CareBots

Solving the problem of meaning in AI

Tom Froese

euCogII 2009

The failure of embodied AI

- [P]erhaps we have all missed some *organizing principle of biological systems*, or some general truth about them.

- Perhaps there is a way of looking at biological systems which will illuminate an inherent necessity in some aspect of the interactions of their parts that is completely missing from our artificial systems.

- I am suggesting that perhaps at this point we simply do not *get it*, and that there is some fundamental change necessary in our thinking.

(Brooks 1997; cf. Brooks 2001)

The failure of embodied AI

- The “big remaining problem” is how to incorporate into embodied AI an account of how we “directly pick up significance and improve our sensitivity to relevance”.

- This ability “depends on our responding to what is significant for *us*” given the current contextual background.
  (Dreyfus 2007)

- Perhaps the enactive paradigm can help? (Froese 2007)

Survival: The mother of all values

- “A robot’s [world] is quite devoid of significance in that there is no sense other than the figurative in which we can say that a robot *cares* about what it is doing” (Di Paolo 2003).

- Why is the sensorimotor approach of embodied AI not the final answer to the problem of meaning?

- “A feedback mechanism may be going, or may be at rest: in either state the machine exists. The organism has to keep going, because to be going is its very existence – which is revocable – and, threatened with extinction, it is *concerned* in existing”.

  Jonas (1966, p. 126):

Kant's "Natural Purposes"

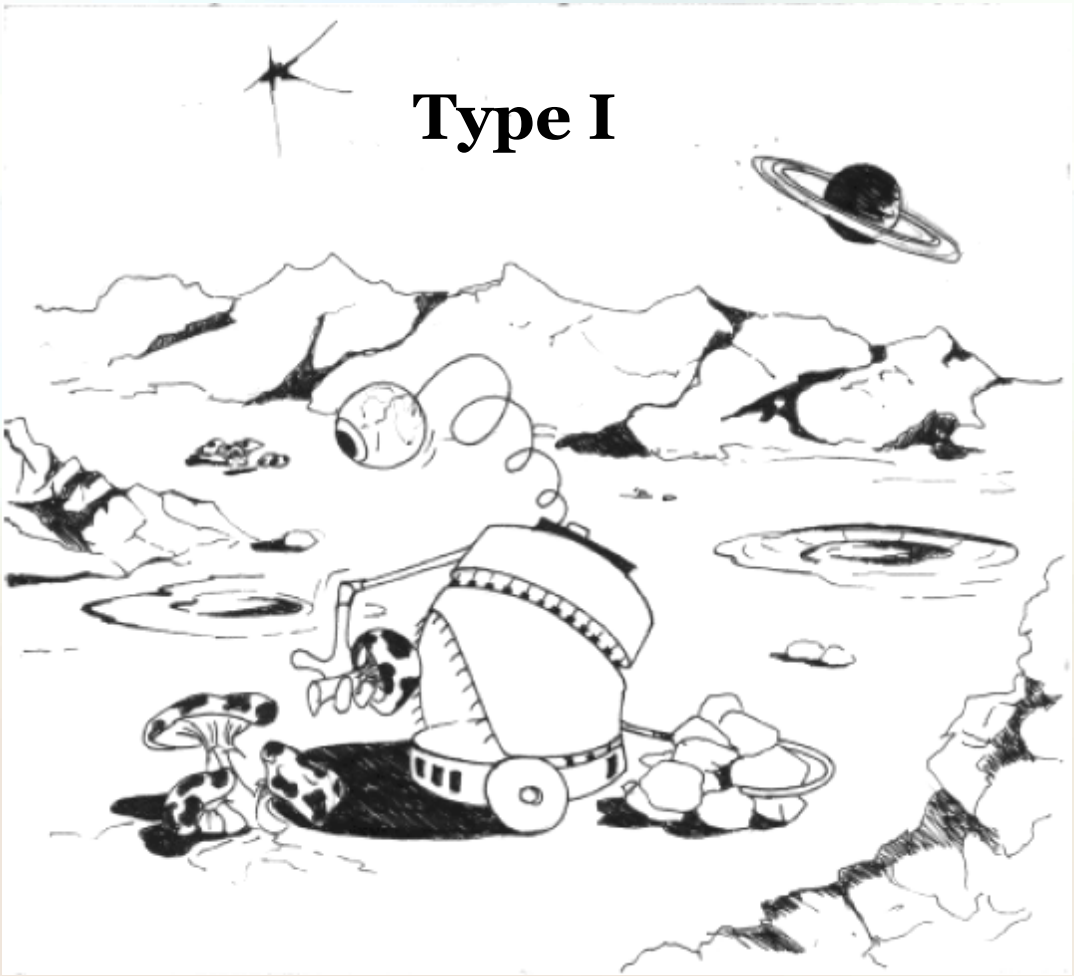
• For Kant a thing exists as a *natural purpose* if it is both cause and effect of itself.

• Only organisms seem to fulfil this criterion: the parts of an organism are there because of the existence of the whole, and the whole is there because of its parts.

• Kant noticed that living systems thus have a peculiar organization that furnishes them with an *intrinsic teleology* (i.e. a self-generated goal-directedness).

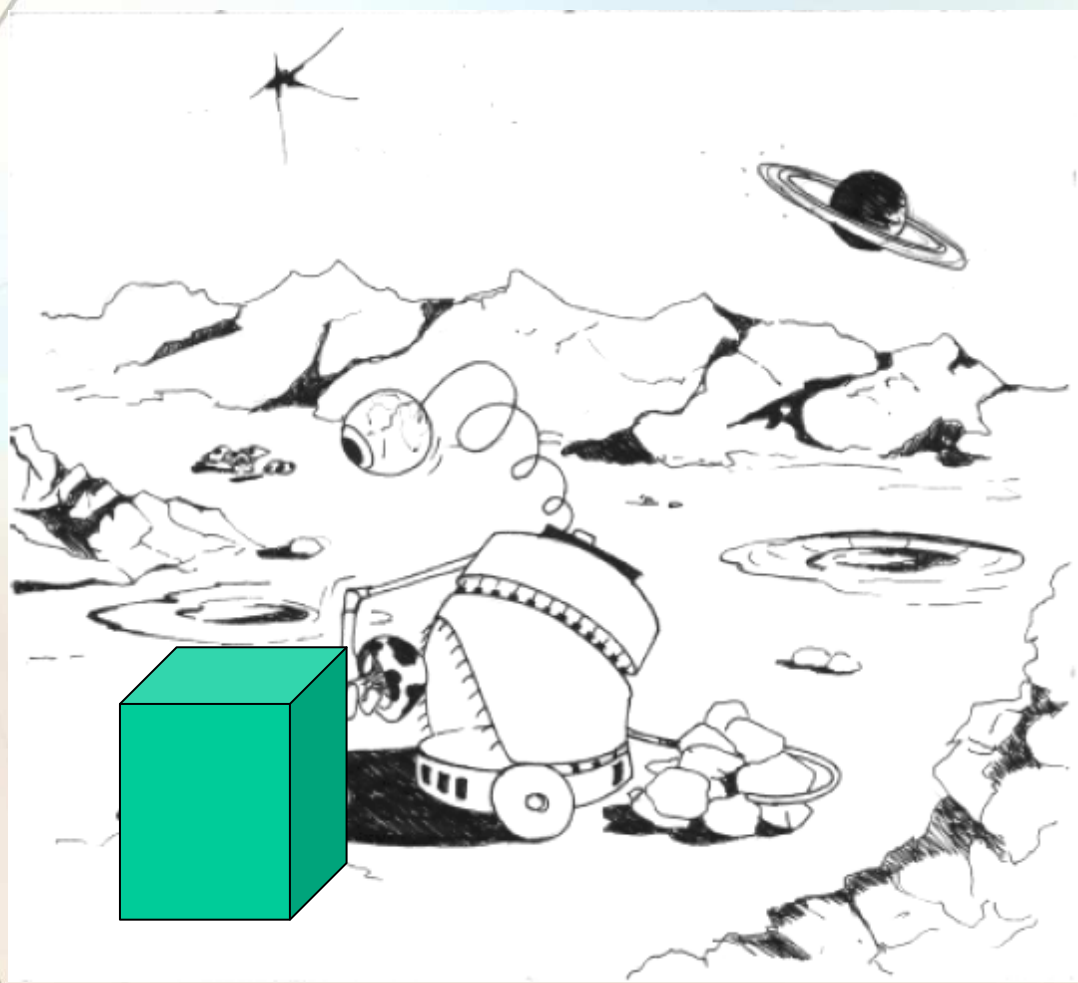
Return of the ‘Fungus Eaters’!

A typography of AI - by Froese and Ziemke (2009)



Pfeifer (1996)

Type Ia – Good Old-Fashioned AI



Pre-determined world

Pre-determined behavior

Pre-determined goals

Pre-determined identity

==

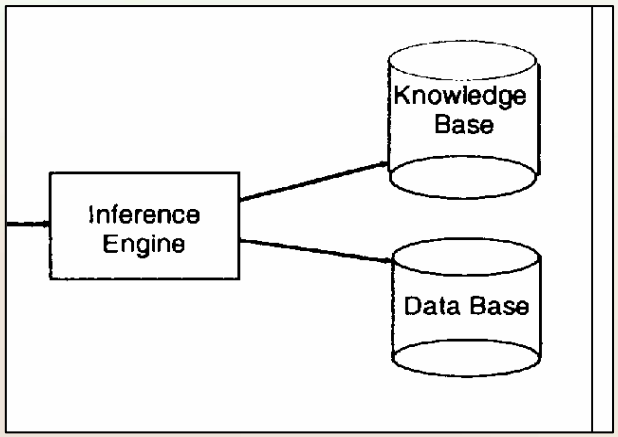
Problem of meaning &

Inter frame problem &

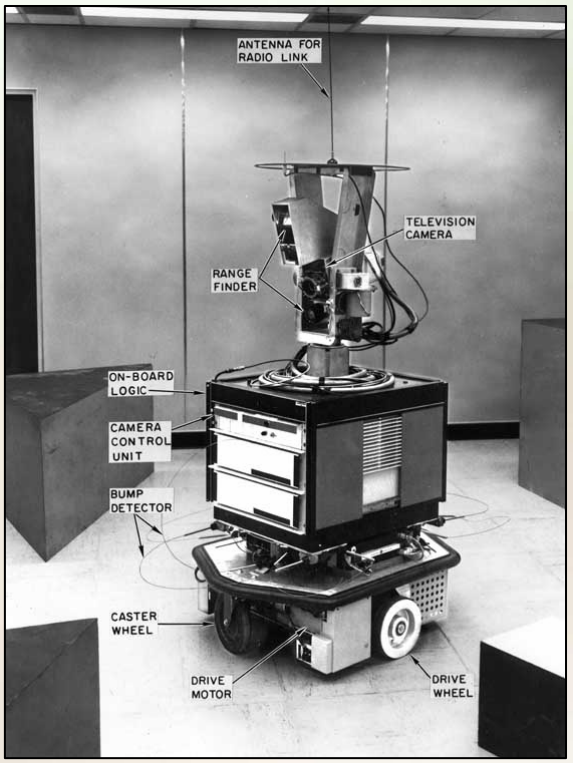
Intra frame problem!

(Wheeler 2009)

Type Ia – Good Old-Fashioned AI



Expert systems, GOFAI, etc.



Type Ib – Embodied-embedded AI

Restricted world

Emergent behavior

Pre-determined goals

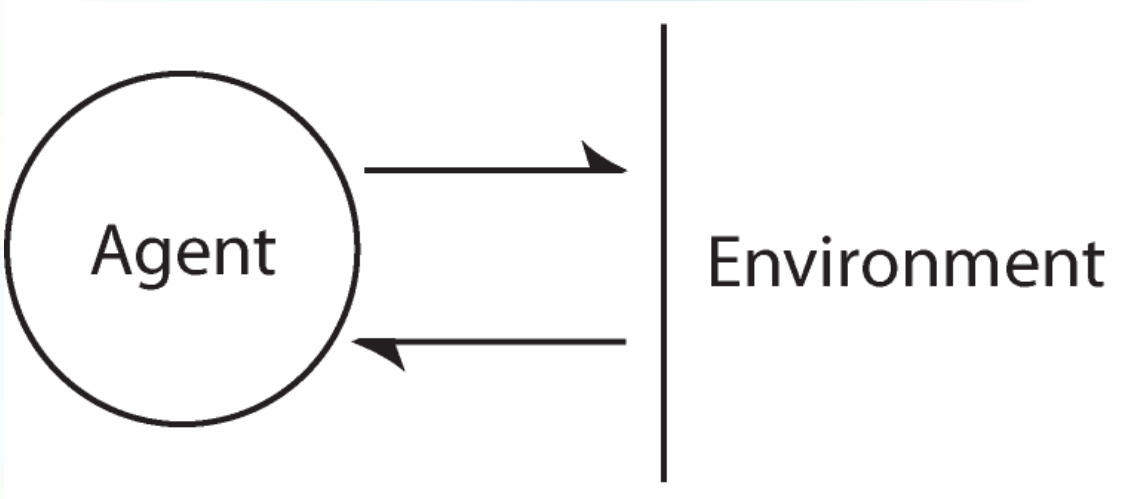
Pre-determined identity

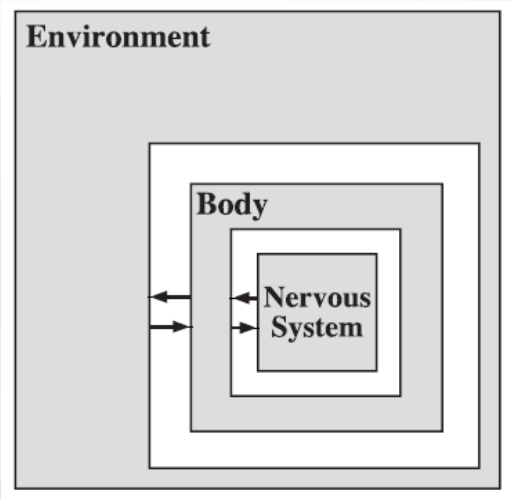
==

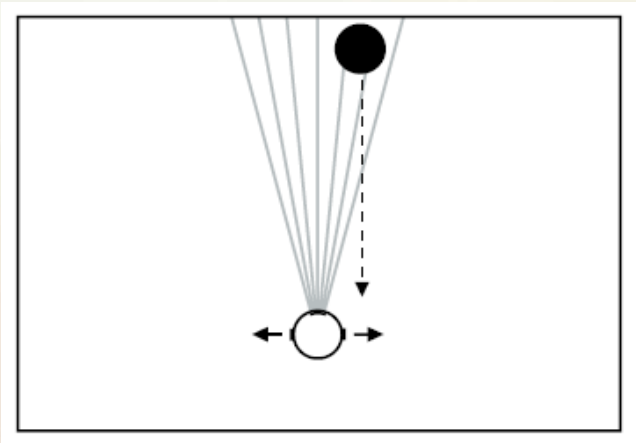
Problem of meaning &

Inter frame problem

Type Ib – Embodied-embedded AI







Pfeifer & Scheier (1999); Beer (2003)

Type Ic – Organismic AI

Restricted world

Emergent behavior

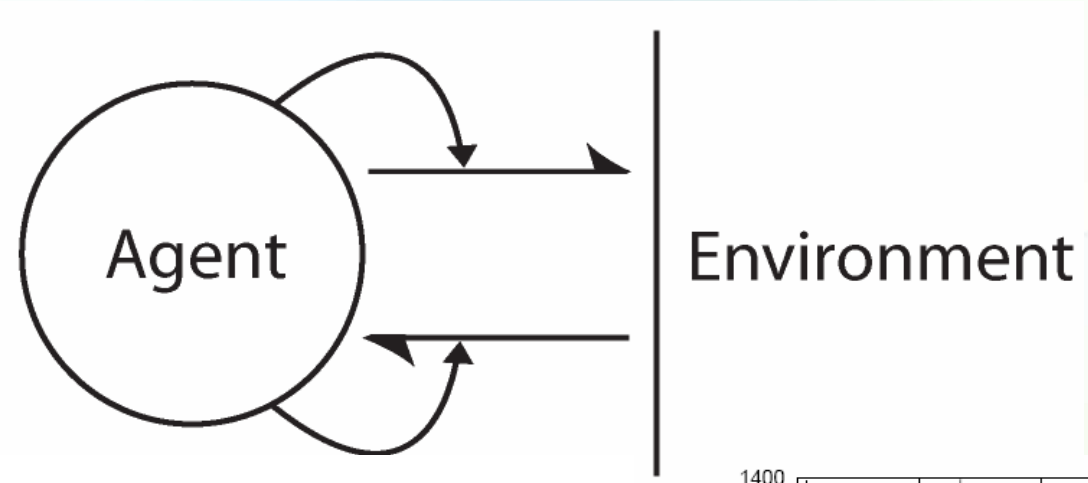
Emergent goals

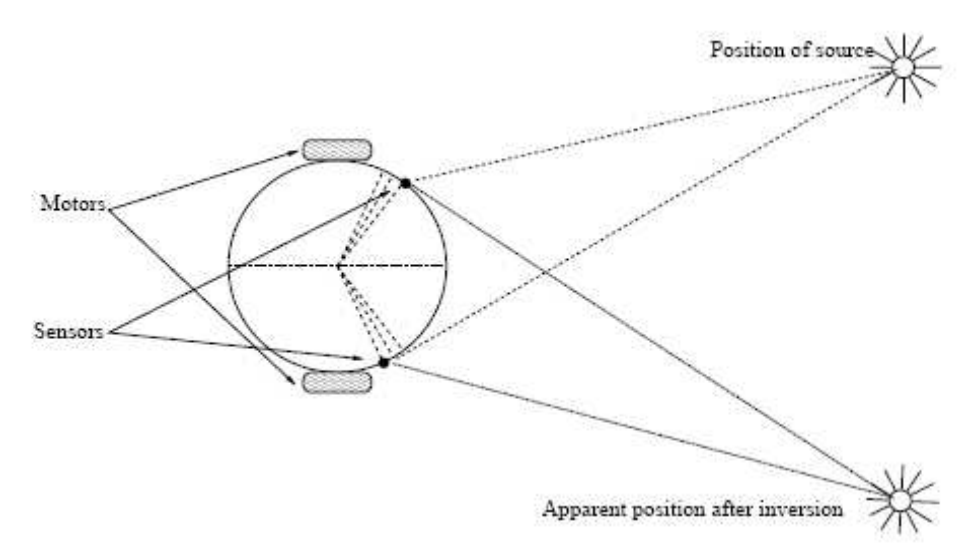
Pre-determined identity

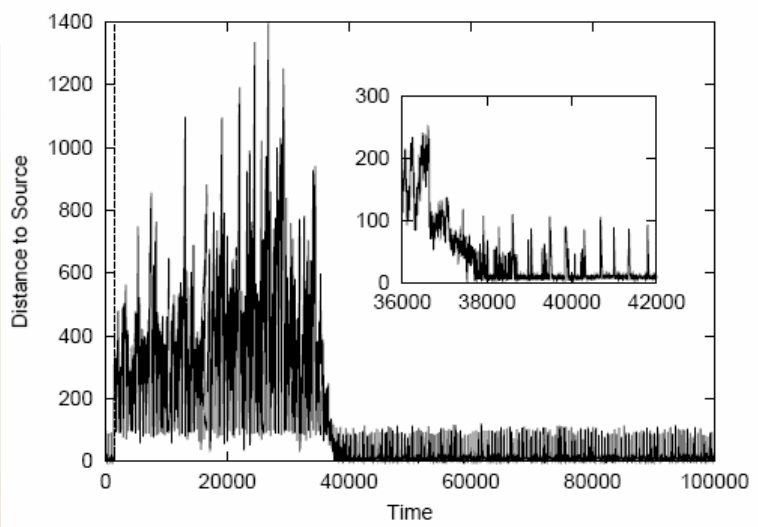
==

Problem of meaning

Type Ic – Organismic AI







Di Paolo (2000); Iizuka & Di Paolo (2007; 2008)

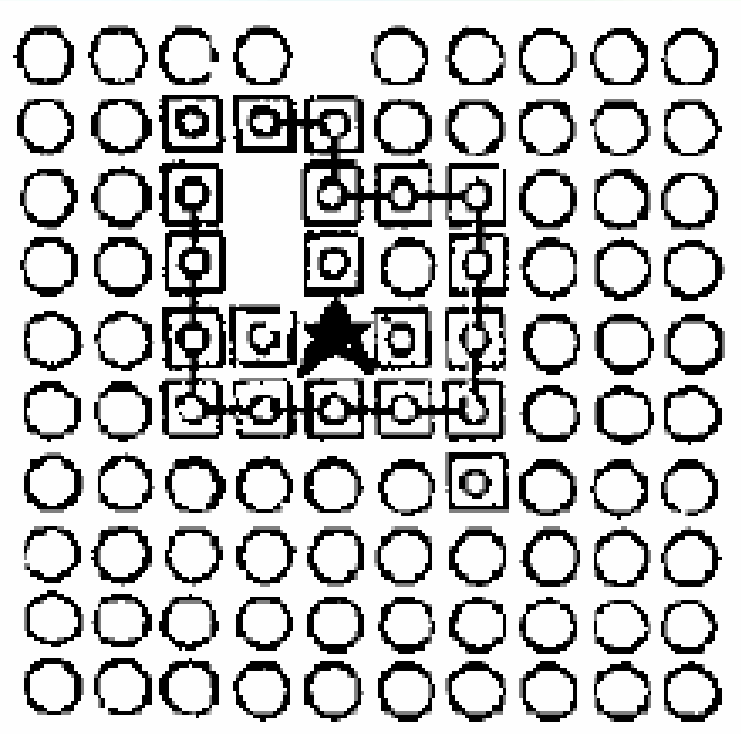
Return of the ‘Tesselation automata’!

A typography of AI - by Froese and Ziemke (2009)

Type II

Why no *robots*?

(Froese, et al. 2007;
Froese & Di Paolo 2008)

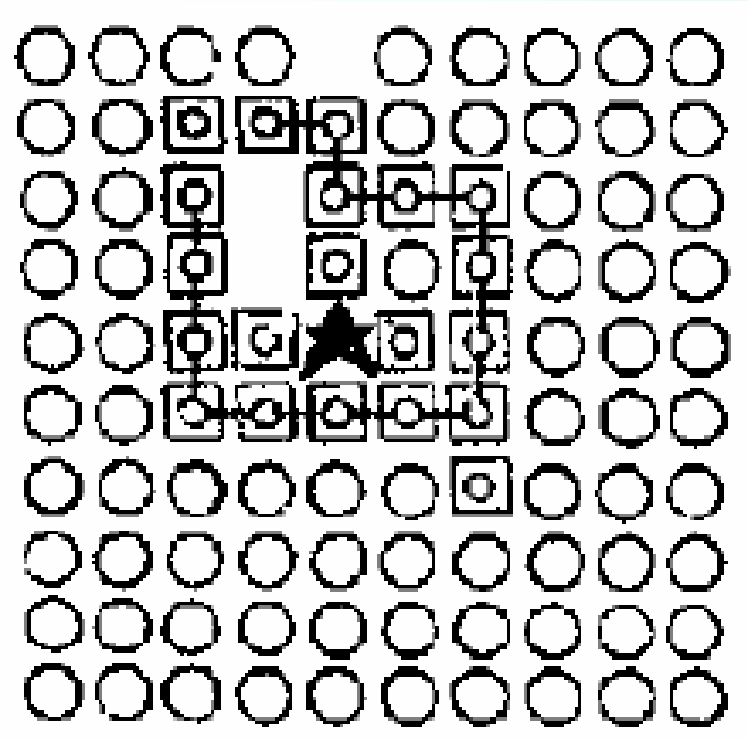


The hard problem of AI:
How to engineer second-order emergence?

(Froese & Ziemke 2009)

Varela, Maturana & Uribe (1974)

Type IIa – Tesselation automaton



No world

No behavior

Singular goal

Emergent identity

==

Restricted meaning &

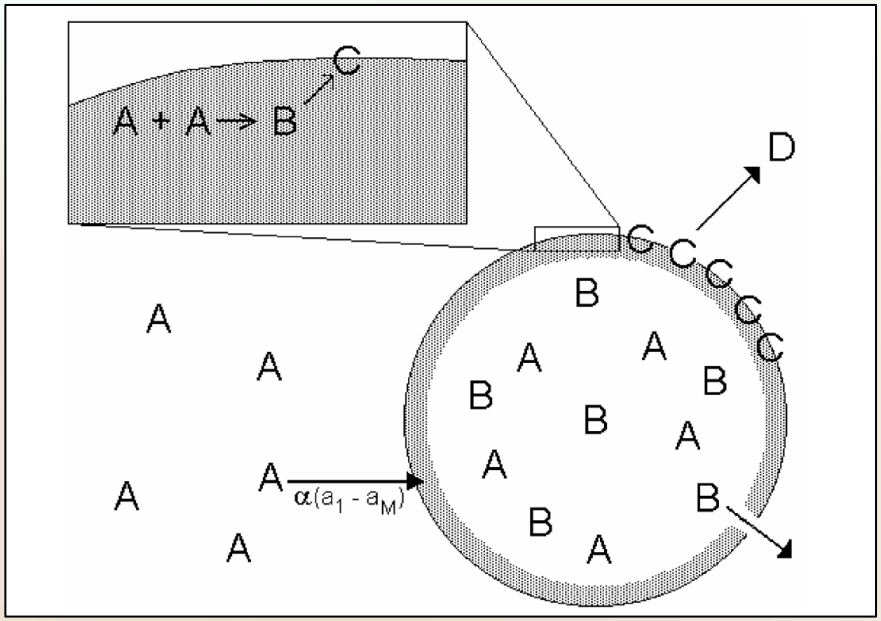
Inter frame problem &

Intra frame problem!

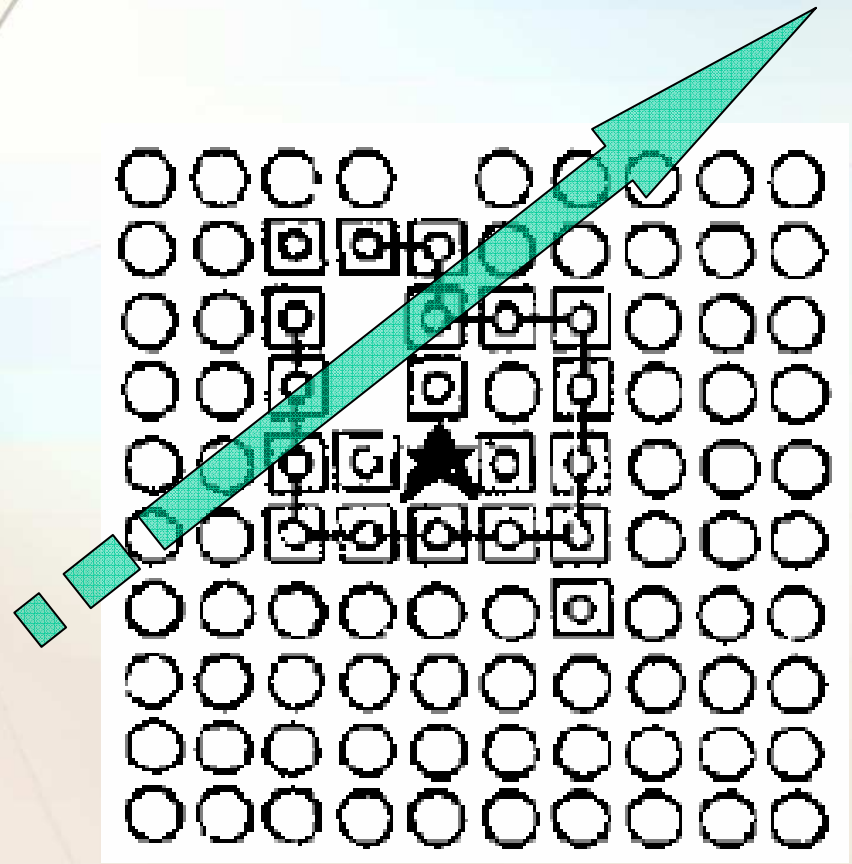
Varela, et al. (1974); Bourgine & Stewart (2004)

A typography of AI: **Type IIa**

Bourgine & Stewart (2004)



Type IIb – Tesselation automaton



Restricted world

Emergent behavior

Singular goal

Emergent identity

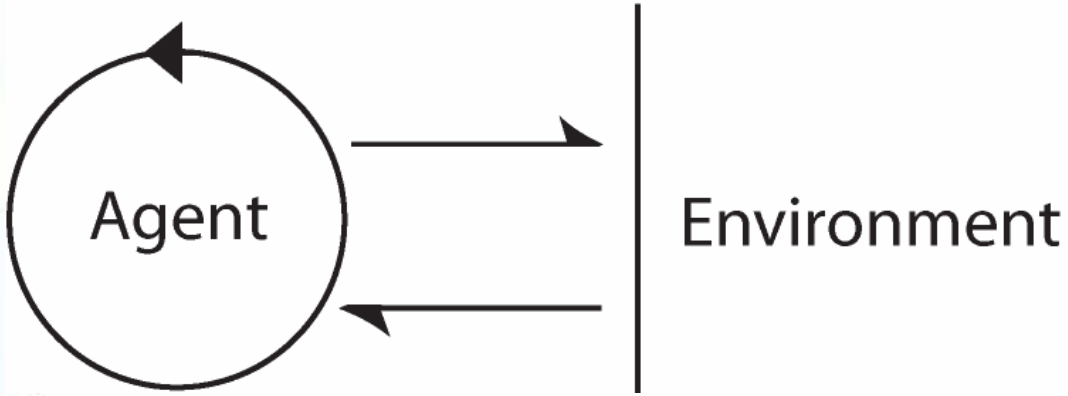
==

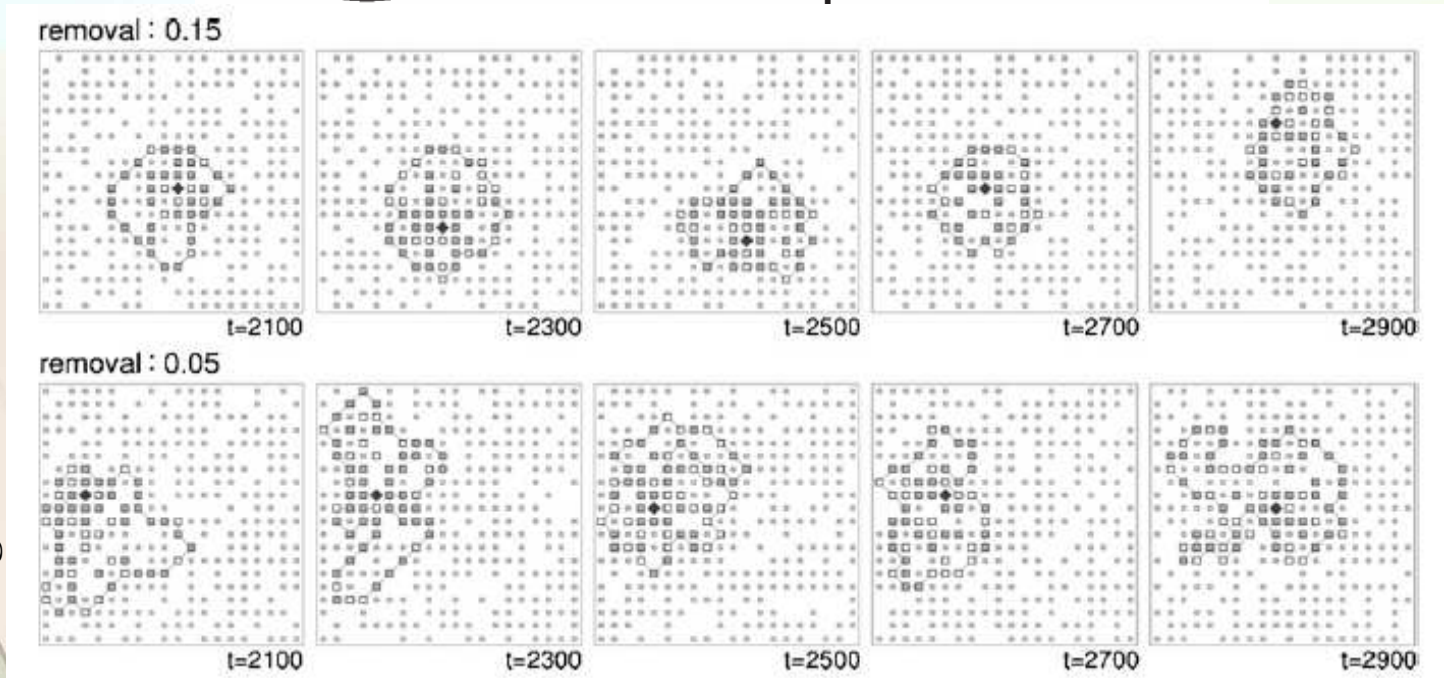
Restricted meaning &

Inter frame problem

Varela, et al. (1974); Bourgine & Stewart (2004); Egbert & Di Paolo (2009)

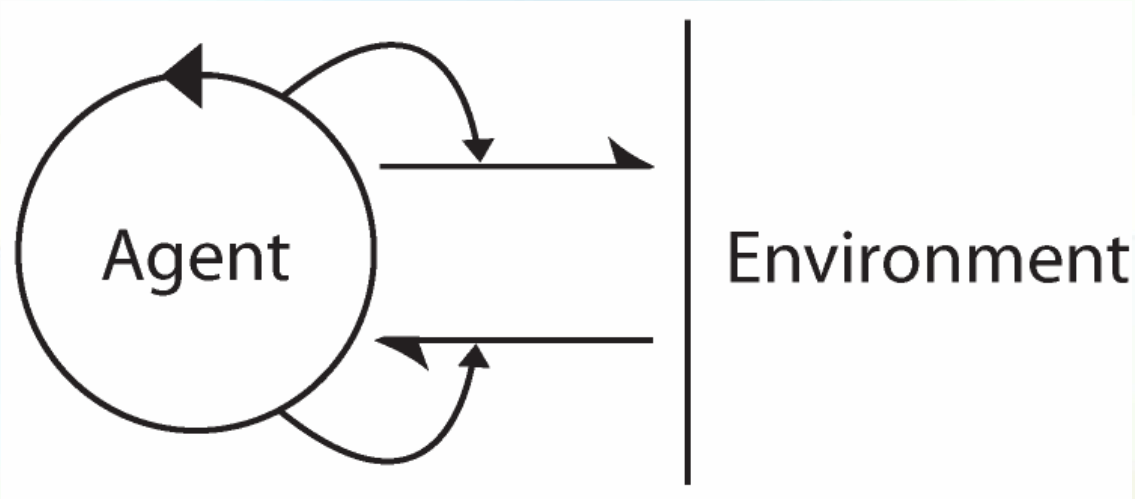
Type IIb – Tesselation automaton

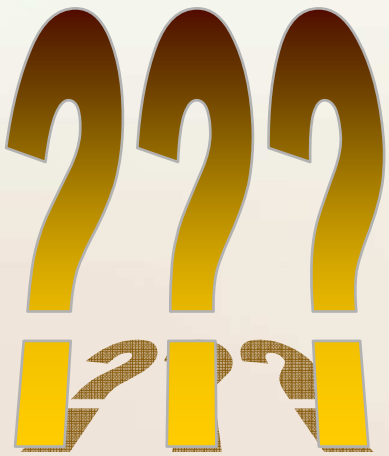




Ikegami & Suzuki (2008)

Type IIc





Beyond chemical autopoiesis?

(Virgo, et al. 2009; Froese & Di Paolo 2008)

References

- Froese, T. & Ziemke, T. (2009), “Enactive Artificial Intelligence: Investigating the systemic organization of life and mind”, *Artificial Intelligence*, 173(3-4), pp. 366-500
- Virgo, N., Egbert, M. & Froese, T. (2009), “The Role of the Spatial Boundary in Autopoiesis”, in: *Proc. of the 10th Euro. Conf. on Artificial Life*, Berlin, Germany: Springer
- Froese, T. & Di Paolo, E. A. (2008), “Can evolutionary robotics generate simulation models of autopoiesis?”, Cognitive Science Research Paper, 598, Brighton, UK: University of Sussex
- Froese, T. (2007), “On the role of AI in the ongoing paradigm shift within the cognitive sciences”, in: M. Lungarella *et al.* (eds.), *50 Years of AI*, Berlin, Germany: Springer, pp. 63-75
- Froese, T., Virgo, N. & Izquierdo, E. (2007), “Autonomy: a review and a reappraisal”, in: F. Almeida e Costa *et al.* (eds.), *Proc. of the 9th Euro. Conf. on Artificial Life*, Berlin, Germany: Springer, pp. 455-464