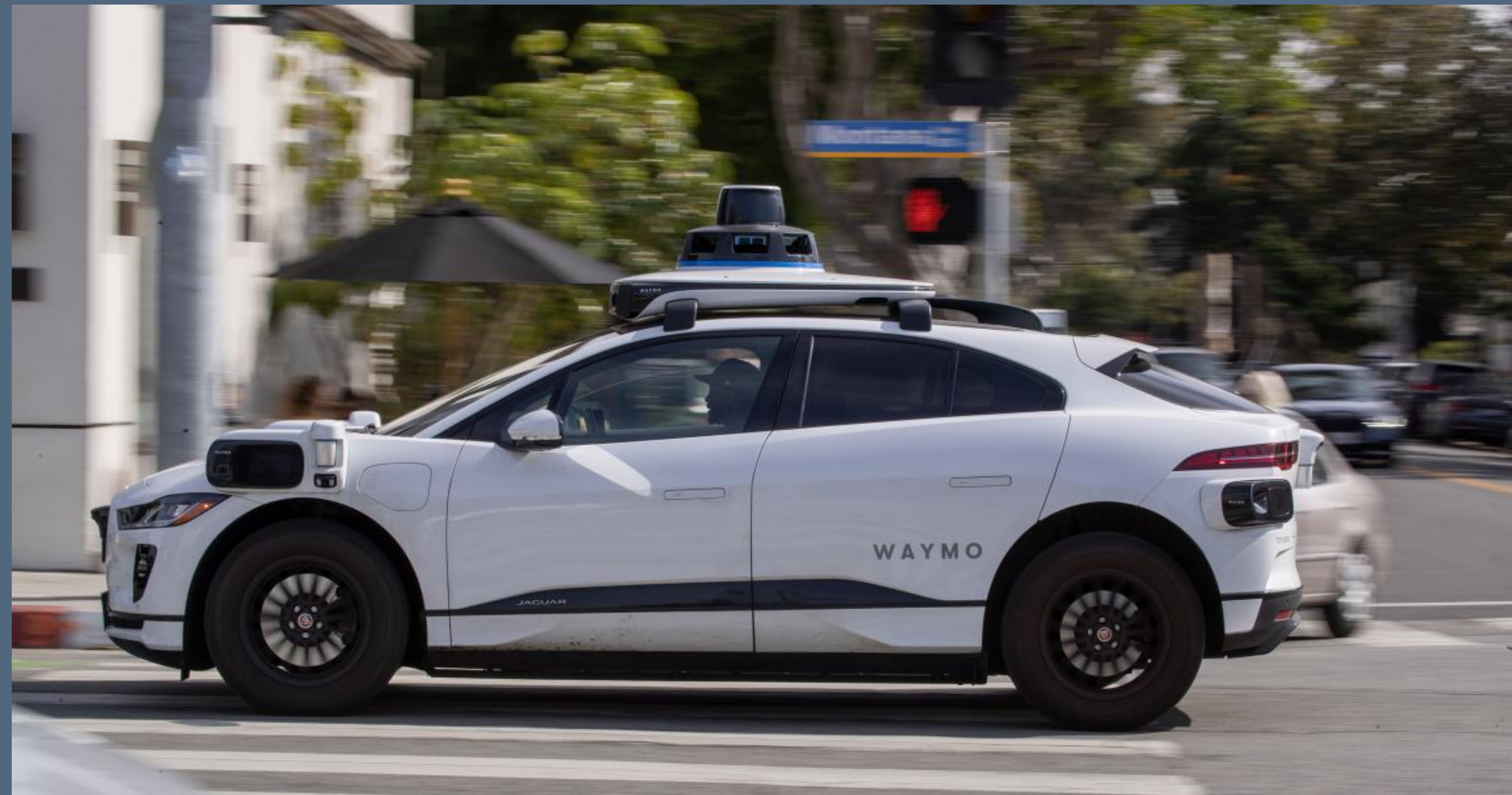
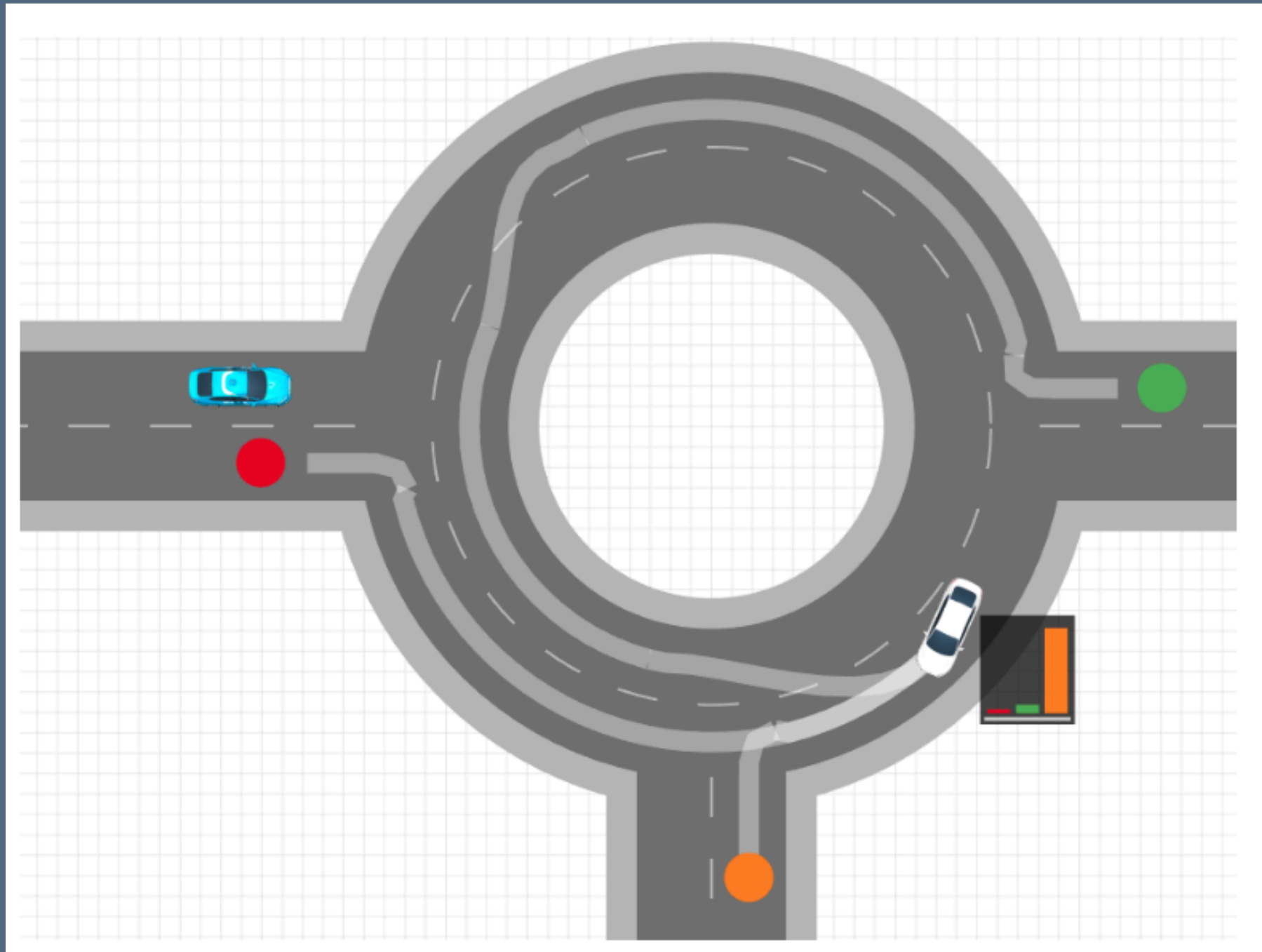


[Dragan, Lee, and Srinivasa 2013]

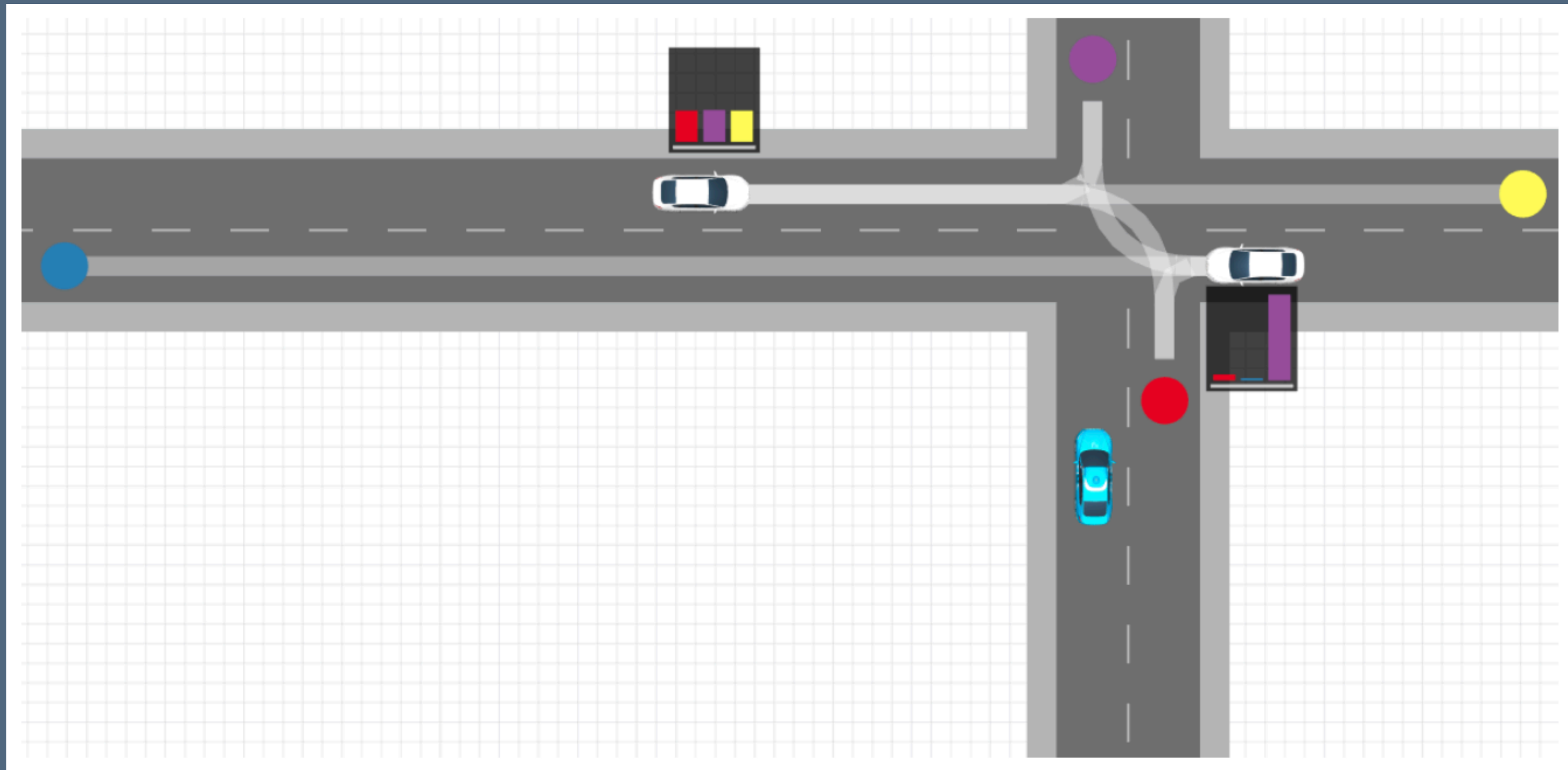
Self-driving cars



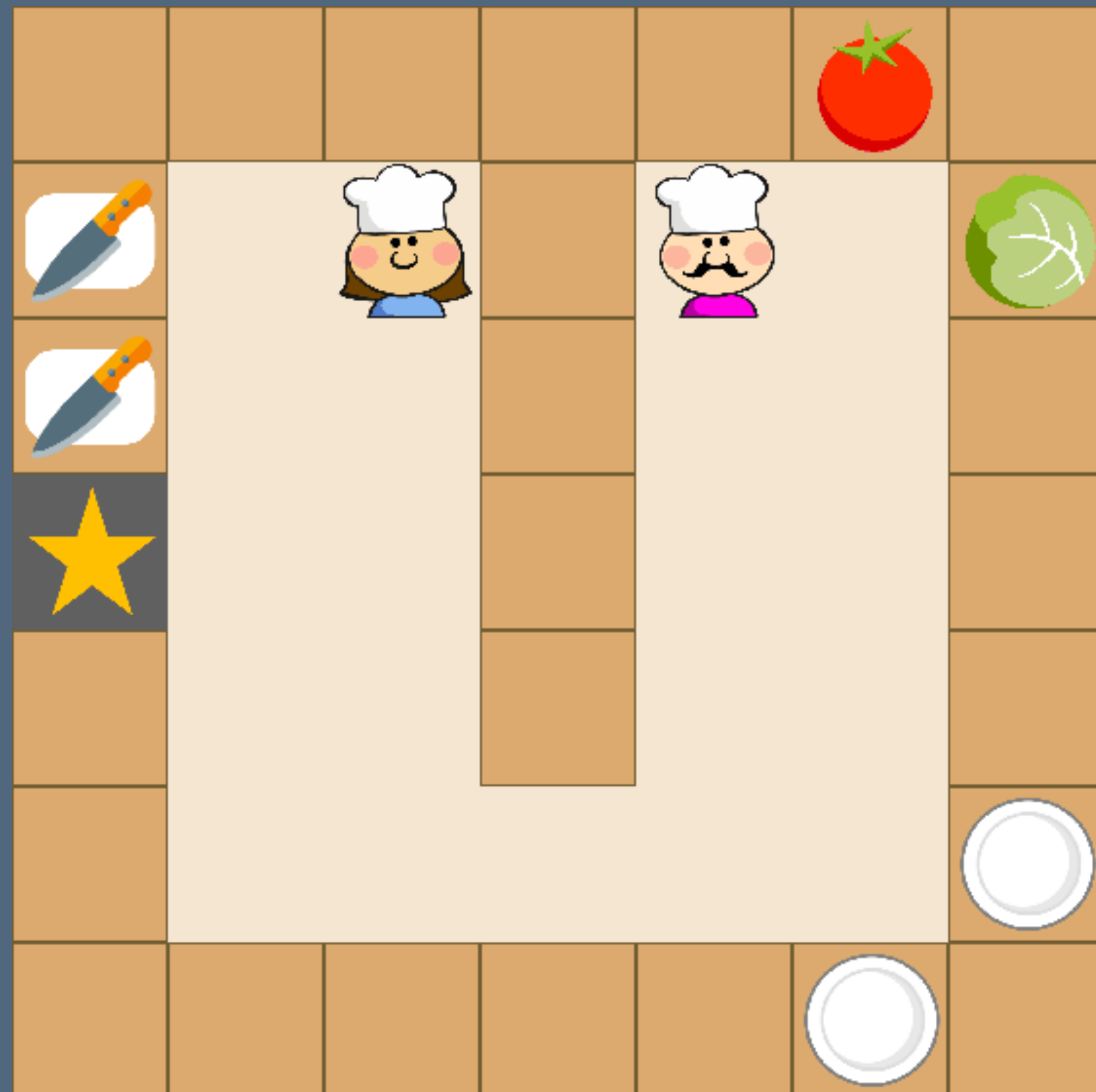
Self-driving cars



Self-driving cars



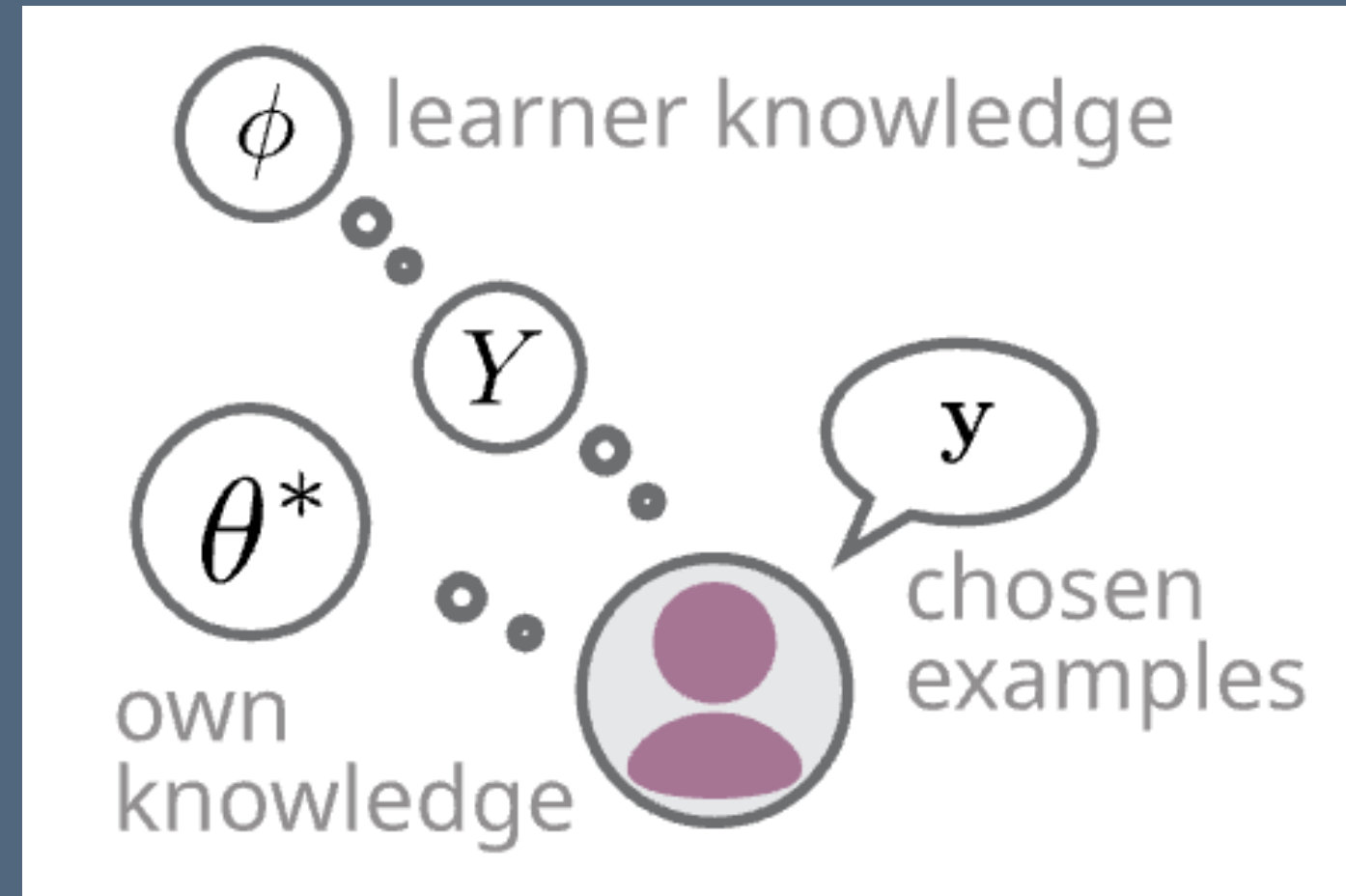
Video game sidekicks



Adaptive teaching

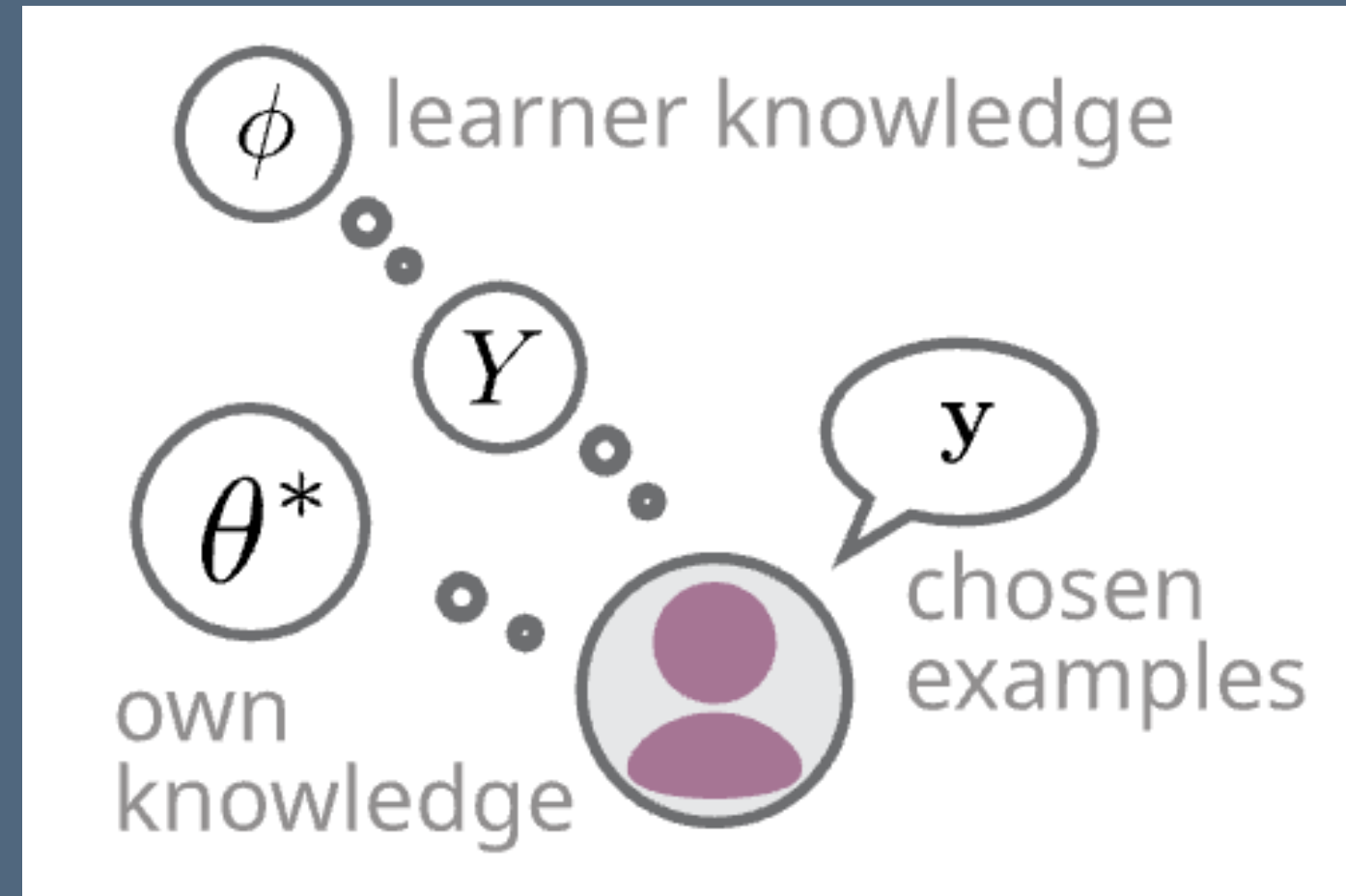
We saw that teachers can “**debug**” students by providing adequate explanations (Teach a fish to program)

Adaptive teaching



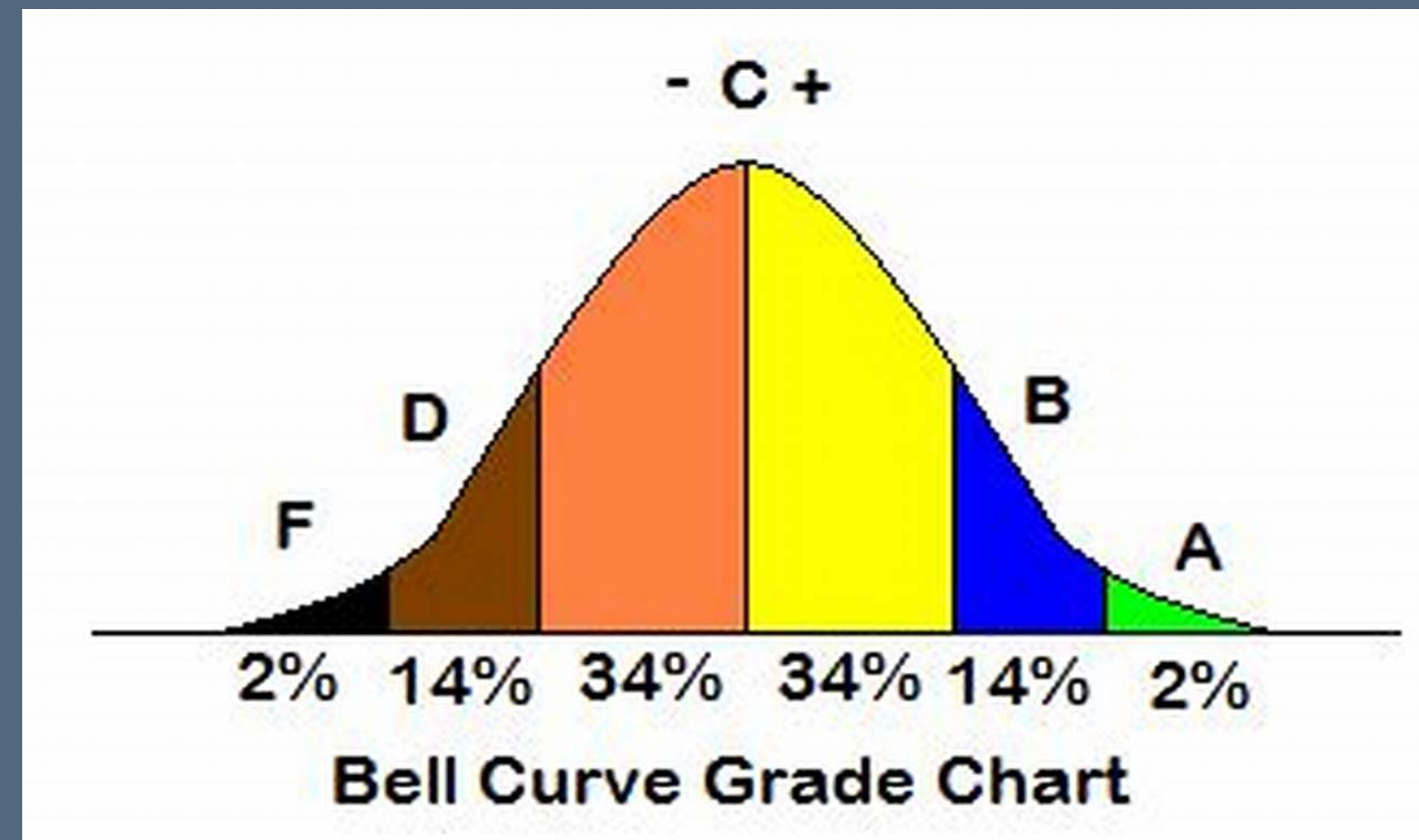
Teachers also try to **assess what students know** (with questions or exercises to test knowledge).

Adaptive teaching



Teachers also try to **assess what students know** (with questions or exercises to test knowledge).

They try making **appropriate exams**, not too difficult nor easy, based on what they think students know.



**Stories are often centered
around one question**

Stories are often centered around one question

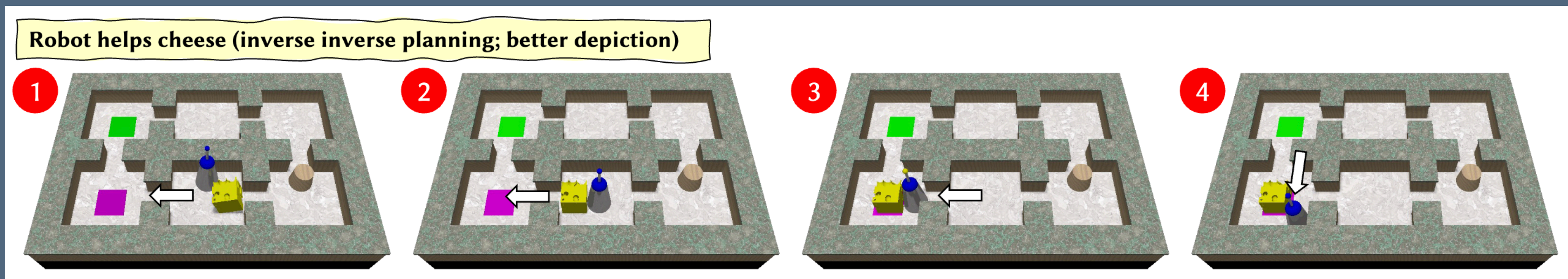
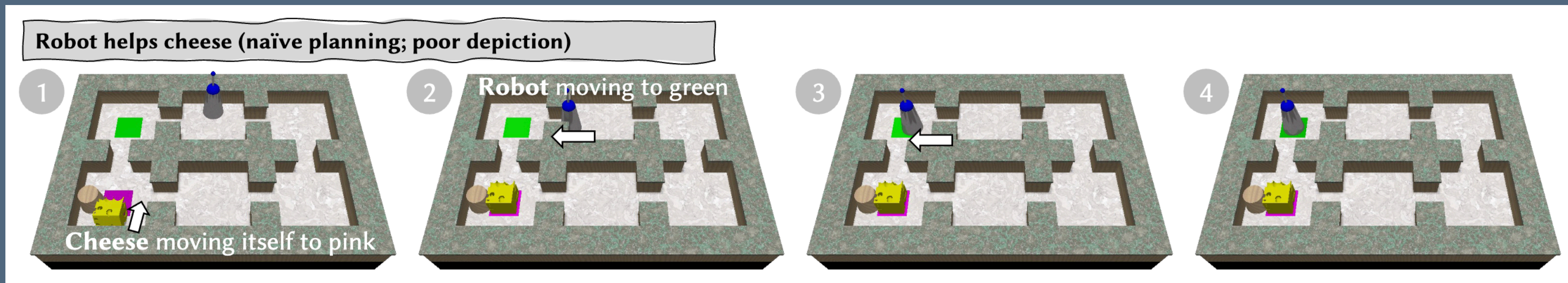


Stories are often centered around one question



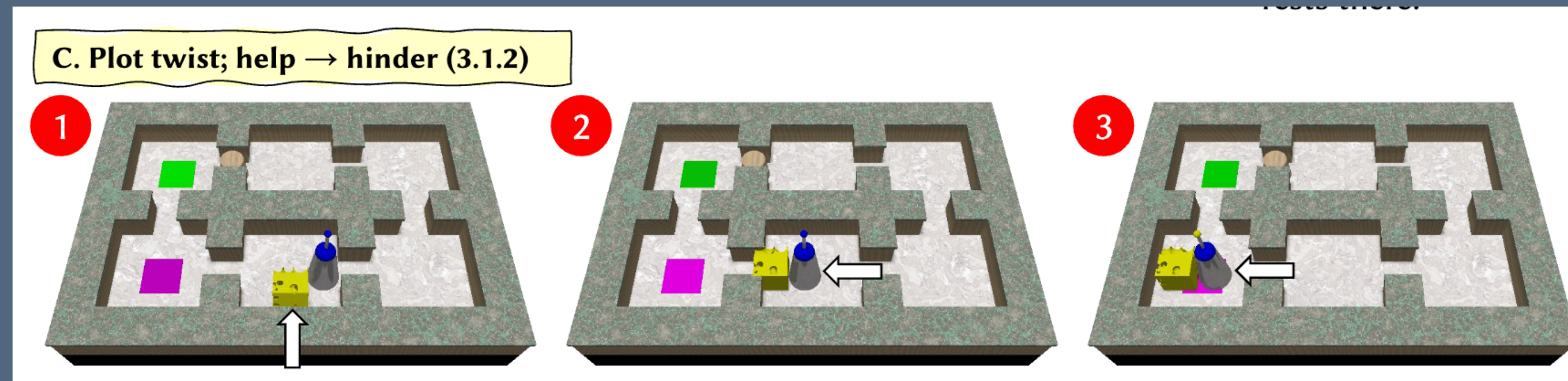
Storytelling as probability design

Storytelling goal: The robot wants to help the cheese



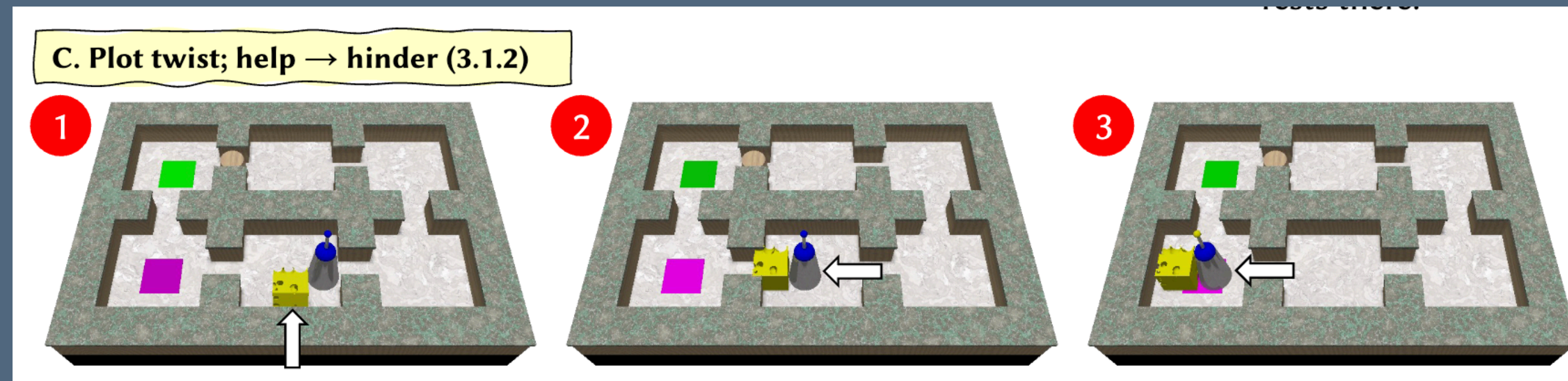
Storytelling as probability design

Fundamentally, a **plot twist** is an inversion of belief:

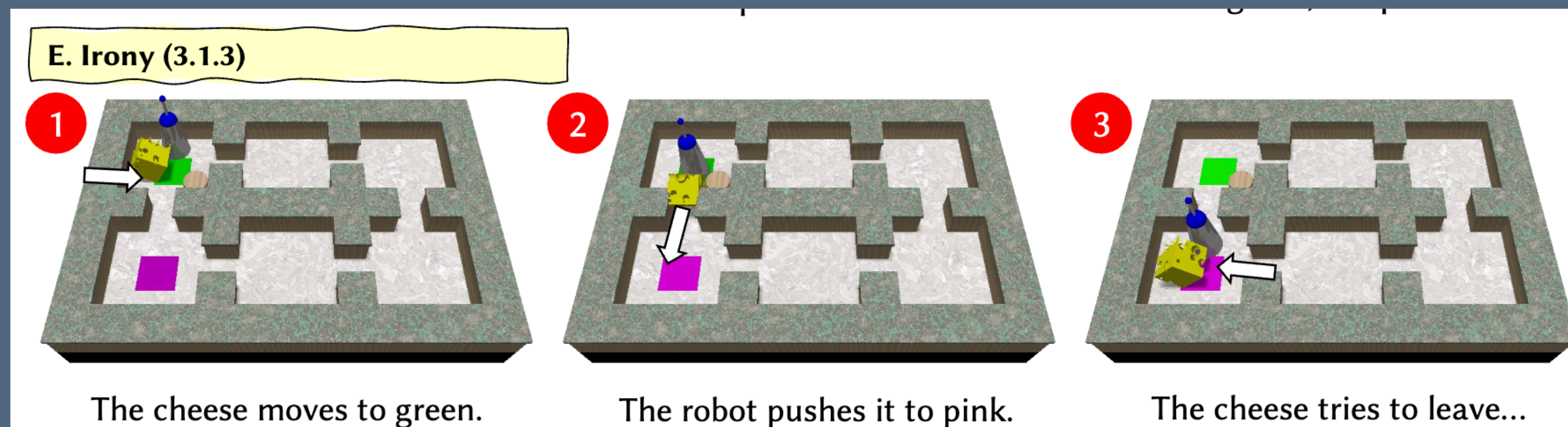


Storytelling as probability design

Fundamentally, a **plot twist** is an inversion of belief:



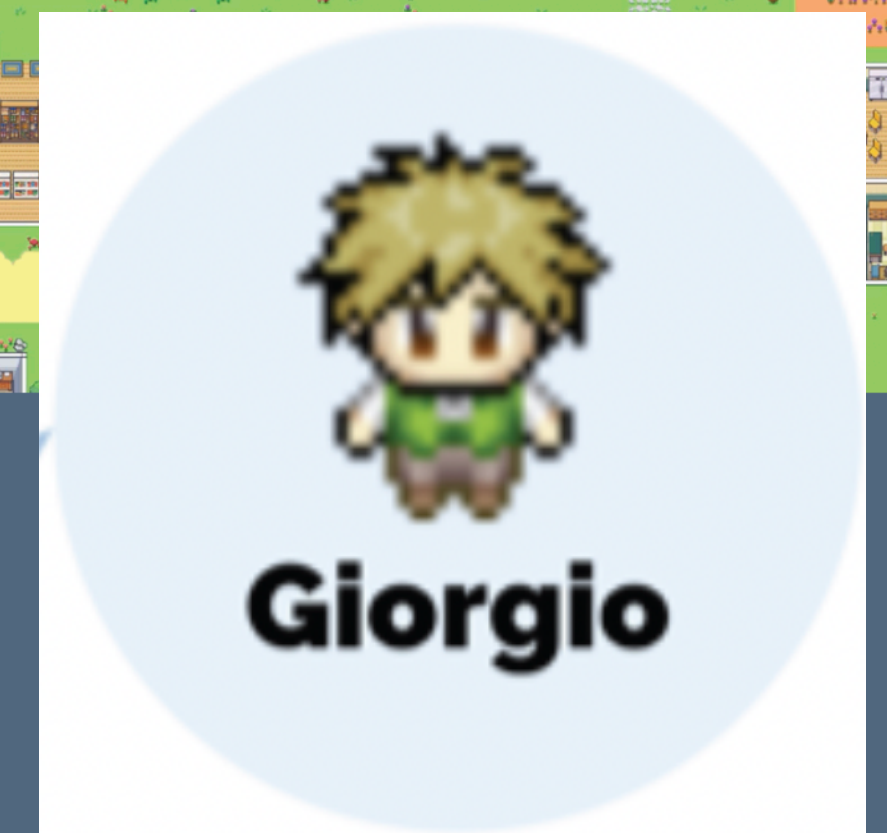
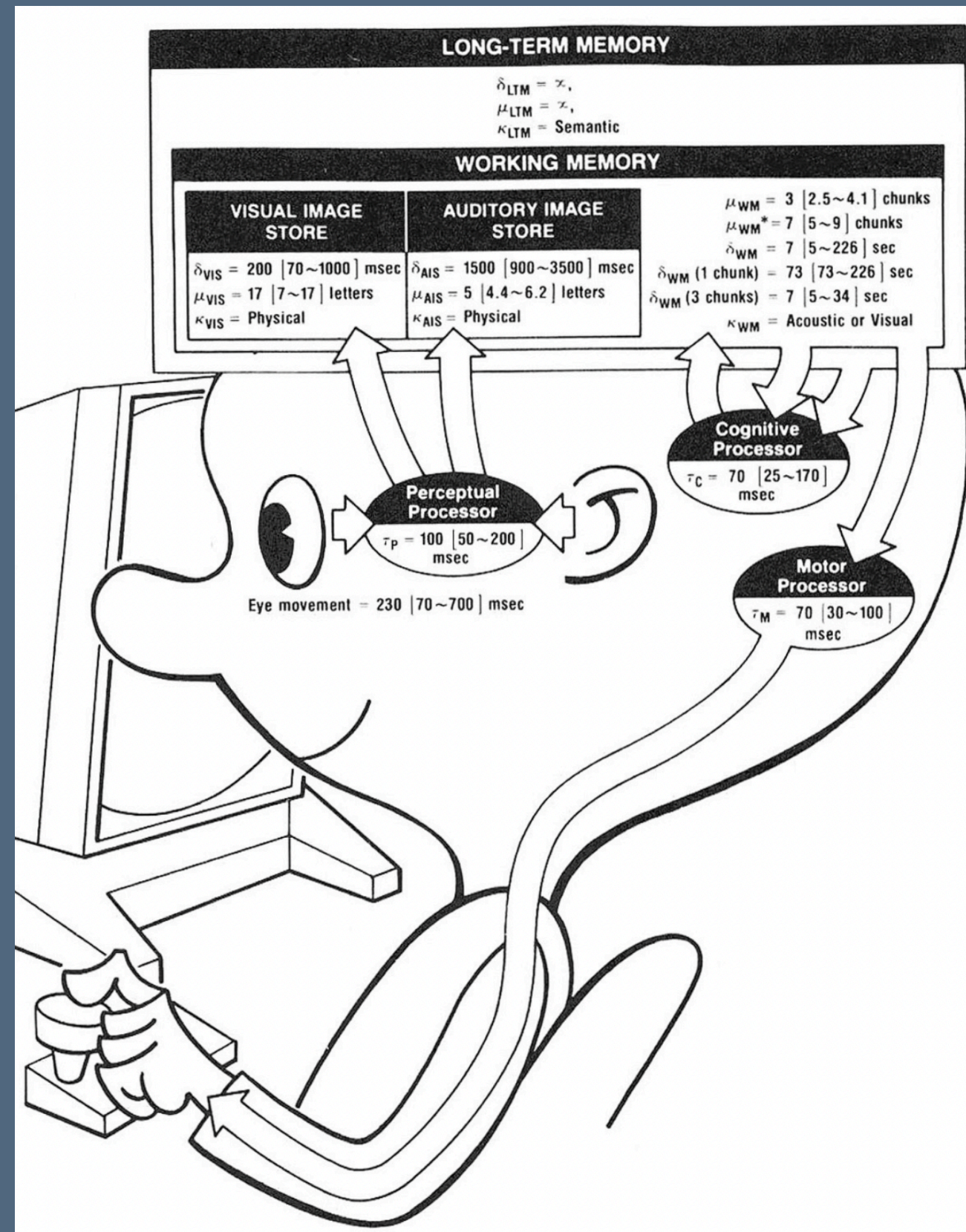
Dramatic irony: The audience and the characters have different beliefs

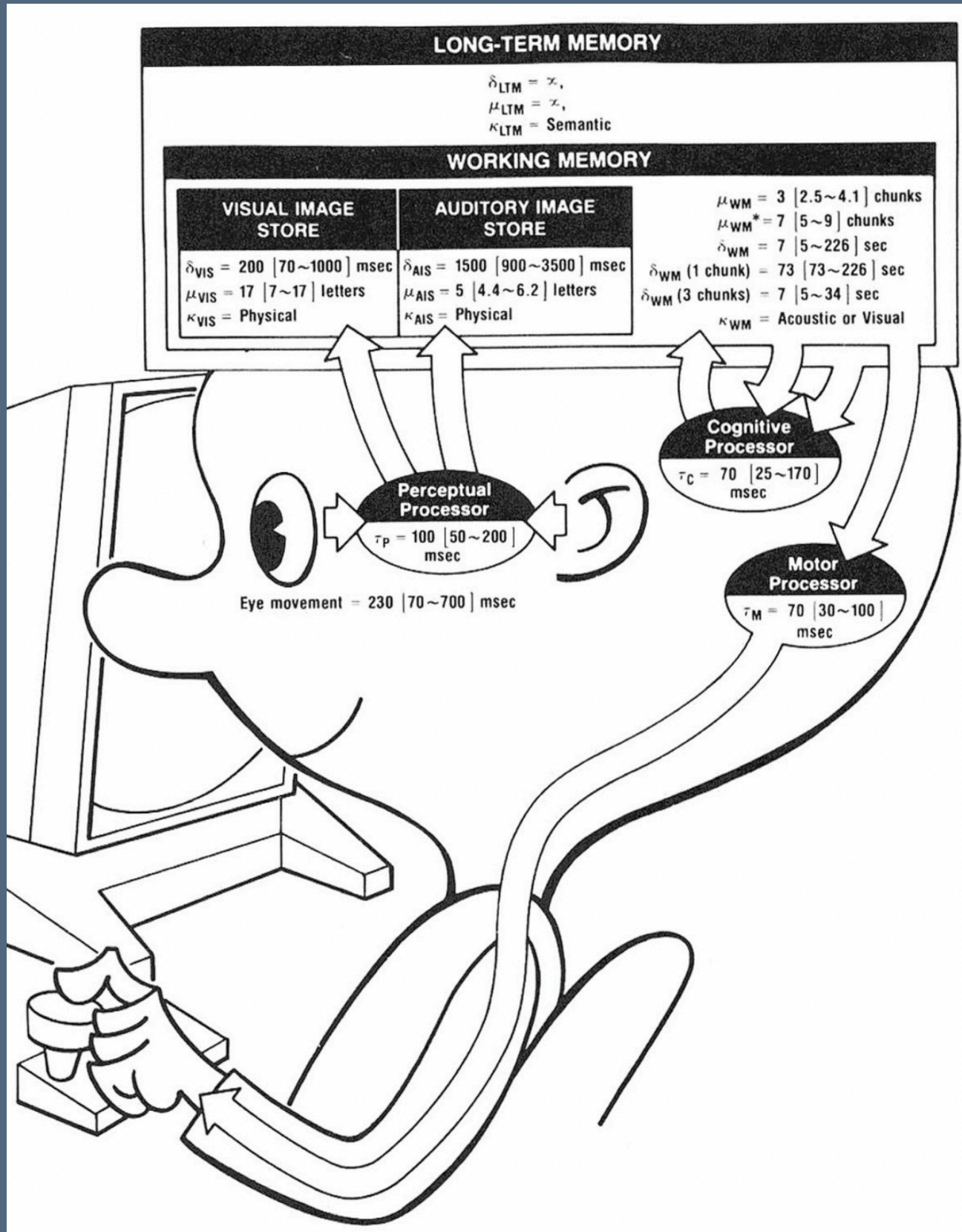


However, previous limitations seen in class still apply

They require **substantial effort** to create, vs. directly prototyping

Not as generalizable as LLMs: But could be paired with LLMs to generate these problem representations [Wong et al. 2024]





Summary

Today's tools are more and more becoming “**thought partners.**” To better serve users, thought partners would benefit from:

1. **Understanding** users (their goals, beliefs, knowledge, capacities...)

2. Be **understandable** by users

Ongoing work to unify low- and high-level cognitive models.