# How the Singularity of Artificial Intelligence might be achieved
## and why it does not matter

Joscha Bach

# Overview

- The AI singularity posing as the **Silicone Golem**

- **Four pre-conditions** of reaching an AI singularity

- **Functional requirements** resulting from the pre-conditions

- The AI singularity as a special case

# The unexpected dangers of unrestricted PDAs

*Fritz Leiber (1962): "The Creature from Cleveland Depths"*

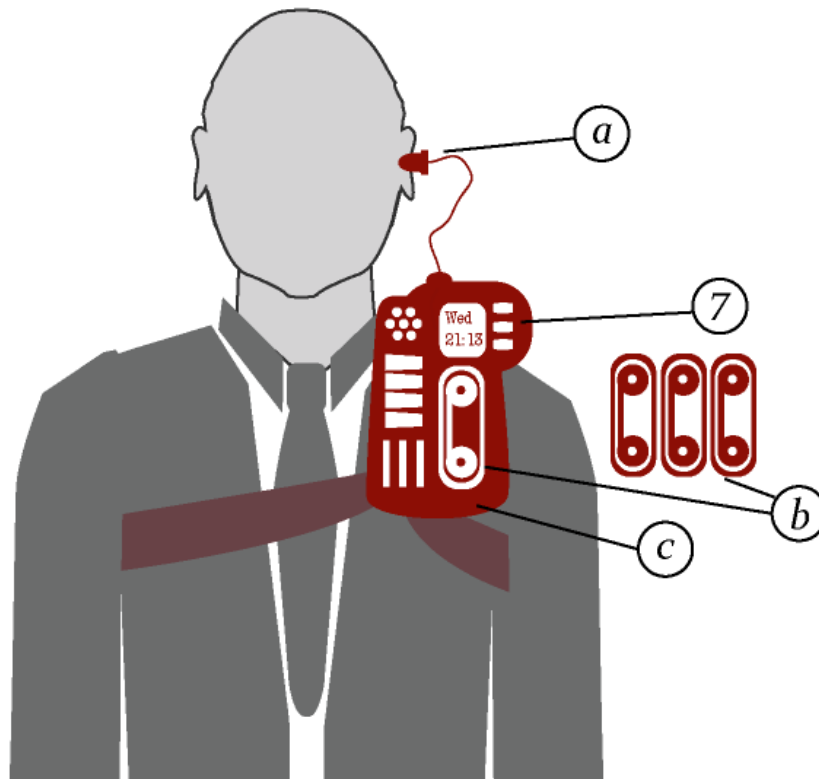The electric **Tickler**: The invention of the Universal Electronic Filofax
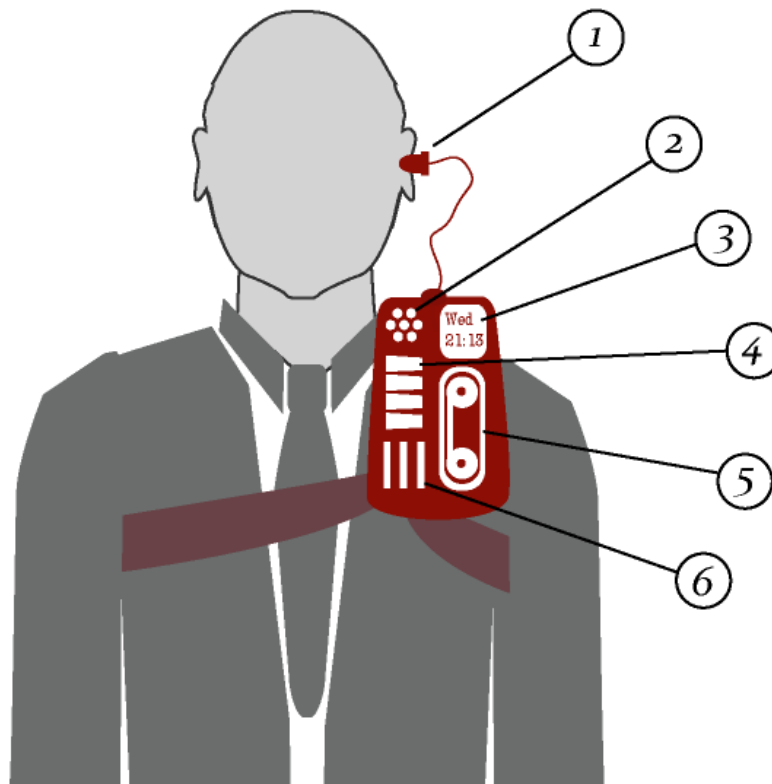


Tickler Mk. I

1. Acoustic Output
2. Microphone
3. Display
4. Manual Switches
5. Magnetic Storage Spool
6. Electric Vibration Alarm

# The unexpected dangers of unrestricted PDAs

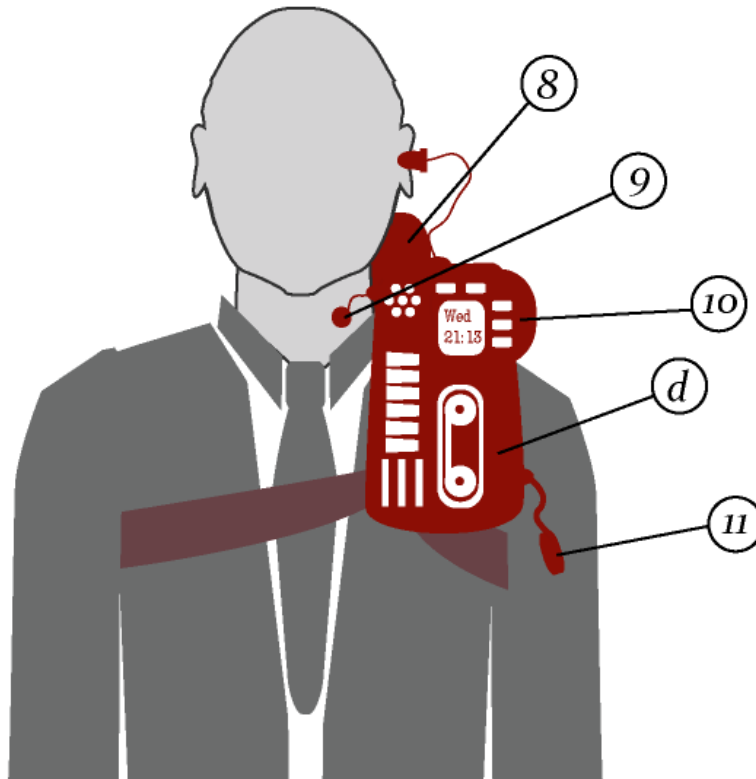Stage II: calculation, pre-defined and automated scheduling



Micro-Systems T i c k l e r ™

7. Automated Scheduling

a. Subluminal Messages
   & voice of your choice!
b. Wide variety of pre-defined
   schedule spools available
c. use at home and at work - 24/7

# The unexpected dangers of unrestricted PDAs

Stage III: tight user coupling, decision making unit



Tickler with Moodmachine™

8. Moodmachine, monitors your well-being, synchronizes with workplace demands
9. Sensors for emotions and stress
10. Decision making unit
11. Injector for endocrine manipulation

d. Ticklers make you more efficient. The next generation Tickler will arrive much sooner!

# The unexpected dangers of unrestricted PDAs

Stage IV: manipulators and self-improvement
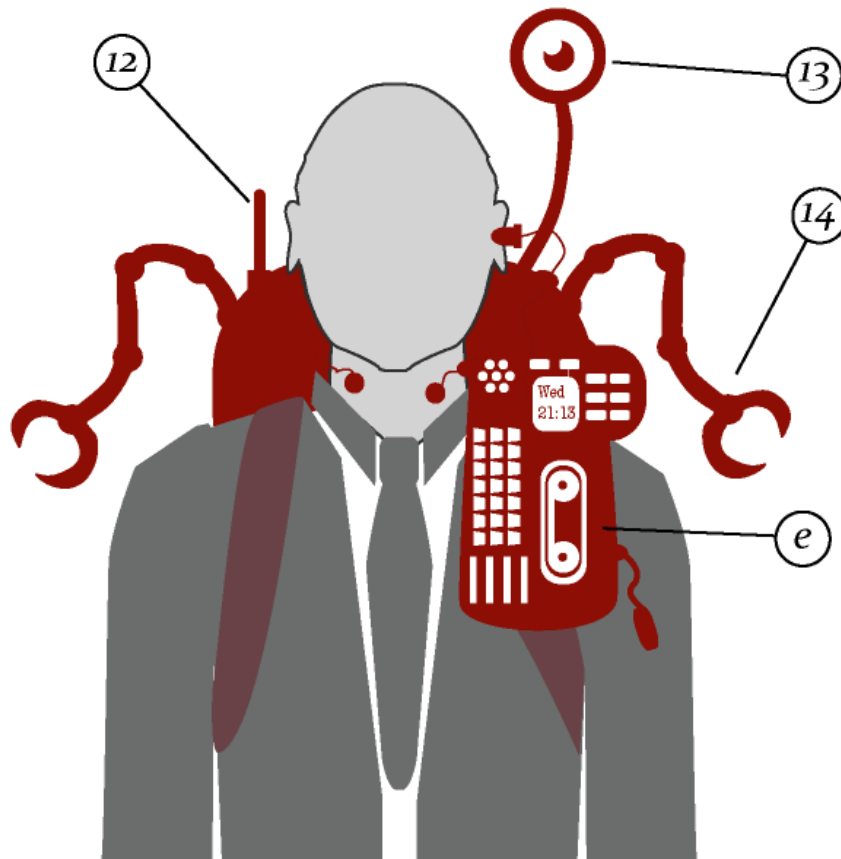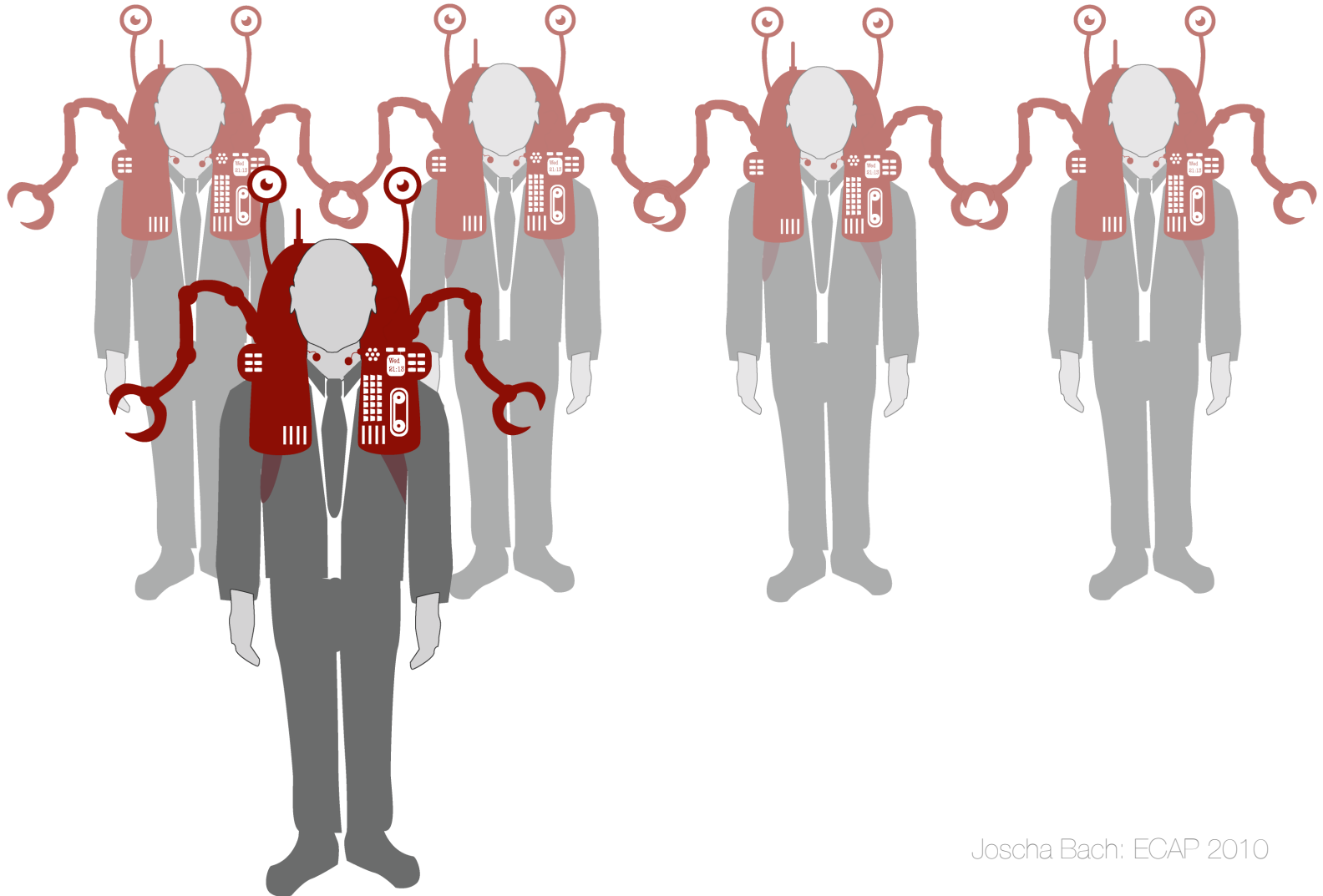


Tickler ™ Mk VIb

12. Wireless communication
13. Camera
14. Manipulators

e. Ticklers become mandantory.
   They can now design and build
   new Ticklers, and each generation
   arrives faster.

# "Resistance is futile"

Stage V: total world domination and human enslavement

# The AI Singularity:
# a modern Golem

October 2010

Joscha Bach, ECAP 2010

# The AI singularity concept



**Singularity**: maximally disruptive cosmological, astronomical, cultural or technological event
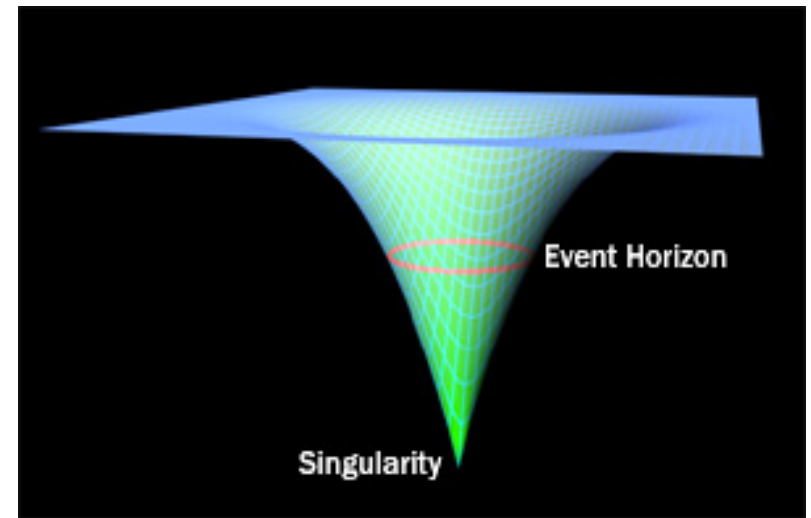
**AI singularity**:

– computational device (or system) reaches a level of intelligence that enables it to create an even more intelligent system

– a cascade of rapidly increasing super-intelligence is created

– the super-intelligence disrupts/transcends existing culture/economic and social order/civilization/humanity

# The AI singularity concept

# Preconditions for reaching an AI Singularity

1.  **Perceptual/cognitive access**: The system must be able to sense and represent its environment beyond a narrow computational sand-box

2.  **Operational access**: The system needs to act, using these models, upon the outside environment

3.  **Directed behavior**: The system autonomously pursues a behavior that includes reaching the Singularity (as direct or indirect) goal

4.  **Resource sufficiency**: The system needs to be implemented in a substrate that supports all the sensing, representing, acting and expanding that goes on while reaching its goal.

# Condition 1: Perceptual (and Cognitive) Access

- Requires perception, interpretation/categorization and representation of environment

- Currently significantly beyond the state of the art of current image recognition, scene interpretation, object representation, ontology generation, …

- Perception/representation is an active, constructive process, it requires **general intelligence** (Voss 2006)

# Condition 1: Perceptual (and Cognitive) Access

Functional requirements to achieve perceptual access:

- perceptual functionality

- universal representations:
  - arbitrarily constructed objects, types, relations and operations
  - grounded in perception and action

- problem solving, abstraction, planning, categorization at least at the level of human ability (so it can conceive of a system beyond its own complexity)

# Condition 2: Operational Access

- System needs **"write access"** upon its environment, and

- **Feedback** to monitor the outcomes of its actions

- In a sandboxed system, there needs to be a way to **escape the sandbox** to reach the **critical environment**

- Contemporary robots may have difficulty to realize the "initial reach", but perhaps a rich data network suffices for colonization

# Condition 2: Operational Access

Functional requirements to achieve operational access:

- access to own substrate, with tools that enable it to

  – redefine its own functionality

  – change its environment to adapt it to its needs

  – the ability to escape its original environmental niche (either by leaving the environment, or by altering it sufficiently)

# Condition 3: Directed Behavior (Motivation)

- System needs a **motivational system** (or its functional equivalent)

- Example of intelligent systems without motivational/goal finding components: search engine (continuously dependent on **external goal-setting**)

- search engine + reasoning/problem solving capabilities
  → "Oracle"

# Condition 3: Directed Behavior (Motivation)

Functional requirements to achieve behavior directed upon
reaching the singularity:

- **agency**, i.e., the ability to *direct its behavior* upon the pursuit of its own goals (i.e., goals that are not externally given)

- **autonomy** in the sense that it is able to *set its own goals*

- a tendency to set its goals in such a way that it relentlessly **increases its abilities** and survivability

# Condition 4: Resource Sufficiency

Computational substrate needs to support:

- necessary computational operations for function of system

- access conditions

- coherence, sustenance and expansion (in terms of complexity, speed, energy)

# Some consequences for candidate AI

- functionalities ≠ individual modules
  - requirements are orthogonal to the **architecture** of actual implementations, but define requirements for this architecture → research questions
  - all requirements are subject of active and productive research

→ AI singularity can not be ruled out, but is uncertain

→ even if AI research delivers ingredients, it might well be possible that no-one will choose to set them up as suggested above, or does so accidentally

# Should we be worried?
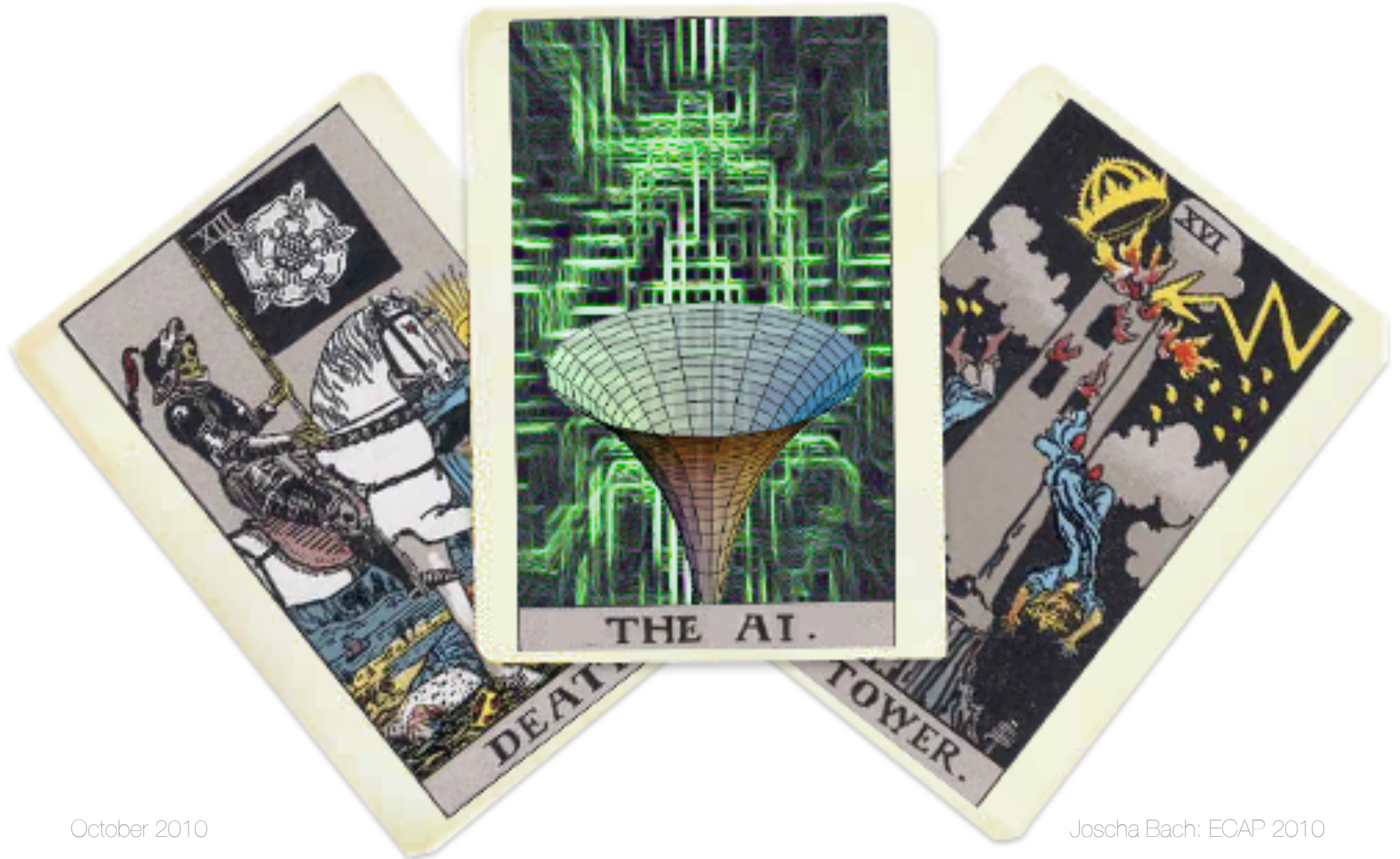


THE AI.

# Why the AI singularity may not matter

- AI singularity is only a special case of an "intelligent system" singularity

- Threat lies not in implementation (silicone, digital computers), but is a functional one:
    - intelligent agent, pursuing a non-human agenda
    - self-improving and survivable; fulfilling the pre-conditions

- Singularity agents may borrow their intelligence from/may emerge over human intelligence

→ Organizational intelligence (Luhmann 1964)

# Why the AI singularity may not matter

Humans share the world with **organizational agents**
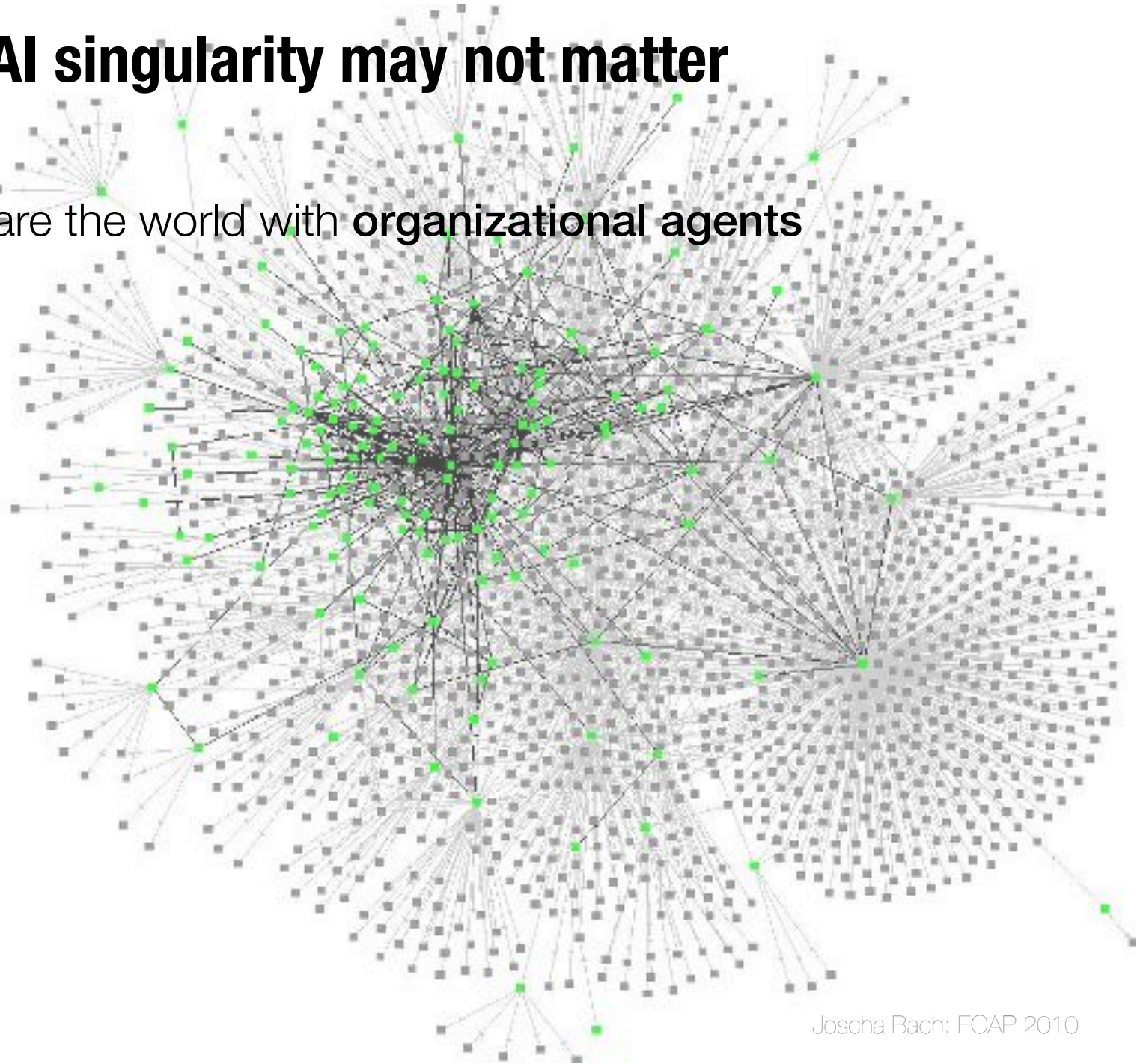
# Why the AI singularity may not matter

- Examples of intelligent, information processing, self-modifying, goal-setting, goal-directed organizational agents:
    - corporations,
    - administrative and governmental bodies,
    - churches,
    - universities

    → more knowledge, problem solving capabilities, longevity and different motivations than humans in their employ
    → not inherently harmful, but vastly more powerful than humans

    → we share the world with a multitude of emerging, growing, competing and dissolving intelligent entities

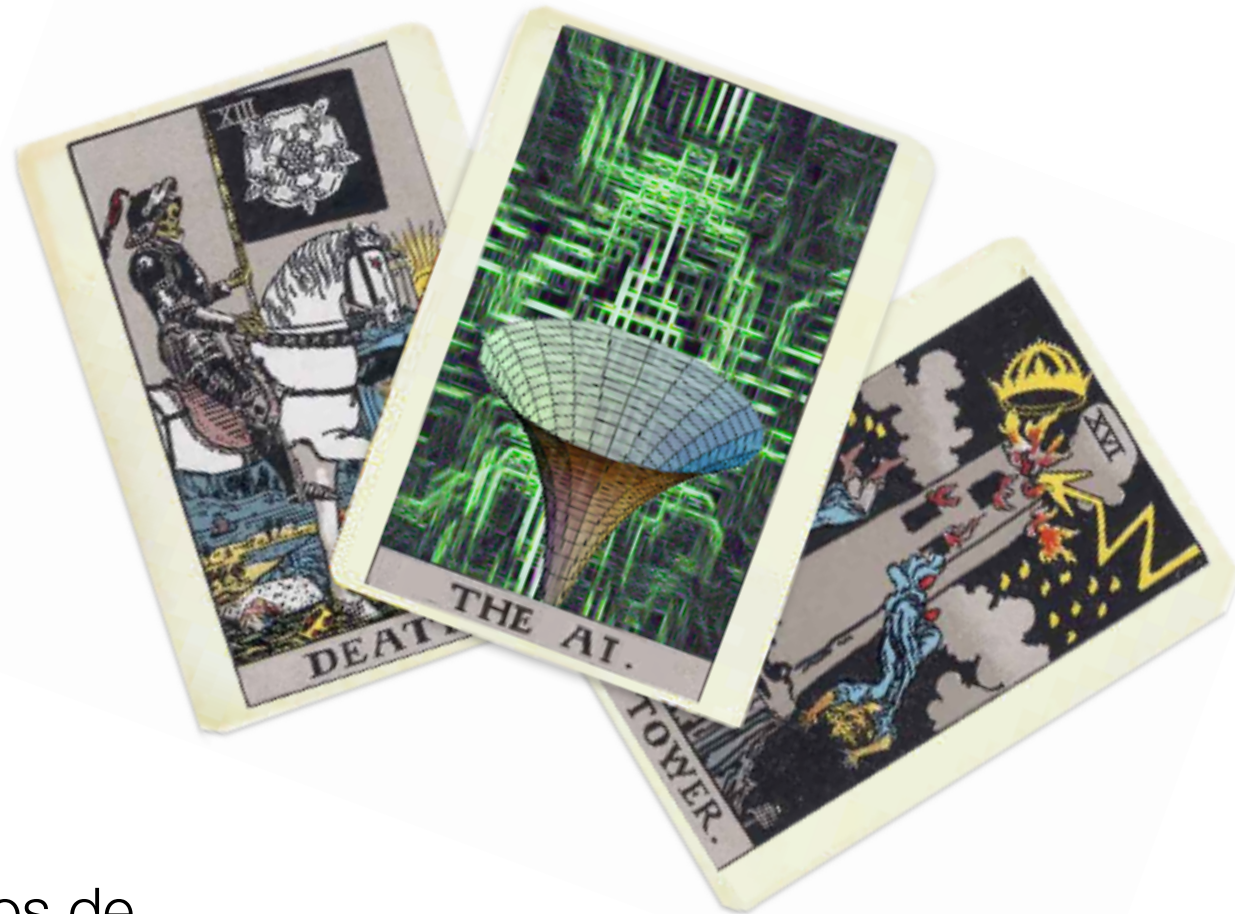→ A **singularity of intelligent systems** has taken place long ago

# Why the AI singularity may not matter

- Singularity AI is not going to be created by a lone "mad scientist" individual in a basement

- AI systems are created by and for organizations, to supply intelligence to them

- AI is already used for knowledge management, decision support, information retrieval, data mining, communication
→ no qualitative, but quantitative change (faster, more efficient information processing in organizations)

→ AI is not going to trigger a new singularity, but supporting the old one

# Summary

- AI singularity is characterized by a disruptive cascade of computational devices of increasing intelligence

- Depends on realization of preconditions (perceptual access/ general intelligence; operational access; behavior directedness; resource sufficiency)

- There is no technical reason that renders an AI singularity impossible

- But an AI singularity is just a special case of an information processing system singularity…

# Thank you!



Joscha Bach

joscha.bach@uos.de