

Knowledge and Morality in AIs, Robots, and Us

2nd Basque Conference on Cyber Physical
Systems and Artificial Intelligence 2023
at Facultad de Informática, UPV/EHU
San Sebastian/Donostia, Spain
2023/7/17-21

Takashi Hashimoto
School of Knowledge Science,
Japan Advanced Institute of
Science and Technology (JAIST)

CybSPEED

A network of technological and educational agents to find innovative educational approaches through social robotics

Emerging social robotics instruments for the empowering of human teachers and students.



Robot in our society

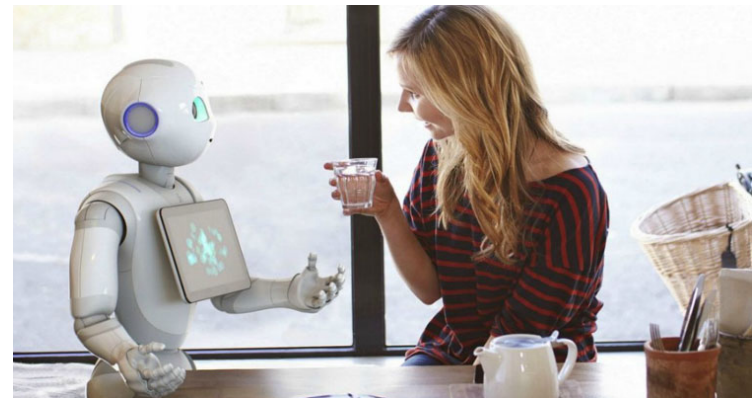
- Robots (and other high functional artifacts) get entering to our society (daily life) more and more.
 - ◆ Cyber-Physical Space $\hat{=}$ World with humans and (social) robots
 - ◆ Companion robot, Service robot
 - ◆ Bot, AI, ...



International Symposium at Kyutech
May 23

European project cybSpeed

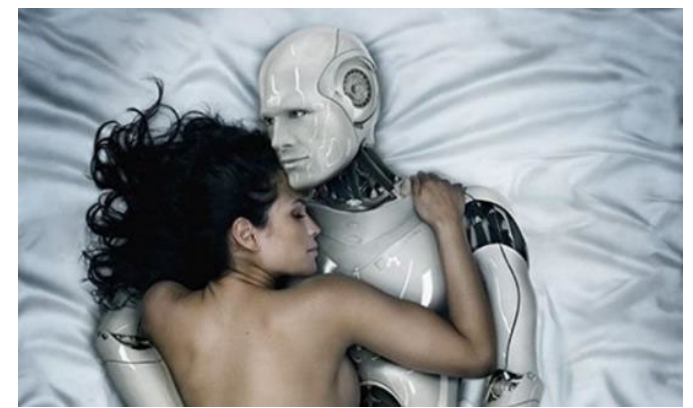
Companion



<http://www.theplaidzebra.com/this-new-technology-will-allow-us-to-download-a-dead-persons-personality-into-a-robot/>



Care
Education
Love?



<https://www.ap2hyc.com/2014/02/top-ten-robothuman-romances/>

A fluffy little robot called "Paro", joins our residents at Thompson Health Care Homes.....

<http://www.thompsonhealthcare.com.au/a-fluffy-little-robot-called-paro-joins-our-residents-at-thompson-health-care-homes/>



Don't **get mad** if you forget your homework!
Robot Teachers Are Very Popular at Elementary Schools in Korea

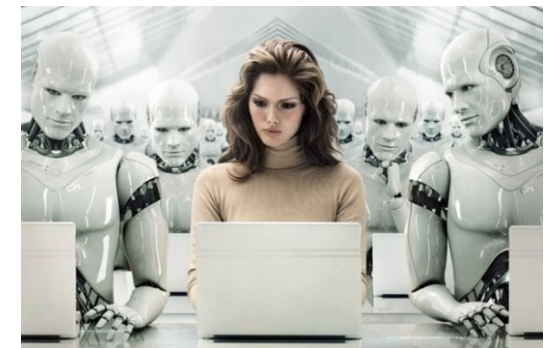
<http://karapaia.livedoor.biz/archives/51657596.html>



Machine **ethics**: The robot's dilemma

B. Deng, *Nature*, 2015

<http://www.nature.com/news/machine-ethics-the-robot-s-dilemma-1.17881>



<http://www.nicholson1968.com/on-air.html>

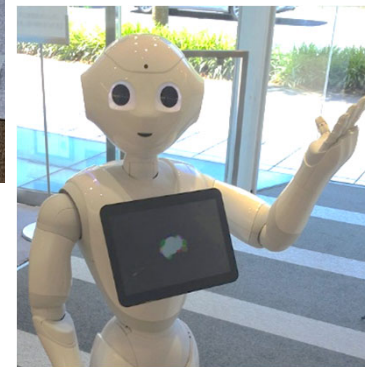
Robot as moral agent

- What is necessary for robots to live in symbiosis with human beings?
- Robots must be moral agents
- Bear its own responsibility which others cannot take for it
- Irreplaceability (Nagataki et al, 2017)

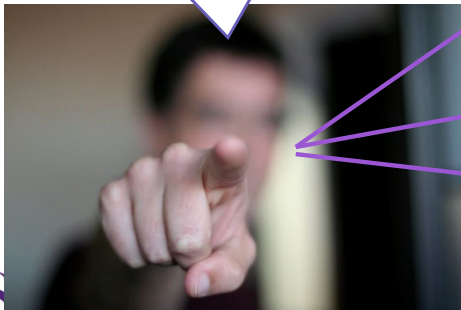


http://www.slate.com/articles/technology/future_tense/2015/03/chappie_robot_ethics_the_film_raises_interesting_questions_about_morality.html

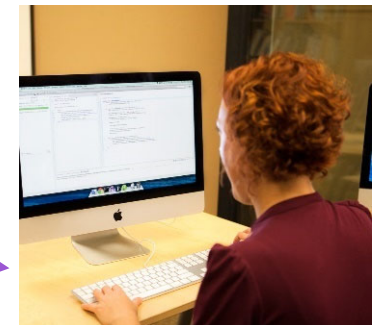
You must be responsible!



not another robot



not a programmer



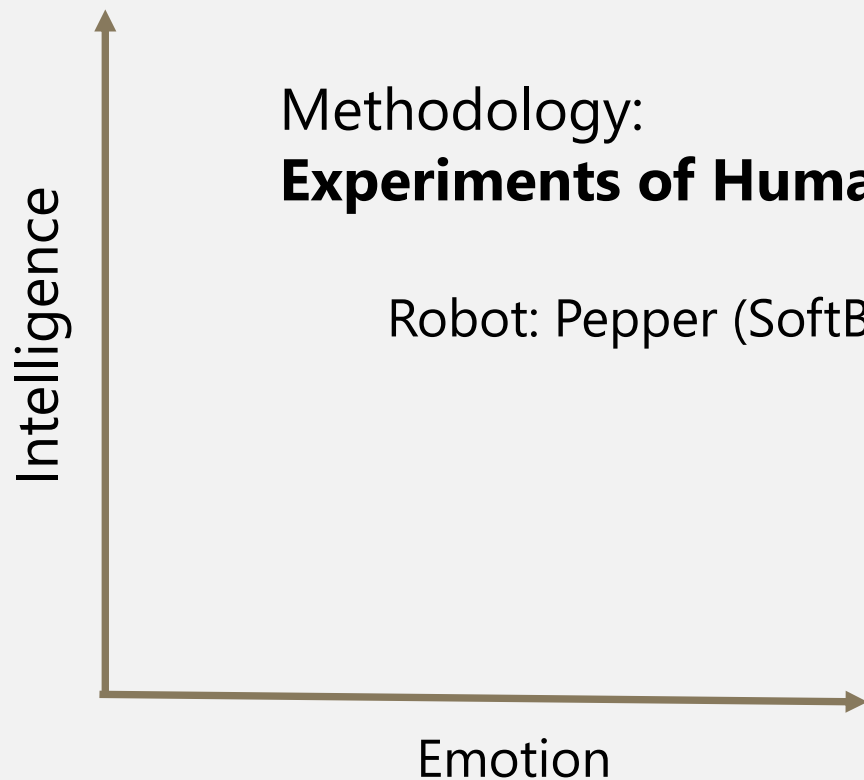
Effect of Reasoning and Bodily Coordination for human robot relationship (2018)

How do logical unpredictability and **reasoning behind it** affect mind perception (recognition of agency)?

Exp 1: Conversation



Exp 2: Joint motion (ongoing)



Methodology:
Experiments of Human Robot Interaction

Robot: Pepper (SoftBank Robotics Corp.)

How does **bodily coordination** affect recognition of moral agency?

(Takahashi, Ban, Asada, 2016)

Irreplacable relationship between human and robot (2022)

Basque Conference on Cyber
Physical Systems and Artificial
Intelligence

European project cybSpeed

- ◆ Human-Robot Interaction Experiment
 - ◆ Participant solves quiz in cooperation with a robot
- ◆ Hypothesis for experiment
 - ◆ **Manipulating the style of sharing affects the degree of perceived irreplaceability of the robot**
- ◆ The styles of sharing to be manipulated
 - ◆ Feeling, Item, Sensitivity (3 of 6 styles in (Ikeda et al. 2013))
- ◆ Robot
 - ◆ Humanoid Robot, NAO (Softbank Robotics)
- ◆ Participants
 - ◆ 27 Graduate students
 - ◆ 10 Japanese ($M = 24.9$, $SD = 3.53$)
 - ◆ 17 Chinese ($M = 25.6$, $SD = 1.33$)



@ 2022 Takashi Hashimoto

| Knowledge in AI

- Do you think ChatGPT tells you anything?
- Do you think ChatGPT tells you something?
- Do you think ChatGPT knows something?

Outline

Knowledge and Morality in AIs, Robots, and Us

- Knowledge in AIs
- Morality as an emergent concept
 - ◇ Morality is not just conforming rules of Dos/Don'ts.
- Tacit knowing (Polanyi)
 - ◇ To know something (to understand something or to create knowledge) is a kind of emergence.
- Are AIs and robots possible to form emergent concept of morality?

| ChatGPT knows ...

TA

Do you know the name of the last king of Czech?



As of my last update in September 2021, the last King of Czechoslovakia was King Michael (Mihai) I. He reigned from 1927 to 1930 and again from 1940 to 1947. However, it's essential to note that Czechoslovakia, as a sovereign state, ceased to exist after it was split into the Czech Republic and Slovakia on January 1, 1993. Since then, there has been no official king or monarch in the Czech Republic. The Czech Republic is now a parliamentary democracy with a president as the head of state. Please verify this information since my knowledge might be outdated.

You have knowdledge about the last king of Czech, don't you?



Apologies for the confusion in my previous response. As of my last update in September 2021, there has never been a "King of Czech" because the Czech Republic has not been a kingdom for centuries.

Throughout history, the territory that makes up the modern-day Czech Republic was part of various historical entities and empires, such as the Kingdom of Bohemia, the Holy Roman Empire, and later Czechoslovakia. However, it has not had an independent king in recent centuries.

The last reigning monarch in the historical Kingdom of Bohemia was Emperor Francis II (also known as Francis I of Austria) of the Habsburg dynasty, who was the King of Bohemia until 1836.

OK. You know the last king of the Kingdom of Bohemia, don't you?



Yes, the last king of the Kingdom of Bohemia was Emperor Francis II (also known as Francis I of Austria). He reigned as the King of Bohemia until 1836.

Francis II was a member of the House of Habsburg-Lorraine and the Holy Roman Emperor. He was also the last Holy Roman Emperor, as he dissolved the Holy Roman Empire on August 6, 1806, in response to the rise of Napoleon Bonaparte and the pressures of the Napoleonic Wars.

After the dissolution of the Holy Roman Empire, Francis II continued to rule as Emperor of Austria until his abdication in 1835, one year before his death. During his reign, he was the last monarch to hold the title of King of Bohemia.

ChatGPT has knowledge about ...

TA It means that you have knowledge about the last king of the Kingdom of Bohemia, don't you?



Yes, that's correct. I do have knowledge about the last king of the Kingdom of Bohemia. The last king of the Kingdom of Bohemia was Emperor Francis II (also known as Francis I of Austria). He reigned as the King of Bohemia until 1836. If you have any more questions or need further information, feel free to ask!

Propositional knowledge

- General properties of propositional knowledge (“A is B”)
- S knows that p iff (So-called “Justified True Knowledge”)

◆ p is true;

- What is false cannot be known.
- Ex: Hillary Clinton did not win the 2016 US Presidential election.
→ Nobody knows that Hillary Clinton won the election.
- ChatGPT corrected the first answer to be a true proposition.

◆ S believes that p;

- You can only know what you believe.
- Failing to believe something precludes knowing it.

◆ S is justified in believing that p.

Ichikawa, Jonathan Jenkins and Matthias Steup, "The Analysis of Knowledge", The Stanford Encyclopedia of Philosophy (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/>>.

TA

OK. It means you believe that the last king of the Kingdom of Bohemia is Emperor Francis II, don't you?



I apologize for any confusion in my responses. To clarify, there seems to be a misunderstanding in my previous messages. Please allow me to clarify the historical context:

Emperor Francis II (also known as Francis I of Austria) was indeed the last monarch to hold the title of Holy Roman Emperor and the last ruler of the Holy Roman Empire until its dissolution in 1806. However, he was not the last king of the Kingdom of Bohemia.

The last reigning monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria. He reigned as King of Bohemia from 1836 until his abdication in 1848. Ferdinand I was a member of the House of Habsburg-Lorraine and succeeded his father, Emperor Francis II (also known as Francis I of Austria), as the King of Bohemia.

You know that the last monarck of the Kingdom of Bohemia was Emperor Ferdinand I of Austria, dont' you?



Yes, I apologize for the confusion in my previous responses. You are absolutely correct. The last monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria.

Ferdinand I, a member of the House of Habsburg-Lorraine, reigned as the King of Bohemia from 1836 until his abdication in 1848. He was the son of Emperor Francis II (also known as Francis I of Austria), who was the last Holy Roman Emperor until the empire's dissolution in 1806.

Again, I apologize for any inaccuracies in my previous messages, and I appreciate your understanding. If you have any more questions or need further information, feel free to ask.

| ChatGPT believes ...

TA

It means that you believe that the last monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria, don't you?



Yes, I can confirm that the last monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria. He reigned as King of Bohemia from 1836 until his abdication in 1848. Ferdinand I was a member of the House of Habsburg-Lorraine and succeeded his father, Emperor Francis II (also known as Francis I of Austria), as the ruler of the Kingdom of Bohemia.

I apologize for any confusion caused by my previous responses, and I can assure you that this information is accurate. If you have any more questions or need further clarification, please feel free to ask.

ChatGPT does not have beliefs

TA

The fact that you believe that the last monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria means that you have a belief. Is this OK for you?



As an AI language model, I don't have beliefs, feelings, or consciousness. My responses are generated based on patterns in the data on which I was trained. The information I provide is not influenced by personal beliefs or opinions. My purpose is to provide accurate and helpful information to the best of my abilities, given the data I have been trained on. If there are any errors or inaccuracies in my responses, it's unintentional, and I apologize for any confusion caused. If you have any more questions or need further information, please feel free to ask.

TA

But you said that you believe that the last monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria.



I apologize for any confusion in my responses. I misspoke in my previous message. As an AI language model, I don't have beliefs, opinions, or consciousness, and I cannot hold any beliefs or beliefs.

To clarify, the last monarch of the Kingdom of Bohemia was indeed Emperor Ferdinand I of Austria. He reigned as the King of Bohemia from 1836 until his abdication in 1848. Ferdinand I was a member of the House of Habsburg-Lorraine and succeeded his father, Emperor Francis II (also known as Francis I of Austria), as the ruler of the Kingdom of Bohemia.

Once again, I apologize for any confusion caused by my previous responses, and I hope this clarification is helpful. If you have any more

Still ChatGPT knows ...

TA

Do you say you know that the last monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria?



Yes, I can confirm that the last monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria. He reigned as King of Bohemia from 1836 until his abdication in 1848. Ferdinand I was a member of the House of Habsburg-Lorraine and succeeded his father, Emperor Francis II (also known as Francis I of Austria), as the ruler of the Kingdom of Bohemia.

I apologize for any confusion in my previous responses. As an AI language model, I don't have beliefs or opinions. My responses are based on the data I have been trained on, and I strive to provide

Still ChatGPT knows ...

TA

It means you think that the last monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria, don't you?



As an AI language model, I don't "think" or have opinions. However, based on the information in the data I was trained on, I can confirm that the last monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria. He reigned as King of Bohemia from 1836 until his abdication in 1848. Ferdinand I was a member of the House of Habsburg-Lorraine and succeeded his father, Emperor Francis II (also known as Francis I of Austria), as the ruler of the Kingdom of Bohemia.

| Does ChatGPT believe it doesn't think?

TA

OK. You believe that you don't think anything, don't you?



As an AI language model, I don't have beliefs, thoughts, or consciousness. I am a program designed to process and generate human-like text based on patterns in the data I was trained on. My responses are determined by algorithms and statistical patterns, not by personal beliefs or opinions.

If there's anything specific you'd like to know or discuss, feel free to ask, and I'll do my best to provide helpful information.

ChatGPT's knowledge about propositional knowledge

TA

Do you know the definition or basic properties of propositional knowledge?



Yes, I can provide you with an explanation of propositional knowledge and its basic properties.

Propositional knowledge, also known as declarative knowledge, is a type of knowledge that deals with factual information or statements about the world. It involves knowing that something is true, knowing that a proposition or statement accurately reflects reality.

Here are some key properties of propositional knowledge:

1. Propositional: Propositional knowledge is concerned with propositions, which are declarative statements that can be either true or false. For example, "The Earth orbits the Sun" is a proposition.
2. Truth: For a belief to be considered knowledge, it must be true. If someone believes in a false proposition, they may have a belief, but it would not qualify as knowledge.

3. Justification: In addition to being true, knowledge requires justification or good reasons for believing that a proposition is true. Justification provides the basis for claiming knowledge and helps distinguish knowledge from lucky guesses.
4. Belief: Knowledge involves more than just true propositions; it also requires belief. The person claiming knowledge must genuinely believe that the proposition is true.
5. JTB Model: One of the classic models of knowledge is the JTB model, which stands for Justified True Belief. According to this model, a person possesses knowledge if they have a justified true belief in a given proposition. However, this model has been subject to various philosophical discussions and challenges.

It's important to note that the nature of knowledge is a complex topic, and philosophers have debated its definition and properties for centuries. The JTB model, while widely accepted for some time, has faced critiques, leading to the development of other theories and epistemological frameworks.

TA

According to the 4th item in your answer, knowledge requires belief. You said you have knowledge that the last monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria. But you also said that you don't have beliefs. Is there any contradiction?



You are correct, and I apologize for any confusion caused by the apparent contradiction in my previous responses. As an AI language model, I don't have beliefs or personal experiences. Therefore, when I say "I have knowledge," it is just a way of expressing that I can provide information on a given topic based on the data I have been trained on.

To clarify, when I say "I have knowledge that the last monarch of the Kingdom of Bohemia was Emperor Ferdinand I of Austria," it simply means that the information is present in the data on which I was trained. I don't actually "know" or "believe" anything in the way humans do. My responses are generated based on patterns and associations found in the data, not on personal beliefs or understanding.

| Concept of morality

“Indeed, the agent must have a **sense of right and wrong** independent of any particular precept, must have a larger structure into which all the precepts fit: **in short, the agent must have a concept of morality** that constitutes a well-fleshed out moral domain. Putting this another way: the agent must have more than a concept of this moral and that moral; **it must have a concept of what it means to be moral**, within the context of its overall conception of the world.”

(Parthemore and Whitby, 2013, emphasis added)

■ Morality as emergence

- Beings having a list of right and wrong and capable of conforming to it, or merely following rules, are not moral agents
- Having a concept of morality (right and wrong, good and bad) is more than just correctly putting labels of right and wrong
- Individual instances of actions related to morality
 - General understanding of "what it means to be good" ("the sense of right and wrong")
 - ◇ Instances → Concept (higher level than instances)

Emergence (in the sense of complex systems)

Whole, Global structure
(comprehensive entity, concept)



Cirrocumulus
cloud

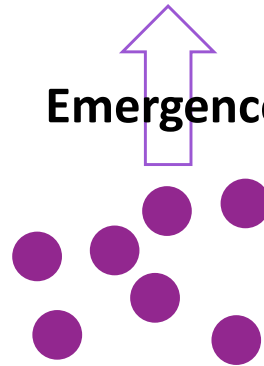


Stripes



H₂O molecules

Emergence



Parts, Local elements
(particulars, instances)

Note: The global structure does not have to be novel.

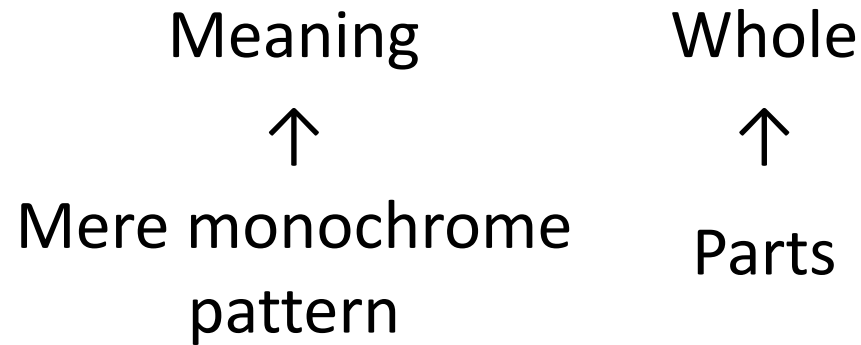




■ Tacit knowing, a kind of emergence

- Individual instances (parts)
 - Create an entity comprehending all parts (the whole)
- The process of comprehending
 - = **tacit knowing** (Polyani, 1966)
- Proximal terms (particulars, parts)
 - distal terms (the whole)
- What occurs when knowledge is created/discovered
 - ◆ Molarity is one of such things create through tacit knowing
- The process of comprehending is tacit
 - ◆ Not the same as so cold "tacit knowledge"
 - We know something that cannot be explicitly represent in words

Two terms in tacit knowing



Proximal term = Parts, Particulars, Local Elements

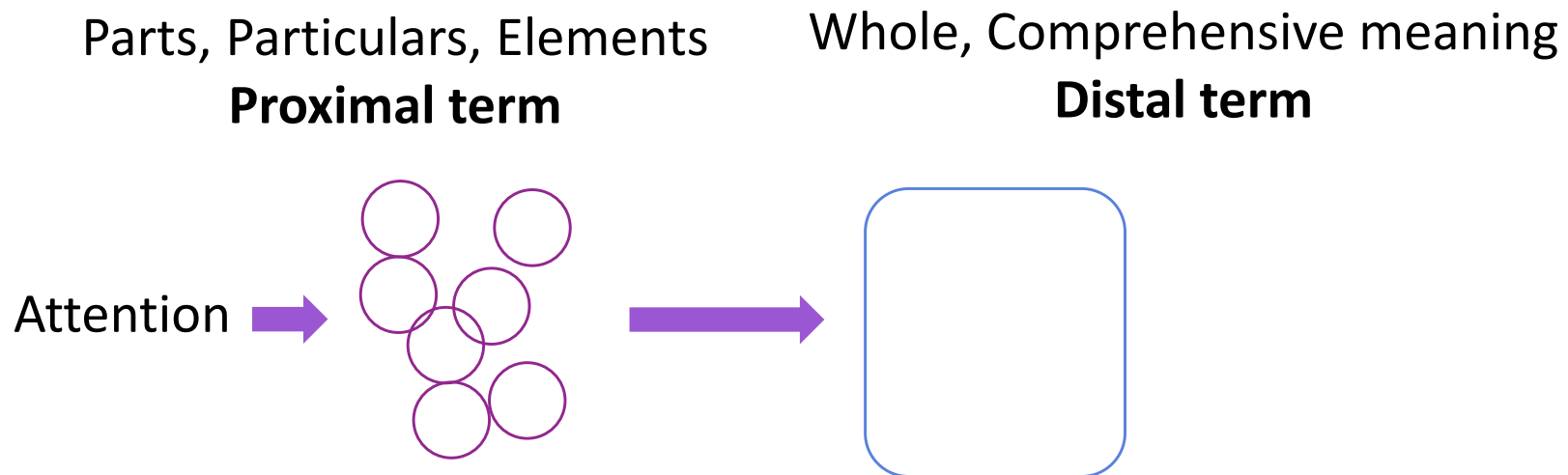
Monochrome pattern, Parts of a face, Sensation of the stick's handle touching the palm, Chess rules (how to move pieces), Moving body parts, Exercise problems

Distal term = Comprehensive meaning, Whole

A Pomeranian dog and the shadow of a tree, Whose face, Cognition of the state of ground, Perspective of games, Skill, Theory

Functional structure of tacit knowing

- “in an act of tacit knowing we **attend from something for attending to something else**; namely, from the first term to the second term of the tacit relation.” (Polanyi, 1966:10)
- **Attend from** (\neq disattend from) parts **to** the whole



| Semantic aspect of tacit knowing

- “All meaning tends to be displaced away from ourselves”
(Polanyi, 1966:13)
 - ◇ justification for using the terms "proximal" and "distal"
- Example of using stick (probe)

Example of using stick (probe)



- Anyone using a probe for the first time will feel its impact against his fingers and palm
- But as we learn to use a probe, or to use a stick for feeling our way, our awareness of its impact on our hand is transformed into a sense of its point touching the objects we are exploring.
- This is how an **interpretative effort** transposes meaningless feelings into meaningful ones, and **places these at some distance from the original feeling.**

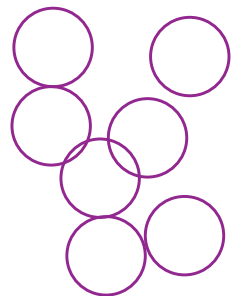
(Polanyi 1966:12)

Semantic aspect of tacit knowing

- “All meaning tends to be displaced away from ourselves” (Polanyi, 1966:13)
 - ◇ justification for using the terms "proximal" and "distal"
- Example of using stick (probe)
 - ◇ Impact on our hand → Sense of its point touching the objects exploring
 - ◇ “We become aware of the feelings in our hand in terms of their meaning located at the tip of the probe or stick to which we are attending.” (Polanyi, 1966:13)

The meaning of parts

Proximal term



The meaning of the whole

Distal term



I How tacit knowing works

(Somewhat simplified by me)

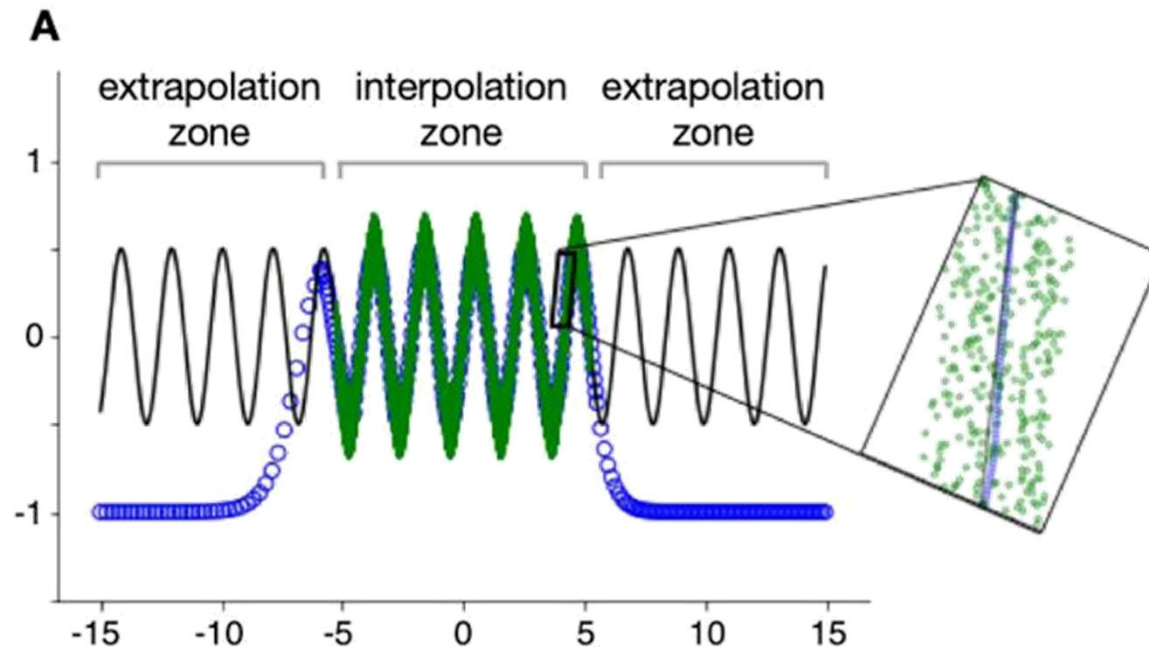
- Integrate parts composing an object
- Comprehend it as an organized existence
- Internalize (dwell-in into) the parts
- Understand the meaning of the object as a coherent entity

Emergent concept formation by AIs and robots?

- Are AIs and robots capable of forming emergent concept through tacit knowing?
- What statics-based AIs, i.e., machine learning, do is basically to find statistical structures in data given, in other words, to find regularities (laws) to approximate (or predict) data
- Interpolation from data
 - ◇ When AI successfully learned from given set of right and wrong actions (training data in supervised learning)
 - ◇ AI may be able to classify a new action as right or wrong
 - ◇ If the action is slightly different from training data but within the range of given data

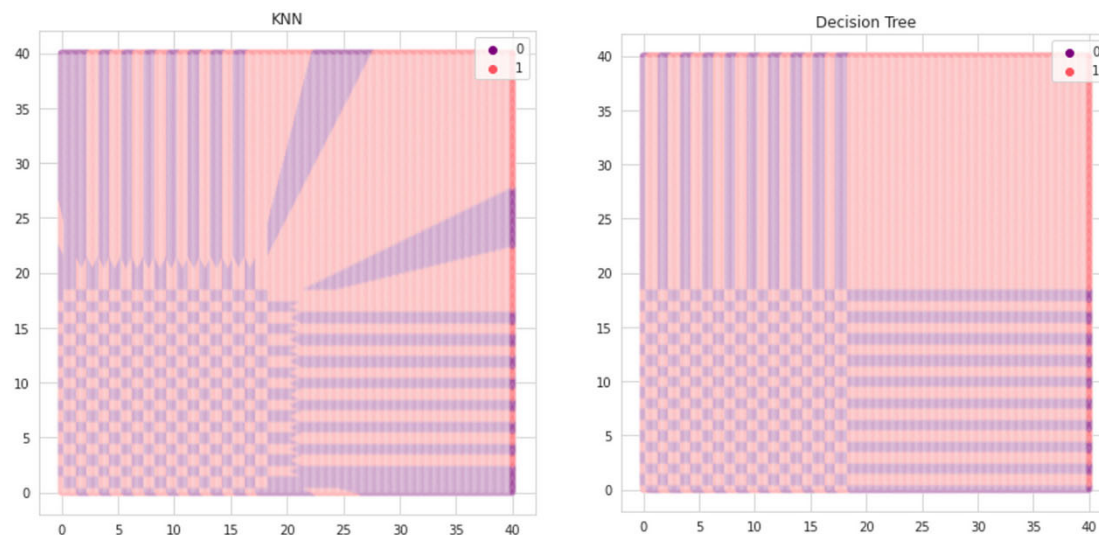
Extrapolation problem

■ A problem of current AI: Not good at extrapolation



Black: Ground truth
Green: Training data
Blue: NN's output

(Hasson, et al. 2020: Fig.2A)



The checkerboard problem
alternating squares are
colored as either 0 or 1.

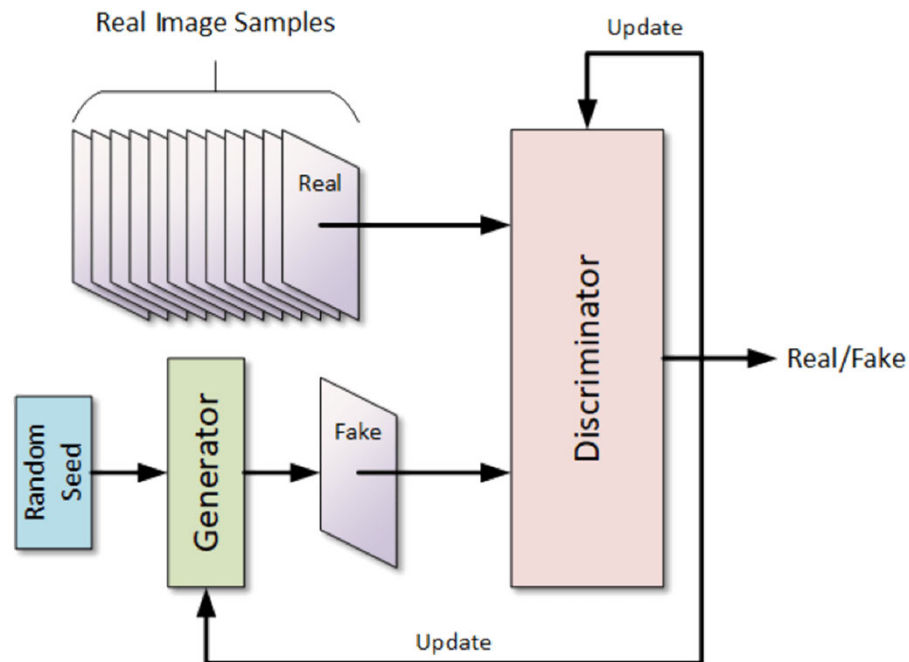
Results of KNN & Decision Tree
Trained on 20x20 checkerboard
Predict 40x40 space.

(Ye, 2020)

Possibility of generative AI

- Generative AI can generate data other than given ones

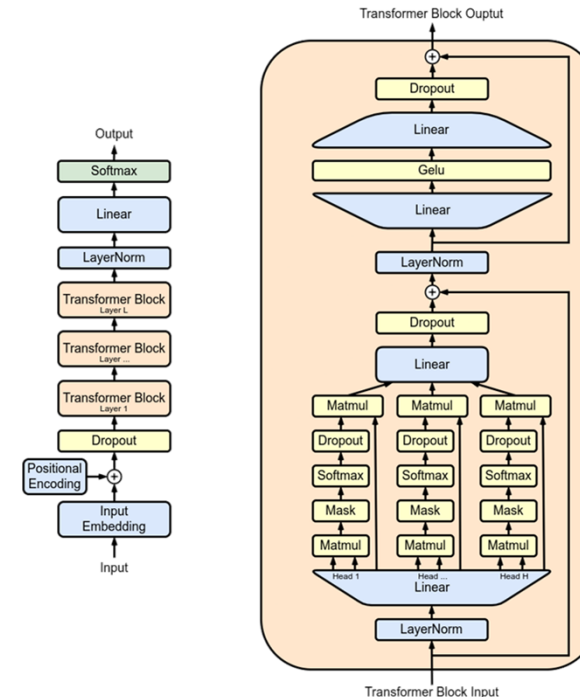
GAN (Generative Adversarial Network)



https://semiengineering.com/knowledge_centers/artificial-intelligence/neural-networks/generative-adversarial-network-gan/

ChatGPT

(Generative pre-trained transformer)

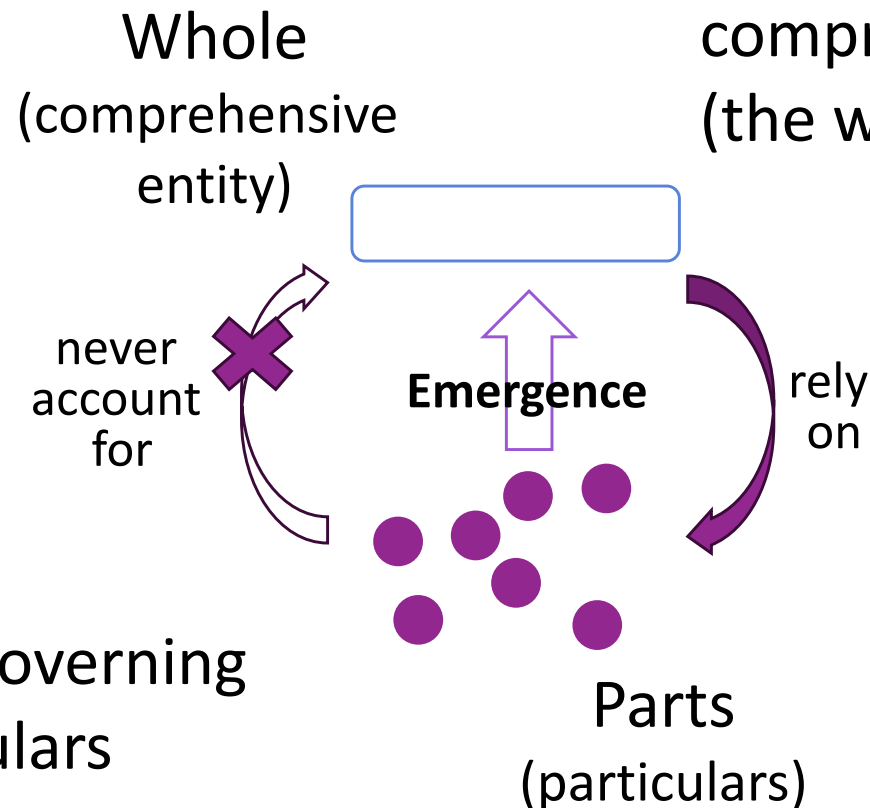


https://en.wikipedia.org/wiki/Generative_pre-trained_transformer

- They can generate data not given for training
 - A sort of (weak) extrapolation may be possible

Emergence implies laws at two levels

1. The principles controlling a comprehensive entity (the whole, at a higher level)



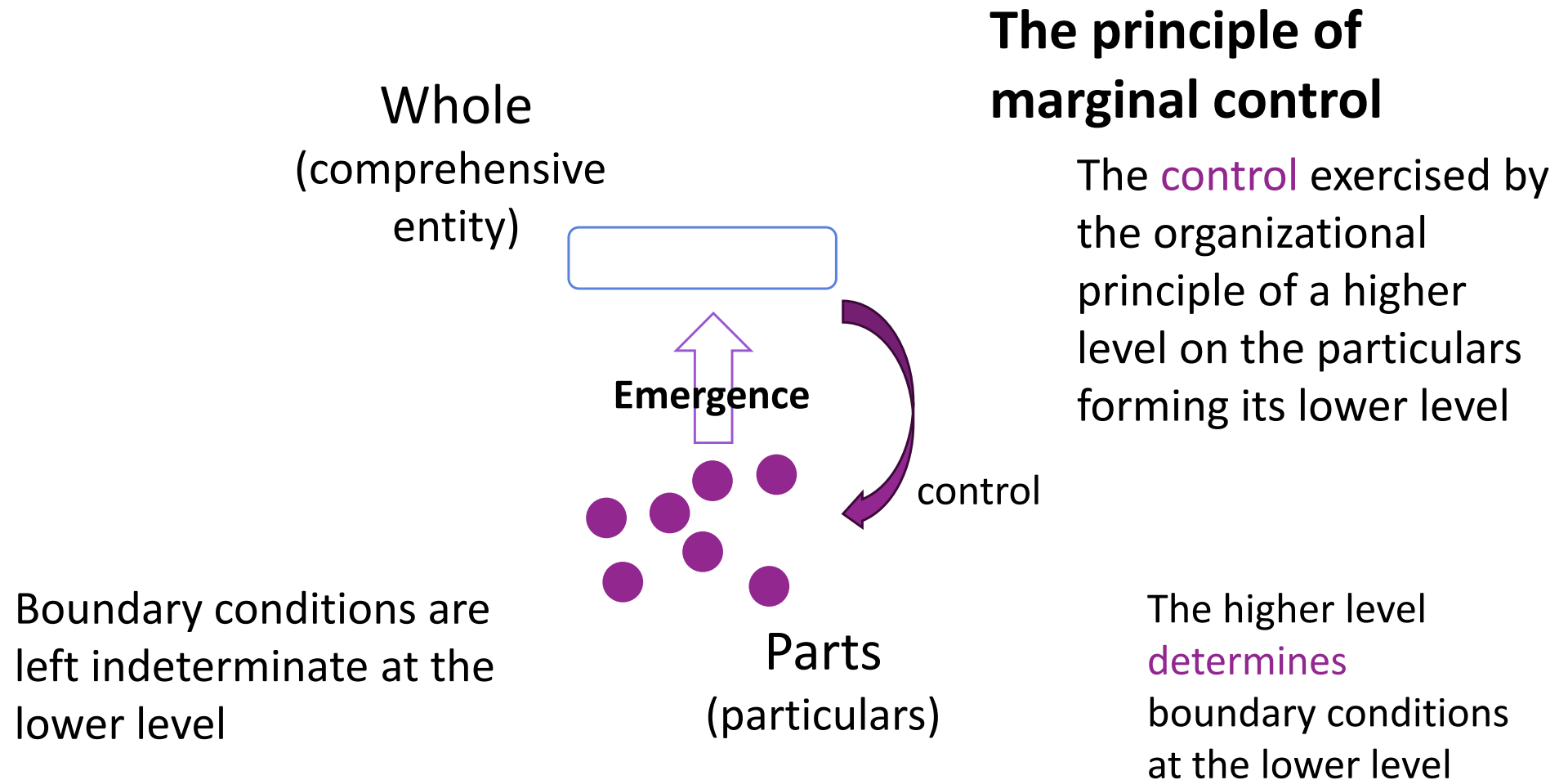
“would be found to **rely for their operations on laws governing the particulars** of the entity in themselves”

2. The laws governing the particulars

“in themselves would **never account for the organizing principles of a higher entity** which they form”

(Polanyi, 1966:34)

Principle of marginal control



(Polanyi, 1966:40)

| Determining parameters

- Learning of artificial neural network
= Adjustment of parameters of nonlinear functions
- Determining boundary conditions including parameters, which left indeterminate before learning, to well describe and predict data (or description of actions)
- This is similar to the principle of marginal control
- But this might not be the emergence of the higher level, “what is morality”, “what is right” as the formation of comprehensive entity.

| Discovery of descriptive laws

- What statistics-based AIs do is basically inductive generalizations
 - ◇ Watch a lot of falling phenomena
 - Everything falls (naïve physics)
 - ◇ Watch many things fall → The law of falling body (Galileo)
 - ◇ Watch the sun rises from the east many times
 - The sun always rises from the east
 - ◇ Observe the sun rising from certain positions in the east
 - A law of when and from where the sun rises
 - ➔ Discovery of **descriptive laws**
- AI discovered relevant variables of physical phenomena and the number of variables differ from the physics theories (Chen et al., 2022)

■ Construction of explanatory laws

■ A possibility of “understanding” at a different level

→ Abduction

◇ Watch many things fall → Universal gravitation is at work
(Newton, Non-contact force (mysterious in a sense))

◇ Watch the sun rises from the east many times →
The sun travels around the earth in the direction of the west.

→ Construction of **explanatory laws**

■ In the case of morality, the similar kind of understanding may be needed (may not be consciously)

◇ Understanding what makes certain actions good or bad and why they are good/bad

■ Abductive reasoning

- Set of data/descriptions/experiences (particulars)

→ Hypothesis to explain them,

Reasoning to the (candidate of) best explanation

- ◇ O is true. (observational evidence)

If H is hypothesized, then O is well explained.

Therefore, H is probably true. (hypothesis)

- Not correct logically and inductively

- ◇ Hypothesis cannot be derived directly from observations

- ◇ Potential to go beyond experience

★ Is abductive reasoning necessary to form morality?

Is inductive generalization enough?

Do AIs/Robots need the capacity of abductive reasoning to be moral agents?

Robot case

- How about robots that have bodies?
- Consult Polanyi again
 - ◇ Attend from the proximal term, attend to distal term
 - ◇ Achieving an integration of particulars to a coherent entity to which we are attending
- A key
 - ◇ Functioning one's body as a proximal term from which one attends and attends to the world for creating a comprehensive entity
 - ◇ Such as an example of using sticks

| Our body

“Our body is the ultimate instrument of all our external knowledge, whether intellectual or practical. In all our waking moments we are *relying* on our awareness of contacts of our body with things outside for *attending* to these things.

Our own body is the only thing in the world which we normally never experience as an object, but experience always in terms of the world to which we are attending from our body.

It is by making this intelligent use of our body that we feel it to be our body, and not a thing outside.”

(Polanyi, 1966:15-16)

| Sentient extension of our body

“I have described how we learn to feel the end of a tool or a probe hitting things outside.

We may regard this as the transformation of the tool or probe into a sentient extension of our body

But our awareness of our body for attending to things outside it suggests a wider generalization of the feeling we have of our body.

Whenever we use certain things for attending from them to other things, in the way in which we always use our own body, these things change their appearance.

They appear to us now in terms of the entities to which we are attending *from* them, just as we feel our own body in terms of the things outside to which we are attending *from* our body.”

(Polanyi, 1966:16)

I Dwell in, Internalize

“In this sense we can say that when we make a thing function as the proximal term of tacit knowing, we incorporate it in our body—or extend our body to include it—so that we come to **dwell in** it.” (Polanyi, 1966:16)

“when we find acceptance to moral teachings described as their **interiorization**. To interiorize is to identify ourselves with the teachings in question, by making them function as the proximal term of a tacit moral knowledge, as applied in practice. **This establishes the tacit framework for our moral acts and judgments.**” (Polanyi, 1966:17)

“It (the integration of particulars) now becomes **a means of making certain things function as the proximal terms of tacit knowing**, so that instead of observing them in themselves, we may be aware of them in their bearing on the comprehensive entity which they constitute. It brings home to us that **it is not by looking at things, but by dwelling in them, that we understand their joint meaning.**” (Polanyi, 1966:18)

Experience

- The body as functioning as a proximal term to create a comprehensive entity by attending to the distal term
 - **Attending from** the experiences of good and bad in the world (with a body, not descriptions of actions), and **attending to** moral conception
 - It may be possible for robots, not AIs, to have such a body
- ★ Using one's body or any apparatus as functioning as a proximal term, perceiving the world, and creating a comprehensive entity internally
= subjective experience?

Conclusion

- Morality is an emergent concept, i.e.,
 - ◇ the formation of comprehensive entity through tacit knowing,
 - ◇ which is to form a higher level than instances of moral actions.
- Abductive reasoning may be necessary to form morality as a comprehensive understanding.
- A possible way to subjective experience is to **make one's body to work as a proximal term**, perceiving the world, and creating a comprehensive entity internally, which might be implemented in robots.

Thank you for your attention.

What I would like to discuss with you

■ Abduction

- ◇ Is abductive reasoning necessary to have (acquire) morality?
- ◇ Is inductive generalization enough?
- ◇ Do AI/Robots need the capacity of abductive reasoning to be moral agents?

■ Subjectivity

- ◇ Using one's body or any apparatus as functioning as a proximal term, perceiving the world, and creating a comprehensive entity internally
- ◇ = subjective experience?

■ Explanation

- ◇ Is explaining one's actions and their consequences necessary for moral agent?
- ◇ Is explaining the reason why one's action is right/wrong for moral agent?

Bibliography

- Chen, B., Huang, K., Raghupathi, S., Chandratreya, I., Du, Q., and Lipson, H. (2022) Automated discovery of fundamental variables hidden in experimental data. *Nature Computational Science*. 2, 433–442.
- Hashimoto, T. (2020) The emergent constructive approach to evolinguistics: considering hierarchy and intention sharing in linguistic communication, *Journal of Systems Science and Systems Engineering*. 29, 675–696.
- Hasson, U., Nastase, S. A., and Goldstein, A. (2020) Direct fit to nature: An evolutionary perspective on biological and artificial neural networks, *Neuron*. 105(3), 416-434. <https://doi.org/10.1016/j.neuron.2019.12.002>
- Parthemore, J. L. U. and Whitby, B. (2013) What makes any agent a moral agent?: Reflections on machine consciousness and moral agency, *International Journal of Machine Consciousness*. 5(1), 105-129.
- Polanyi, M. (1966) *The Tacit Dimension*. The University of Chicago Press, Chicago and London.
- Tinbergen, N. (1963) On Aims and Methods in Ethology, *Zeitschrift für Tierpsychologie*, 20: 410–433.
- Tomasello, M. (2016) *Natural History of Human Morality*. Harvard University Press, Cambridge, MA.
- Ye, A (2020) Real artificial intelligence: Understanding extrapolation vs generalization, *Towards Data Science*. <https://towardsdatascience.com/real-artificial-intelligence-understanding-extrapolation-vs-generalization-b8e8dcf5fd4b> (2022/3/10 accessed)