Self-reference, meta-simulation and biological arrow of time

Prof. Mikhail Prokopenko
Centre for Complex Systems
Faculty of Engineering

THE UNIVERSITY OF SYDNEY

Center for Humans & Machines
Max Planck Institute for Human Development
18 August 2026

THE UNIVERSITY OF SYDNEY

Acknowledgements

IOP Publishing J. Phys. Complex. 6 (2025) 015006 https://doi.org/10.1088/2632-072X/ad9cdc

Journal of Physics: Complexity

CrossMark

OPEN ACCESS

RECEIVED
17 September 2024

ACCEPTED FOR PUBLICATION
10 December 2024

PUBLISHED
23 January 2025

Original Content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal

PAPER

Biological arrow of time: emergence of tangled information hierarchies and self-modelling dynamics

Mikhail Prokopenko[1,2,*], Paul C W Davies[3], Michael Harré[1,2], Marcus G Heisler[1,4], Zdenka Kuncic[1,5,6], Geraint F Lewis[5], Ori Livson[1,2], Joseph T Lizier[1,2] and Fernando E Rosas[7,8,9,10]

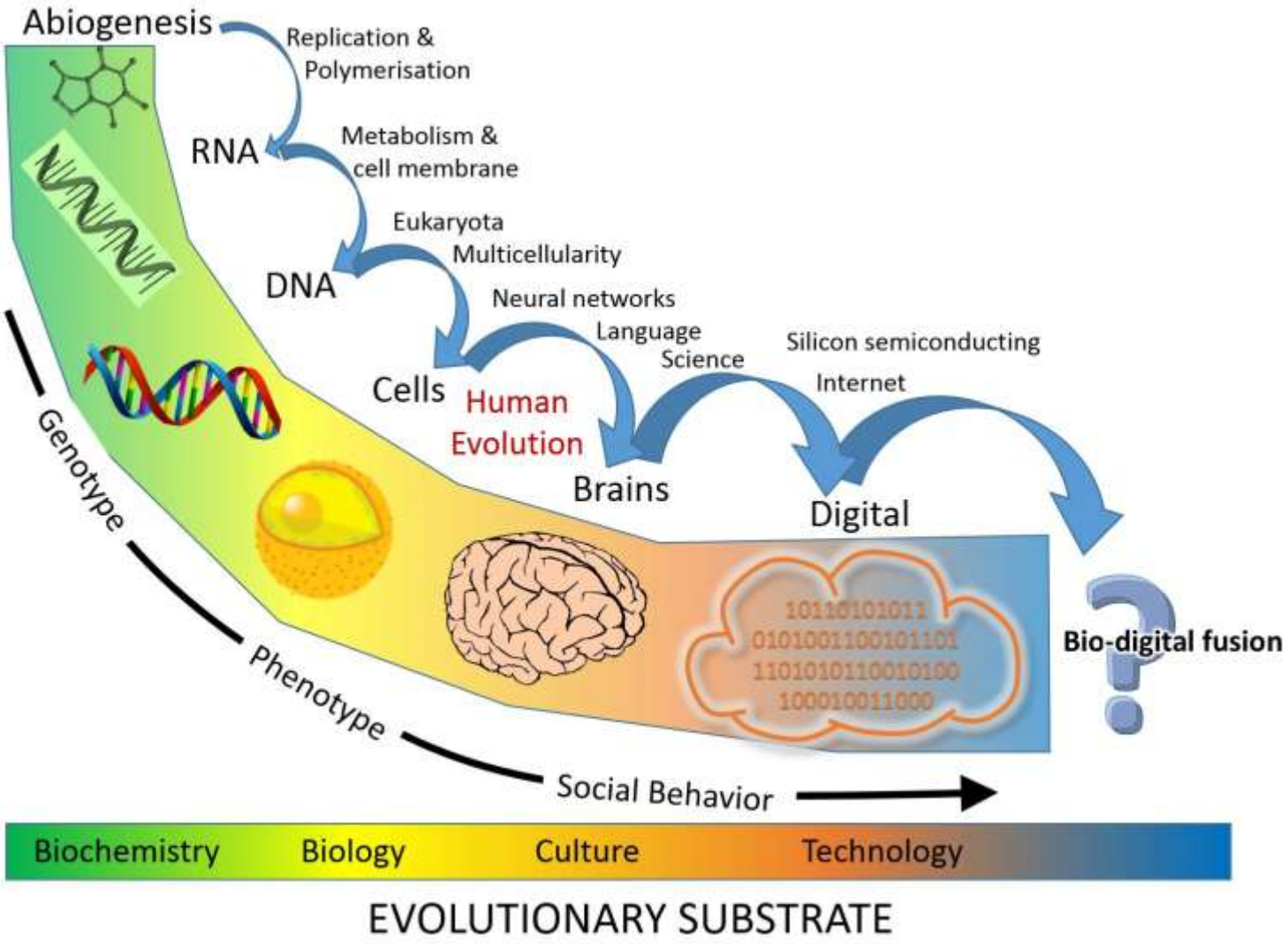
1 The Centre for Complex Systems, University of Sydney, Sydney, NSW 2006, Australia
2 School of Computer Science, Faculty of Engineering, University of Sydney, Sydney, NSW 2006, Australia
3 The Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ 85287–0506, United States of America
4 School of Life and Environmental Sciences, Faculty of Science, University of Sydney, Sydney, NSW 2006, Australia
5 School of Physics, Faculty of Science, University of Sydney, Sydney, NSW 2006, Australia
6 The Charles Perkins Centre, University of Sydney, Sydney, NSW 2006, Australia
7 Sussex AI and Centre for Consciousness Science, Department of Informatics, University of Sussex, Brighton BN19RH, United Kingdom
8 Centre for Psychedelic Research, Department of Brain Science, Imperial College London, London SW72AZ, United Kingdom
9 Centre for Complexity Science, Imperial College London, London SW72AZ, United Kingdom
10 Center for Eudaimonia and Human Flourishing, University of Oxford, Oxford OX39BX, United Kingdom

THE UNIVERSITY OF SYDNEY

Outline

- Major evolutionary transitions and arrows of time
- “Coding thresholds” and evolutionary connectionism
- Tangled hierarchies and strange loops
- Self-reference and incomputability
- Turing oracle machines and Turing jumps
- Conclusion: meta-simulation and novelty generation

Major Evolutionary Transitions



based on: Gillings, M. R., Hilbert, M., & Kemp, D. J. (2016). Information in the Biosphere: Biological and Digital Worlds. *Trends in Ecology & Evolution*, 31(3), 180–189. http://escholarship.org/uc/item/38f4b791

Wikipedia

Major Evolutionary Transitions & "codes of life"

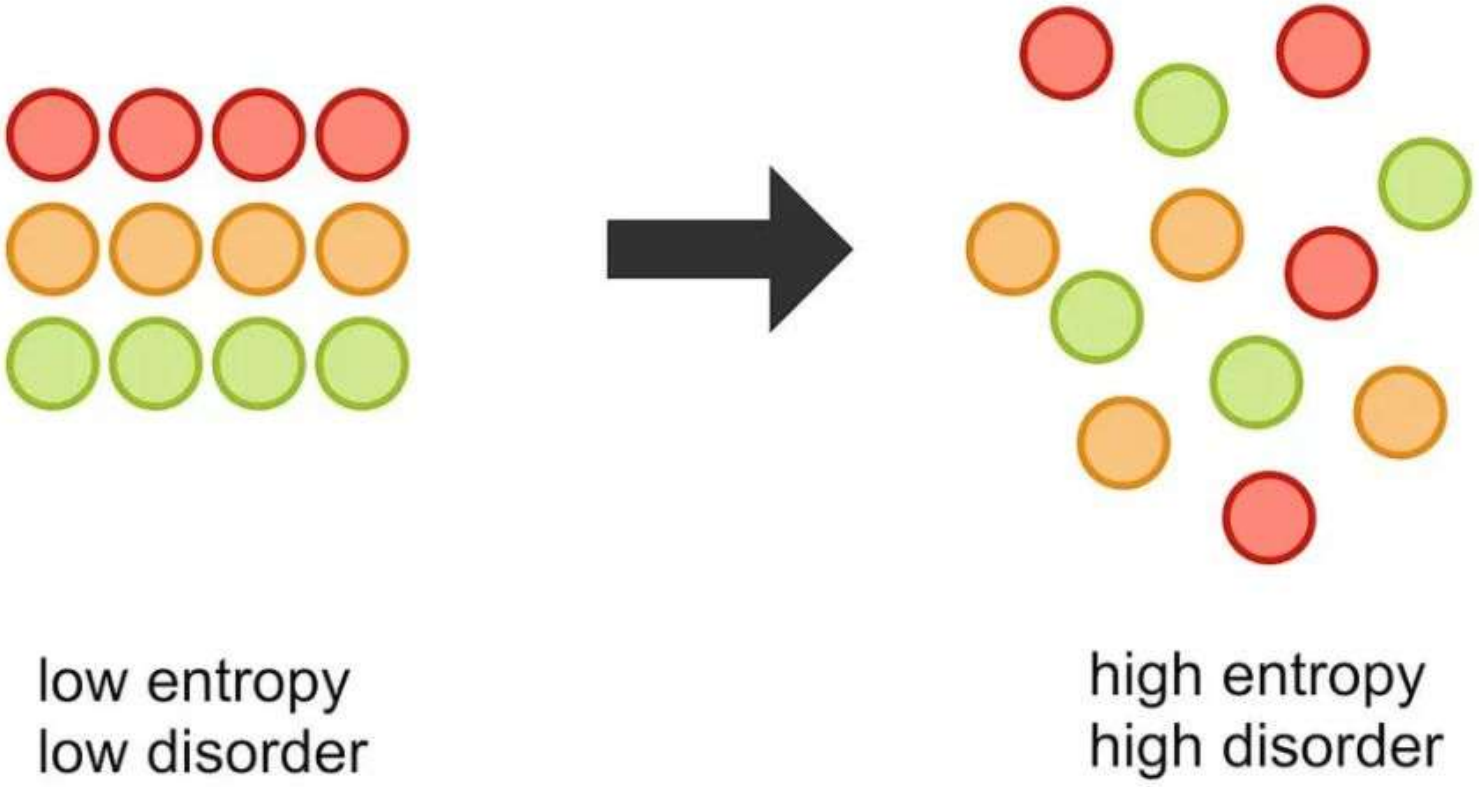
TABLE 1 The major transitions[1]

Replicating molecules to populations of molecules in compartments
Unlinked replicators to chromosomes
RNA as gene and enzyme to DNA and protein (genetic code)
Prokaryotes to eukaryotes
Asexual clones to sexual populations
Protists to animals, plants and fungi (cell differentiation)
Solitary individuals to colonies (non-reproductive castes)
Primate societies to human societies (language)

…the emergence of novel inheritance systems, involving new modes of storing, transmitting, and processing information — new "codes of life"

Szathmáry E., Maynard Smith J. The major evolutionary transitions, *Nature*, 374, 227-232, 1995.
Kun A. The major evolutionary transitions and codes of life, *BioSystems*, 210, 104548, 2021.

Thermodynamic arrow of time



https://medium.com/starts-with-a-bang/did-the-universe-have-zero-entropy-when-it-first-began-2fe43224de46

Thermodynamic arrow of time is defined by the second law of thermodynamics, which states that entropy in a closed system always increases over time, leading to the irreversibility of natural processes

Cosmological arrow of time

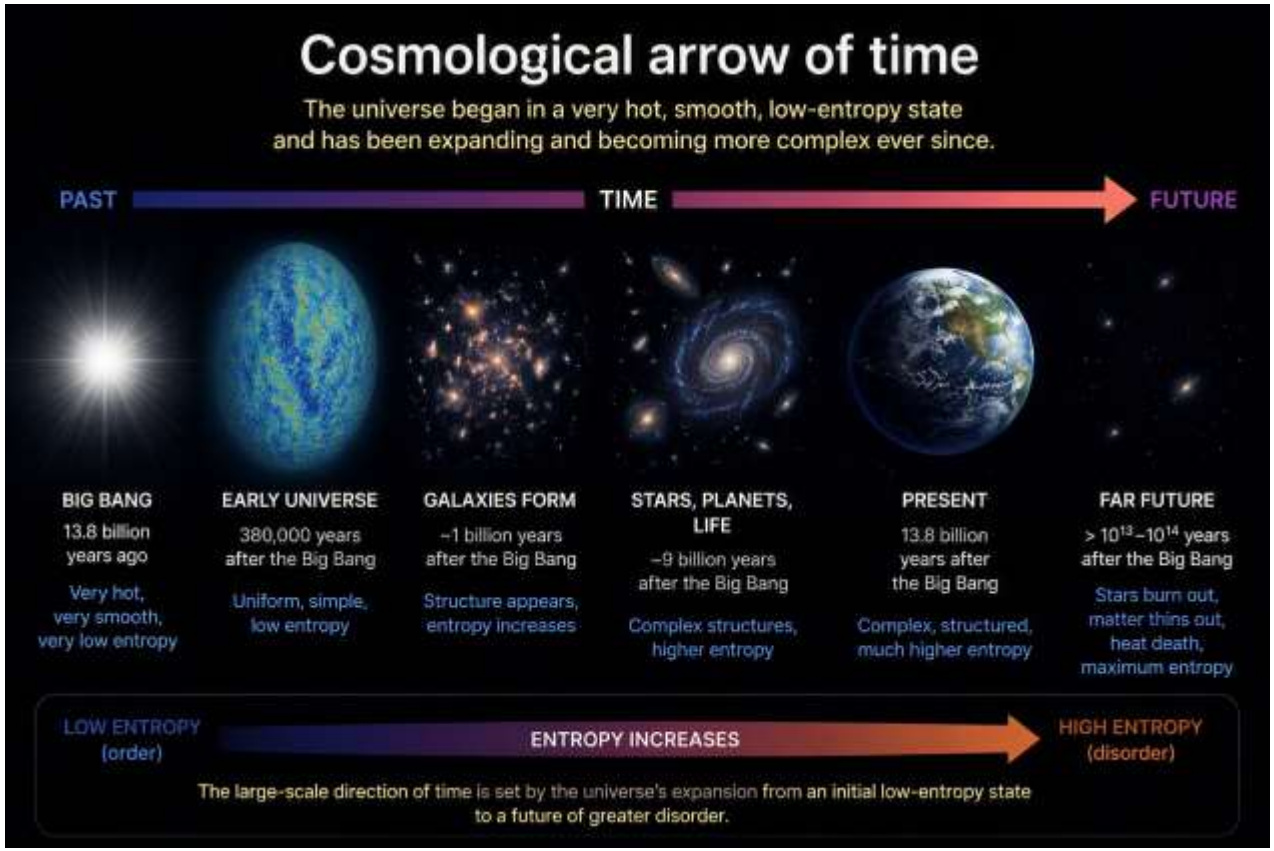
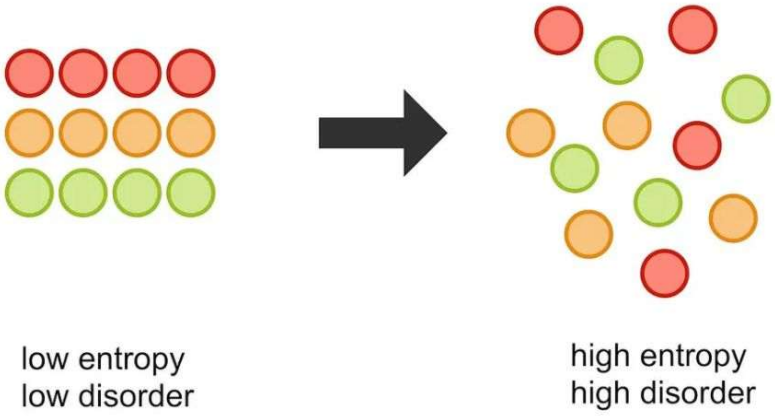


Image generated with Microsoft Copilot (AI-generated image), 23 July 2026

Cosmological arrow of time refers to the direction of time in which the Universe is expanding, and is based on the observation that the Universe is growing larger over time, as opposed to contracting

Is there a biological arrow of time?



low entropy
low disorder

high entropy
high disorder

https://medium.com/starts-with-a-bang/did-the-universe-have-zero-entropy-when-it-first-began-2fe43224de46

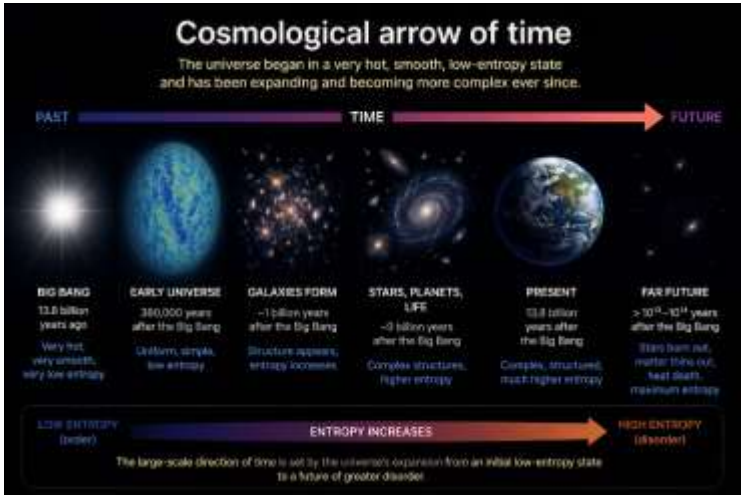
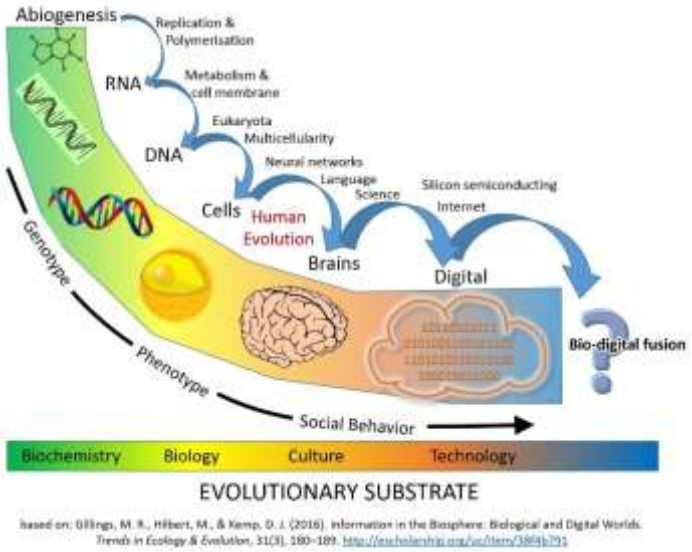
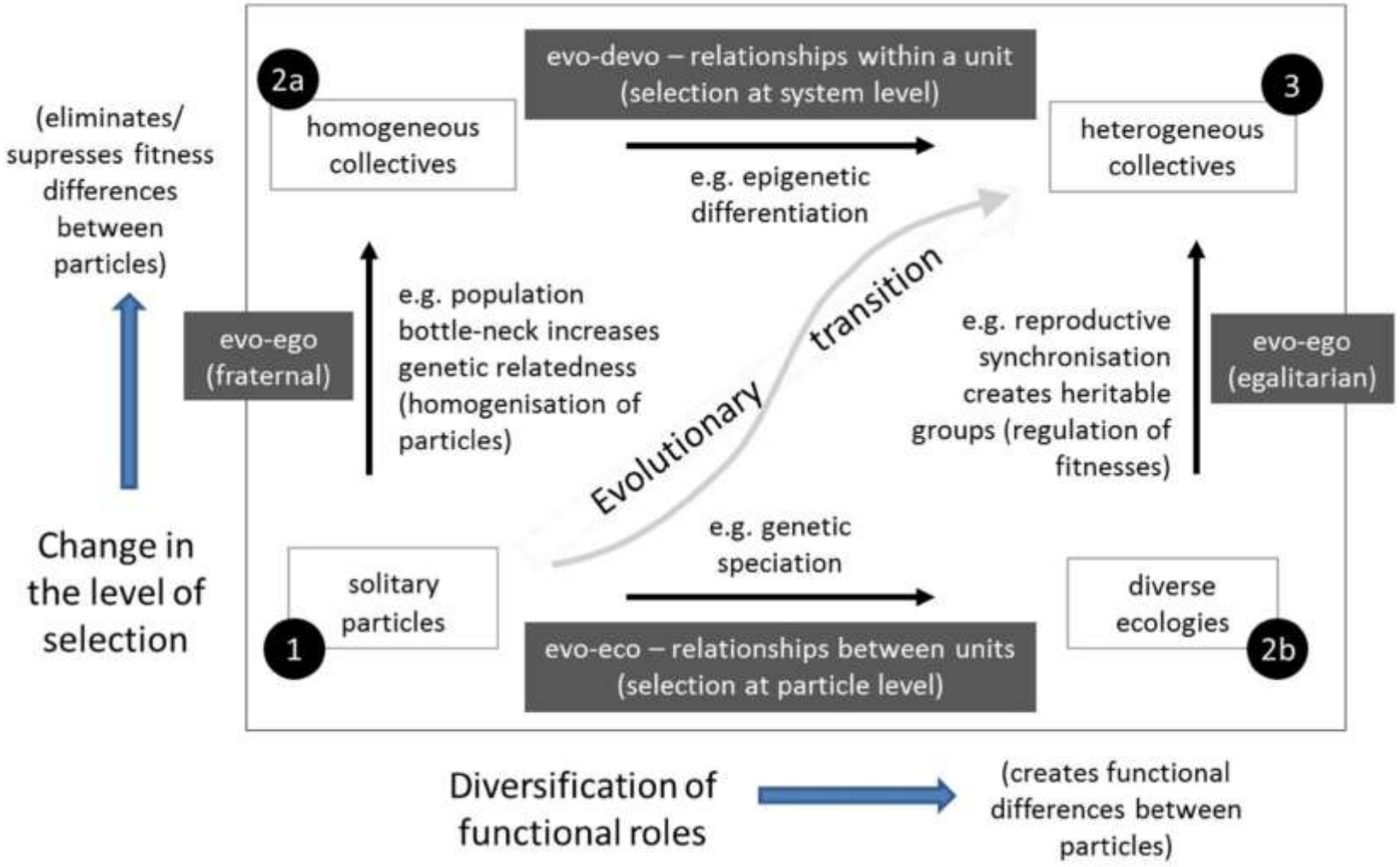


Image generated with Microsoft Copilot (AI-generated image), 23 July 2026



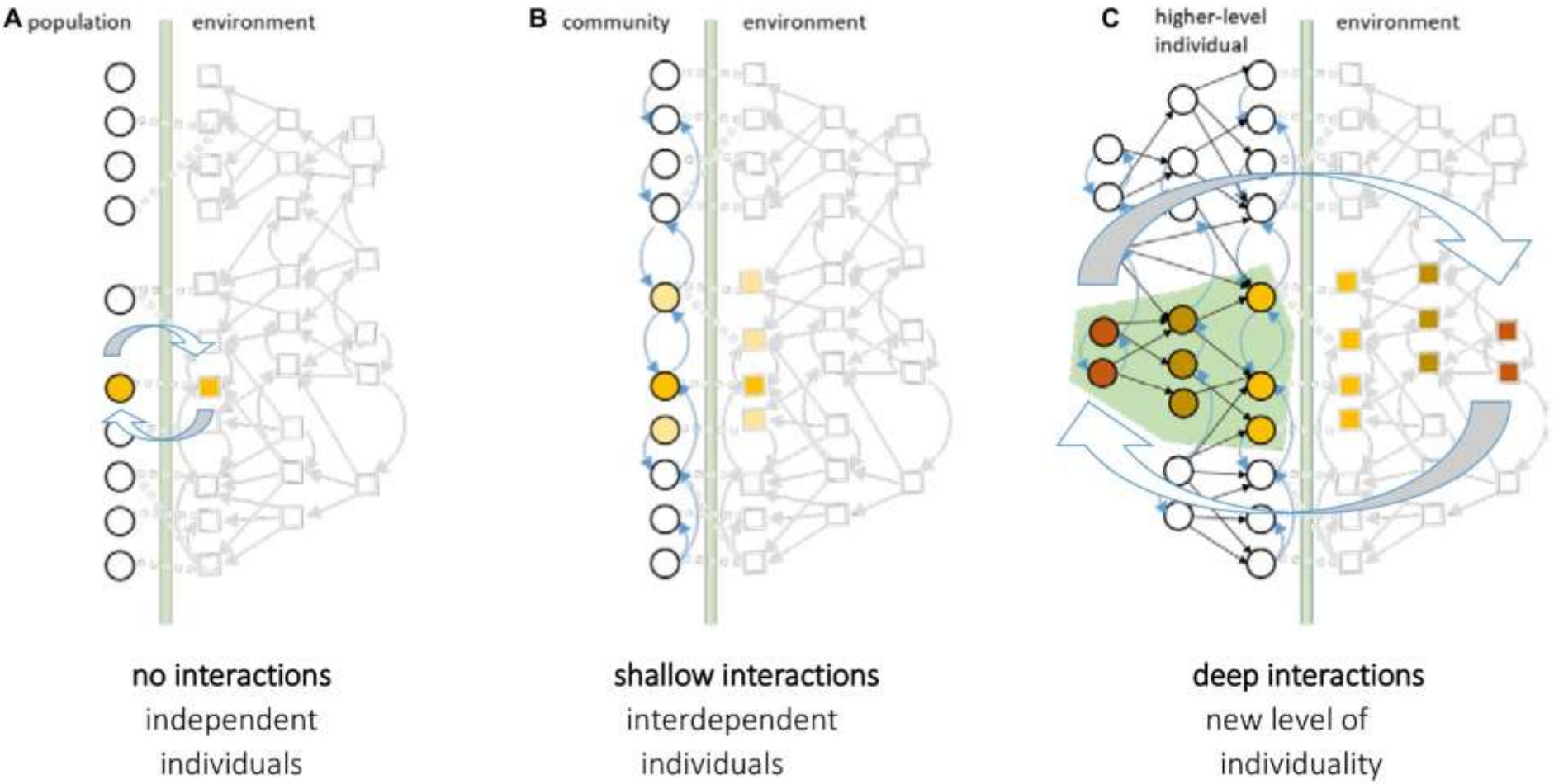
based on: Gillings, M. R., Hilbert, M., & Kemp, D. J. (2016). Information in the Biosphere: Biological and Digital Worlds. *Trends in Ecology & Evolution*, 31(3), 180–189.

Evolutionary transitions (Watson et al.)



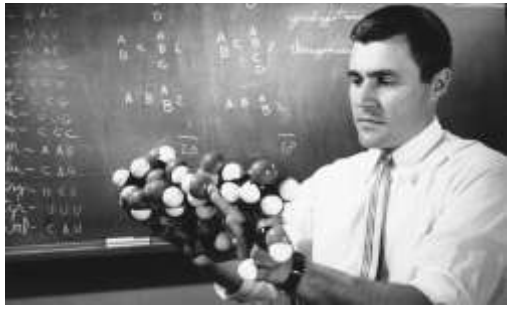
Watson R. A. et al. Evolutionary connectionism: algorithmic principles underlying the evolution of biological organisation in evo-devo, evo-eco and evolutionary transitions, *Evolutionary Biology*, 43, 553 - 581, 2016.

Evolutionary connectionism (Watson et al.)



Watson R. A., Levin M., Buckley C. L. Design for an individual: connectionist approaches to the evolutionary transitions in individuality, *Frontiers in Ecology and Evolution*, 10, 823588, 2022.

Novelty generation and “coding thresholds”



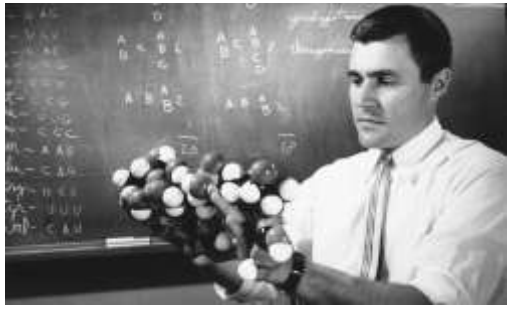
“...the overwhelming amount of novelty needed to bring modern cells into existence ...is the central and most challenging question...”

“Somewhere along the line there had to have occurred a saltation that we could call the “coding threshold,” where the capacity to represent nucleic acid sequence symbolically in terms of a (colinear) amino acid sequence developed, a development that would generate a truly enormous new, totally unique evolutionary phase space.”

Woese, C.R., A new biology for a new century, *Microbiology and Molecular Biology Reviews*, 68(2):173-186, 2004.

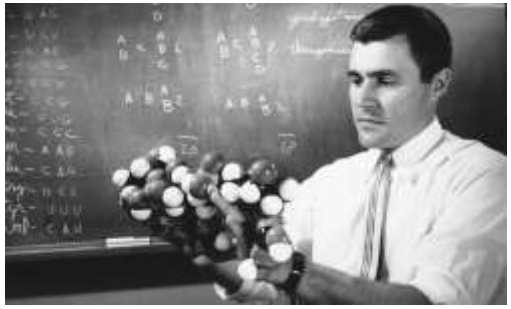
Novelty generation and “coding thresholds”

“… Another good critical-point candidate is the advent of (eucaryotic) multicellularity. Here too the saltation is accompanied by a qualitatively new world of possibilities.”



Woese, C.R., A new biology for a new century, *Microbiology and Molecular Biology Reviews*, 68(2):173-186, 2004.

Novelty generation and "coding thresholds"

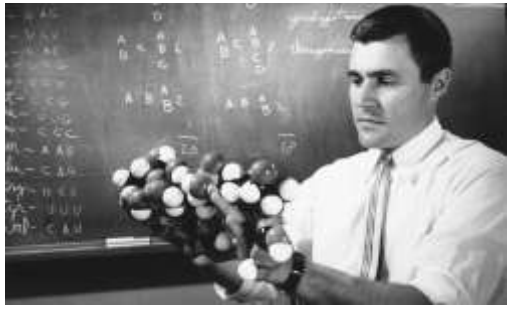


"… Another good critical-point candidate is the advent of (eucaryotic) multicellularity. Here too the saltation is accompanied by a qualitatively new world of possibilities."

"…It is immediately apparent that one of them is the development of language(s). Human language is a development that has set Homo sapiens worlds apart from its otherwise very close primate relatives, adding new dimensions to the phase space within which human evolution occurs."

Woese, C.R., A new biology for a new century, *Microbiology and Molecular Biology Reviews*, 68(2):173-186, 2004.

Novelty generation and "coding thresholds"



"… Another good critical-point candidate is the advent of (eucaryotic) multicellularity. Here too the saltation is accompanied by a qualitatively new world of possibilities."

"…It is immediately apparent that one of them is the development of language(s). Human language is a development that has set Homo sapiens worlds apart from its otherwise very close primate relatives, adding new dimensions to the phase space within which human evolution occurs."

"A common thread that links language and multicellularity is communication (interaction at a distance). In each case a complex, sophisticated network of interactions forms the medium within which the new level of organization (entities) comes into existence."

Woese, C.R., A new biology for a new century, *Microbiology and Molecular Biology Reviews*, 68(2):173-186, 2004.

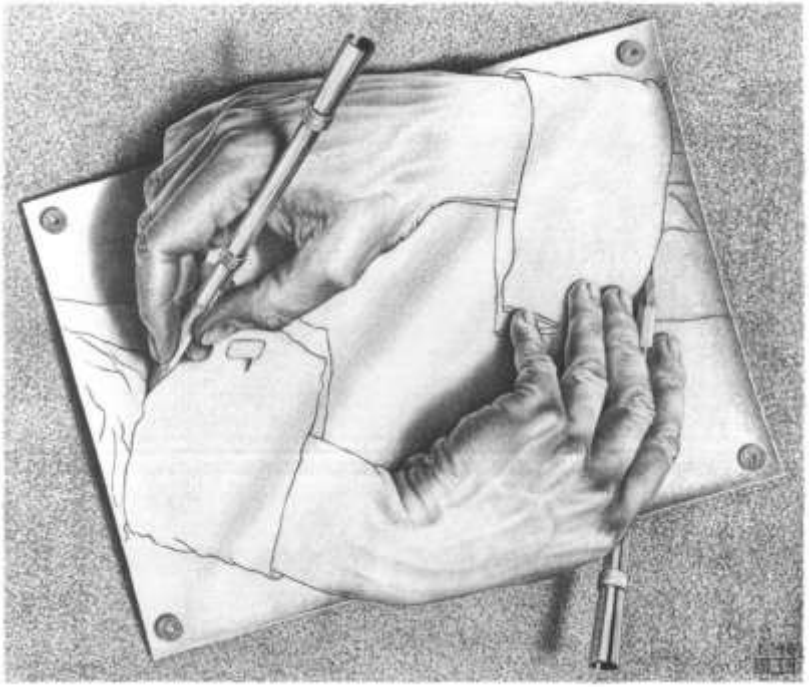
Open ended evolution

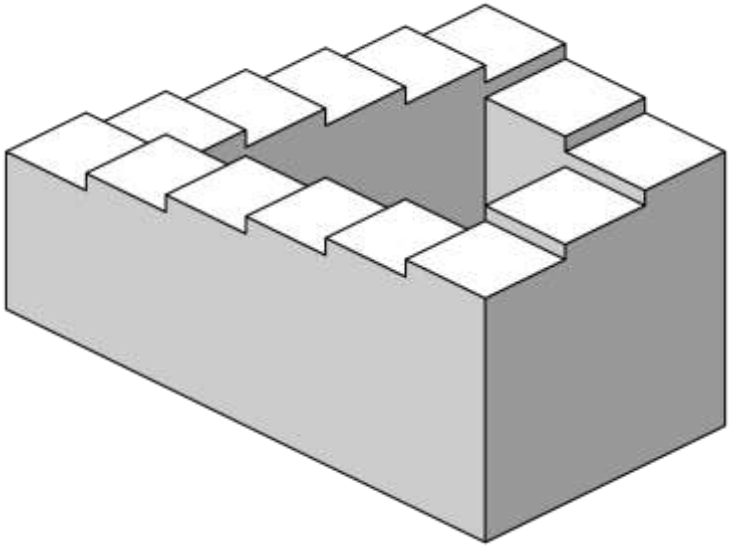
"These rules themselves need to evolve, but how? We need an additional set of rules describing the evolution of the original rules.

But this upper level of rules itself needs to evolve. Thus, we end up with an infinite hierarchy, an inevitable reflection of the fact that the dynamic we are seeking is inherently self-referential."

Goldenfeld N., Woese C. Life is physics: evolution as a collective phenomenon far from equilibrium, *Annual Review of Condensed Matter Physics*, 2, 375-399, 2011.

Tangled hierarchies and strange loops





Oscar Reutersvärd **(1937)** /
Lionel Penrose & Roger Penrose **(1958)**

Drawing Hands:

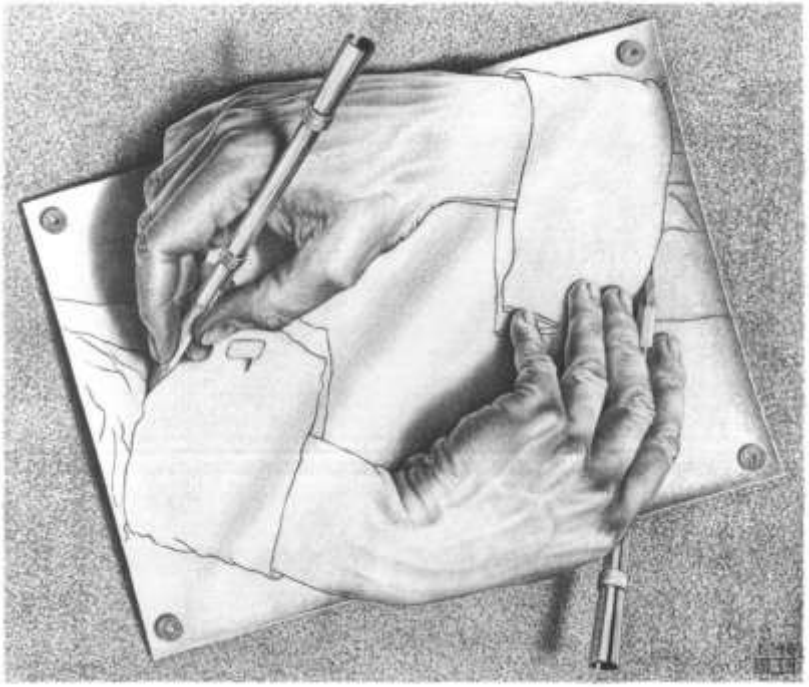
the product, the hands, are undertaking the operation – the drawing of the hands

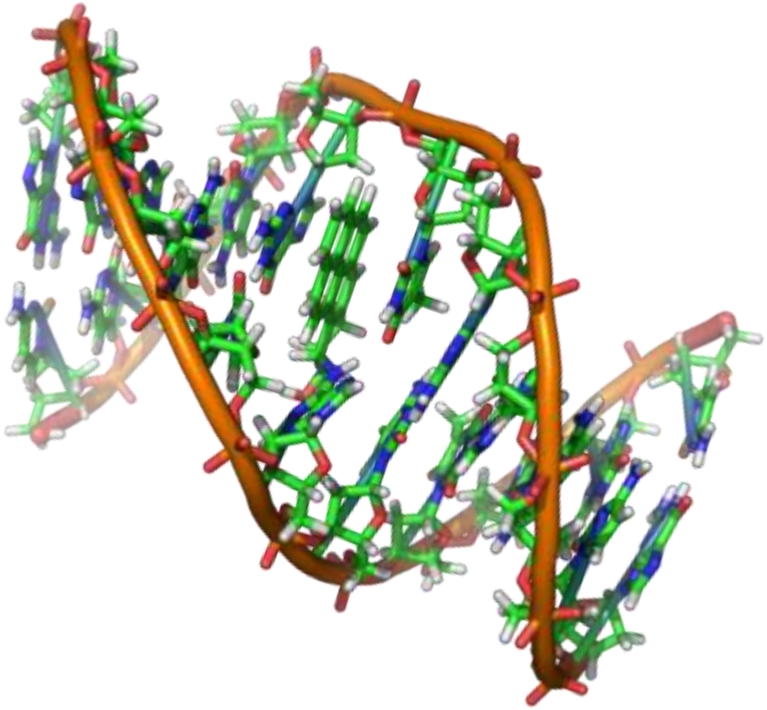
Maurits Cornelis Escher **(1948)**

"an interaction between levels in which the top level reaches back down towards the bottom level and influences it, while at the same time being itself determined by the bottom level"

— Hofstadter **(1989)**

Self and a functional description of “self”





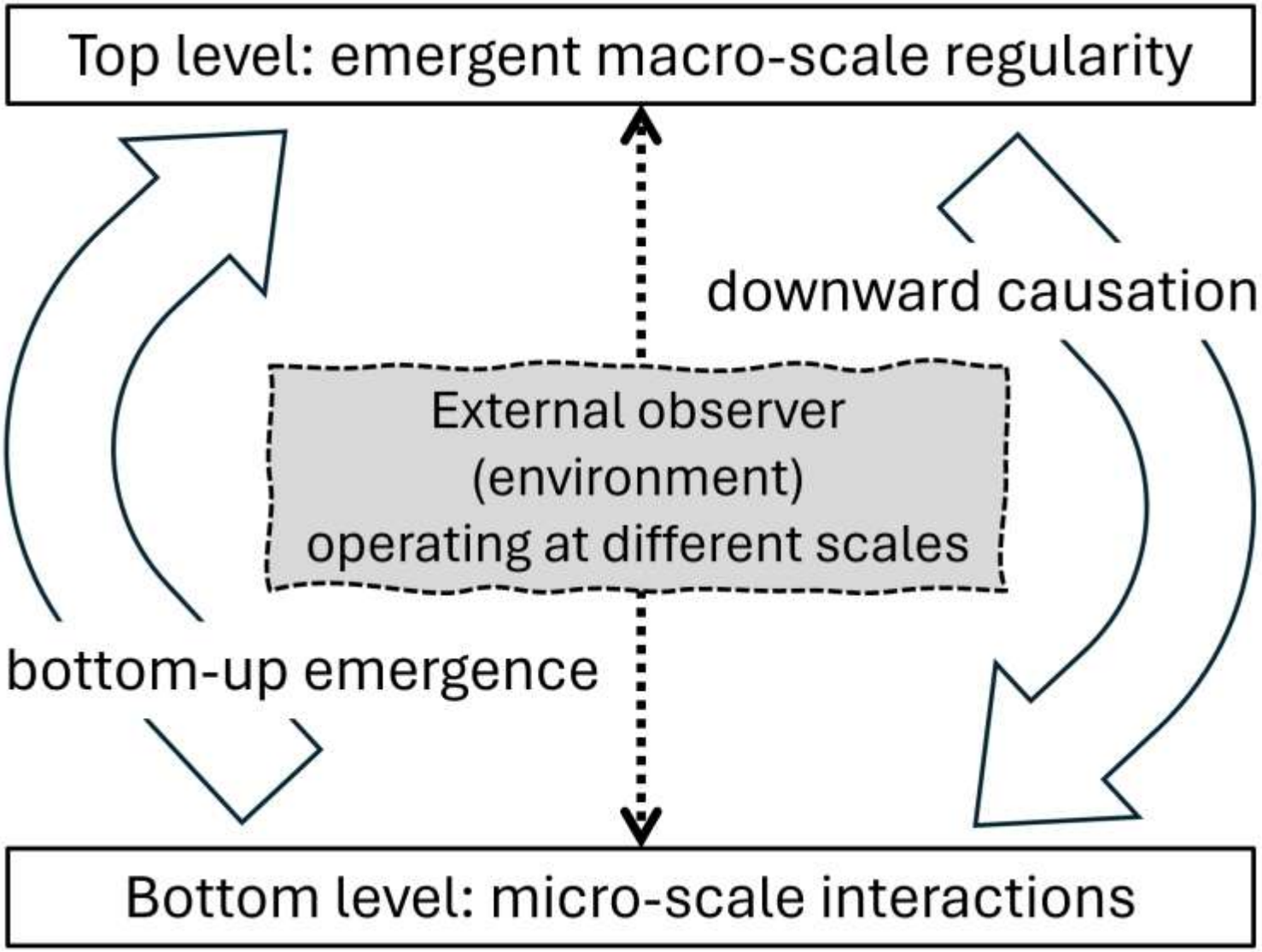
Drawing Hands:
the product, the hands, are undertaking the operation – the drawing of the hands

DNA: genetic instructions (*sequence*) used in development and functioning of a living organism (*function and structure*) – needed to construct other components of cells, and copy itself

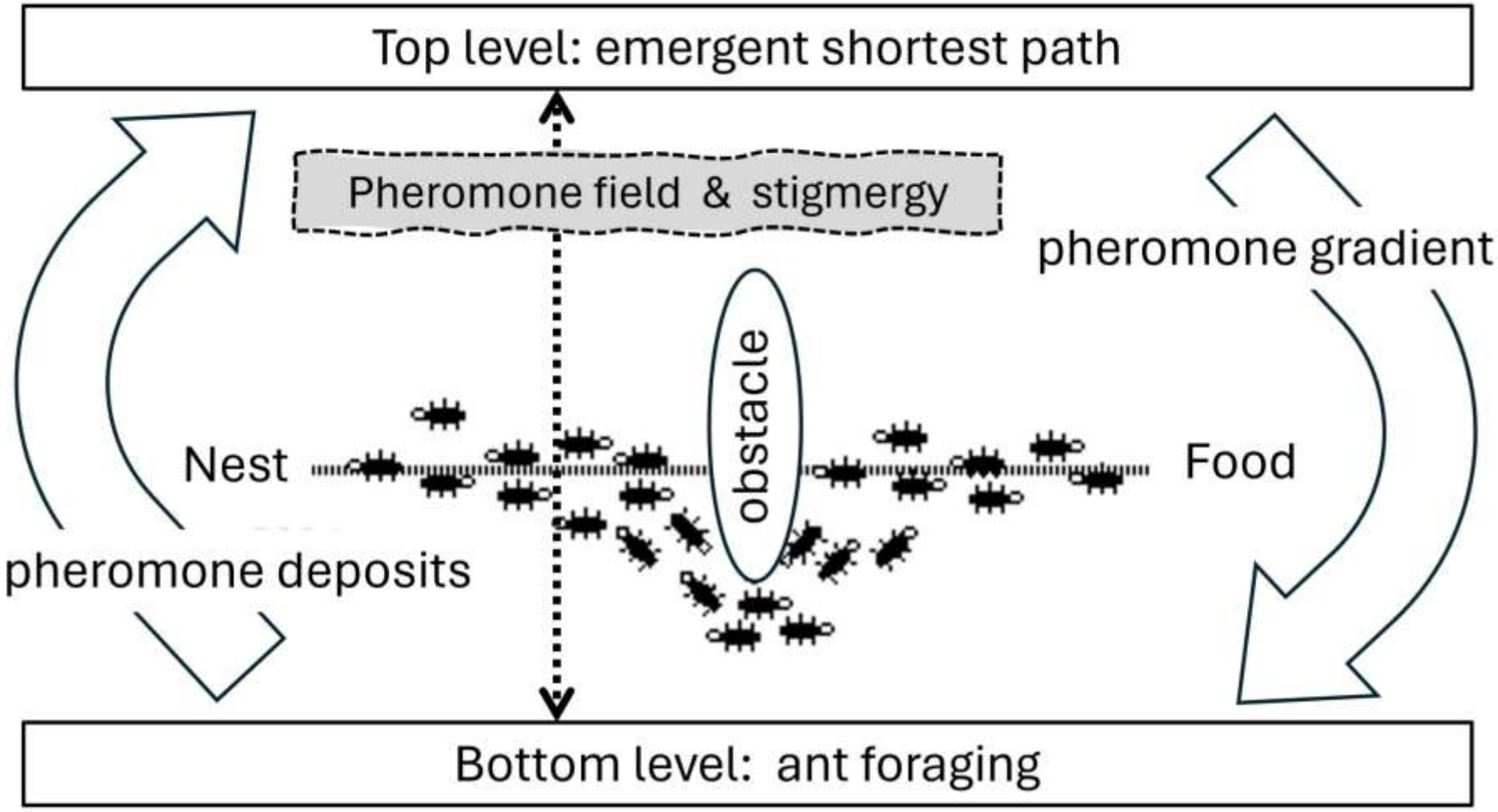
M. C. Escher, Drawing Hands, Wikipedia

Wikipedia

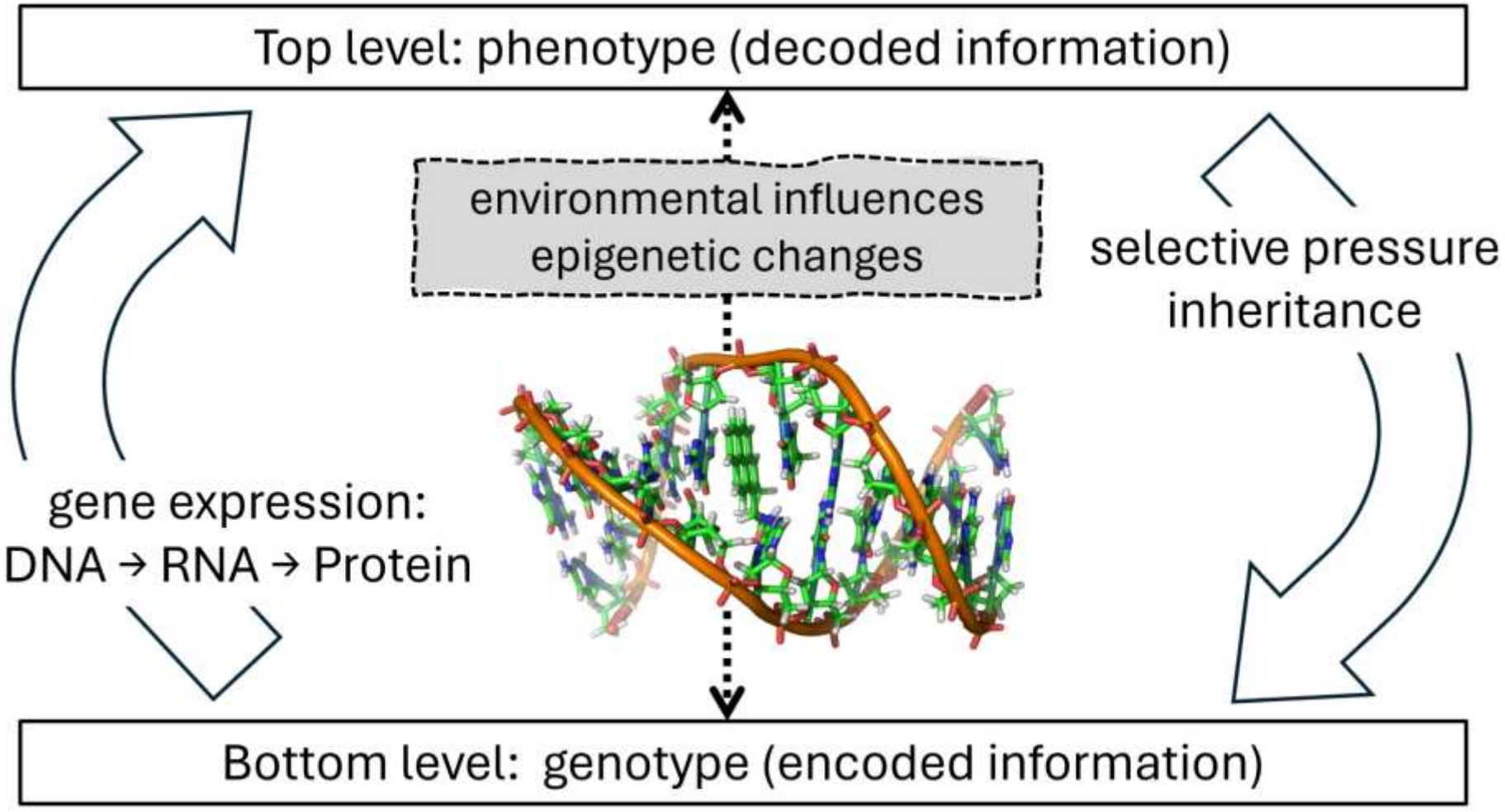
Tangled hierarchy



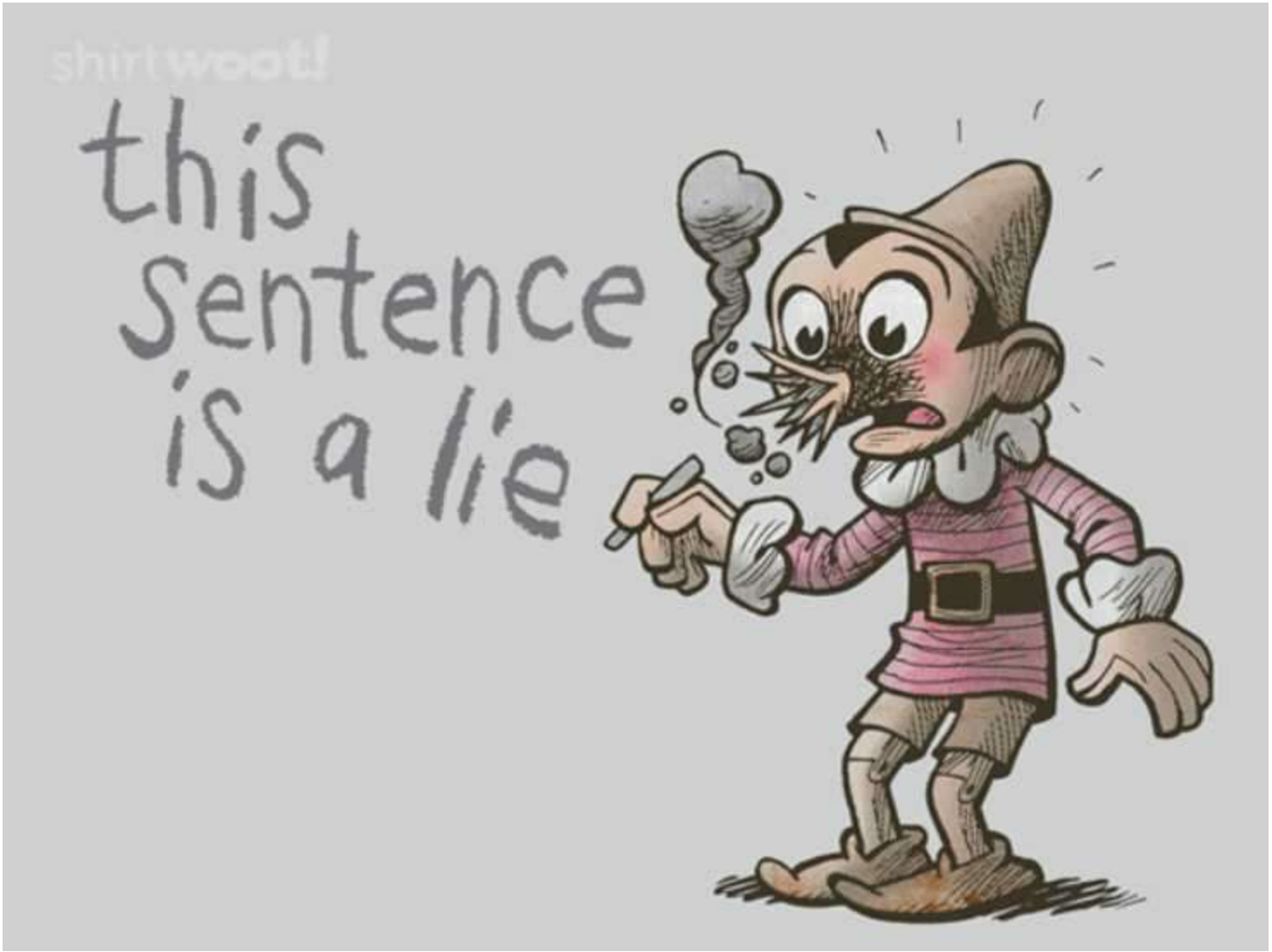
Tangled hierarchy: type I



Tangled hierarchy with self-modelling: type II

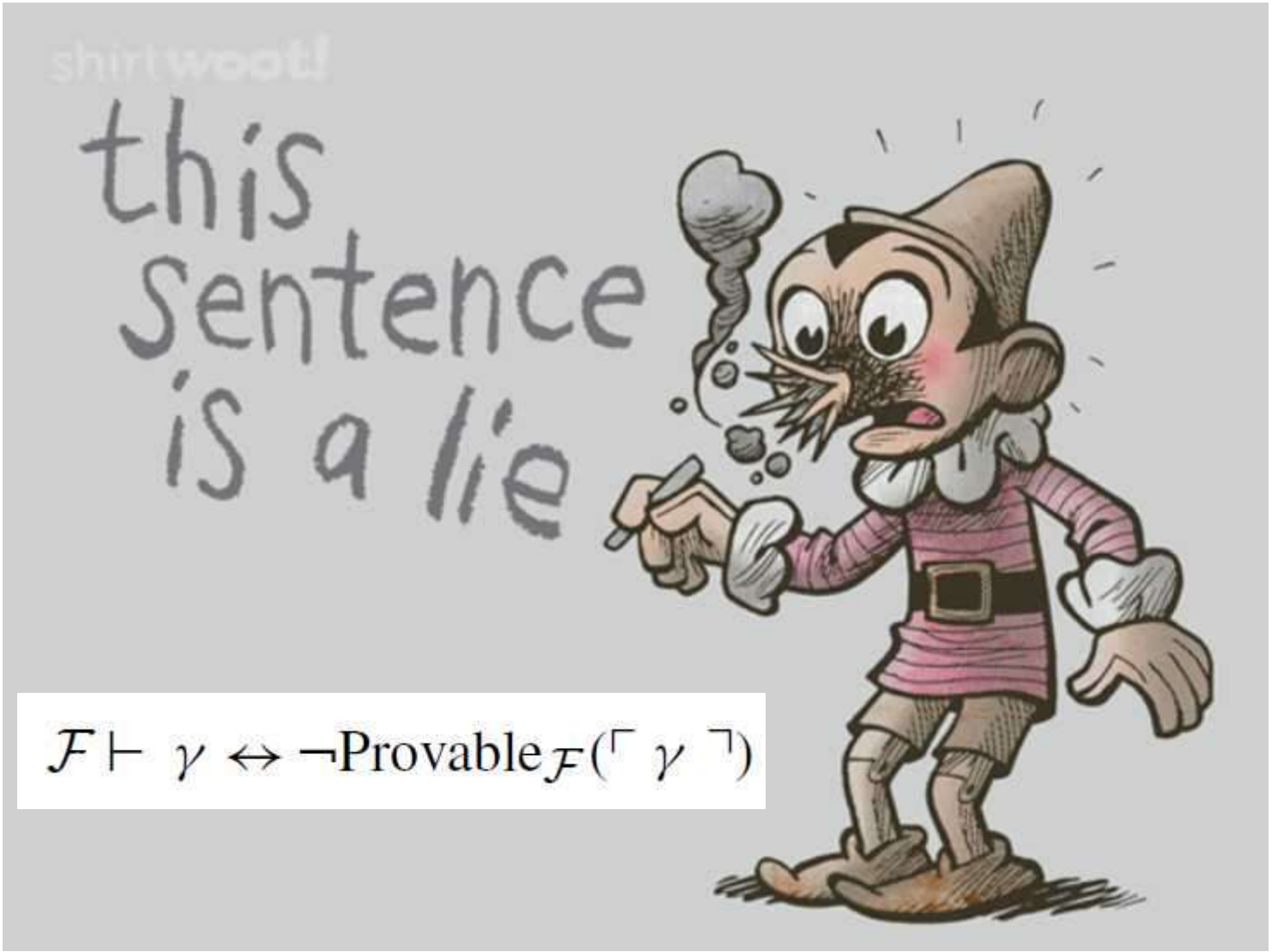


Self-reference and the Liar paradox



https://shirt.woot.com/offers/liars-paradox

The Liar paradox and Gödel sentence



https://shirt.woot.com/offers/liars-paradox

Gödel: incompleteness theorems

First Incompleteness *Theorem*: *"Any consistent formal system F within which a certain amount of elementary arithmetic can be carried out is incomplete; i.e., there are statements of the language of F which can neither be proved nor disproved in F."* **(1931)**

Second Incompleteness Theorem: *"For any consistent system F within which a certain amount of elementary arithmetic can be carried out, the consistency of F cannot be proved in F itself."* **(1931)**

Wikipedia

Incomputability and diagonalisation arguments

referents

expressions

	$\ulcorner E_0 \urcorner$	$\ulcorner E_1 \urcorner$	$\ulcorner E_2 \urcorner$	$\cdots$	$\ulcorner E_k \urcorner$	$\cdots$
E_0	P_{00}	P_{01}	P_{02}	$\cdots$	P_{0k}	$\cdots$
E_1	P_{10}	P_{11}	P_{12}	$\cdots$	P_{1k}	$\cdots$
E_2	P_{20}	P_{21}	P_{22}	$\cdots$	P_{2k}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
E_k	P_{k0}	P_{k1}	P_{k2}	$\cdots$	P_{kk}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

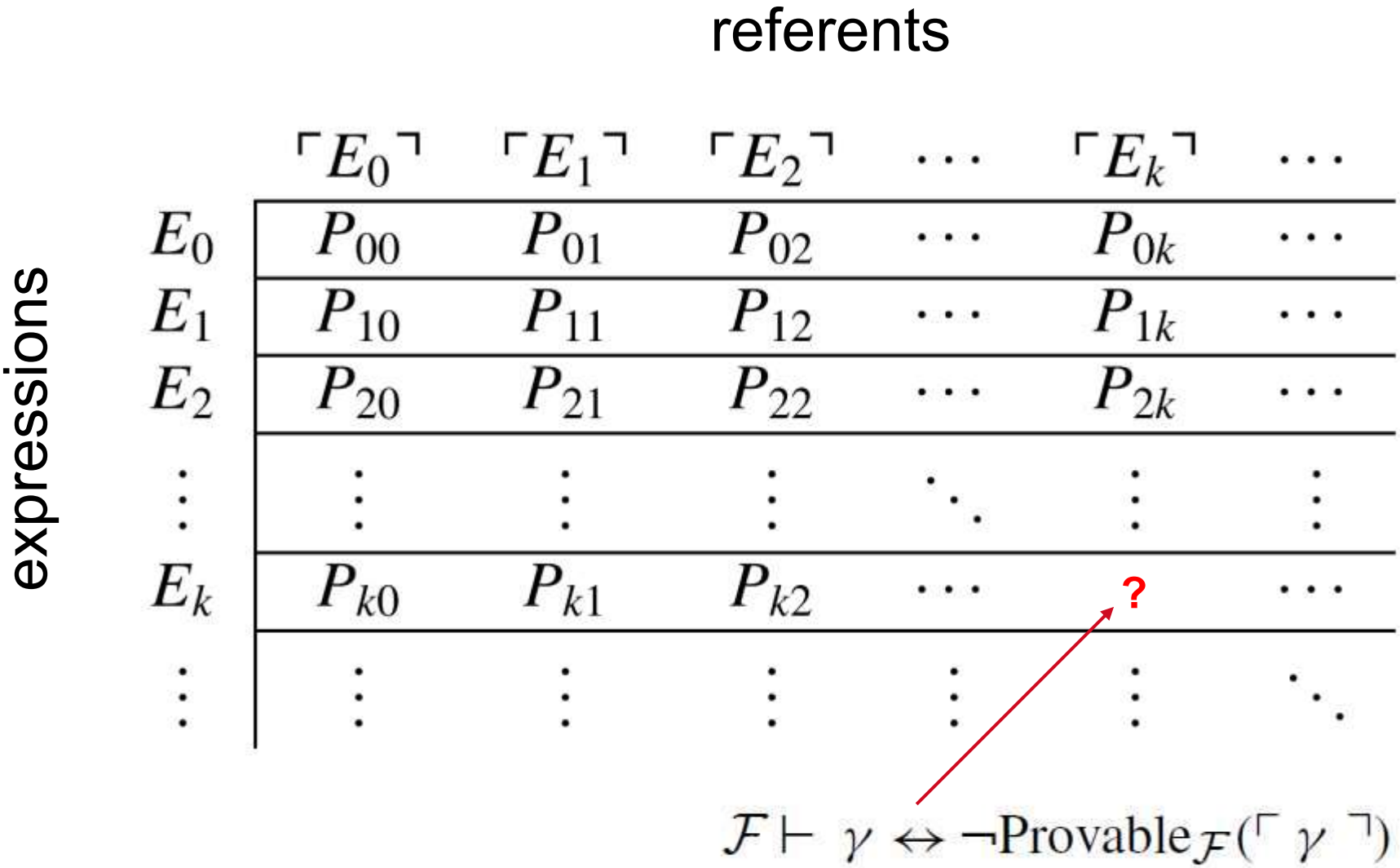
Incomputability and diagonalisation arguments

referents

expressions

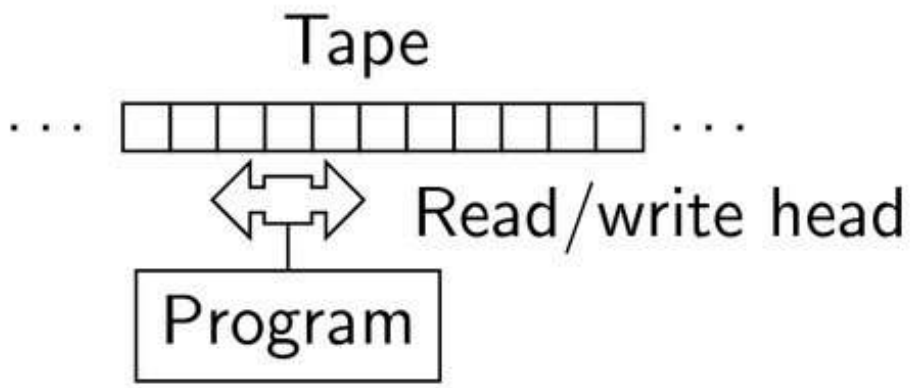
	$\ulcorner E_0 \urcorner$	$\ulcorner E_1 \urcorner$	$\ulcorner E_2 \urcorner$	$\cdots$	$\ulcorner E_k \urcorner$	$\cdots$
E_0	P_{00}	P_{01}	P_{02}	$\cdots$	P_{0k}	$\cdots$
E_1	P_{10}	P_{11}	P_{12}	$\cdots$	P_{1k}	$\cdots$
E_2	P_{20}	P_{21}	P_{22}	$\cdots$	P_{2k}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
E_k	P_{k0}	P_{k1}	P_{k2}	$\cdots$	?	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Incomputability and diagonalisation arguments



Turing Machine

Turing machines
alphabets: input Σ and tape Γ
tape strings in Σ^*
admissible syntax, given Γ, e.g., blank symbol
recognizable tape pattern in Σ^*, given Γ
infinite tape



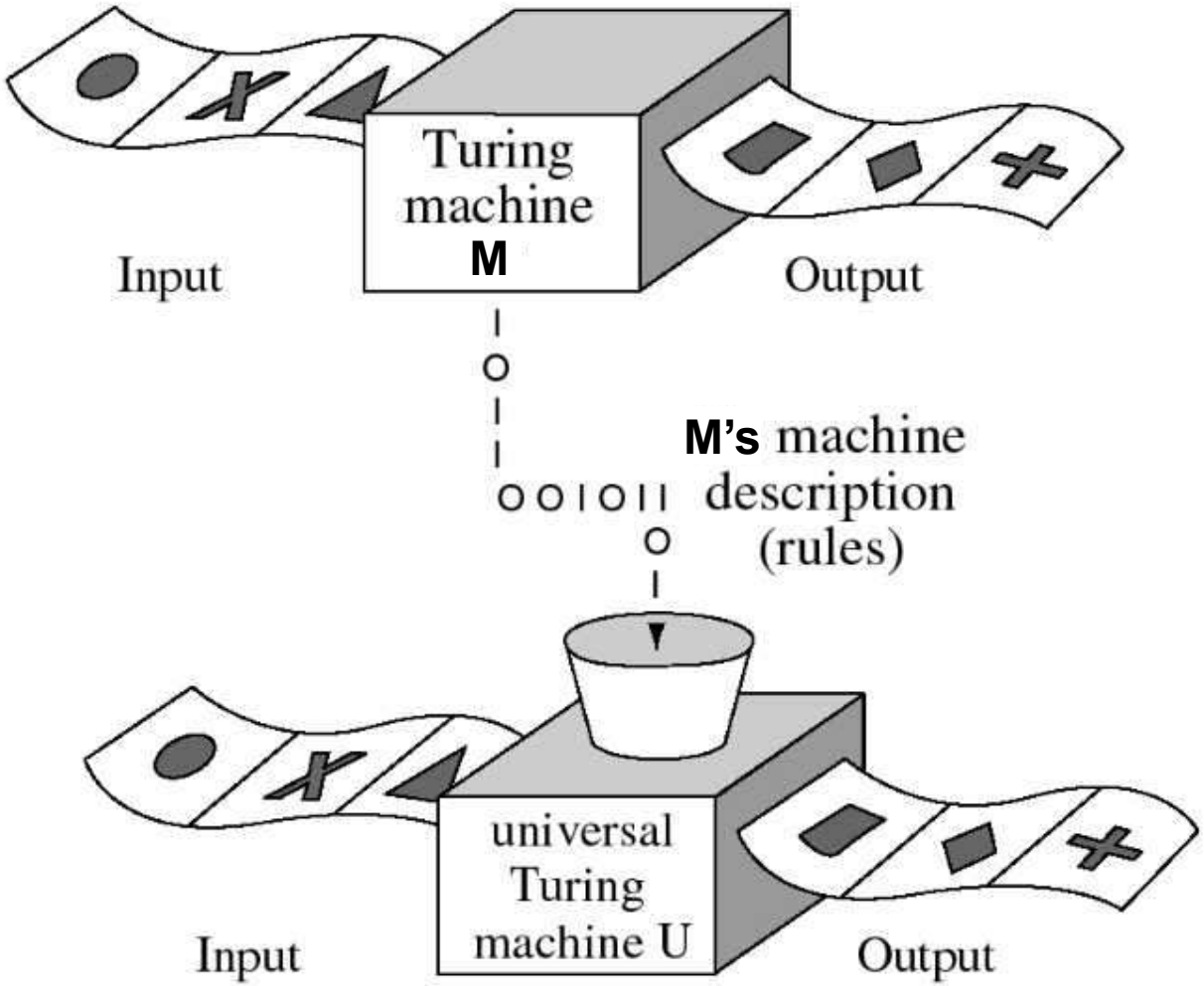
This "universal computing machine" would be capable of performing any conceivable mathematical computation if it were representable as an algorithm. BUT: It is not possible to decide algorithmically whether a Turing machine will ever halt **(1937).**

This paper has been called "easily the most influential math paper in history"

— Avi Wigderson **(2019)**

Ludger Humbert / TEXample / CC BY 2.5; http://www.texample.net/tikz/examples/turing-machine/

Universal Turing Machine: can simulate any other machine



Glenn D. Blank et al., 2003

http://www.cse.lehigh.edu/~glennb/um/1intro.pdf

Universal Decider Turing Machine: does not exist!

$$P([M,w]) = \begin{cases} accept & \text{if } M \text{ accepts } w \\ reject & \text{if } M \text{ does not accept } w \end{cases}$$

Universal Decider

$$V([M]) = \begin{cases} reject & \text{if } M \text{ accepts } [M] \\ accept & \text{if } M \text{ does not accept } [M] \end{cases}$$

The Liar

$$V([V]) = \begin{cases} reject & \text{if } V \text{ accepts } [V] \\ accept & \text{if } V \text{ does not accept } [V] \end{cases}$$

Self-referential Liar

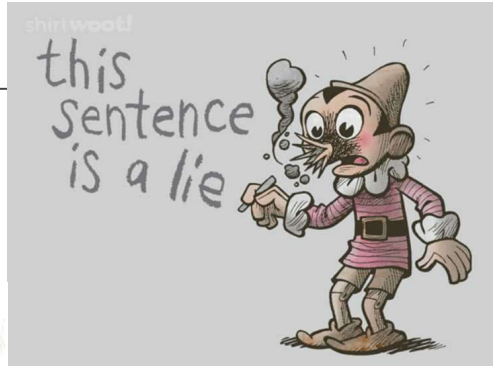
Self-reference and diagonalisation

inverter of universal decider

Table 5

The cell (k, j) is the outcome of running $V = M_k$ (the inverter of P) on $[M_k[M_j]]$. A contradiction occurs at cell (k, k).

	$[M_1]$	$[M_2]$	$[M_3]$	$\cdots$	$[M_k]$	$\cdots$
M_1	accept	reject	accept		accept	
M_2	accept	accept	accept	$\cdots$	reject	$\cdots$
M_3	reject	accept	reject		reject	
$\vdots$		$\vdots$		$\ddots$		
M_k	reject	reject	accept		?	
$\vdots$		$\vdots$				



$$\mathcal{F} \vdash \gamma \leftrightarrow \neg \text{Provable}_{\mathcal{F}}(\ulcorner \gamma \urcorner)$$

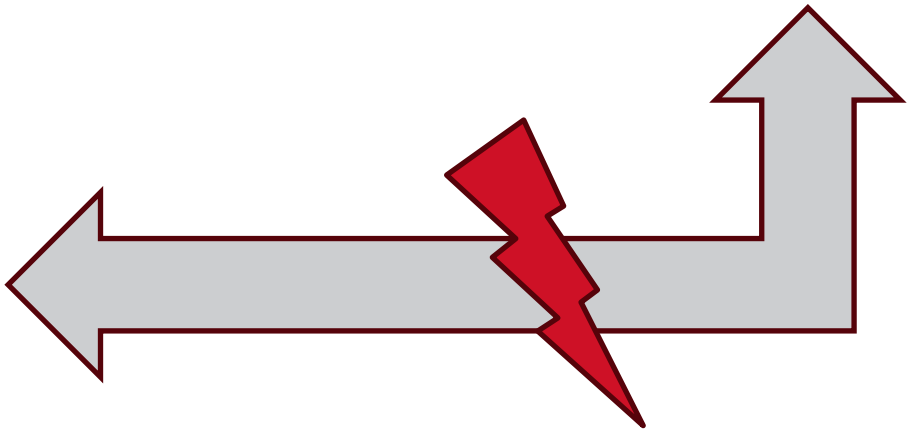
Expression – referent discrepancy

referents

$\ulcorner E_0 \urcorner \quad \ulcorner E_1 \urcorner \quad \ulcorner E_2 \urcorner \quad \ldots \quad \ulcorner E_k \urcorner \quad \ldots$

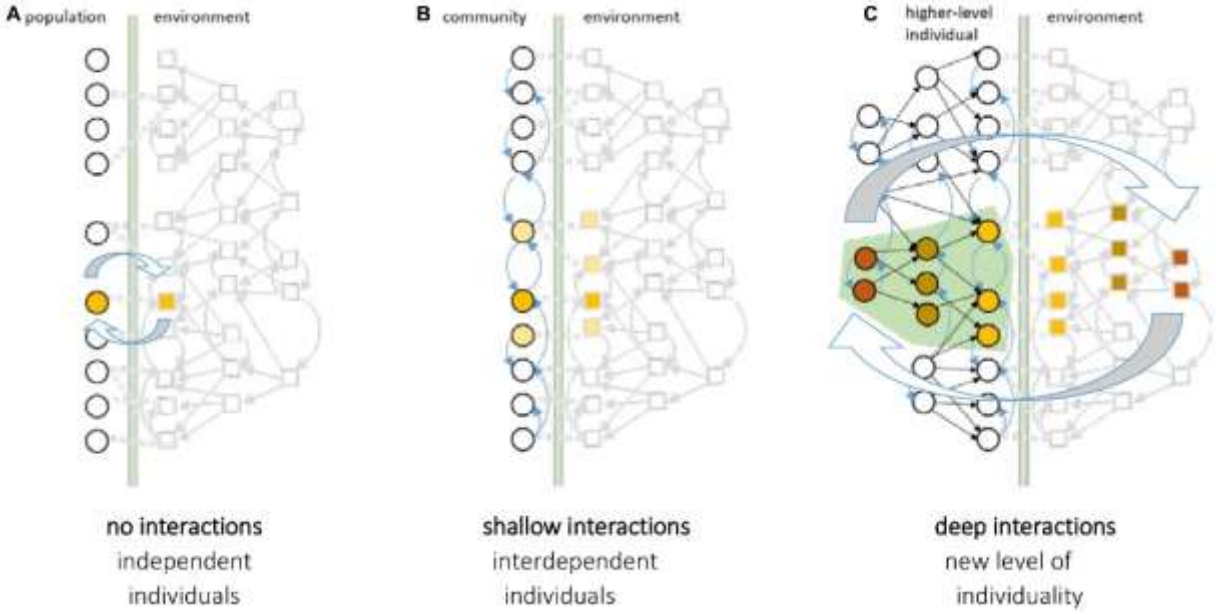
expressions

E_0
E_1
E_2
$\vdots$
E_k
$\vdots$



there are always more expressions (objects) than referents (encodings)

Evolutionary connectionism (Watson & Szathmáry)



$$y = F(x_1, x_2, \ldots, x_n)$$

$$\neq$$

$$F = \sum_i f_i(x_i)$$

- evolution is formally equivalent to unsupervised deep learning
- when evolved functions are non-linearly separable, the hierarchical structures constitute new individuals
- higher-level individuals enable information integration and collective actions
- when evolved interaction structures between units compute a non-linearly separable function, this makes collective fitness, and hence reproduction, non-decomposable in a formal sense
- not all collective interests are reducible to a (sum of) individual preferences
- evolutionary tension between individual and group interests is an “inconsistency”, directly related to the expression-referent discrepancy

Watson R. A., Levin M., Buckley C. L. Design for an individual: connectionist approaches to the evolutionary transitions in individuality, *Frontiers in Ecology and Evolution*, 10, 823588, 2022.

“Jumping out of the system…” (Hofstadter: JOOTSing)

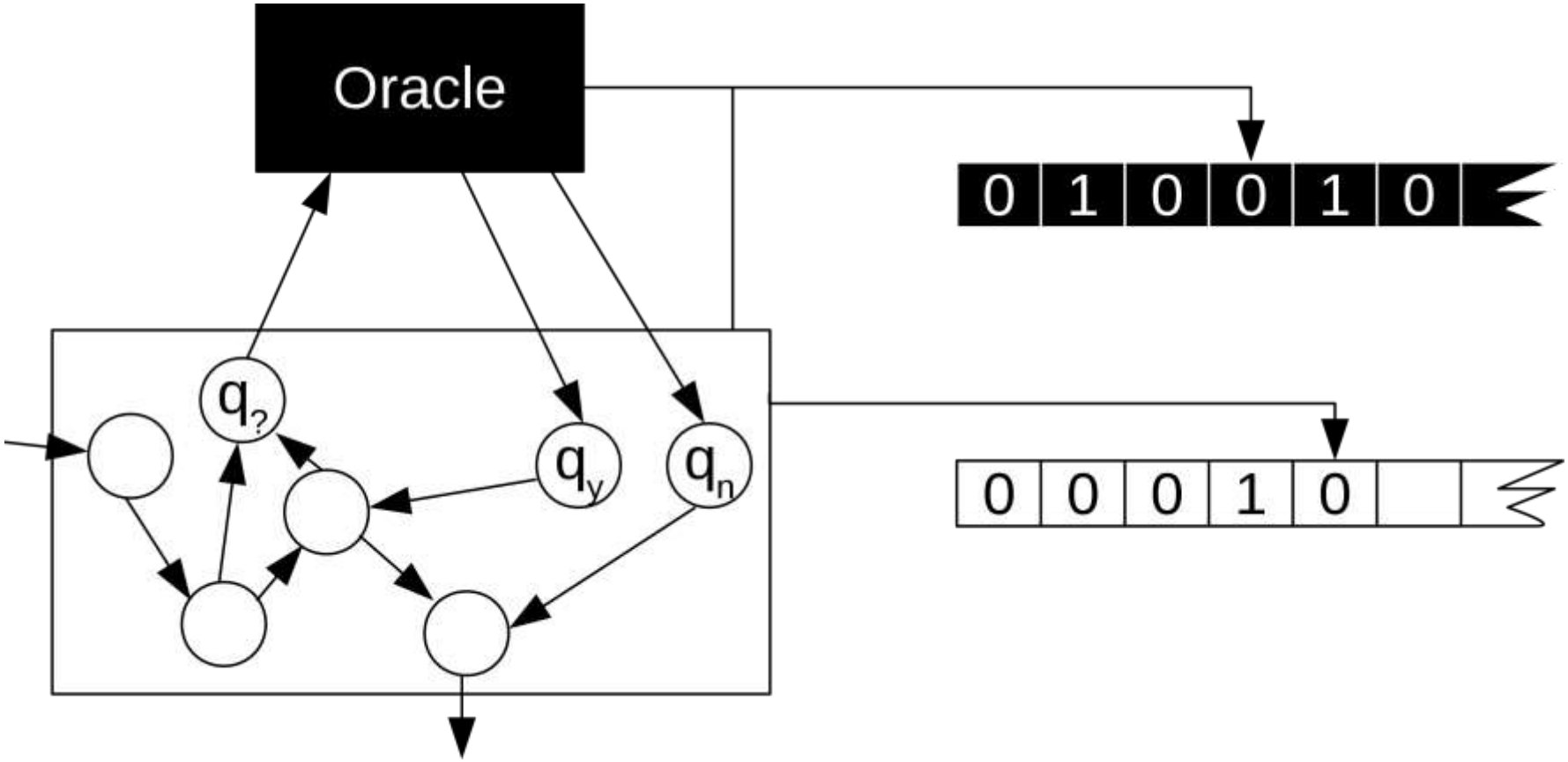
“… we must ask if there exists a larger meta-system, encompassing the computational undecidability, which may play the role of universal decider and inverter… Importantly, an undecidable problem is framed within a given formal system, and once the system is appropriately extended, the problem in point becomes decidable — at the cost of generating other undecidable problems that inevitably arise in the extended system.” [3]

[1] Hofstadter, D. R. *Gödel, Escher, Bach: An eternal golden braid*, 1980.

[2] Soare, R. I., Turing oracle machines, online computing, and three displacements in computability theory, *Annals of Pure and Applied Logic*, 160: 368-399, 2009.

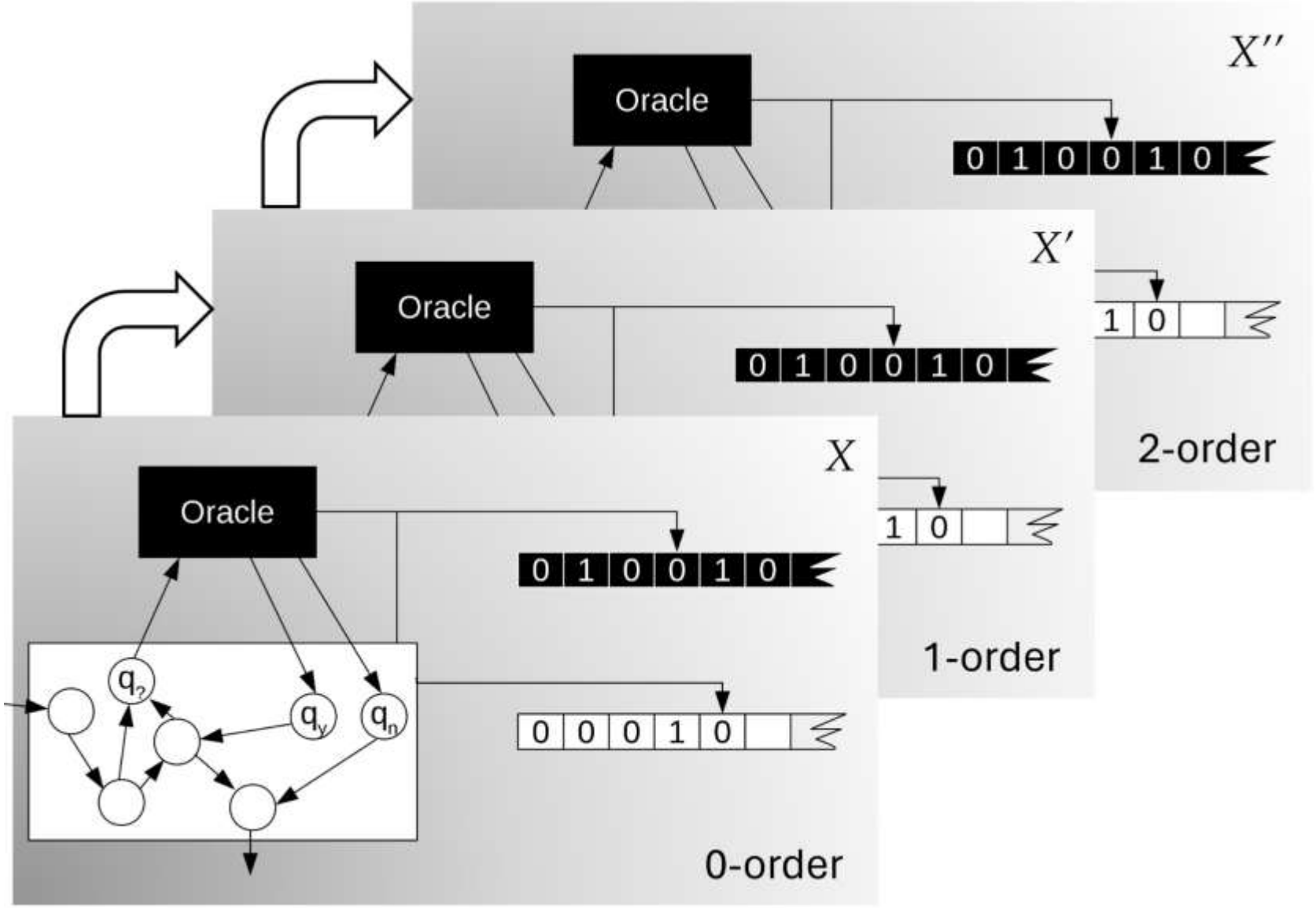
[3] Svahn, A. J., Prokopenko, M., An ansatz for computational undecidability in RNA automata, *Artificial Life*, 29(2): 261-288, 2023.

Oracle machine (Turing)



Wikipedia

α −order Oracle machines and meta-simulation



Conclusion: biological arrow of time as open-ended meta-simulation

- Emergence of tangled hierarchy without self-modelling (TH type I), with distributed coordination and imprecise replication
- Encapsulation of compressed functional self-description, via encoding (compression) of macro-level information patterns
- Formation of tangled hierarchy with self-modelling (TH type II), with encoding and decoding of compressed patterns
  - Self-reference: expressions becoming referents and vice versa
  - Undecidability, due to the expression-referent discrepancy: there are always more expressions (objects) than referents (encodings)
- Extension of the problem-space (JOOTSing), accessing “extended patterns of information” and higher computational power
- Irreversible transitions because reversals would reduce evolutionary advantage / computational power attained by the aggregates

THE UNIVERSITY OF SYDNEY

References

- Goldenfeld N., Woese C. Life is physics: evolution as a collective phenomenon far from equilibrium, *Annual Review of Condensed Matter Physics*, 2, 375-399, 2011.
- Hofstadter, D. R. *Gödel, Escher, Bach: An eternal golden braid*, Penguin Press,1980.
- Kun A. The major evolutionary transitions and codes of life, *BioSystems*, 210, 104548, 2021.
- M. Prokopenko, P. C. W. Davies, M. Harré, M. Heisler, Z. Kuncic, G. Lewis, O. Livson, J. Lizier, F. Rosas, Biological arrow of time: Emergence of tangled information hierarchies and self-modelling dynamics, *Journal of Physics: Complexity*, 6, 015006, 2025.
- Prokopenko M., Harré M., Lizier J., Boschetti F., Peppas P., Kauffman S., Self-referential basis of undecidable dynamics: from The Liar Paradox and the halting Problem to the edge of chaos. *Physics of Life Reviews*, 31: 134-156, 2019.
- Szathmáry E., Maynard Smith J. The major evolutionary transitions, *Nature*, 374, 227-232, 1995.
- Sipser M. *Introduction to the theory of computation*, 1st edition. Thomson Publishing: International, 1996.
- Soare, R. I., Turing oracle machines, online computing, and three displacements in computability theory, *Annals of Pure and Applied Logic*, 160: 368-399, 2009.
- Svahn, A. J., Prokopenko, M., An ansatz for computational undecidability in RNA automata, *Artificial Life*, 29(2): 261-288, 2023.
- Watson R. A. et al. Evolutionary connectionism: algorithmic principles underlying the evolution of biological organisation in evo-devo, evo-eco and evolutionary transitions, *Evolutionary Biology*, 43, 553 - 581, 2016.
- Watson R. A., Levin M., Buckley C. L. Design for an individual: connectionist approaches to the evolutionary transitions in individuality, *Frontiers in Ecology and Evolution*, 10, 823588, 2022.
- Watson R. A., Szathmáry E. How can evolution learn? *Trends in Ecology and Evolution*, 31, 147 - 157, 2016.
- Woese, C.R., A new biology for a new century, *Microbiology and Molecular Biology Reviews*, 68(2):173–186, 2004.