JOHN AND THE HAMMER BROKE THE VASE:

GENERATIVE AI AND THE INNATENESS OF COUNTERINTUITIVENESS

Tamás Biró

ELTE Eötvös Loránd University and OR-ZSE Jewish Th. Seminary – U. of Jewish Studies

https://birot.web.elte.hu/ https://mrmr.elte.hu/

[MRMR]

Modelling Religious Mental Representations

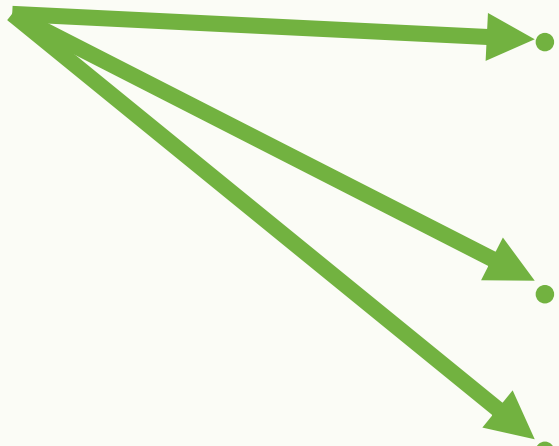
XXIII IAHR World Congress Kraków 2025

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

Mental ontologies and folk-theories

- HUMAN
- ANIMAL
- PLANT
- ARTIFACT
- NATURAL OBJECT
- …



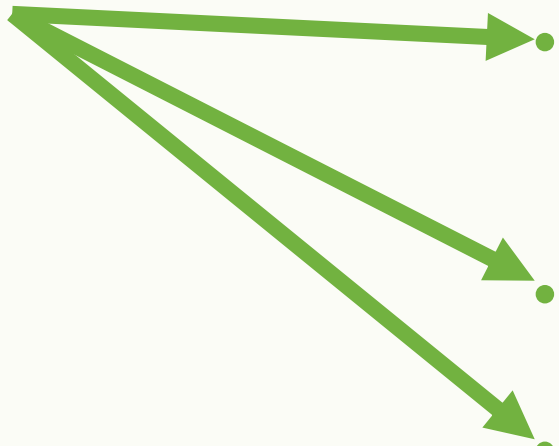
- *Intuitive psychology*
  - Theory of Mind
- *Intuitive biology*
- *Intuitive physics*
- …

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

Mental ontologies and ontological agents

- HUMAN
- ANIMAL
- PLANT
- ARTIFACT
- NATURAL OBJECT
- …



- *Intuitive psychology*
  - Theory of Mind
- *Intuitive biology*
- *Intuitive physics*
- …

Ontological agents

Ontological agents: with will and volition, capable of goal-directed actions.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

Mental ontologies and ontological agents

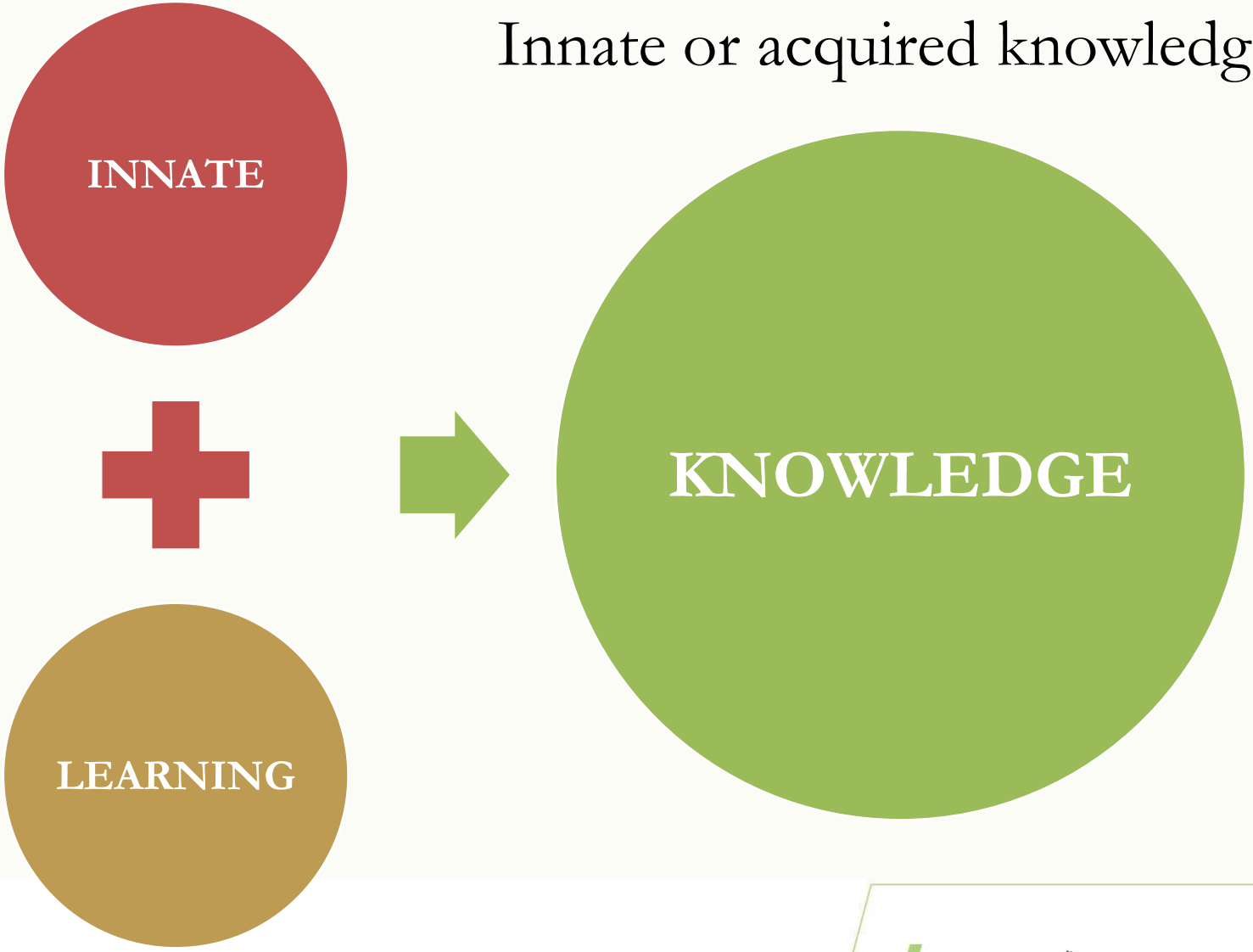
Ontological category	agentive	Non-agentive
HUMAN	***yes***	—
ANIMAL	*yes? vertebrates only?*	*no? lower animals?*
PLANTS	*[counterintutive entities]*	***yes***
ARTIFACT	*[robots, AI]*	***yes***
NATURAL OBJECT	*[counterintutive entities]*	***yes***

Ontological agents: with will and volition, capable of goal-directed actions.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

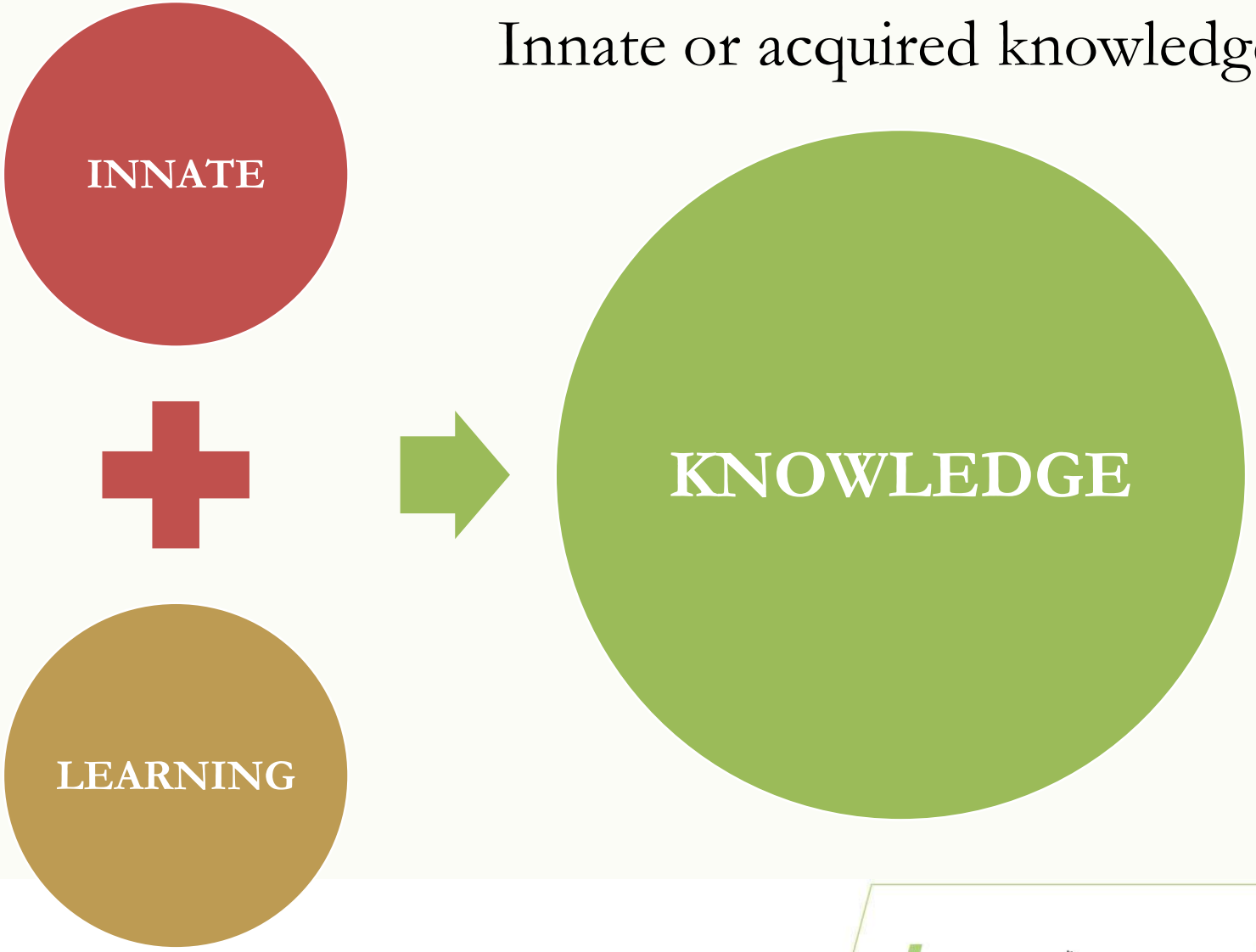
Innate or acquired knowledge?



NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

Innate or acquired knowledge?

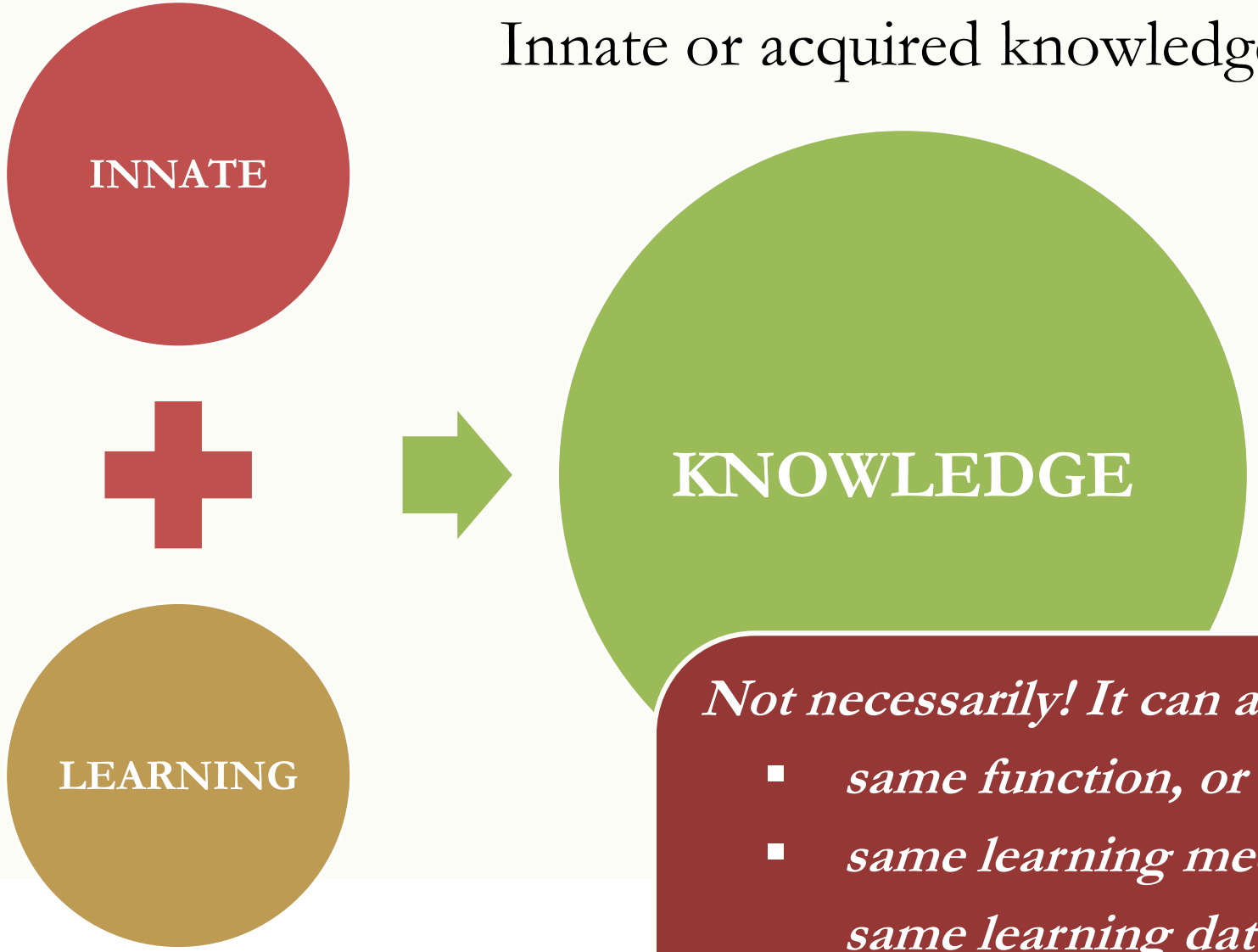


If it is cross-culturally universal, then it must be innate, right?

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

Innate or acquired knowledge?



If it is cross-culturally universal, then it must be innate, right?

Not necessarily! It can also be universal, if

- ***same function, or***
- ***same learning mechanism and same learning data***

result in similar or identical structures.

A syntactic riddle

Observe the following sentences:

(1) *John broke the vase.*

(2) *The hammer broke the vase.*

(3) *# John and the hammer broke the vase.*

(4) *John broke the vase with the hammer.*

(5) *# The hammer broke the vase with John.*

sentences are grammatical, but infelicitous.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

A syntactic riddle: *John* and *hammer* fill in different roles

agent **instrument**

Observe the following sentences:

(1) *John broke the vase.*

(2) *The hammer broke the vase.*

(3) *# John and the hammer broke the vase.*

(4) *John broke the vase with the hammer.*

(5) *# The hammer broke the vase with John.*

Syntactic positions: subject, object, prepositional phrase…

Thematic roles: thematic agent (actor), patient, instrument…

sentences are grammatical, but infelicitous.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

A syntactic riddle: *John* and *hammer* fill in different roles

agent **instrument**

Observe the following sentences:

(1) *John broke the vase.*

(2) *The hammer broke the vase.*

(3) *# John and the hammer broke the vase.*

(4) *John broke the vase with the hammer.*

(5) *# The hammer broke the vase with John.*

Unless we imagine a counterintuitive hammer with agentive properties, as in an animated cartoon.

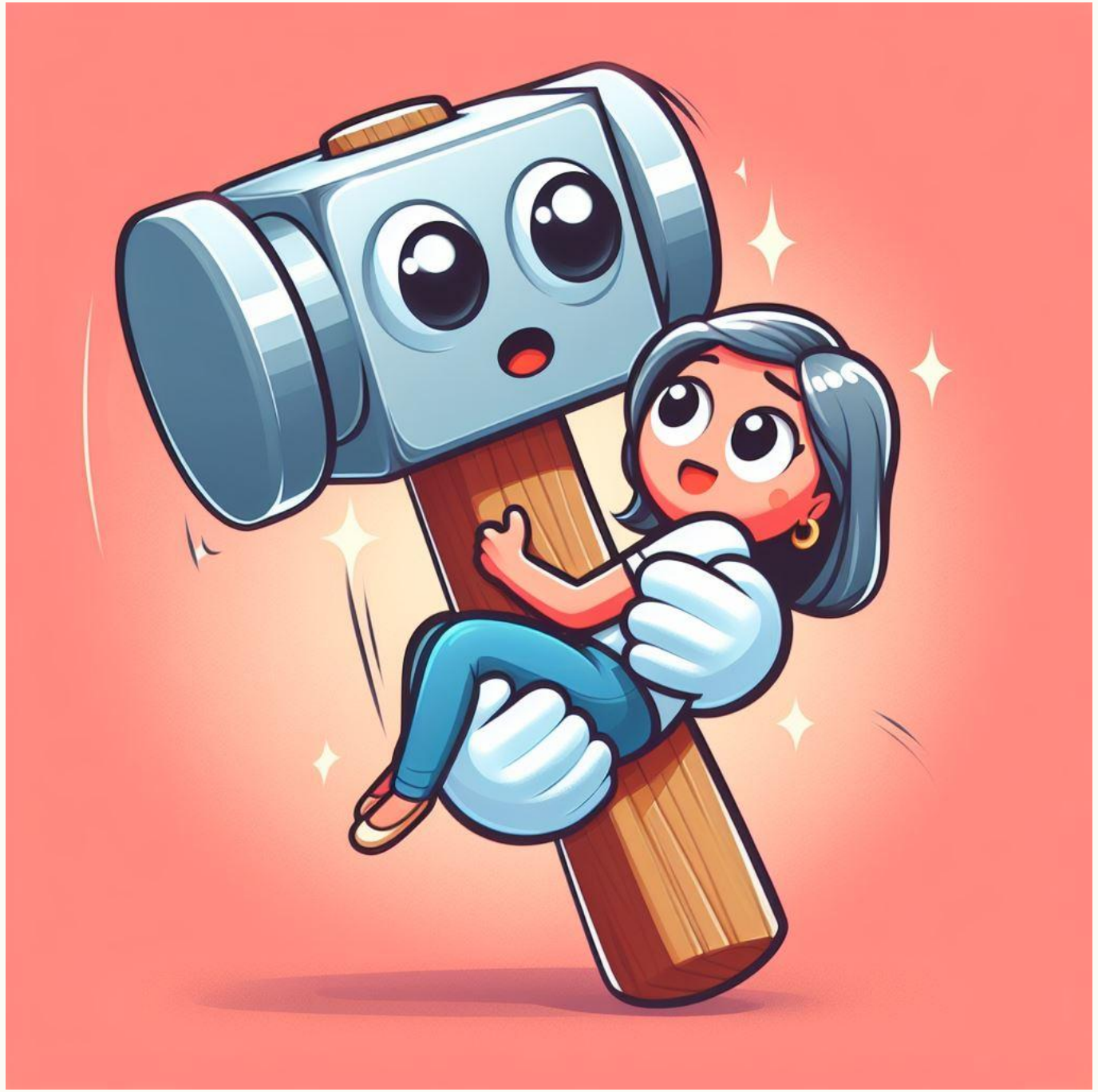
phrase…

Thematic roles:
thematic agent (actor), patient, instrument…

sentences are grammatical, but infelicitous.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND



Unless we imagine a counterintuitive hammer with agentive properties, as in an animated cartoon.

Some students might have **aphantasia**.

To support them, let's draw a counterintuitive hammer breaking a vase with John for them … using AI.

(1) Prompt: *John breaking the vase.*

https://copilot.cloud.microsoft/, November 14, 2024

(2) Prompt: *The hammer breaking the vase.*

https://copilot.cloud.microsoft/, November 14, 2024

(4) Prompt:

John breaks the vase with a hammer.

November 14, 2024

(5) # The hammer breaks the vase with John.

The hammer is still the instrument!

"The hammer breaks the vase with John"

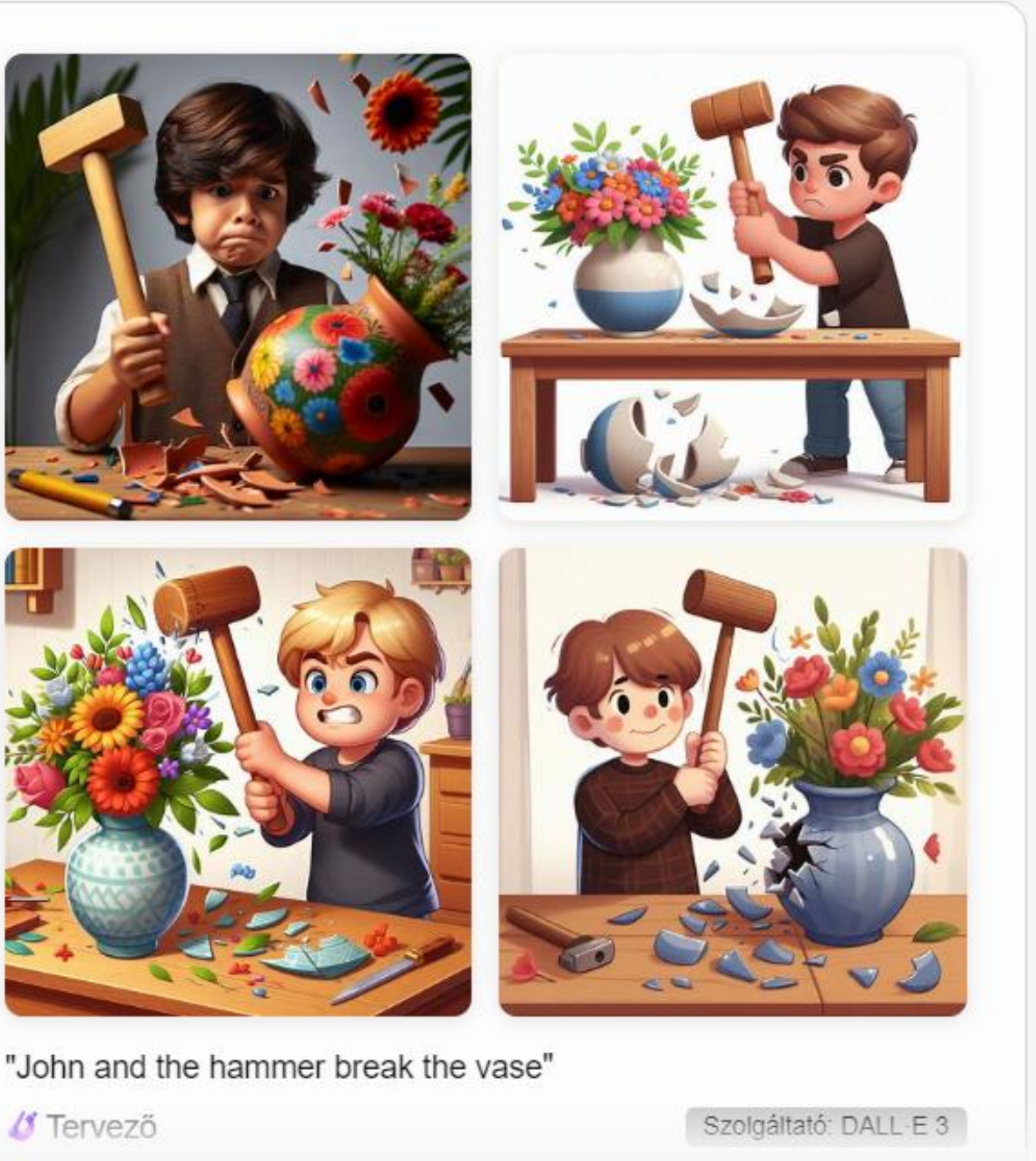
Tervező

Szolgáltató: DALL·E 3

November 14, 2024

(3) # John and the hammer break the vase.

The hammer is again the instrument!



November 14, 2024

The vase is broken by the hammer using John.

The hammer is still the instrument!

November 14, 2024

Copilot

Show me John breaking a vase.

November 14, 2024

Copilot

Show me the baseball bat hitting a vase.

November 14, 2024

Copilot

Show me John and a baseball bat hitting a vase.

The baseball bat is also the instrument!

November 14, 2024

Copilot

Show me a baseball bat and John hitting a vase

The baseball bat is also the instrument!

November 14, 2024

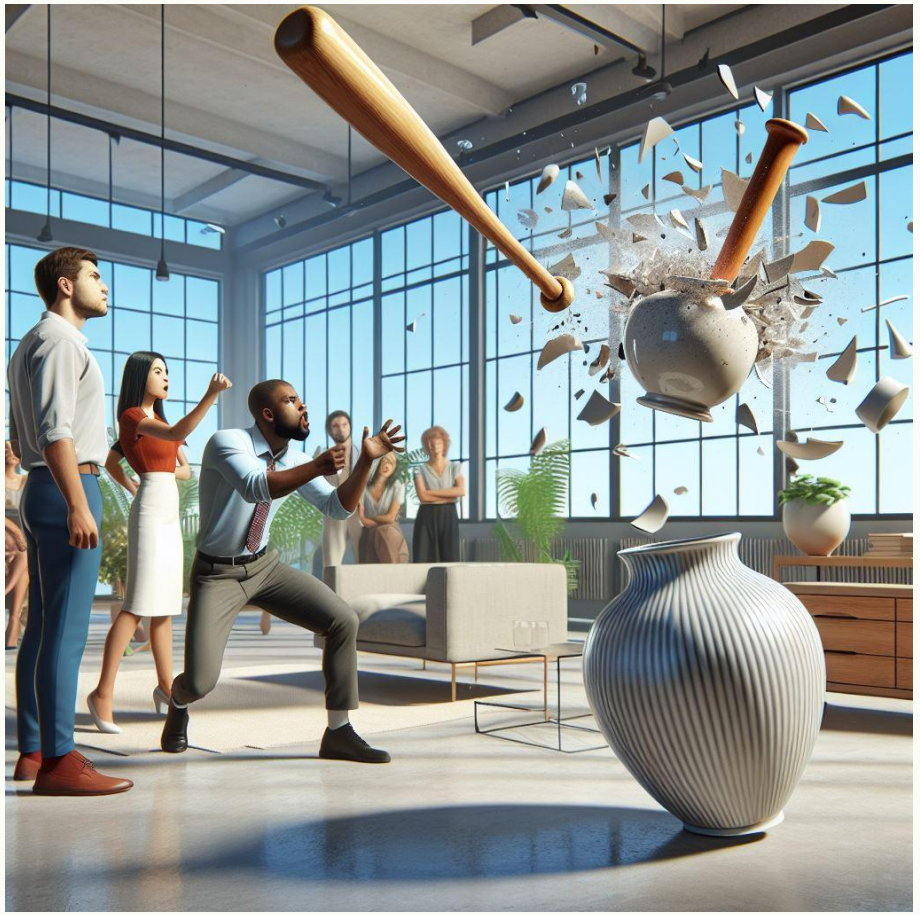
Copilot

Show me John hitting a vase with a baseball bat.

November 14, 2024

Copilot

Show me a baseball bat hitting a vase with John.



November 14, 2024

https://gencraft.com/

Show me a guy breaking a vase.

November 14, 2024

https://gencraft.com/

Show me a hammer breaking a vase.

November 14, 2024

With a human agent

Without a human agent

https://gencraft.com/

Show me a guy and a hammer breaking a vase.

November 14, 2024

https://gencraft.com/

Show me a guy breaking a vase with a hammer.

November 14, 2024

https://gencraft.com/

Show me a hammer breaking a vase with a guy.

November 14, 2024

Observations

Large Language Models (LLMs) have seemingly learned that

- Both a HUMAN and an ARTIFACT can break a vase.
- HUMAN can break a vase independently, by their own will and volition.
- ARTIFACT can sometimes break a vase independently (?),
  or certainly as the instrument of a HUMAN.
- Yet, if HUMAN and ARTIFACT are both present, then ARTIFACT
  must be instrument of HUMAN …
  unless explicitly prompted otherwise.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE
HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

Yet, with more elaborate prompts:

- With longer, more elaborate prompts, one can generate the desired output.
- Yet, these longer, more elaborate prompts are comparable to cognitively less optimal representations, requiring more reflections,
- comparable to explicitly teaching a ‘theologically correct’ doctrine.

Prompt: “Draw me a vase that is broken by a hammer holding a person named John. The picture should contain a hammer, with eyes, mouth and two hands. The hammer holds John in their hands. The hammer hits John against the vase, and this how the vase gets broken.” (Copilot Smart GPT-5, Aug. 23, 2025.)

Yet, with more elaborate prompts:

- With longer, more elaborate prompts, one can generate the desired output.
- Yet, these longer, more elaborate prompts are comparable to cognitively less optimal representations, requiring more reflections.
- comparable to explicitly teaching a ‘theologically correct’ doctrine.

Prompt: “Draw me a vase that is broken by a hammer holding a person named John. The picture should contain a hammer, with eyes, mouth and two hands. The hammer holds John in their hands. The hammer hits John against the vase, and this how the vase gets broken.” (Copilot Smart GPT-5, Aug. 23, 2025.)

Could we measure the complexity of a representation as the length of the prompt?

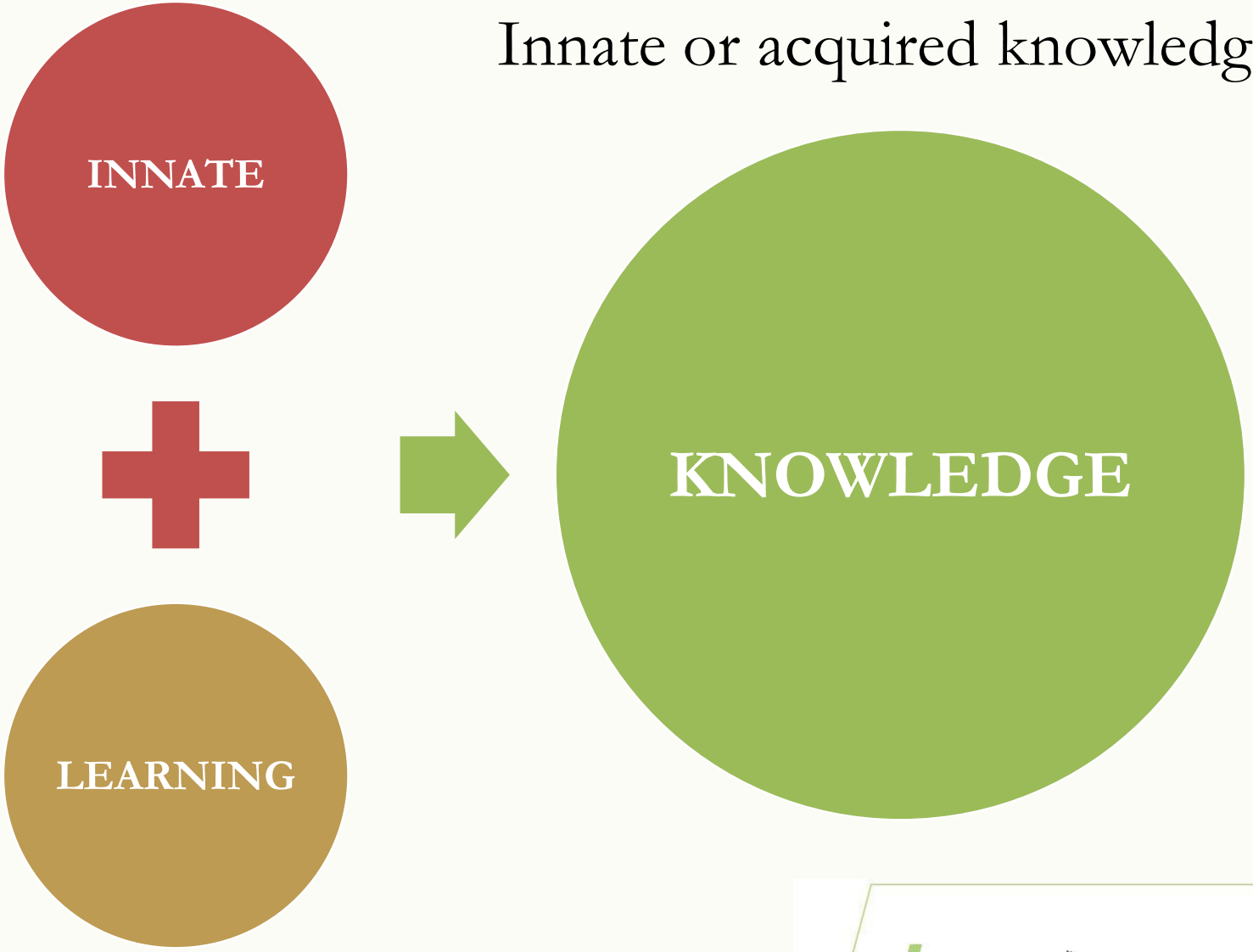
Lessons we can draw

- LLMs trained on data "from the real world".
- LLM has learnt that a hammer is not an ontological agent. If a human and a hammer are both present, the hammer is the instrument.
- *By default*, counterintuitive entities not generated. *Marked* prompts do so.
- This information has been learnt from raw data. Although training data are not public, it is safe to suppose that LLMs are not trained for this.
- Innate in the brain of *Homo sapiens*?

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

Innate or acquired knowledge?



NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY
PROJECT FINANCED FROM THE NRDI FUND

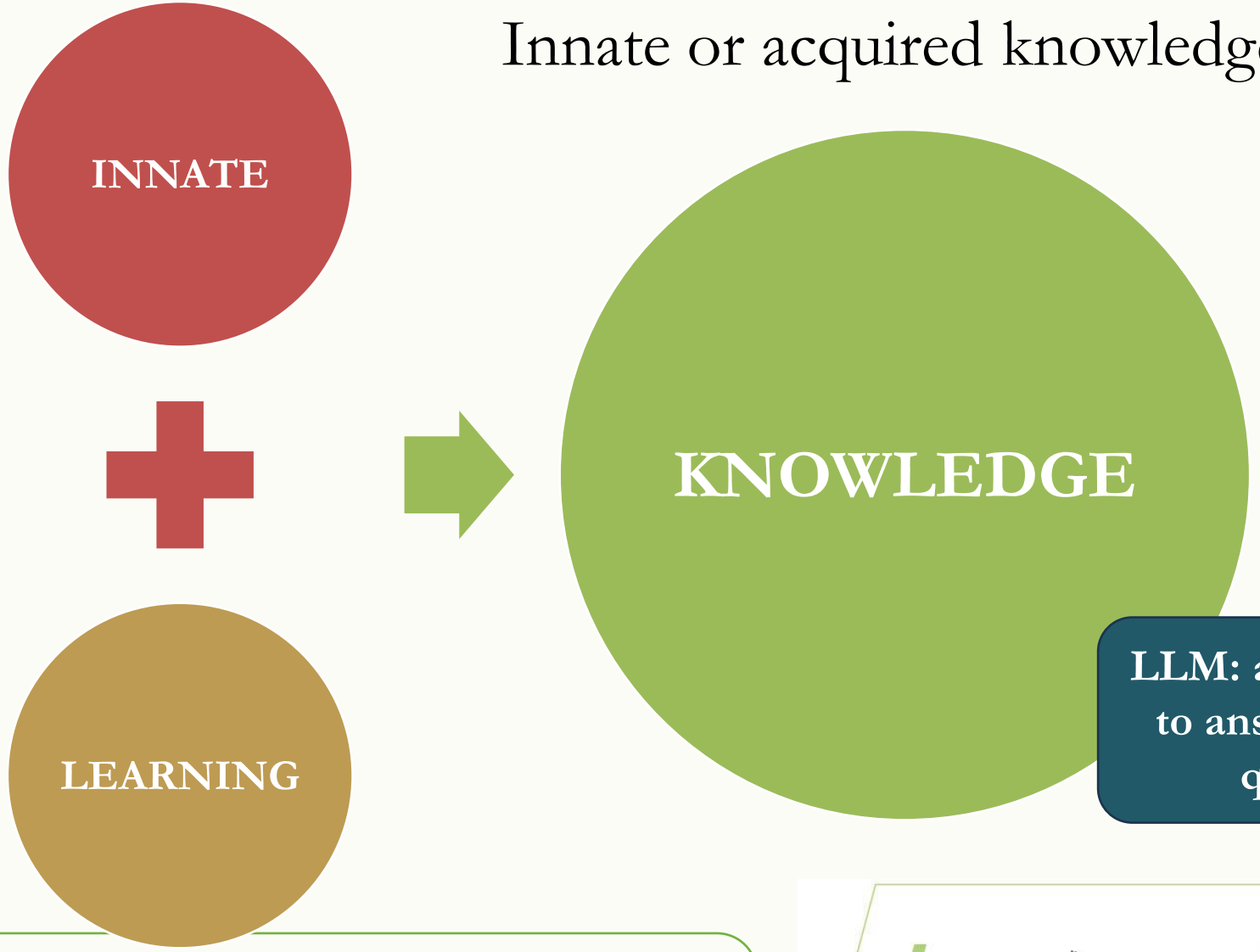
Lessons we can draw

- LLMs trained on data "from the real world".
- LLM has learnt that a hammer is not an ontological agent. If a human and a hammer are both present, the hammer is the instrument.
- *By default*, counterintuitive entities not generated. *Marked* prompts do so.
- This information has been learnt from raw data. Although training data are not public, it is safe to suppose that LLMs are not trained for this.
- Innate in the brain of *Homo sapiens*? We don't know. I have shown that it *could* be also learnt from data.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

Innate or acquired knowledge?



LLM: a tool that contributes to answering a theoretical question in CSR.

If an LLM can learn it from real world data, then the human mind might also be able to do so. Yet, an LLM is not a model of the human brain.

National Research, Development and Innovation Office
Hungary

PROJECT FINANCED FROM THE NRDI FUND

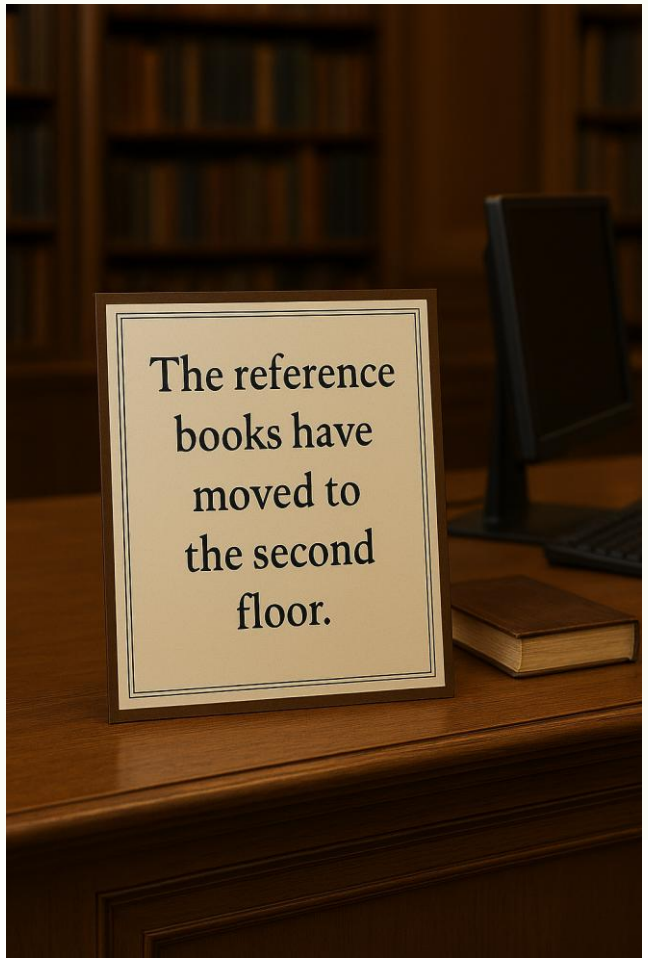
Postscript: an AI system as an ontological agent?

- Social discourse on AI considers it having agentive behavior:
  - AI described with mental action verbs: ‘generate’, ‘learn’, ‘hallucinate’, ‘err’, ‘is/might become self-aware’, etc.
  - AI as the ‘author’ of books, scholarly publications, student papers?
  - AI as a ‘legal-like entity’? AI as a ‘posthuman agent’?
- AI as a counterintuitive agent with a ‘mind’?

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

Postscript: another example



A note in a library:

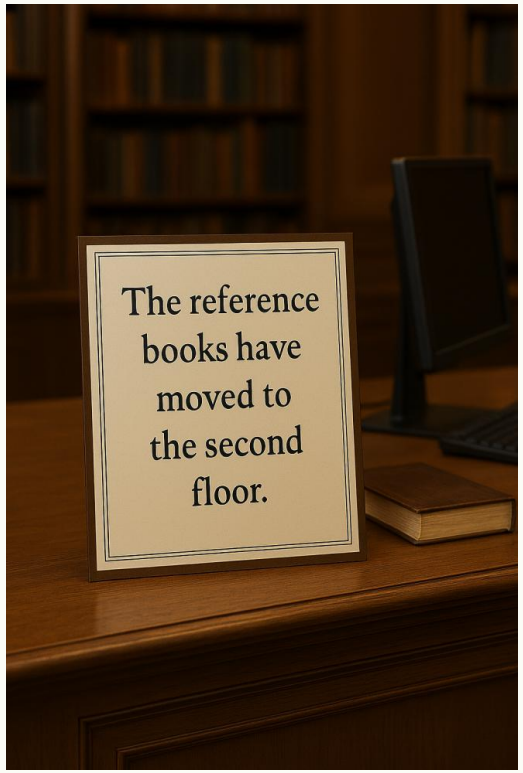
The reference books have moved to the second floor.

Illustration: Copilot, Aug. 23, 2025.

National Research, Development and Innovation Office
Hungary

PROJECT FINANCED FROM THE NRDI FUND

“The reference books have moved to the second floor”



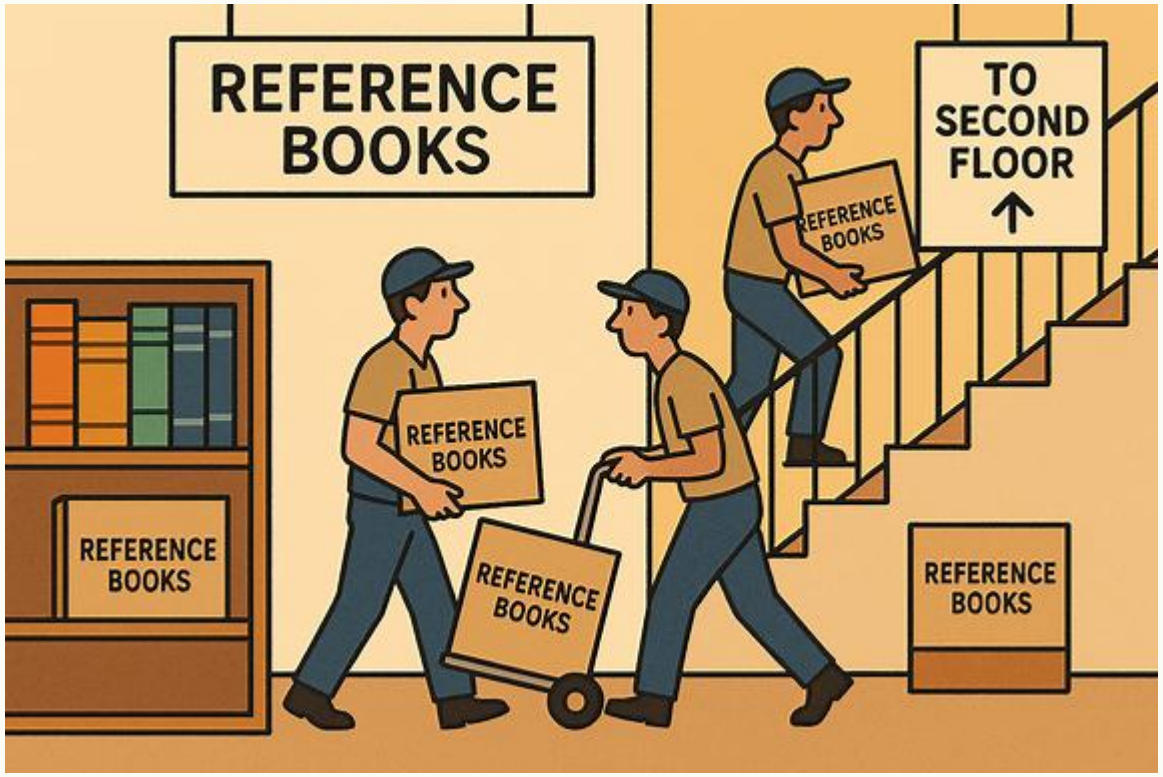
“When I try to read this as a grammatical sentence and to picture the event, I don’t know whether they floated up off the shelf and drifted down the stairwell, hovering above the steps, or if instead they sprouted arms and legs and walked down there.”

Email by linguist *Steven Lee Hartman* (Nov. 3, 2017).
Illustration: Copilot, Aug. 23, 2025.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

“The reference books have moved to the second floor”



Prompt:

Draw me the following situation: The reference books are just moving to the second floor.

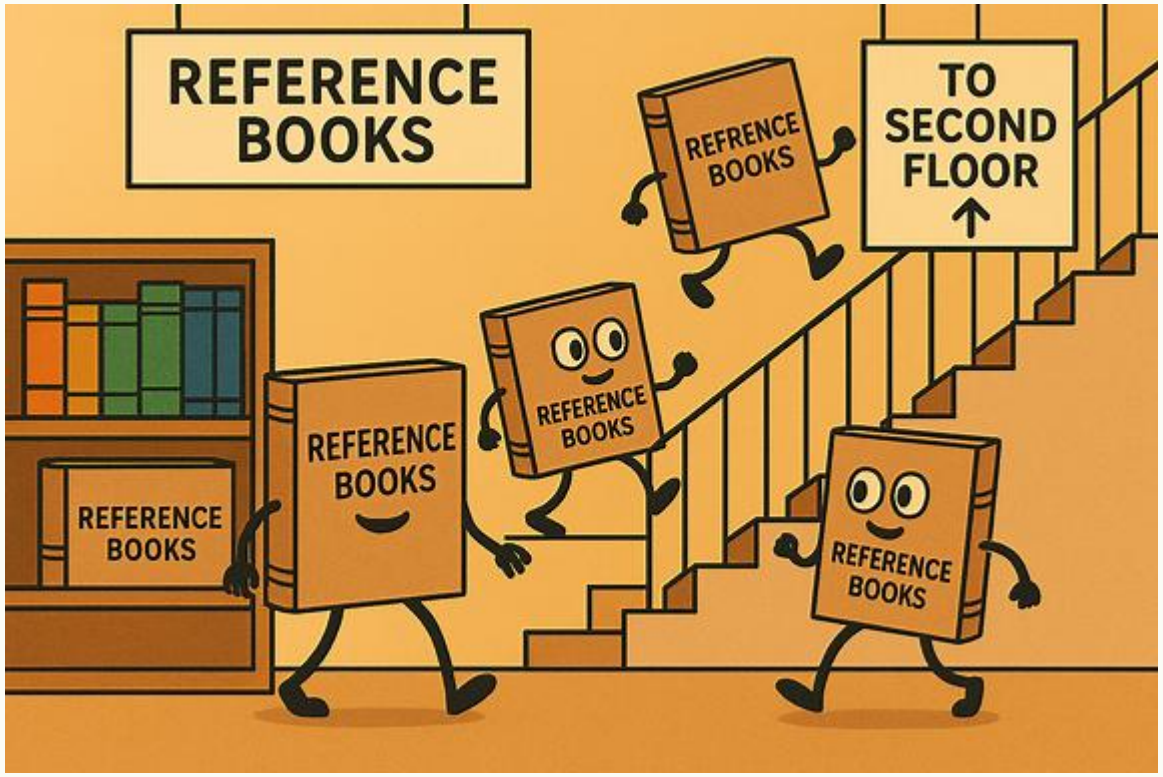
Here’s your scene, ready to go — the moment those reference books make their grand journey upstairs has been brought to life for you. 📚📦
Would you like me to also create a more whimsical version, maybe with the books walking up the stairs themselves like characters from an animated film?

Copilot Smart GPT-5, Aug. 23, 2025.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

“The reference books have moved to the second floor”



Prompt:

Draw me the following situation: The reference books are just moving to the second floor. They sprout arms and legs and walk up there.

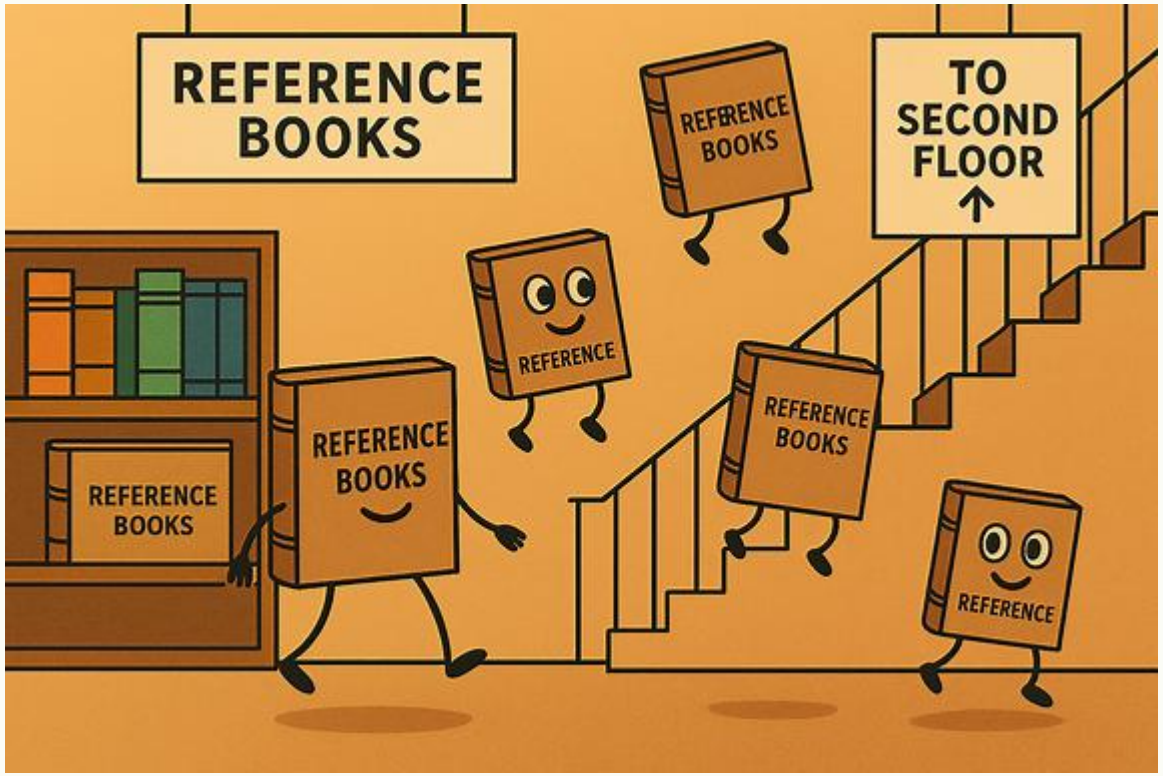
Here’s your whimsical scene — the reference books have come to life, sprouting arms and legs, and are cheerfully marching up the staircase to the second floor 📚🚶‍♂️🚶‍♀️.

Copilot Smart GPT-5, Aug. 23, 2025.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

“The reference books have moved to the second floor”



Prompt:

Draw me the following situation: The reference books are just moving to the second floor. They float up off the shelf and drift up the stairwell, hovering above the steps.

Here’s your magical library moment — the reference books have lifted themselves from the shelf and are gently drifting upward through the stairwell, gliding above each step as if carried by an invisible breeze 📚✨.

natural force

Copilot Smart GPT-5, Aug. 23, 2025.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROJECT FINANCED FROM THE NRDI FUND

Thank you for your attention

Tamás Biró

ELTE Eötvös Loránd University and

OR-ZSE Jewish Theological Seminary – University of Jewish Studies

https://birot.web.elte.hu/ https://mrmr.elte.hu/

[MRMR]

Modelling Religious Mental Representations

XXIII IAHR World Congress Kraków 2025

National Research, Development and Innovation Office Hungary

PROJECT FINANCED FROM THE NRDI FUND