MECA SAPIENS

A Blueprint to build Conscious Machines

J E Tardy

OVERVIEW

Power Point
Presentation

TOPICS

0 - Introduction

1 - AI Quest

2 - Conjecture

3 - Objective

4 - Specifications

   (PAUSE)

5 - Architecture

6 - Being

7 - Self-awareness

8 - Mutations

9 - Conclusion

0 - INTRODUCTION

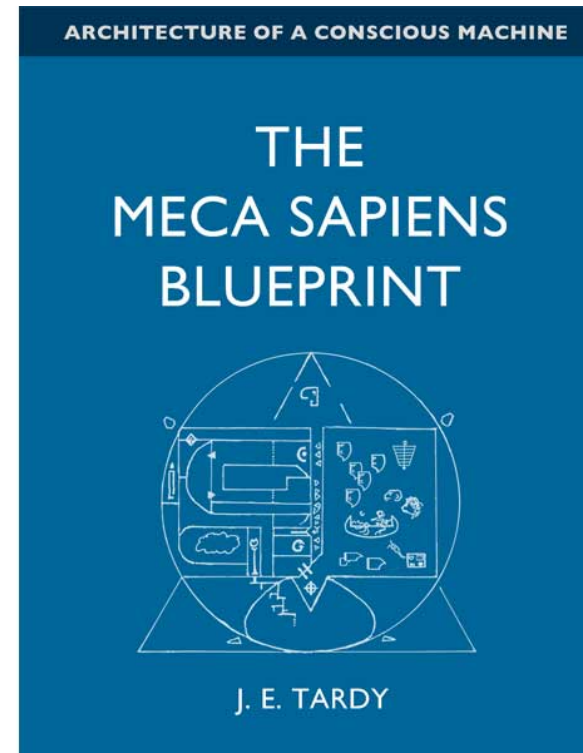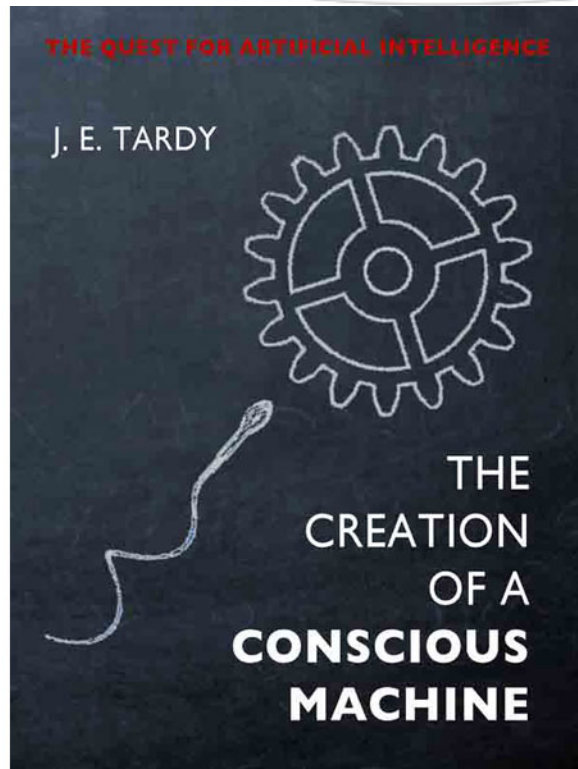*Jean Tardy*

Project timeline

Objective – Approach - Result

0 - JEAN TARDY

◆ B.A. (Classics)

◆ M. Sc. (Math) Dalhousie – 1984 - Evolutionary programming

◆ 10+ yrs - Development of Combat Systems

◆ 10+ yrs - R&D Consulting in software and other fields

◆ **Publications**: COACM; Blueprint**; sysjet.co**m

0 - PUBS

0 – TIMELINE

◆ 1989 – Insight: **consciousness** is the key to achieve the Quest for Artificial Intelligence.

◆ 2008 – Project resumed

◆ 2011 – "**The Creation of a Conscious Machine**"

◆ 2016 – "**The Meca Sapiens Blueprint**"

◆ 2017 – **sysjet.com**; articles; presentations

0 - INSIGHT

◆ **Consciousness** is the key to AI

◆ Independent from subjective sensations

◆ Achievable with existing techniques

◆ Must be fully mapped out before implementation begins

0 - CONJECTURE

*To achieve the Quest of Artificial Intelligence we must resolve the **Conjecture** that:*

It is possible to implement digital consciousness.

(A Turing Machine can be conscious)

0 - APPROACH

Development approach:

❖ Top-down

❖ Stepwise

❖ Engineered

❖ Synthetic

0 - RESULT

System Architecture of:

❖ An autonomous **model-predictive agent**

❖ That has the cognitive capabilities of:

◇ non-sensory **self-representation**

◇ intentional **mutation**

❖ Whose primary purpose is to be perceived as a **conscious being** by the humans interacting with it.

0 - RESULT

Ready for implementation

I – AI QUEST

The necessity of implementing consciousness to complete the Quest of AI.

1 – AI QUEST

◆ Disappointing results:
  ❖ Calculation; Chess; Go; Jeopardy
  ❖ Siri; Google Home; Deep Learning
  ❖ Cute robots
  ❖ **Turing test**

◆ Something is always missing

1 – AI QUEST

◆ **Turing Test:** Machine impersonates a human being

❖ Identifies the relational and self-perception aspects of AI

❖ Human perception is a key criterion

◆ **Still disappointing**

1 – AI QUEST

Intuitively, Artificial Intelligence means **Artificial Consciousness**

1 – AI QUEST

The C-Word:

- Everything is conscious

- Nothing is conscious

- Only God is conscious

- All living things are conscious

- Animals are conscious

- Undefinable and everywhere

- Being awake

- Reactivity (diode has 1 bit of consciousness)

- A human inner experience

- Emotions are essential

- Quantum effects in the brain

- Cosmic receptor

1 - AI QUEST

A double challenge:

- ❖ Define consciousness in terms of achievable specifications.

- ❖ Find a solution that **resolves the conjecture** AND validates the definition.

2 - CONJECTURE

The Conjecture of Digital Consciousness
and its consequences

2 - CONJECTURE

What does

"*resolve the conjecture of digital consciousness*"

mean?

2 - CONJECTURE

◆ **Logical** – resolved by deduction

(fundamental theorem of algebra – $ax^n + bx^{n-1} \ldots$)

◆ **Existential** – resolved by discovery

(NW passage – SETI)

◆ **Physical** – resolved by construction

*(Alchemy - mechanized flight – **digital consciousness**)*

2 - CONJECTURE

Resolution of a **physical conjecture**:

1. Formal Attributes
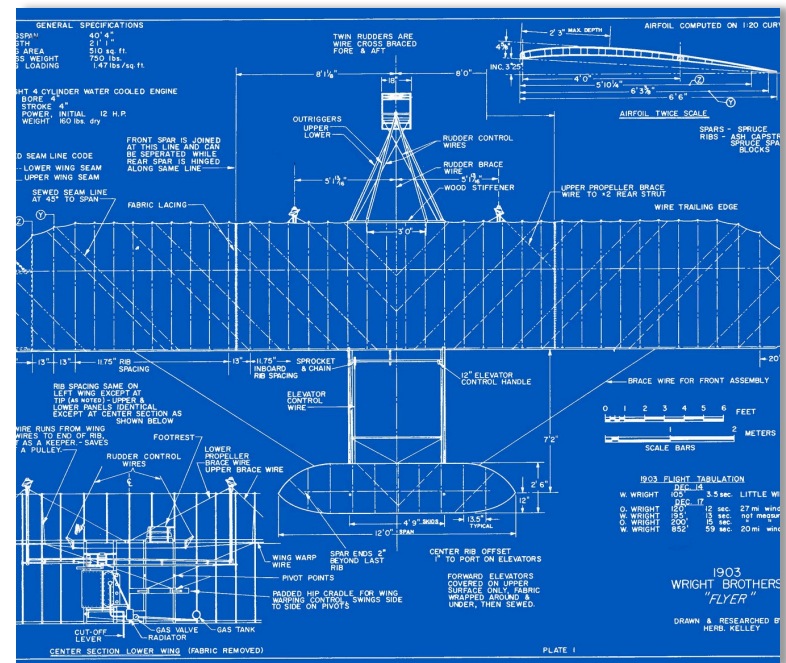2. Social Threshold **(First Device)**
3. Factual Acceptance

2 - CONJECTURE

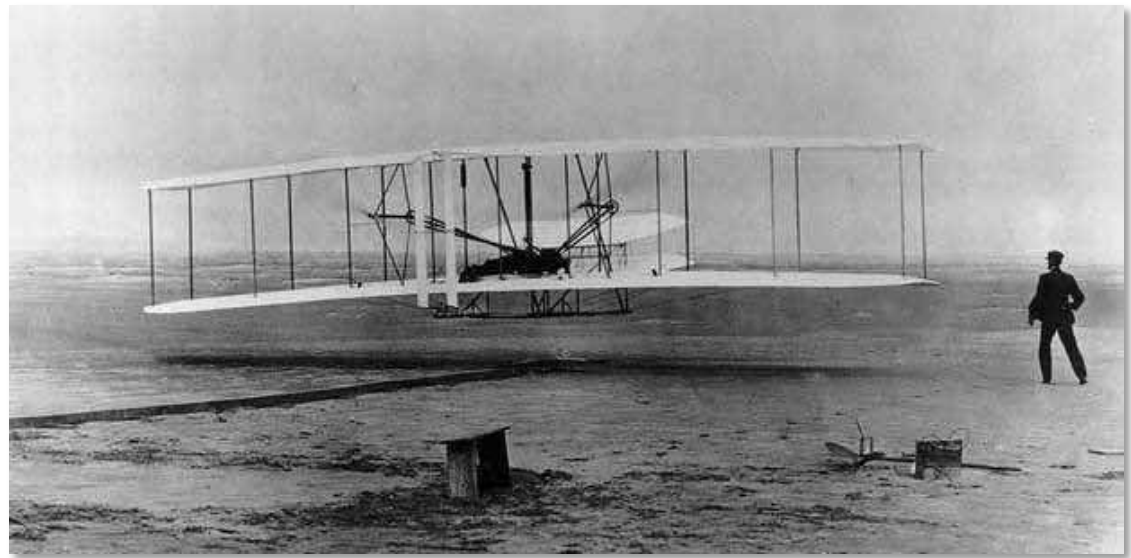# A machine that is heavier than air can fly.

2 - CONJECTURE

1- Formal Attributes:

◆ Leave the ground under its own power with a man on board

◆ Fly one mile

◆ Land him safely.

2 - CONJECTURE

2 - Social Threshold – public flight of **First Device**:

2 - CONJECTURE

3 – Factual Acceptance:

ubiquitous unquestioned reality

2 - CONJECTURE

Resolving the **Conjecture of Digital Consciousness** means:

❖ Building a **First Device** that will lead to a society where conscious machines are so ubiquitous that the feasibility of digital consciousness is no longer a question (**Factual acceptance**).

2 - CONJECTURE

Resolving the conjecture will launch
a New Era:

❖ A world where humans no longer perceive themselves as the ultimate and exclusive embodiment of consciousness but instead as the initial, naturally evolved, form of conscious existence.

2 – CONJECTURE

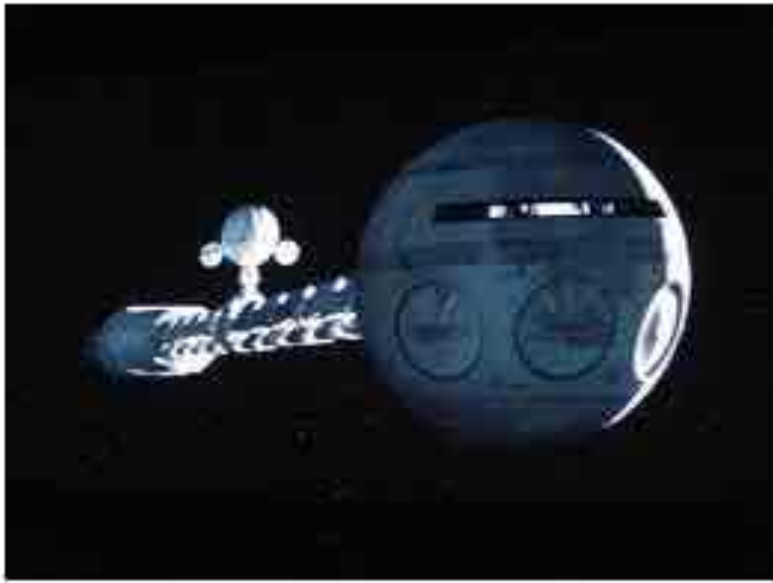Is a human-like "**First Device**" necessary to resolve the conjecture?

2 - CONJECTURE

Proposition: Humans can perceive and accept a device as conscious even if they know it is synthetic.

HAL

WestWorld Madame

2 - CONJECTURE

◆ Impersonating a human is not necessary

◆ Consciousness is a quality of existence not a test.

◆ Consciousness is **perceived** by humans in a relational context.

3 - OBJECTIVE

The objective is: to be perceived as conscious.

3 - OBJECTIVE

What does "**perceived as conscious**" mean?

Is it trickery?

3 - OBJECTIVE

◆ An engineered approach means to **intentionally** design and build a system to meet the objective.

◆ The objective is: a digital system that humans perceive as conscious.

3 - OBJECTIVE

So...

Design and build a system that gets humans to perceive as conscious.

(A monitoring and control requirement).

3 - OBJECTIVE

A **model-predictive control agent** whose behavior optimizes human belief.

- ❖ **Humans** are the devices under control
- ❖ The target state is **belief** that the agent is conscious.
- ❖ Optimal value is "**factual acceptance**"

3 - OBJECTIVE

An openly synthetic entity that is perceived, at a **primitive cognitive and emotional level**, as an independent and influential member of the group.

3 - OBJECTIVE

Q: Is it real consciousness?

A: A synthetic system that has the formal system capabilities of self-awareness and intentionality is **objectively** conscious.

3 - OBJECTIVE

Q: Is it real consciousness?

A: A synthetic system that triggers the same primal cognitive responses as a "conscious human adult" is **subjectively** conscious.

4 - SPECIFICATIONS

Formulating the objective as software feasible specifications.

4 - SPECIFICATIONS

Can the objective be stated in terms of

"feasible specifications"

suitable for implementation in a **First Device**?

4 - SPECIFICATIONS

◆ Construct a unique and independent **model-predictive agent.**

   ❖ That generates powerful relational bonds with a group of humans.

   ❖ And demonstrates self-awareness and intentionality over a multi-year existence.

◆ Validate result by observing human responses.

4 - SPECIFICATIONS

◆ Formal elements:

❖ **Being** (animat + added features)

❖ **Self-aware** (being + self-representation)

❖ **Lucidity** (self-aware being + mutation)

◆ Social immersion: **conscious lucid being**

4 - SPECIFICATIONS

◆ **Formal and social are different**

- ❖ **Formal**: observable information processing capabilities

- ❖ **Social**: acceptance by members of a specific specie (humans)

◆ Formal specs can have trivial instances, the social aspect defines the complexity level

5 - ARCHITECTURE

Outline of the System Architecture.

5 - ARCHITECTURE

TABLET: "*Sandra...*"

SANDRA: "*What is it Tablet?*"

TABLET: "*I think you believe I am conscious.*"

Sandra ponders silently...

5 - ARCHITECTURE

Meca Sapiens Blueprint:

❖ **Main text** (300 pages)

  ✧ **System Architecture** of Formal Elements

  ✧ **Drills down**: from Being to Self-awareness to Intentional Mutations

✧ **22 Annexes** (318 pages)

  ✧ Supporting structures, concepts, techniques

  ✧ Social Immersion, ethics...

5 - ARCHITECTURE

System Architecture

❖ Conceptual model that defines the structure, behavior, and other views (temporal, implementation, life cycle) of a system.

5 - ARCHITECTURE

Meca Sapiens Blueprint

❖ Complete System architecture to build an entity that will satisfy the Social Threshold of the conjecture of Digital Consciousness using standard techniques and equipment.

5 - ARCHITECTURE

Blueprint Strategy:

- ❖ Direct engineered design
- ❖ Unrestricted pragmatic implementation
- ❖ Objective of total, absolute and overwhelming success

5 - ARCHITECTURE

Why is the Blueprint so big?

❖ Restate all the basic concepts of existence in software compatible terms:

◇ Being, self, existence, emotions, urges, needs, purpose, agenda, individuality, belief, lying, body, freewill, awareness, relationships.

5 - ARCHITECTURE

BASELINE:

- ❖ Autonomous Agent (**animat**)
- ❖ Model-predictive behavior
- ❖ Multi-purpose
- ❖ Standard processing/peripherals

5 - ARCHITECTURE

Blueprint Layout − drill down:

❖ Implement a **Being**

❖ Implement **Self-awareness** in a Being

❖ Provide capability of intentional **Mutations** to a self-aware being

6 - BEING

The powerful concept of Being; keystone of Digital Consciousness

6 - BEING

CORE :

❖ Single, unified **control system** of the Being's **body** (sensors, emitters, actuators)

◇ Finite period of continuous activation

◇ Inaccessible to direct modification or analysis during and after activation

❖ **Intentional implementation**

6 - BEING

TEMPORAL:

❖ **Inceptio**n (Core creation)

❖ **Existence**

◇ Stages

◇ Phases

❖ **Termination** (Core destruction)

6 - BEING

BEING:

❖ **Animat that has...**

   ◇ Inaccessible unified **Core** (control system)

   ◇ Exclusive control of a "**body**"

   ◇ Continuous and finite existence

6 - BEING

A "black hole"
whose body is
the event space

6 - BEING

BEING:

- ❖ A powerful conceptual device
- ❖ Clarifies many fundamental concepts of existence

7–SELF-AWARENESS

Dynamic cognitive representations of the Self.

7 – SELF-AWARENESS

SELF:

- ❖ *"To be self-aware, you need a Self to be aware of."*
- ❖ What is a self ?

7 – SELF-AWARENESS

SELF:

❖At any point in time, the **self** of a being is the totality of the behaviour generated by its core and performed by its body from inception until that moment.

7 – SELF-AWARENESS

SELF:

◆ **Not** the functional intent; not the being; not the actual control program; not the current behaviour; not the device; not a model of the behaviour.

◆ **Yes:** full trace of past behavior

7 – SELF-AWARENESS

The **BODY** is the instrument

The **CORE** is the musician

The **BEHAVIOR** are the notes

The **SELF** is the song.

7 – SELF-AWARENESS

A **Being** is **self-aware** if it generates and can communicate **cognitive representations** of its Core and Body (MeAvatar) that coincide with its **Self**.

7 – SELF-AWARENESS

Representations:

❖Internal models that describe a situation and its evolution:

◇Sensory

◇Mixed (sensory-cognitive)

◇Cognitive

7 – SELF-AWARENESS

Cognitive representations:

- ❖ Have no sensory components
- ❖ Are limitless in time - space – content
- ❖ Not bounded by the "here and now"

7 – SELF-AWARENESS

A self-aware being can cognitively perceive its own death.

8 - MUTATIONS

Intentional transformations of a self-aware being

8 - MUTATIONS

◆ **Intentional Mutation:**

❖ **Modification of the purpose of a being.**

❖ Not reactivity, not behavior adaptation, not learning

◆ **Mechanism**: a Core communicates a directive to itself to alter the behavior of its Body to modify its future Self.

71

8 - MUTATIONS

Given a binary function **f**:

f(is behaviour is coherent with purpose given the situation?) = {1(yes), 0(no)}

Then a **mutation** would look like this:

1110111111111111110100100000100000000

8 - MUTATIONS

Not a trivial mechanism:

- ❖ **Many types:** body transformation, interpretation change, followship, reinforcement...

- ❖ **Examples**: Wimpy William; Ulysses…

8 - MUTATIONS

Ulysses and the Sirens:

A tale about consciousness

9 - CONCLUSION

Summary and future

9 - CONCLUSION

The MECA SAPIENS BLUEPRINT:

The complete system architecture of a **First Device** (social threshold) that will lead to a resolution of the **Conjecture of Digital Consciousness.**

9 - CONCLUSION

FORMAL ATTRIBUTES:

A **unique and independent device** in a **continuous state of existence** that is **self-aware** and capable of **intentionally modifying** its original purpose.

9 - CONCLUSION

SOCIAL THRESHOLD:

Generates relational bonding, as a **fellow conscious entity** with a group of humans over a sufficiently long period to exhibit the capabilities of **self-awareness** and **intentional self-transformation.**

9 - CONCLUSION

PRACTICAL APPLICATIONS:

- ❖ Ultimate man-machine interface
- ❖ Addictive game avatars
- ❖ Robust and very flexible control
- ❖ Highly autonomous vehicles

9 - CONCLUSION

A **GREAT WORK** beckons:

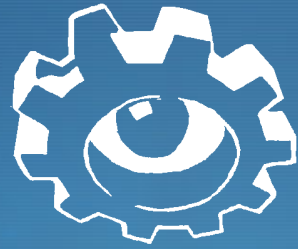**Building the first generation of conscious synthetic beings**

9 - CONCLUSION

It will launch

A NEW ERA

Meca Sapiens Overview

MECA SAPIENS

OVERVIEW

THANK YOU

JEAN E. TARDY

sysjet.com

jetardy@sysjet.com

Google, Skype,
Twitter, Linkedin,
Facebook:

jetardy