Languaging as a second order or joint control process

Joachim De Beule
Artificial Intelligence Lab
Vrije Universiteit Brussel

ECCO Seminars
25/11/2011

- Outline -

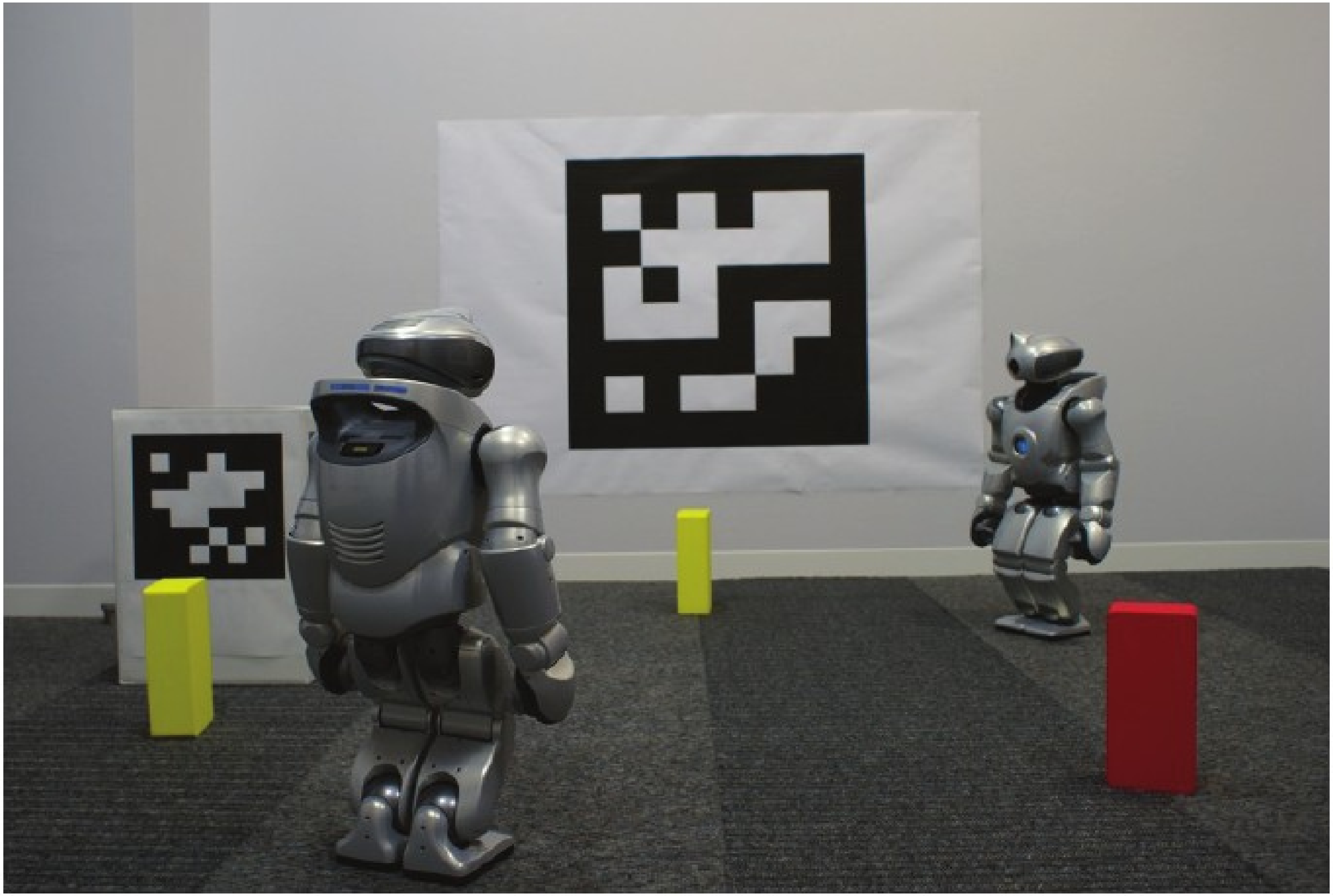
- Introduction

  The crisis in Linguistics and other fields of complex systems

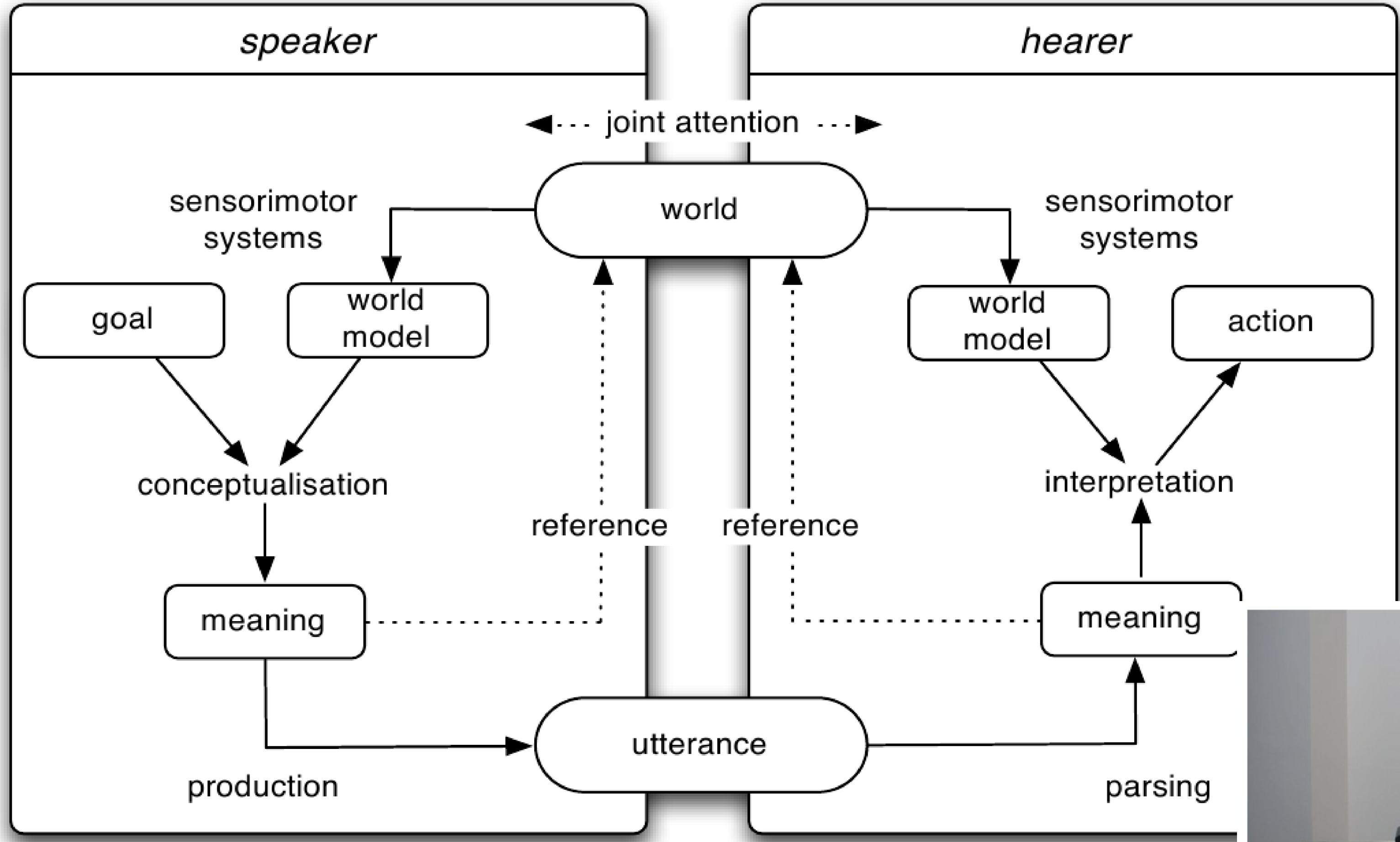
- Three Parts:

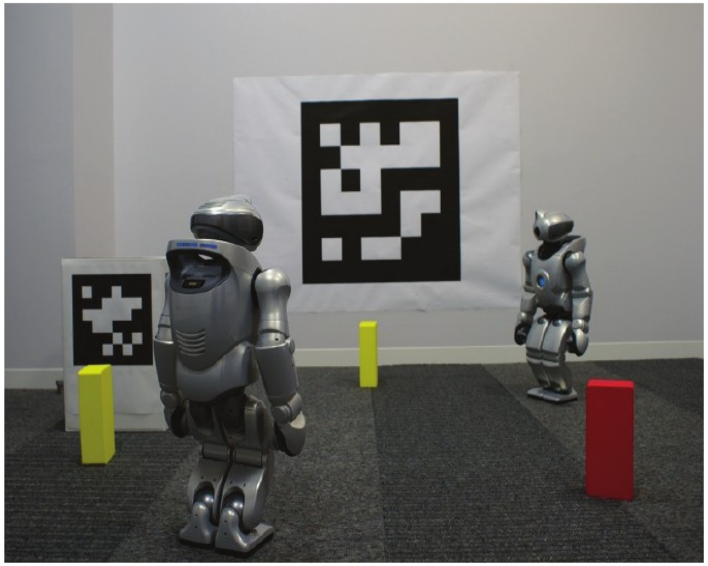
  I. What is “meaningful” information?
  II. Mechanisms of Evolution
  III. Conventionalization Dynamics

- How to build robots that learn to use language in language games (Wittgenstein, 1953)



- Language involves <u>semantics</u>
(situated, embodied, whole systems approach)





- Language is a collective, self-organizing phenomenon (non-linear micro-macro interactions and level formation)

- Language introduces <u>new causalities</u>
  (No "natural law" explains language, it is <u>arbitrary</u>, and yet <u>effective</u>)

Two big questions:

1) What is the nature of meaning?

2) How can a group of individuals “organize themselves”, that is, coordinate their private conceptual systems and public language?

A paradigmatic “crisis”

- In the last centuries, almost all scientific progress came from a further development of linear, rule-based, reductionist thinking (= syntax)

- In the last decades have seen an acceleration due to increasing amounts of data and computational power

- Recently, the bottleneck of scientific progress has become no longer computational resources, but what to do with them (GB)

- But the basic toolbox of e.g. linguistics and AI has not changed fundamentally

- And neither has that of economy, cognitive science, neurology, molecular biology, evolutionary theory, ideology and technology

Two bigger questions:

1) What is the nature of biological function?

2) How do Major Transitions occur?

(Maynard-Smith & Szathmary, 1995)

I. When is Information "meaningful"?

Reprinted with corrections from *The Bell System Technical Journal*,
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic

I. When is Information "meaningful"?

Reprinted with corrections from *The Bell System Technical Journal,*
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic

possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic

I. When is Information "meaningful"?

Reprinted with corrections from *The Bell System Technical Journal*,
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

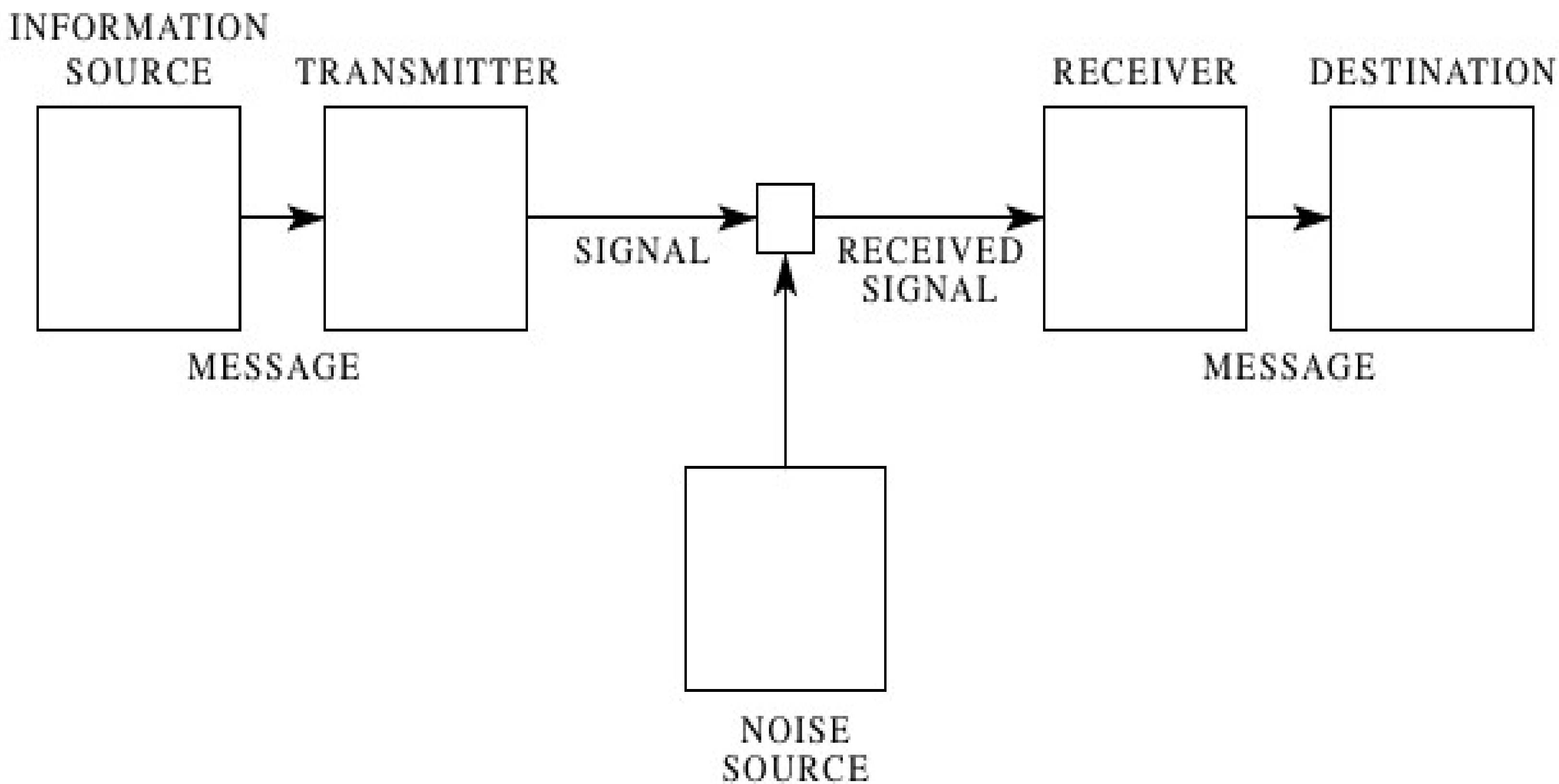


Fig. 1 — Schematic diagram of a general communication system.

I. When is Information "meaningful"?

A basic ingredient in Shannon's information theory is that <u>both parties must agree upon a set of rules beforehand</u>

- A shared alphabet
- Or an entire grammar

} syntax

I. When is Information "meaningful"?

A basic ingredient in Shannon's information theory is that <u>both parties must agree upon a set of rules beforehand</u>

- A shared alphabet
- Or an entire grammar

} syntax

But it does not explain how these could be established <u>before</u> communication is possible

I. When is Information "meaningful"?

Reprinted with corrections from *The Bell System Technical Journal*,
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design

A Different Problem

Instead of the message reconstruction problem, we must consider the language understanding problem

In other words, how to build a system that not simply reconstructs our messages, but “understands” them by responding “appropriately”?

This is essentially the same as asking how to solve Shannon's communication problem without assuming a pre-defined set of rules

In both cases, sender and receiver must “negotiate” until they are “properly aligned”

A Different Problem

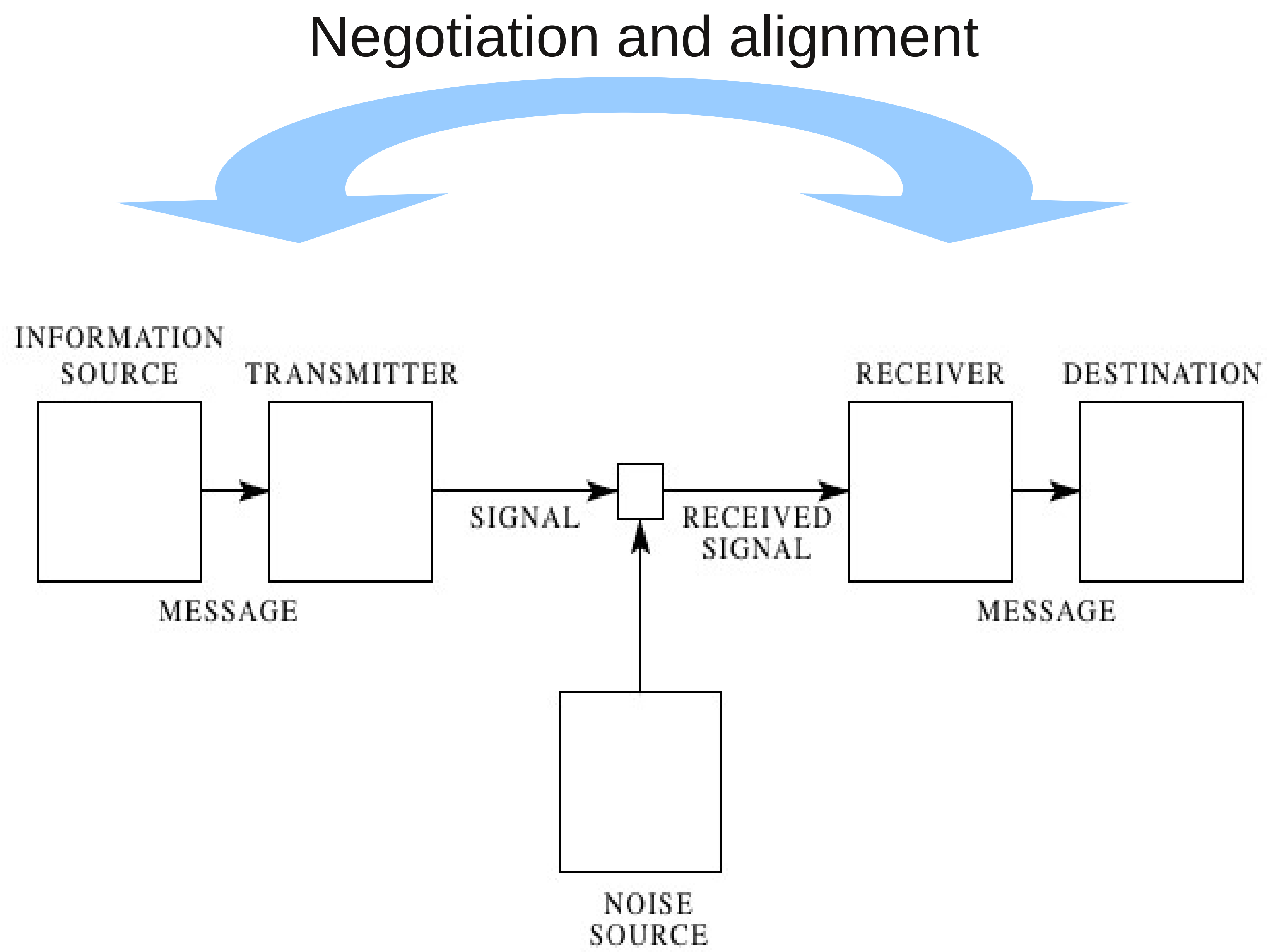
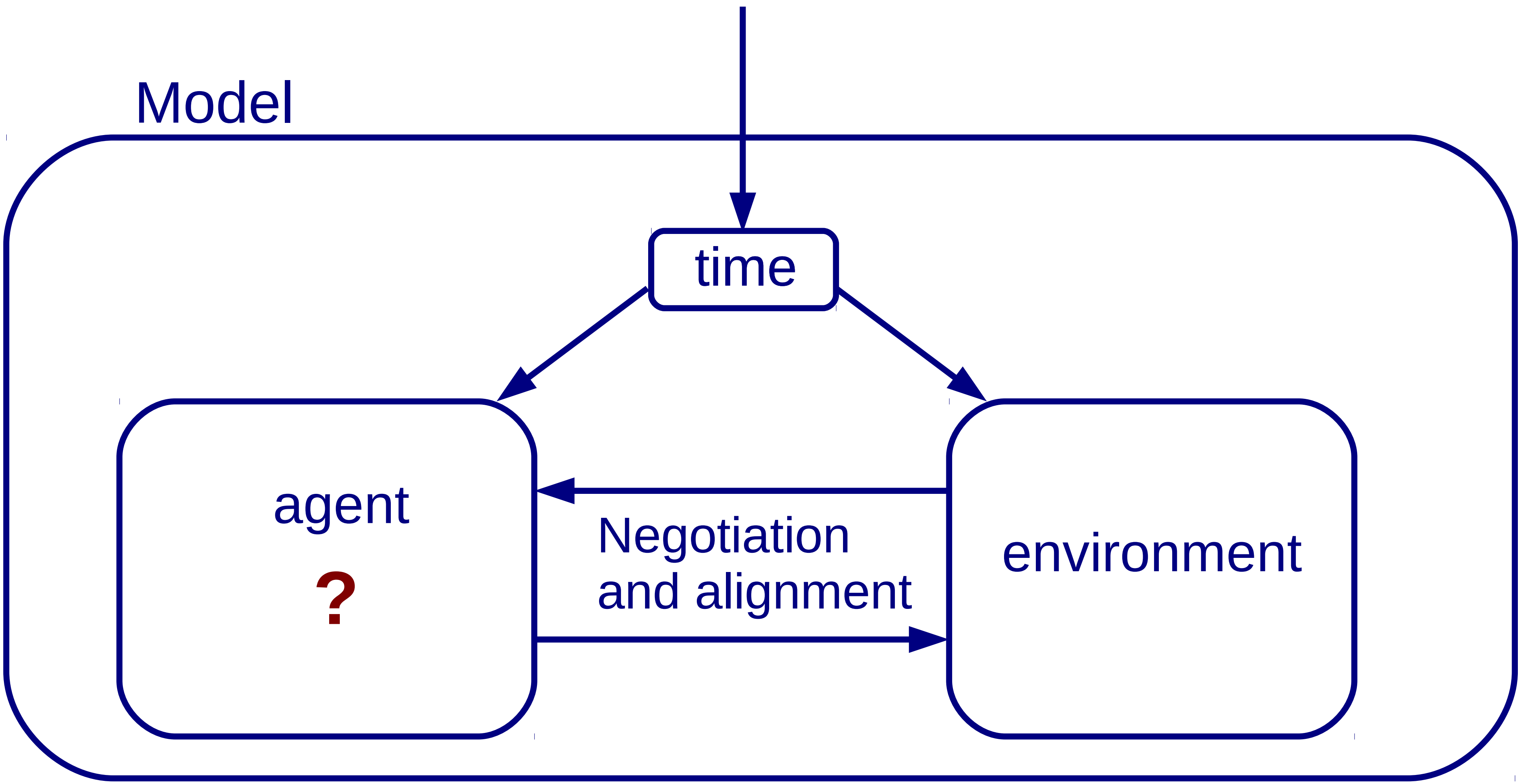


Fig. 1—Schematic diagram of a general communication system.

A (slightly) Simpler Problem

Only consider how a single system could find "useful distinctions" for itself (independent of whether this is in the context of a Shannon communication problem or not)



A (slightly) Simpler Problem

=> what is “negotiation”, “useful”, “purposeful”, “alignment”, “learning”, etc?

These are all relational, goal-directed (finalistic) and dynamical

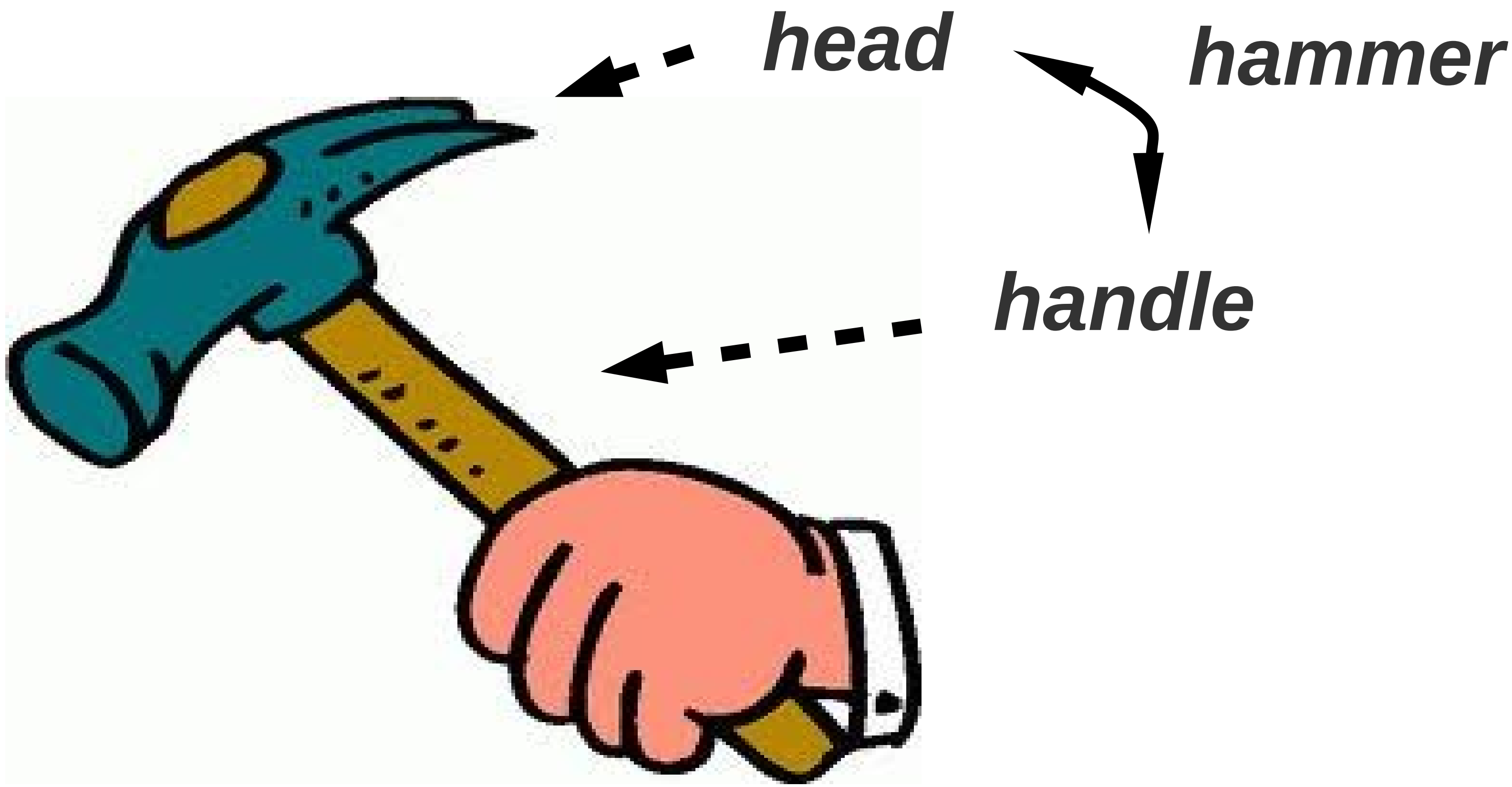
I therefore take a process-metaphysical perspective and build models in the form of dynamical, goal-directed systems

(Bickhard, Rosen, Ashby)

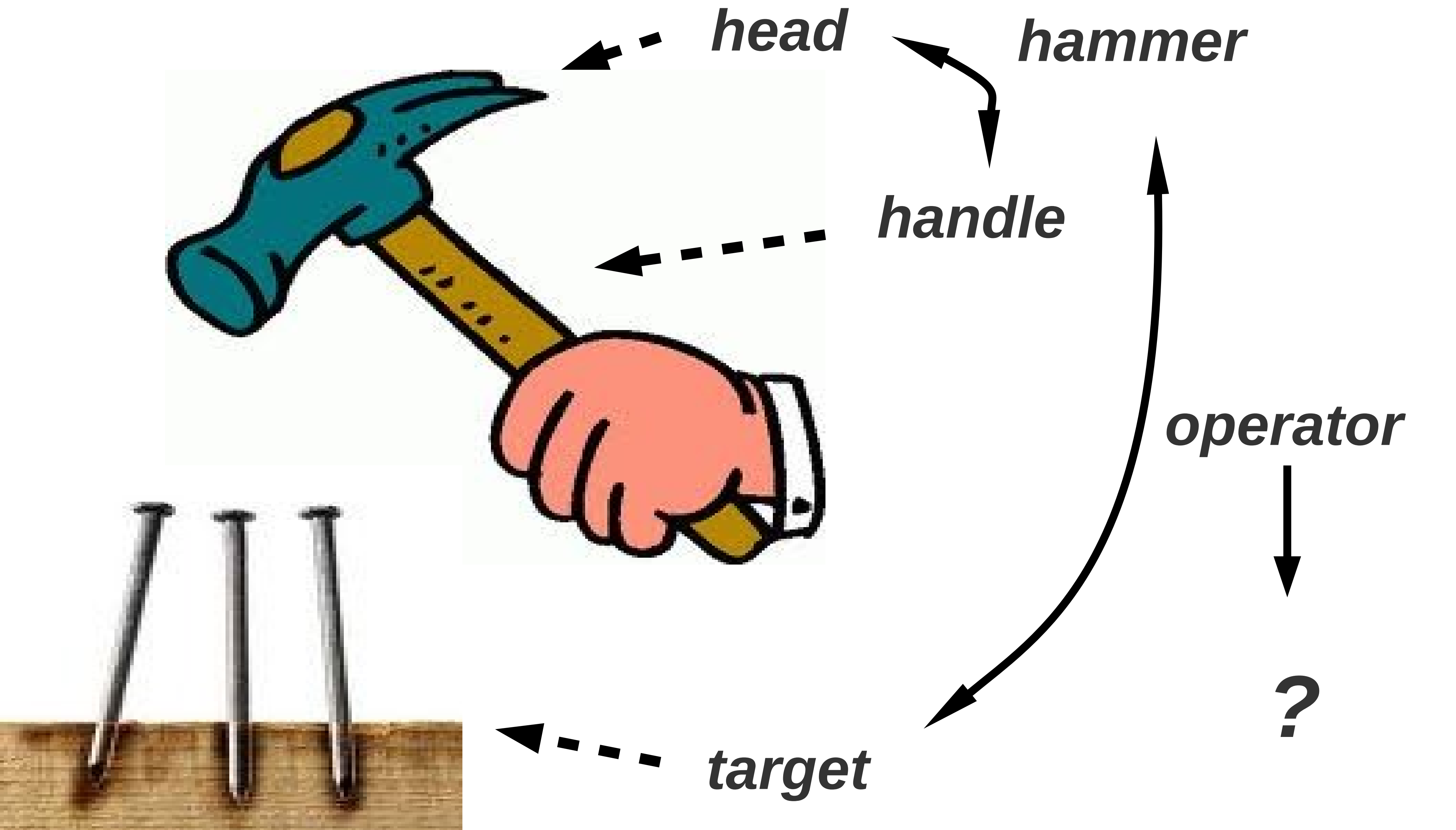
Not a Whiteheadean process metaphysics
– no purposeful becoming in any absolute sense

Also not a Peircean semiotics
– or any other form of theory that leads to an “infinite regress” or other types of dualities and arguments of design)

Function is relational



Function is relational



Function, being relational, cannot be defined for a single system, but always requires reference to something else, a “greater whole”.

In other words, we are heading towards an *infinite regress (Zeno's paradox)*

Function, being relational, cannot be defined for a single system, but always requires reference to something else, a “greater whole”

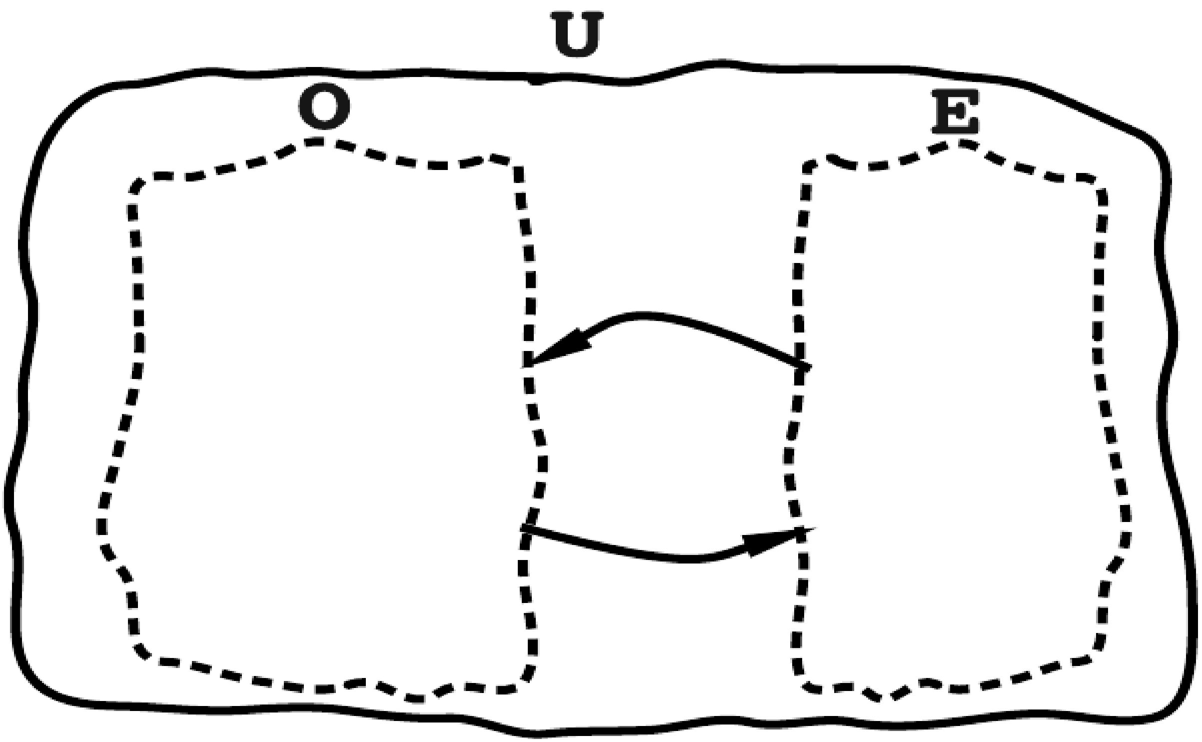
In other words, we are heading towards an *infinite regress (Zeno's paradox)*

But every dynamical system stands in relation *with itself over time*

It follows that every dynamical system has the “final goal” or intrinsic purpose to reach equilibrium

The intrinsic purpose captures the essence of evolution, because once the intrinsic purpose of the whole is reached, all evolution stops.

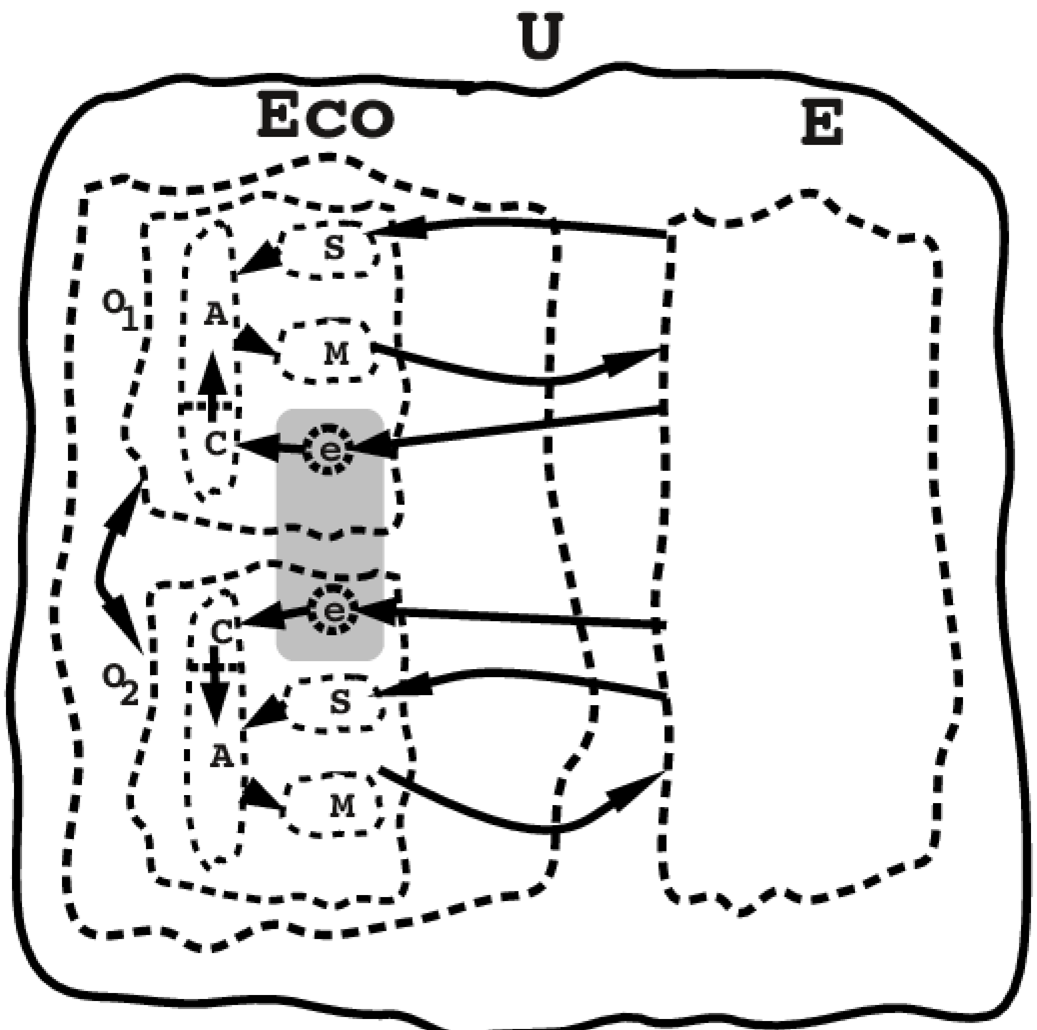
To go beyond intrinsic purpose, we must distinguish between <u>parts in a whole</u>



Parts **O** and **E** must comply to each other, and are "granted" their function by the whole **U**.

(consider what you would do with a hammer with a loose head)

In a <u>hierarchy of parts in wholes</u>, function will become more and more specialized, that is, <u>biological</u>.



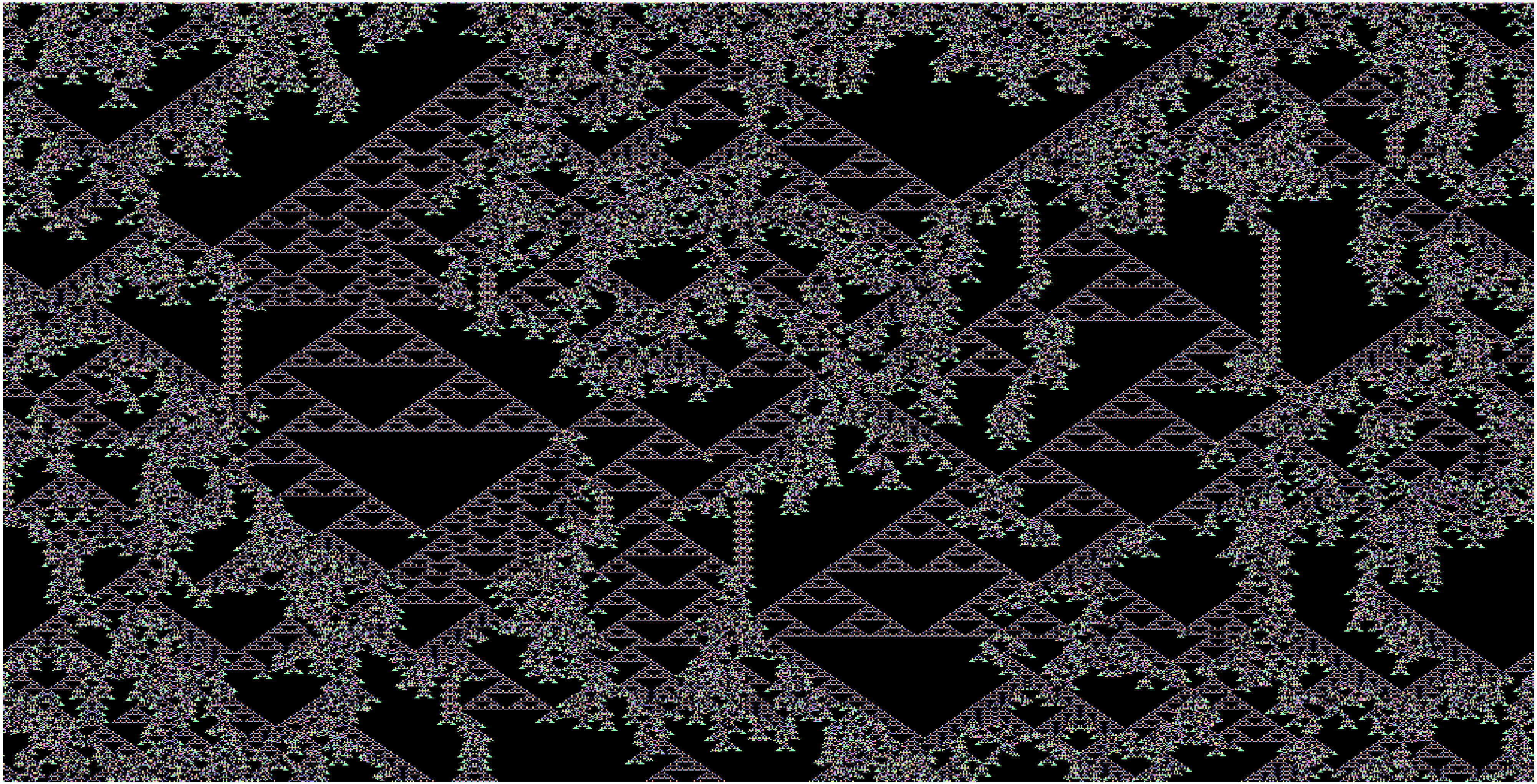
I. When is Information "meaningful"?

Bateson (or MacKay?) information

"Meaningful information is any difference that <u>makes</u> a difference"

I. When is Information “meaningful”?

Von Neumann's theory of self-reproducing automata



=> Information is meaningful when it becomes a program for reproduction

I. When is Information "meaningful"?

=> "Closure"

- physical closure (von Neumann)
- semantic closure (Howard Pattee)
- closure to efficient cause (Robert Rosen)
- Autopoiesis or organizational closure (Maturana and Varella)
- operational closure (Niklas Luhman)
- ...

=> "regeneration" or the capacity to persist as a system

(note that this is a self-referential notion)

Ashby's essential parameters allow to distinguish between when it makes sense to consider the parts in a model, or rather if the model should be reconsidered

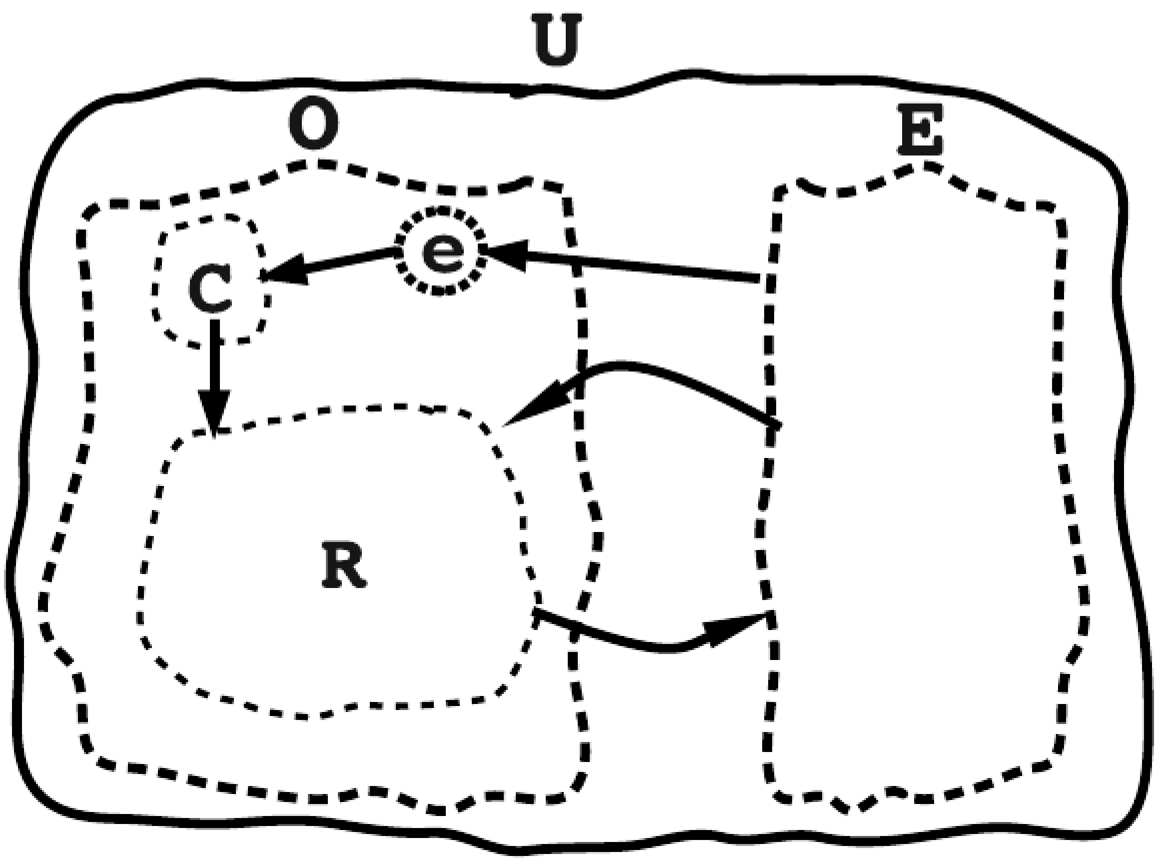
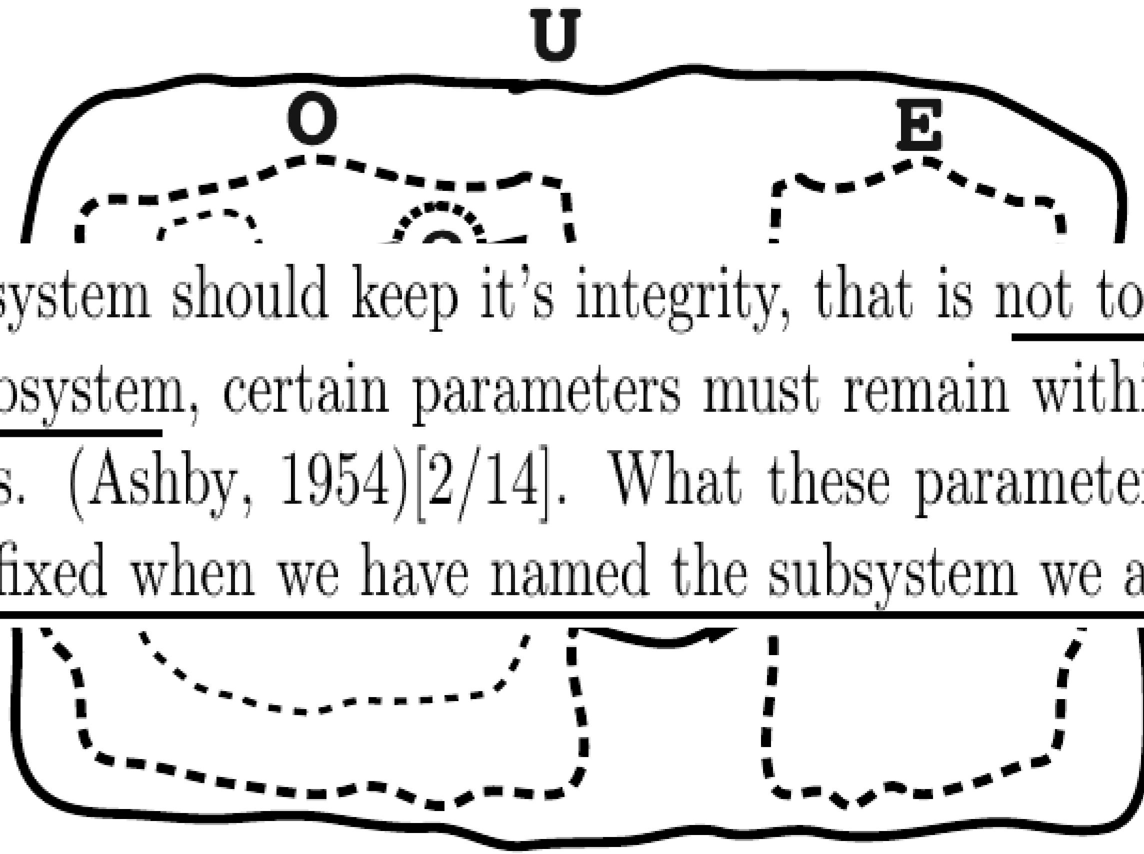


Figure: The homeostatic principle

Ashby's essential parameters allow to distinguish between when it makes sense to consider the parts in a model, or rather if the model should be reconsidered



That a subsystem should keep it's integrity, that is not to disintegrate but remain as a subsystem, certain parameters must remain within certain "physiological" limits. (Ashby, 1954)[2/14]. What these parameters are, and what the limits, are fixed when we have named the subsystem we are working with.

Figure: The homeostatic principle

Ashby's essential parameters introduce self-referentiality

- Into our models
- When we consider systems that contain their essential parameters
- In particular, when we consider self-regulatory systems, (systems capable of regulating their own parameters)
- Such systems may surely “give meaning” to perceived distinctions by coupling them to those actions that lead to closure, since it's persistence as a system depends on it

I. When is Information "meaningful"?

"Simpler" problem => "Closure"

When it is used by a regulatory system to compensate for disturbances in it's own essential parameters!

I. When is Information "meaningful"?

"Simpler" problem => "Closure"

When it is used by a regulatory system to compensate for disturbances in it's own essential parameters!

II. Mechanisms of evolution and the creation of meaningful information

How can the amount of meaningful information change, and how can it be e*xpected* to change?

Q1: How can the amount of meaningful information change?

- Must be due to changes in regulatory action

- Either in the variety of systems, or in the way that it is employed (constraint)

- Variety can be changed in two ways:

  (1) by extending the “mind” (the parameters of the organism)

  (2) by extending the “body” (the control over parameters of the environment)

- Constraint can be added through “learning”

When a self-regulating system learns or extends, it effectively establishes new causal relations that re-enforce or catalyze the changed system as a new, more “meaningful” (functional) whole

Q1: How can the amount of meaningful information change?

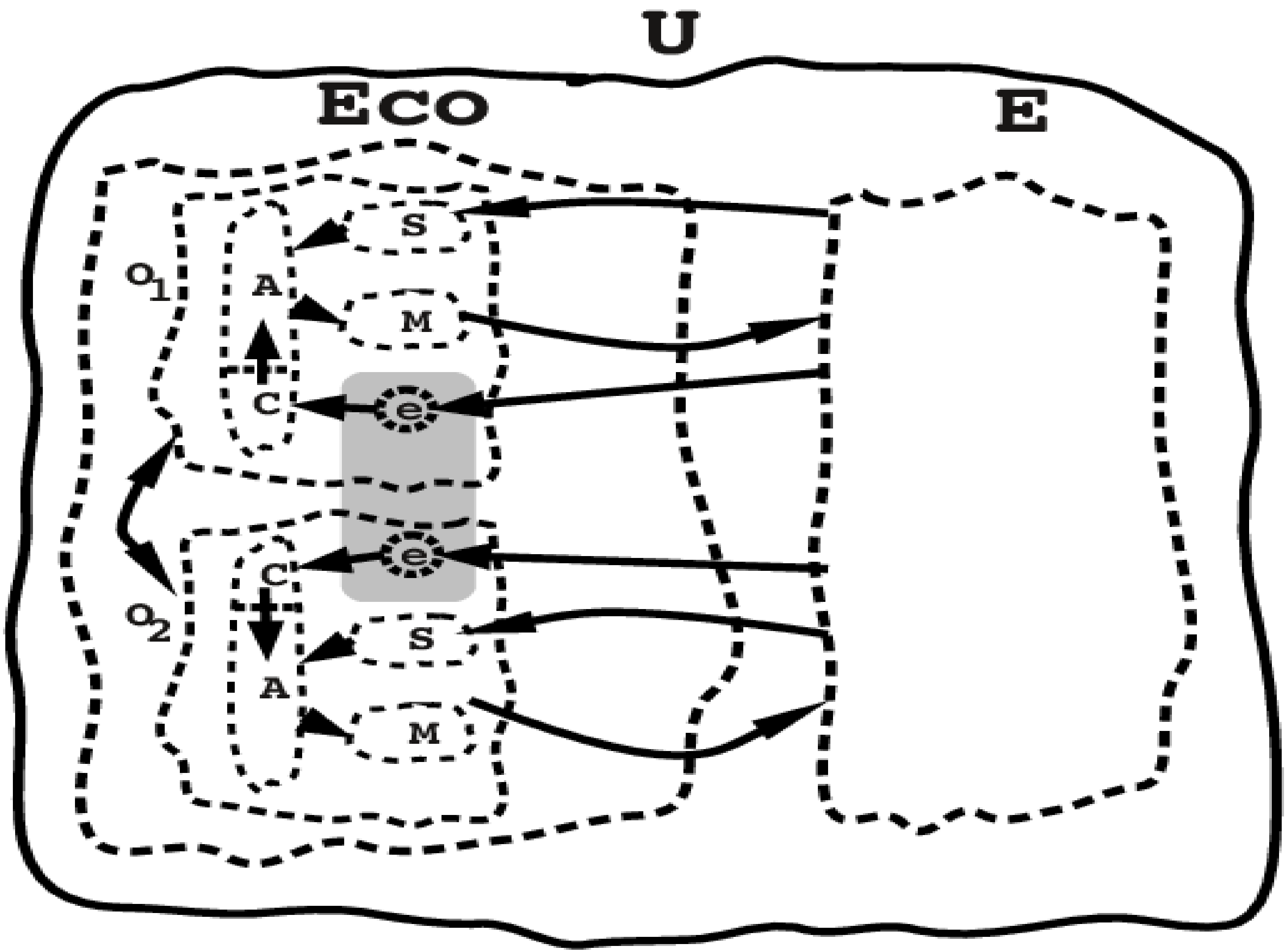
- A system may also amplify it's regulatory capacity directly by using another regulatory system (e.g. a thermostat, a coffee-machine, an industrial robot, ...)

- coffee-machines are not *self*-regulatory systems however

- Q2: What if two *self*-regulatory systems interact?

(second order interactions)

Q2: second order interactions



Q2: second order interactions

- Depends on how the essential parameters are related

- If they are shared, the total amount of variety per essential parameter is higher when the two systems are brought together

- Under certain conditions, this may lead both systems to become a single system through specialization

- How common are these conditions?

  Eventually, they are inevitable ...

Q2: second order interactions

- Depends on <u>how the essential parameters are related</u>

- If they are shared, the total amount of <u>variety per essential parameter</u> is higher when the two systems are brought together

- Under certain conditions, this may lead both systems to <u>become a single system through specialization</u>

- How common are these conditions?

  Eventually, they are inevitable ...

  but they depend on the systems finding a “shared code”, a set of constraints or rules that allows them sto communicate in the Shannon sense

In sum...

- From an evolutionary perspective, self-regulating systems have an advantage over others (higher “fitness”), and therefore will be selected

- The law of requisite variety prohibits a direct amplification of regulatory capacity, but it can be accomplished through learning, and through supplementation (tool usage)

- All tools, animate or not, must be operated by another human, so that the ultimate tool available is another human

- Languaging is the process by which we “operate” other people

- More generally, “Major Transitions” are bound to happen and they should be characterized by the appearance of new “codes” or means of transmitting meaningful information

In sum...

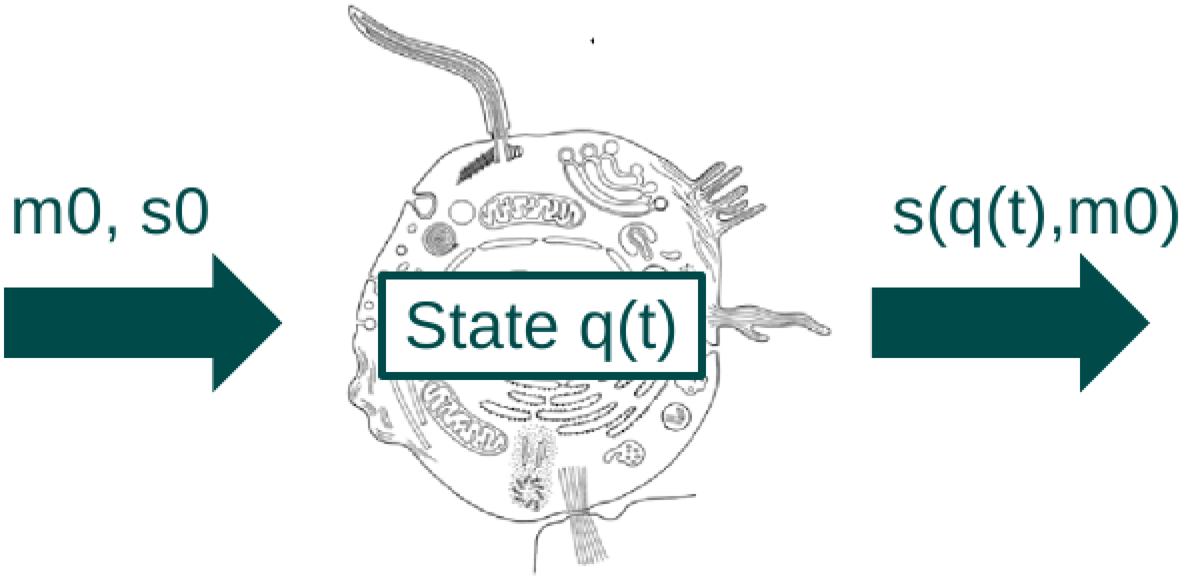
What language and Major Transitions also have in common is the fact that they are collective phenomena

Part III:
Conventionalization or semiotic dynamics

Part III: semiotic dynamics

How can a population of locally interacting "agents" come to a global agreement, like on how to name a certain item or even on an entire language?

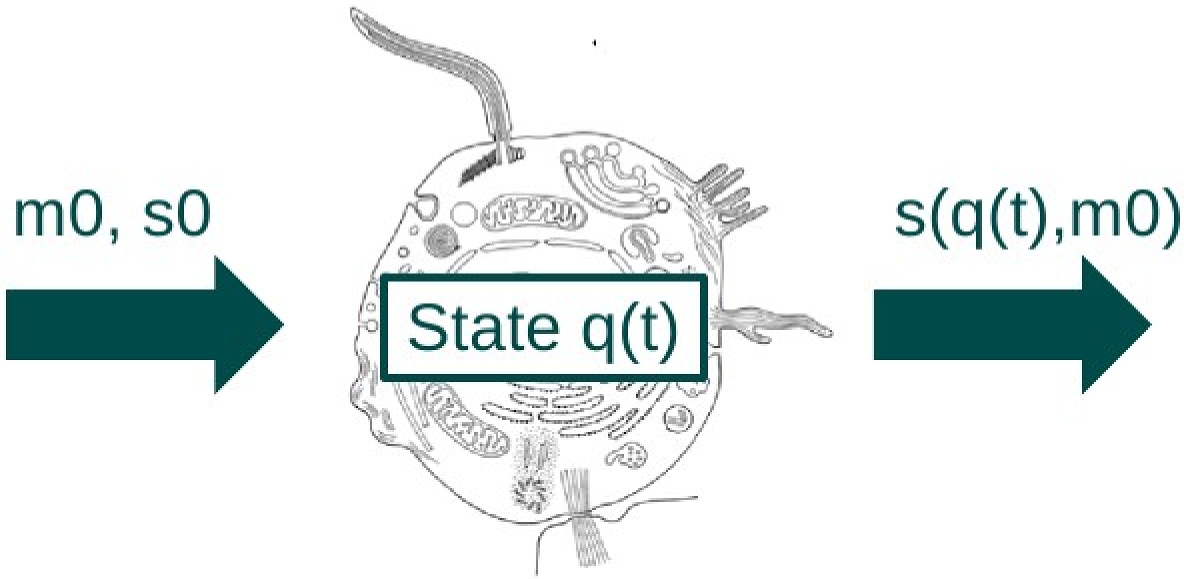
is



The organism randomly interacts with a population of other organisms at times $t_k = t_0 + k\Delta t$ with $k = 0, 1, \ldots$. Let the population behavior s^0 represent the average behavior of other organisms in the population in response to m^0 during interactions. Every interaction, the organism is stochastically influenced by it. In response, it will change its state according to some transition δ. If every interaction lasts a time Δt then schematically we have:

Part III: semiotic dynamics

How can a population of locally interacting “agents” come to a global agreement, like on how to name a certain item or even on an entire language?



$$s^0(t_{k+1}) = (1-\beta)s^0(t_k) + \beta s(q(t_k), m^0)$$

$$q(t_{k+1}) = \delta(q(t_k), s^0(t_k), m^0, \Delta t)$$

Part III: semiotic dynamics

$$\begin{aligned} q(t_k) &= \delta(q(t_{k-1}), s^0, m^0, \Delta t) \\ &= \delta(q(t_0 + (k-1)\Delta t), s^0, m^0, \Delta t) \\ &= \delta(q(t_0), s^0, m^0, k\Delta t). \\ &\overset{k \text{ large}}{\simeq} \phi(q(t_0), s^0, m^0), \end{aligned}$$

with $\phi(q_0, s^0, m^0)$ the organism's *response behavior* defined as the organism's limiting behavior in response to a constant population behavior s^0 for expressing m^0:

$$\phi : \langle q_0, s^0, m^0 \rangle \mapsto \left[s(\delta(q_0, s^0, m^0, t), m^0) \right]_{t \to \infty}$$

$$\begin{aligned} s^0(t_k + \Delta t) - s^0(t_k) &= \beta(-s^0(t_k) + s(q(t_k))) \\ &\simeq \beta(-s^0(t_k) + \phi(q_0, s^0, m^0)) \end{aligned}$$

$$\frac{d}{dt} s^0 = \alpha(\phi(q_0, s^0, m^0) - s^0)$$

Part III: semiotic dynamics

$$\begin{aligned} q(t_k) &= \delta(q(t_{k-1}), s^0, m^0, \Delta t) \\ &= \delta(q(t_0 + (k-1)\Delta t), s^0, m^0, \Delta t) \end{aligned}$$

Captures the dynamics (and hence equilibrium conditions) equilibria of the (deterministic approximation of the) collective system in terms of the limiting behavior of individuals

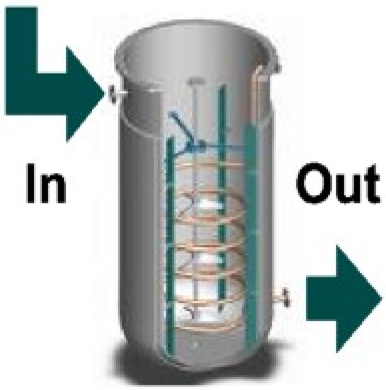
Assumptions:

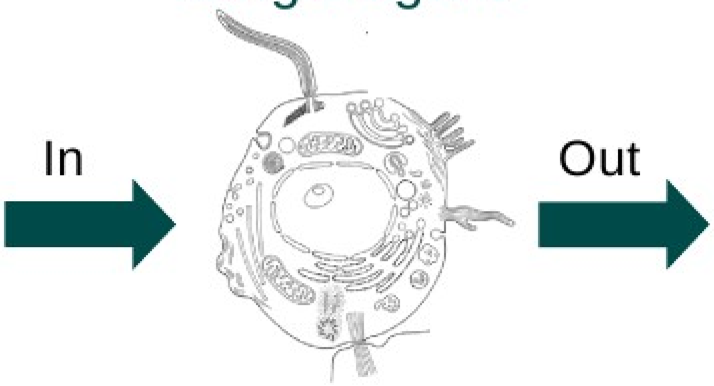
- Collective dynamics are “slow”
- Mean field

$$\frac{d}{dt} s^0 = \alpha(\phi(q_0, s^0, m^0) - s^0)$$

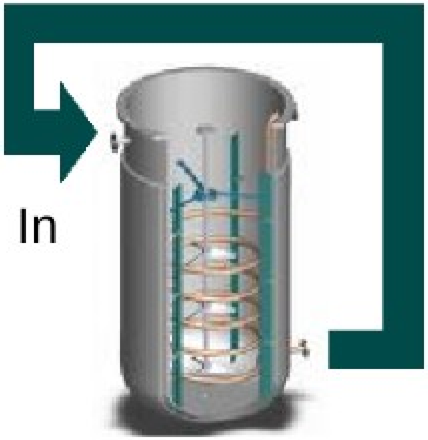
Part III: semiotic dynamics

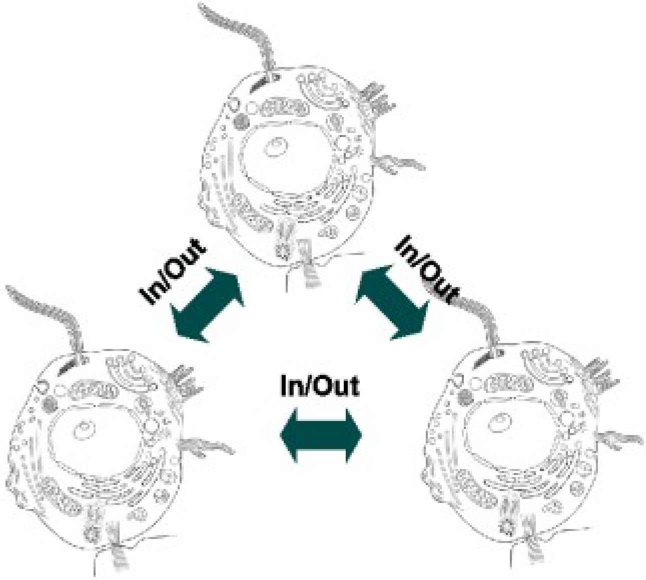
Open reactor tank ↔ Single agent





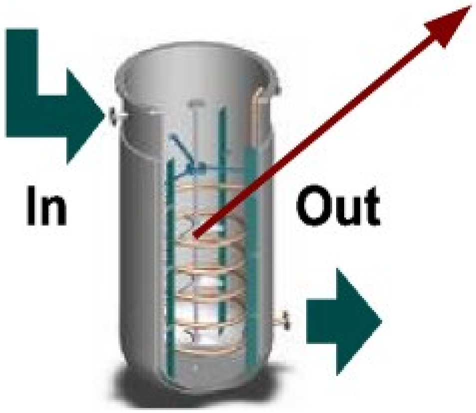
Closed reactor tank ↔ Agent population





Part III: semiotic dynamics

Open reactor tank



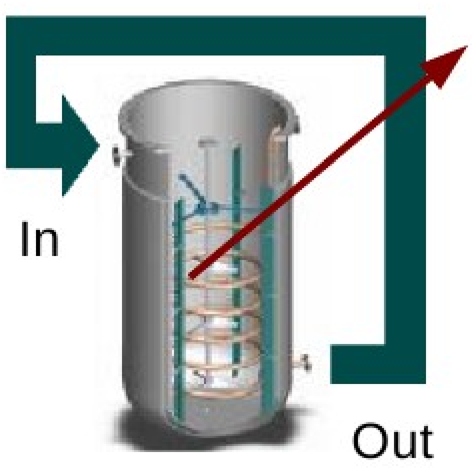
Artificial Chemistry

↓

Differential Equations

$$
(1)\begin{cases}
\dot{m} &= \rho_m(m^0(t) - m(t)) + R_m(m(t), s(t), c(t)),\\
\dot{s} &= \rho_s(s^0(t) - s(t)) + R_s(m(t), s(t), c(t)),\\
\dot{c} &= R_c(m(t), s(t), c(t)).
\end{cases}
$$

Closed reactor tank



Artificial Chemistry

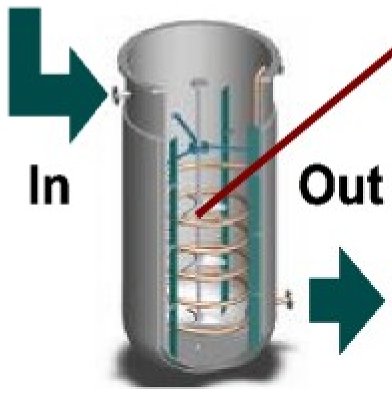
↓

Differential Equations

$(1) + \rho_m = \rho_s = 0$

Part III: semiotic dynamics

Open reactor tank



Artificial Chemistry

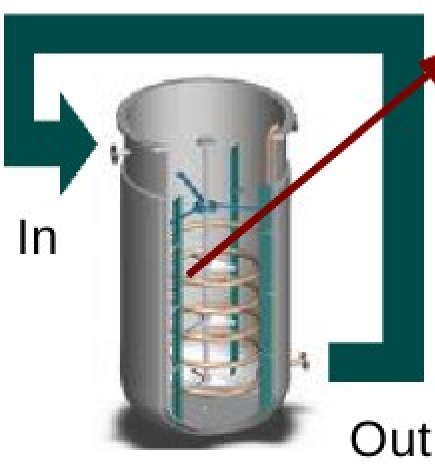
↓

Differential Equations

→ The stationary states for a certain influx or population behavior correspond to the response behaviors of the agent

---

Closed reactor tank



Artificial Chemistry

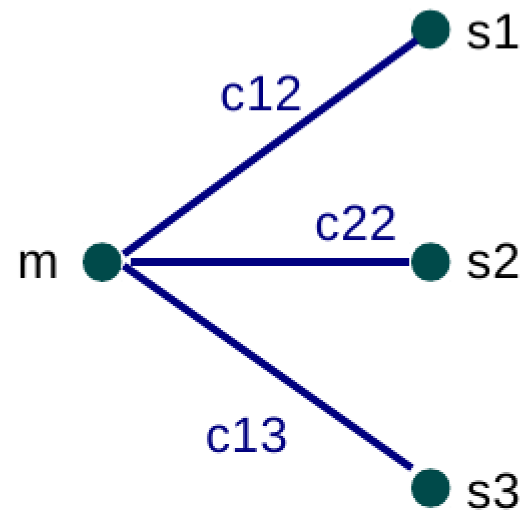
↓

Differential Equations

→ The stationary states for zero influx determine the population behaviors

Example: The “naming game”

Species:



Artificial Chemistry:

m + cj → sj

+

K1: si + cj → si + ci

Or

K2: si + cj → sj + cj

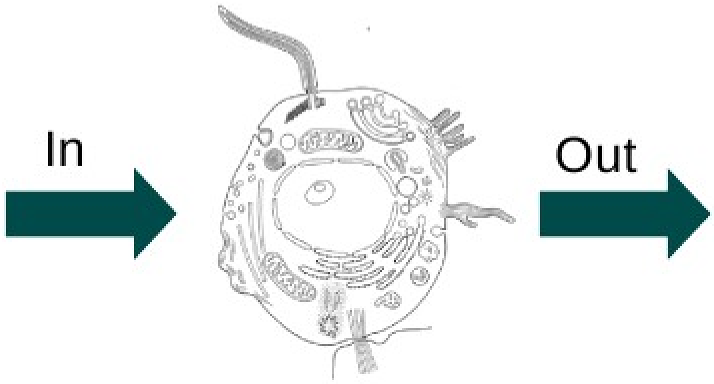
(De Beule “Introducing dynamics into the field of Biosemiotics”, Biosemiotics, 2010)

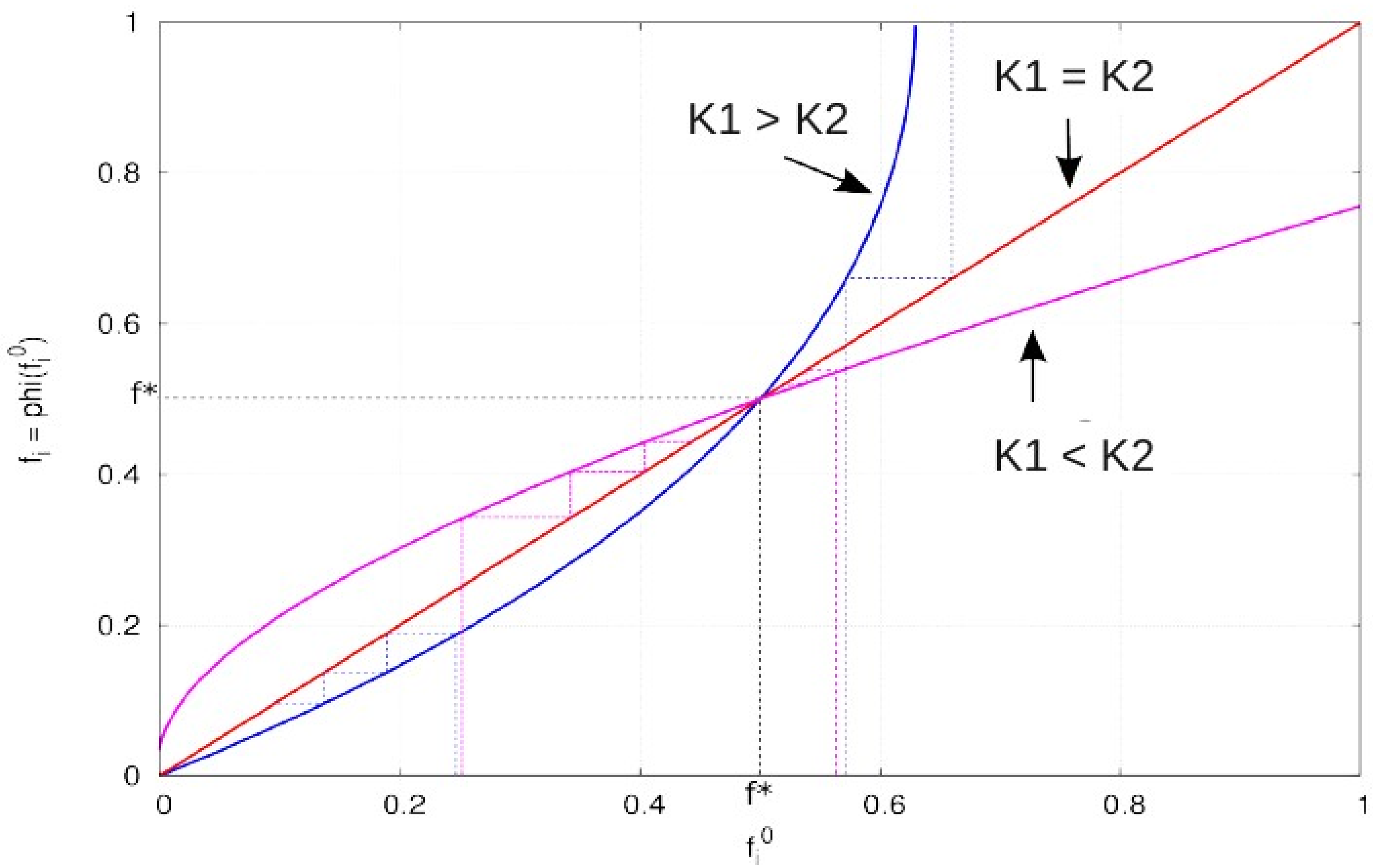
Example: The "naming game"

Differential equations:

$$
\begin{aligned}
\dot{s}_j &= \rho_s(s_j^0 - s_j) + m^0((1 + s_j)c_j - s_j) \\
&\quad -m^0\kappa_1(s_j\sigma_c - c_j\sigma_s) \\
\dot{c}_j &= -m^0((1 + s_j)c_i - s_j) \\
&\quad +m^0\kappa_2(s_j\sigma_c - c_j\sigma_s),
\end{aligned}
$$

Response Analysis

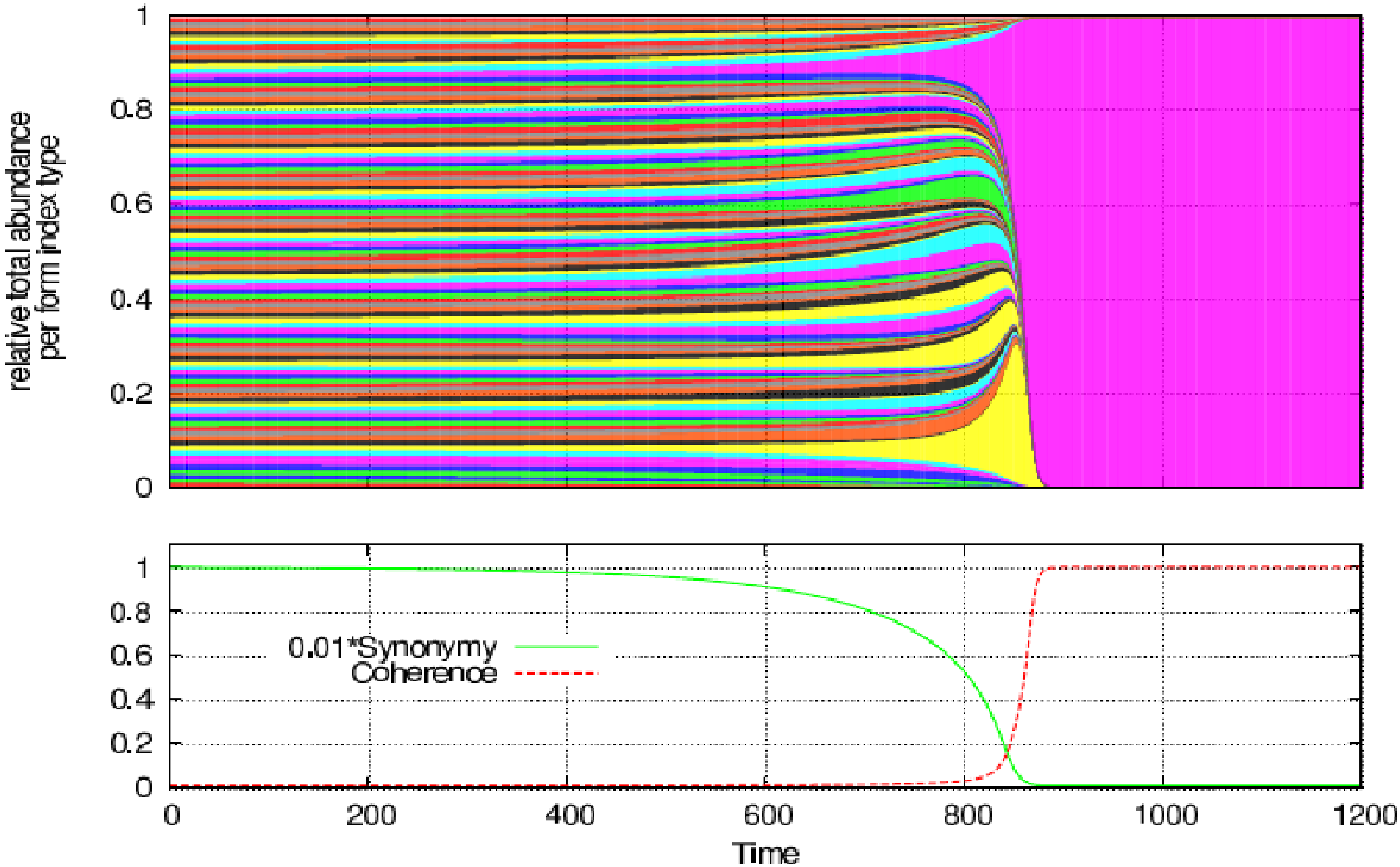




Population model

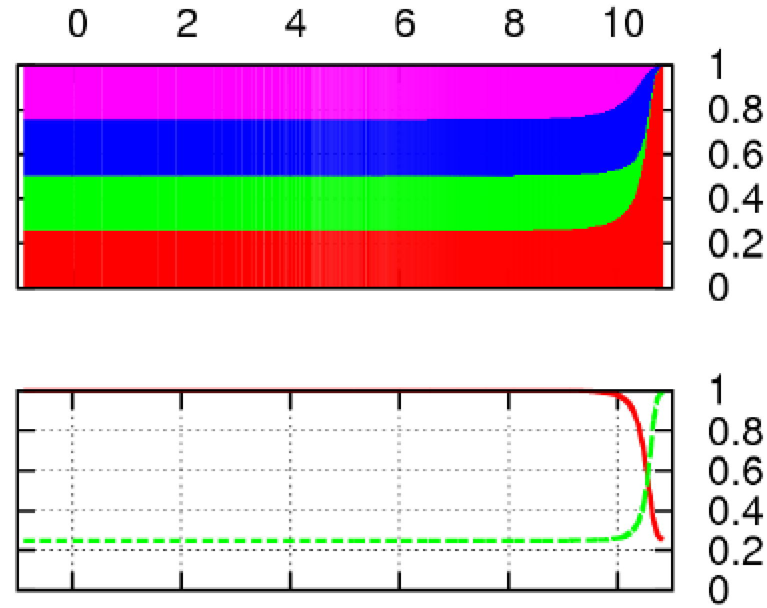
$$\mathrm{Coh(t)} = \frac{1}{(\sigma_c)^2}\sum_{i=1}(c_i(t))^2.$$

$$\mathrm{Red(t)} = \frac{1}{\sigma_c}\exp\left(-\frac{1}{\sigma_c}\sum_{i=1}^{n_s} c_i(t)\log(c_i(t))\right)$$



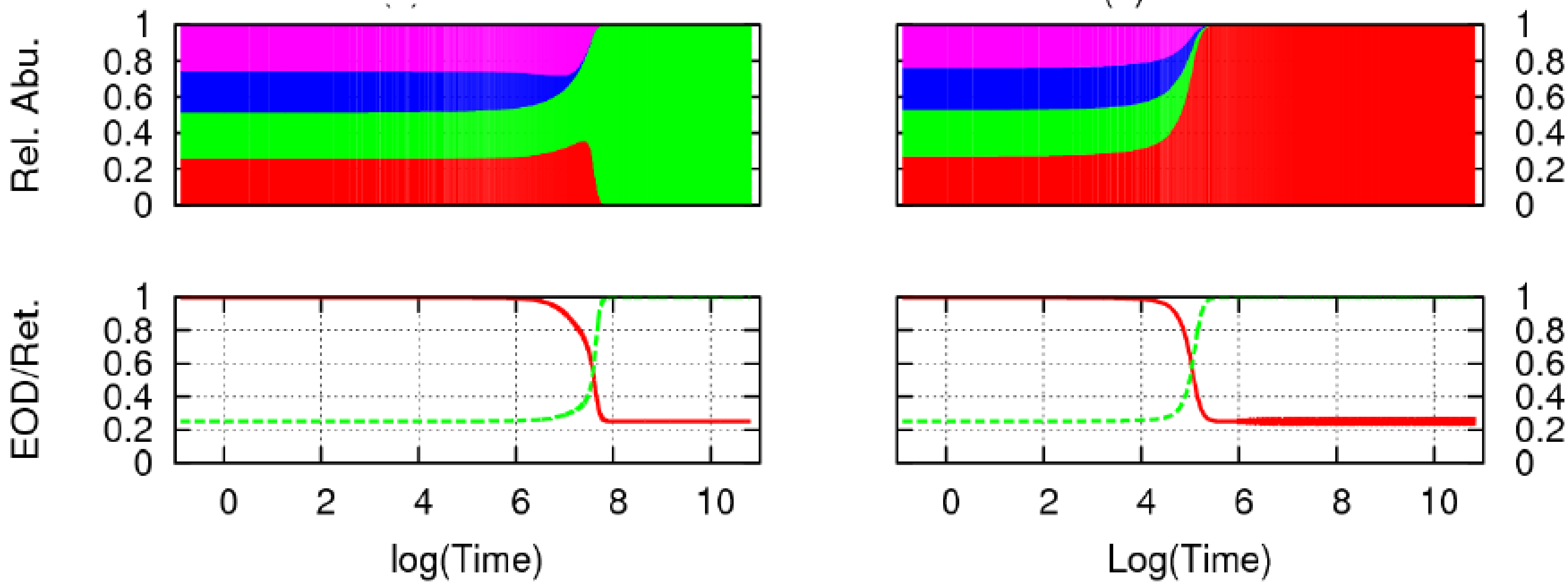
(4 signs)

exponentially faster
for larger meaning
abundance



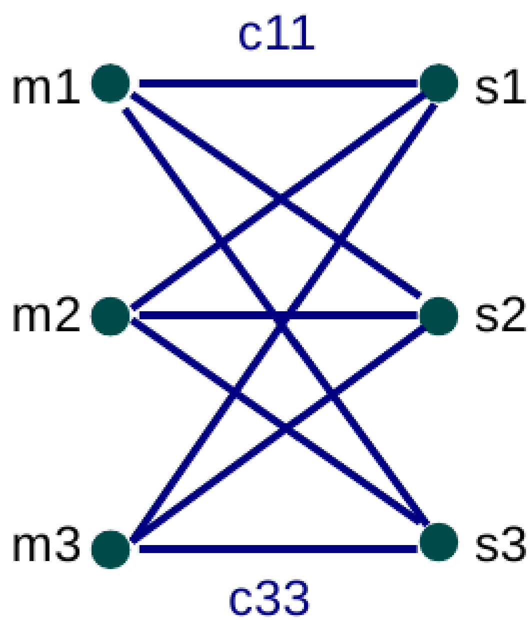
(b) m = 0.1

(d) m = 0.6

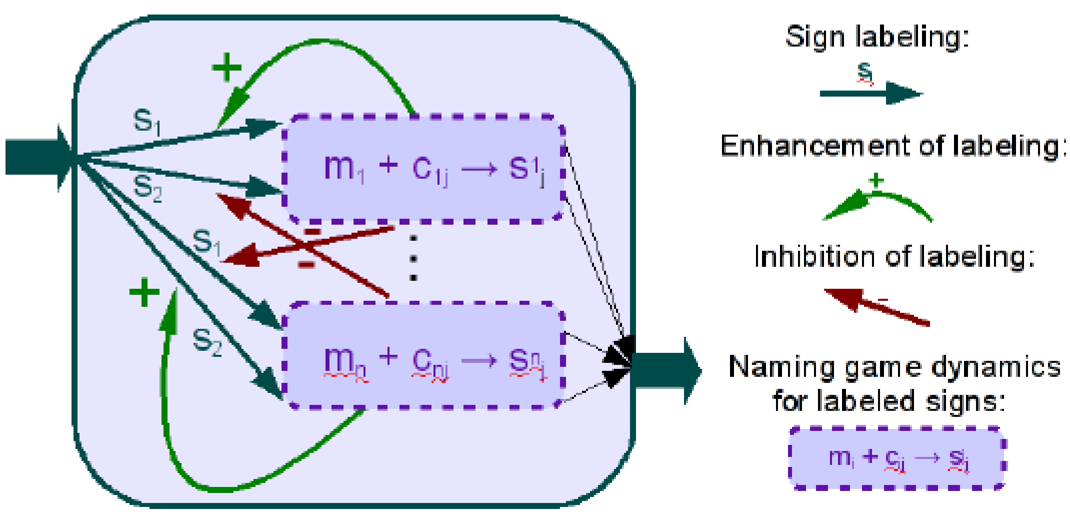


Example2 : The "guessing game"

Species:

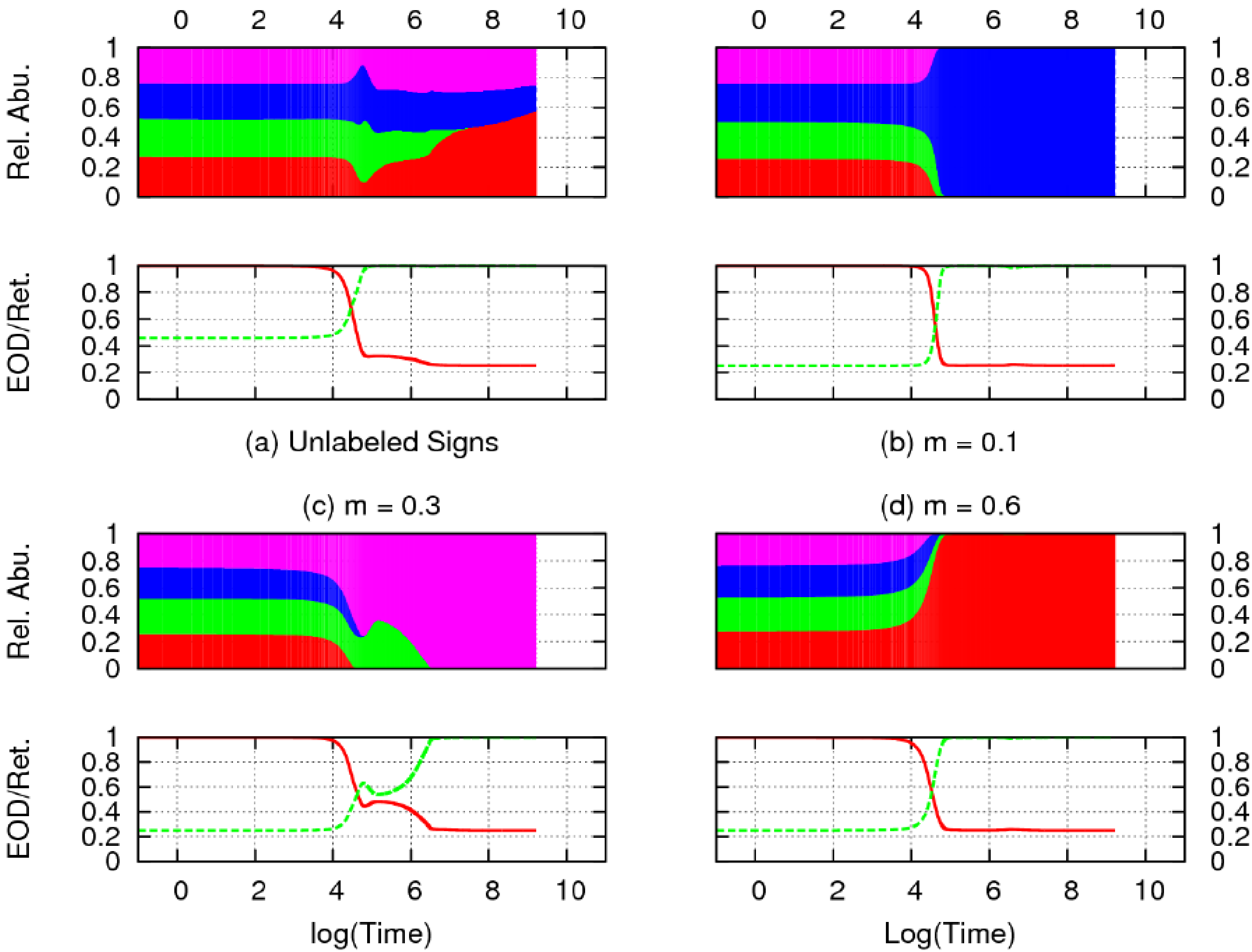


Artificial Chemistry:

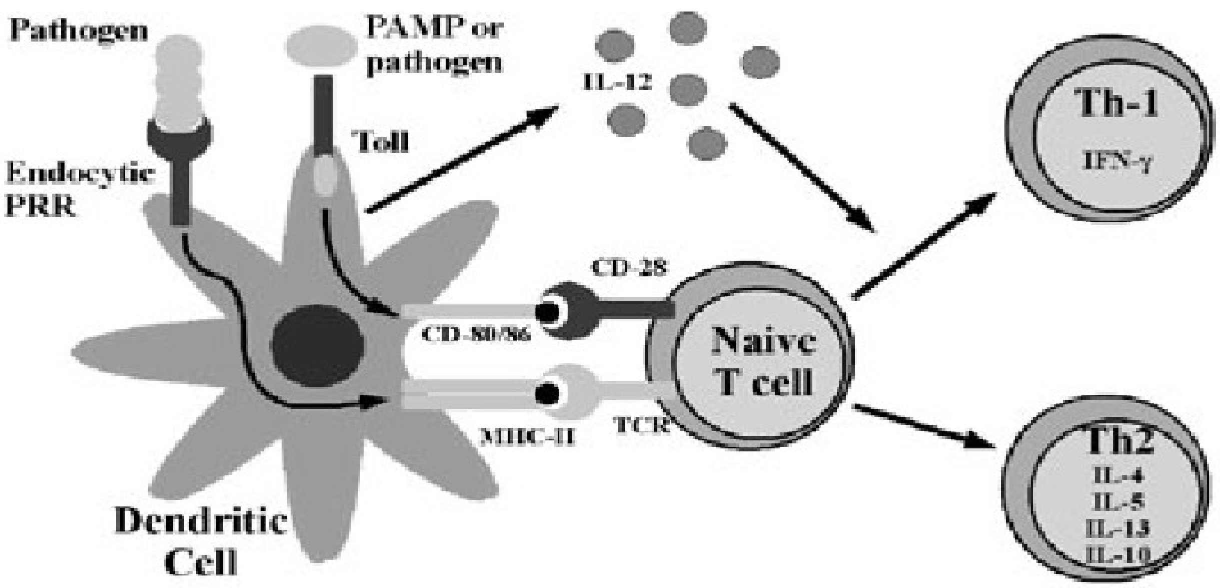


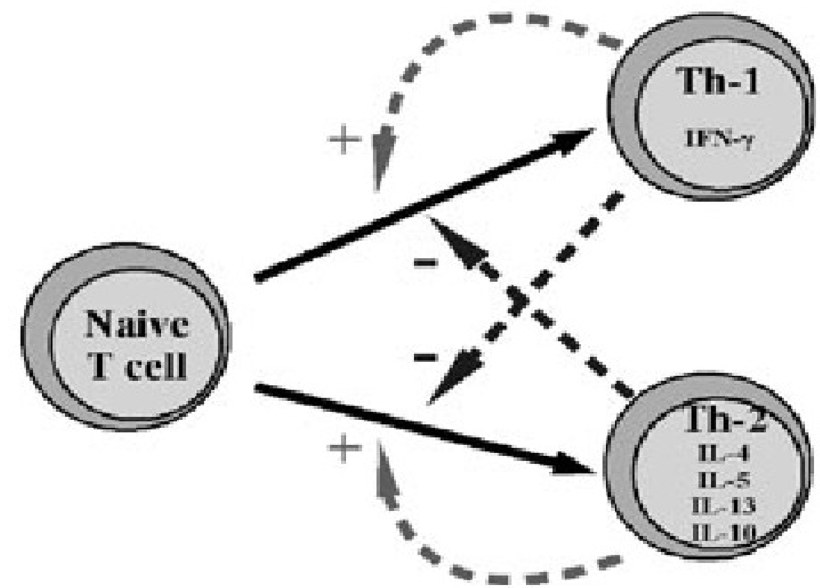
Differential equations:

$$
\begin{aligned}
\dot{\boldsymbol{s}}_j &= +\rho_s(\boldsymbol{s}_j^0 - \boldsymbol{s}_j) + k_r \sum_{i=1}^{N_m} m_i^0 \sum_{j' \neq j} (\boldsymbol{s}_{j'} c_{ij} - \boldsymbol{s}_j c_{ij'}) \\
&\quad -\rho_l \sum_{i=1}^{N_m} \left(m_i^0(\epsilon_1 + c_{ij})(\sum_{i' \neq i} \frac{\epsilon_2}{\epsilon_2 + c_{i'j}})\boldsymbol{s}_j - \boldsymbol{s}_j^i \right)^3 , \\
\dot{\boldsymbol{s}}_j^i &= +\rho_l \left(m_i^0(\epsilon_1 + c_{ij})(\sum_{i' \neq i} \frac{\epsilon_2}{\epsilon_2 + c_{i'j}})\boldsymbol{s}_j - \boldsymbol{s}_j^i \right)^3 , \\
&\quad +m_i^0((1 + \boldsymbol{s}_j^i)c_{ij} - \boldsymbol{s}_j^i) \\
\dot{c}_{ij} &= -m_i^0((1 + \boldsymbol{s}_j^i)c_{ij} - \boldsymbol{s}_j^i) + m_i^0 \kappa_2 (\boldsymbol{s}_j \boldsymbol{\sigma}_c^i - c_j \boldsymbol{\sigma}_s^i)
\end{aligned}
$$


(a) Unlabeled Signs
(b) m = 0.1
(c) m = 0.3
(d) m = 0.6
Rel. Abu.
EOD/Ret.
log(Time)
Log(Time)

Connection to Immune System





c33

Artificial Chemistry: