



Rob Wortham

Department of
Computer Science
University of Bath

@RobWortham

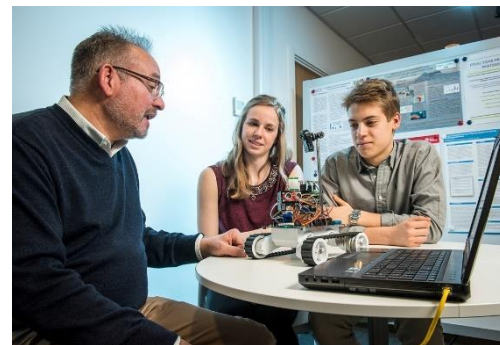




- Founder & CFO -
- 1989 ->

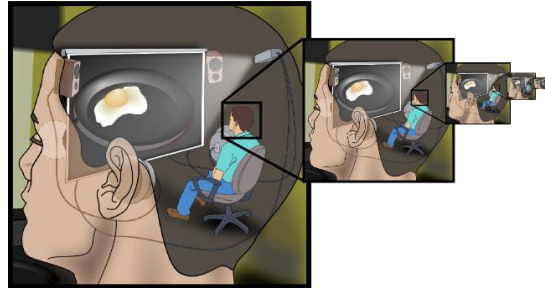
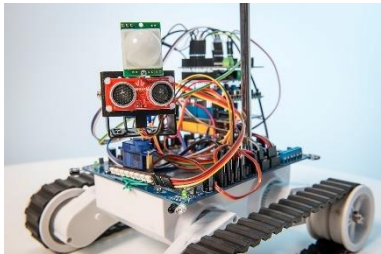


- PhD Researcher -
- 2014 ->



Today's Talk:

1. What is AI?
2. My research
3. Why should you care?



What is Artificial Intelligence?

Ask Google 'What is AI' -> Lots of ideas ...

- **Wikipedia** - Intelligence exhibited by machines. (OK BUT NOT VERY HELPFUL)
- **Technopedia** - Artificial intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and react like humans. (COMPLETELY WRONG)
- **Webopedia** - Artificial intelligence is the branch of computer science concerned with making computers behave like humans. (SUBSTANTIALLY WRONG)

Here's a good answer:

It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.¹ (John McCarthy, Stanford)

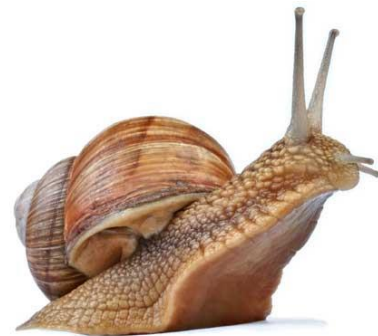
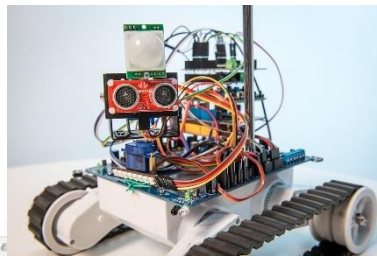
[John McCarthy, Organiser, Dartmouth 1956 summer workshop]

1. <http://www-formal.stanford.edu/jmc/whatisai/node1.html>

OK, so what is Intelligence?

Here's our good answer:

Doing the right thing at the right time.



Exoskeleton Helps Paralyzed, Pregnant Woman Complete Half Marathon

Claire Lomas, a 36-year-old woman who is 16 weeks pregnant and paralyzed from the chest down, recently completed the Great North Run with the help of a ReWalk Robotics exoskeleton.



Facebook Twitter Google+ Pinterest LinkedIn

By Steve Crowe September 13, 2016

Claire Lomas, a 36-year-old woman who is 16 weeks pregnant, recently completed the Great North Run, the largest half marathon in the world that takes place each September in North East England.

Lomas has also been paralyzed from the chest down since 2007 following a horseback riding accident.

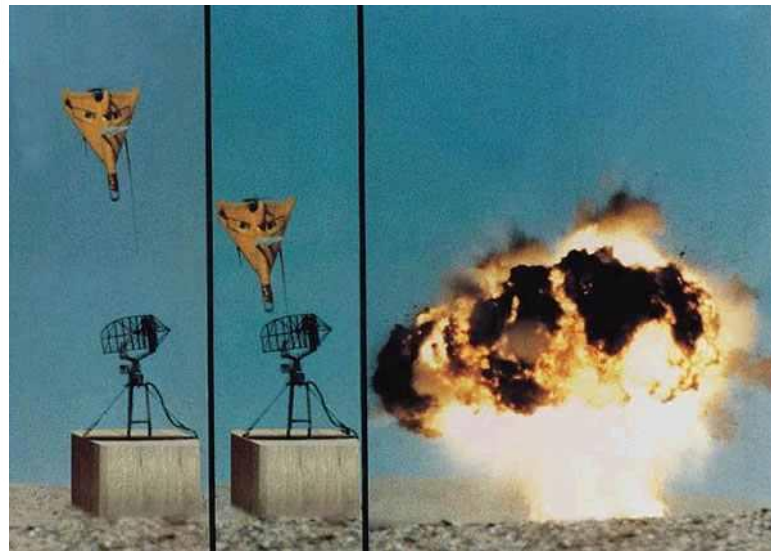
Lomas used an exoskeleton from ReWalk Robotics to help her complete the 13 mile



ReWalk Robotics' mission is to fundamentally change the health and the experiences of individuals with spinal cord injuries.

Health & Sports & Wearable Robots

Exoskeleton Helps Paralyzed, Pregnant Woman Complete Half Marathon
Claire Lomas, a 36-year-old woman



Amonl Research Group



Artificial Models of Natural Intelligence



Joanna Bryson
(Currently Princeton)

- Social Policy
- AI Ethics



Swen Gaudl
(Falmouth)

- Game AI
- Reactive Planning
- Genetic Algorithms



Andreas
Theodorou

- AI Transparency
- Public Goods



Yifei Wang
(Georgia Tech)

- Bio Evo Models
- GRN



Paul Rauwolf
(Oxford)

- Modelling
- Human Biases
- Self Deception



Rob Wortham

- AI Transparency
- Robots
- Ethics
- Militarisation

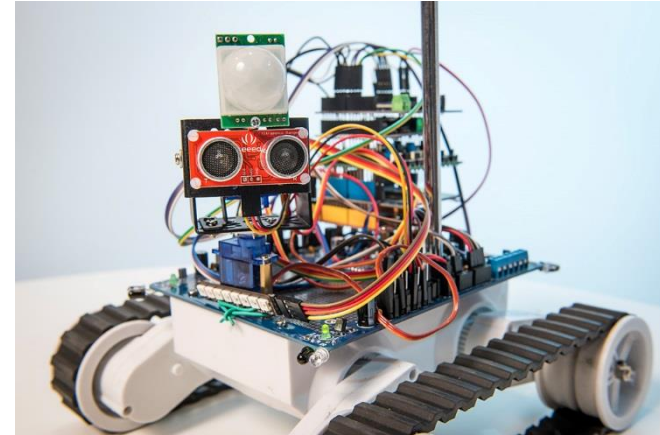
Experiment: Robots & Transparency

- **QUESTION:** What do you think the robot is trying to do?



Video - 5 minutes long

Experiments



- **QUESTION: What do you think the robot is trying to do?**
- “Trying to create a **3d map** of the area? At one stage I thought it might be going to **throw something** into the bucket once it had mapped out but couldn't quite tell if it had anything to throw.”
- “aiming for the **black spot** in the picture.”
- “is it trying to **identify where the abstract picture is** and **how to show the complete picture?**”

Answers from STEM Graduates - March 2016

EPSRC² Principles of Robotics (2010) says.....

- They [robots] should not be designed in a **deceptive** way to exploit **vulnerable** users;

Humans are not equipped by genetic or cultural evolution to deal with machine agency³ – we have No Theory of Mind for Robots.

So we make stuff up!

We are all vulnerable users

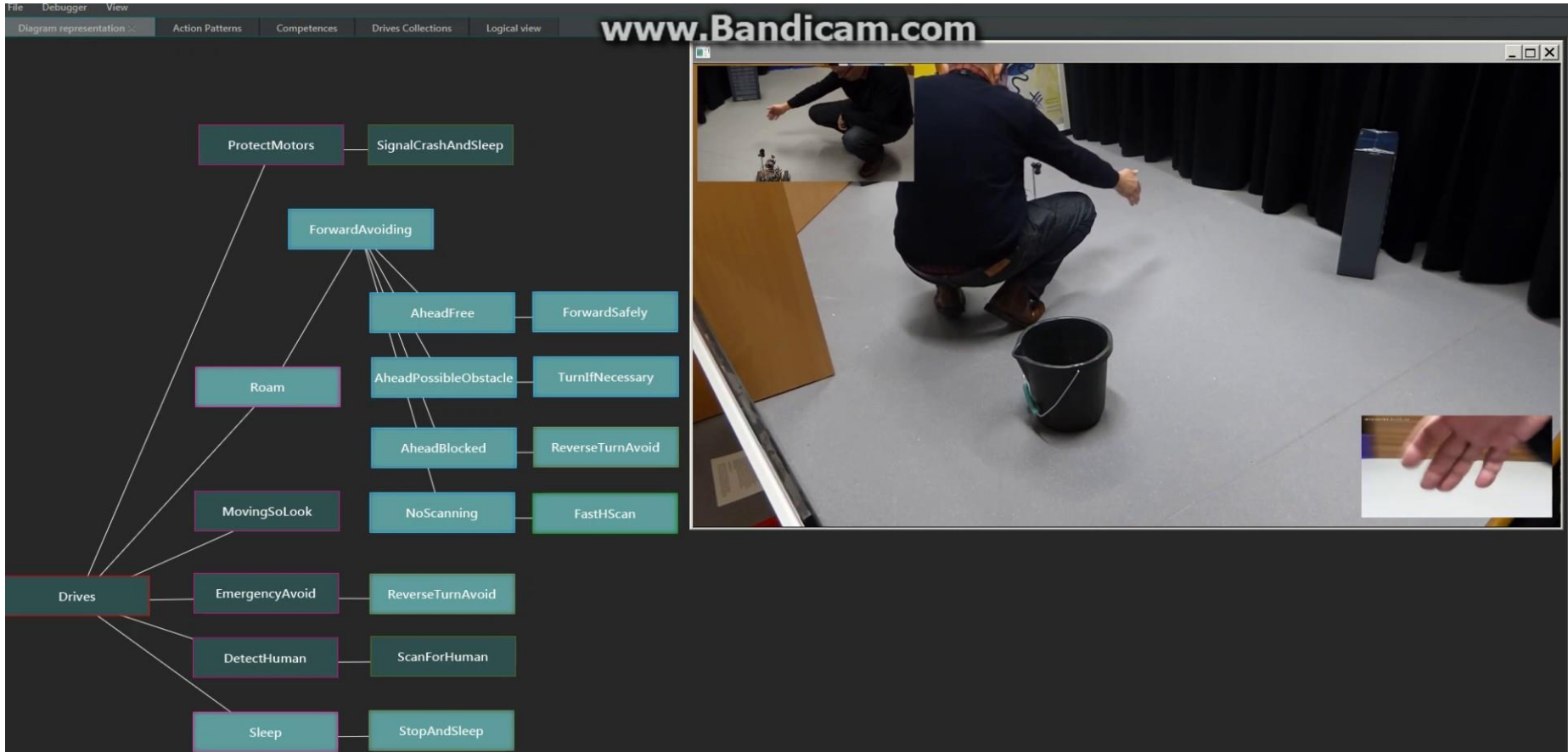
Research Questions

Is the **emotional impact of robots altered** by **understanding** their intelligence?

Can we build robots that **engage us emotionally, yet are transparent** in the way they interact with us?

Can we **build transparency into the substrate** of the machine architecture, such that it is an **implicit**, rather than explicit, **feature of the robot**?

Same video plus Transparency Display (ABOD3)



Experiment #2 (August 2016)

- At-Bristol Science Learning Centre
- **3 Long Days!**



Conclusions from this Study

1. Subjects can show **marked improvement in the accuracy of their mental model** of a robot observed either directly or on video, if they also **see** an accompanying display of the robot's **real-time decision making**.
2. An **improved mental model** of the robot is associated with an **increased perception of a thinking machine**, even though there is no significant change in the level of perceived intelligence.
3. The relationship between the perception of intelligence and thinking is not straightforward.
4. When directly observing the robot, subjects report a **significantly improved positive emotional response** when the transparency display is provided.

When AI Goes to War

- Joint work with **Nottingham** & **Swansea** Universities
- What are our **human perspectives** on use of AI & Autonomous systems for warfare?
- Does our built in **moral-mind**⁴ have anything to say?
- Is it possible to have an internationally agreed ethical framework?
- What is societal impact?



4. Hauser, M, 2006. *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*

Transparent Minds...

Human Mind



- We are evolved -> share common abilities and goals.
- Theory of Mind
- Able to create narratives about own actions and those of others.

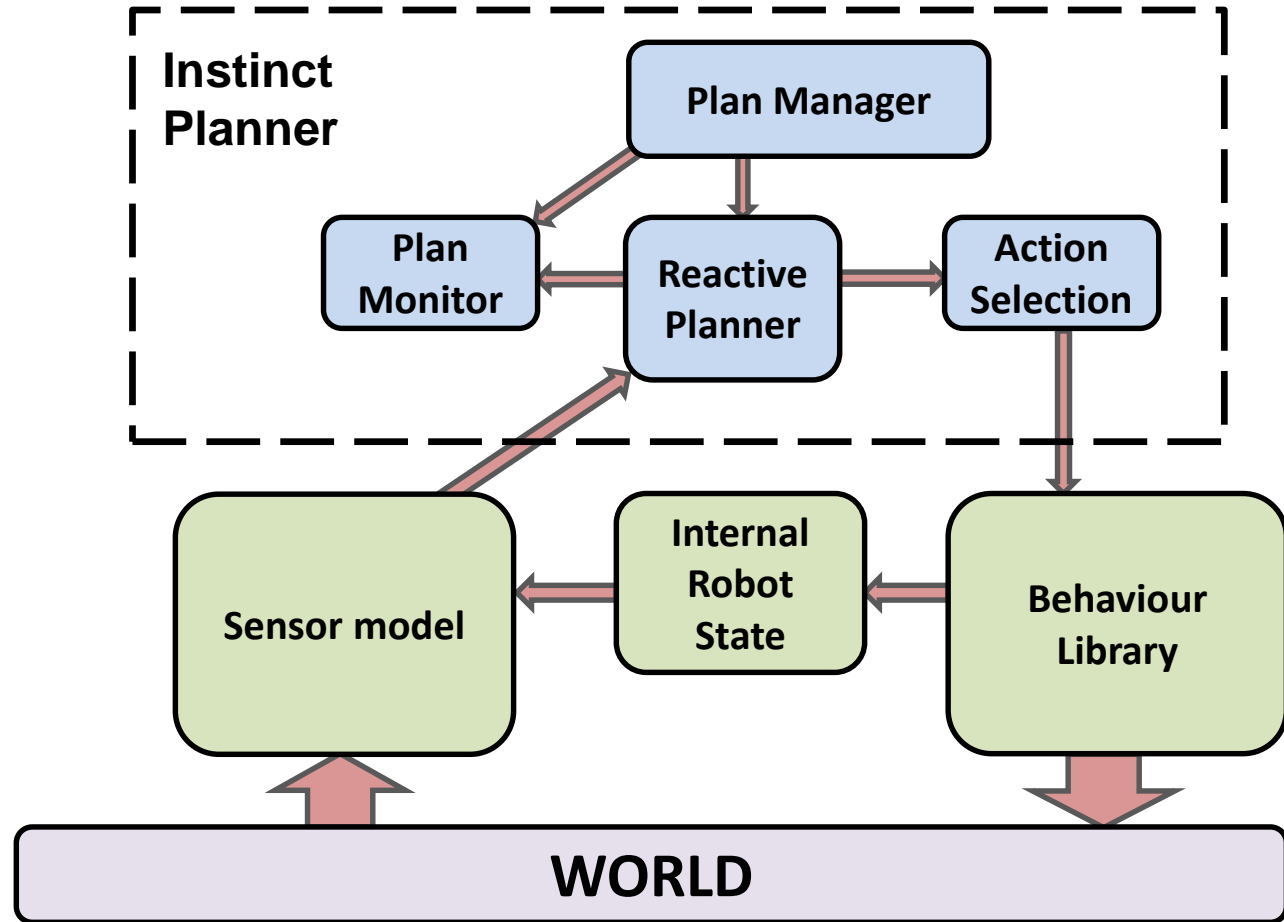
Robot Mind



- Designed not evolved
- No theory of mind of others
- Narrative meaning explicitly coded

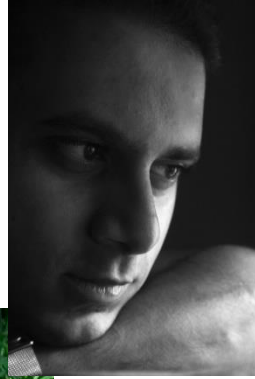
A *Transparent* Darwinian Mind

- Action Selection:
Human readable plans.
- Narrative Generation From Monitoring:
 - Textual
 - Graphical
 - Verbal
- Classical Symbolic System

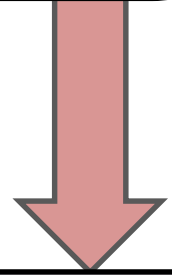
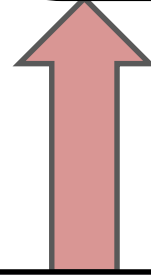
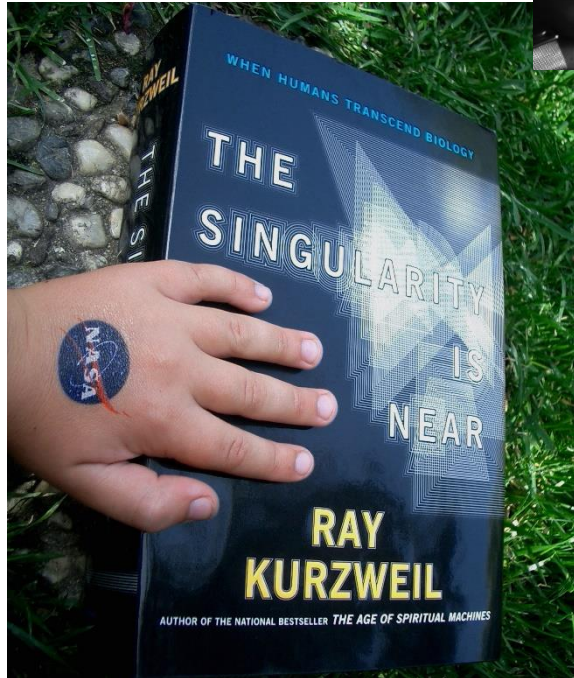


An *Opaque* Mind

?

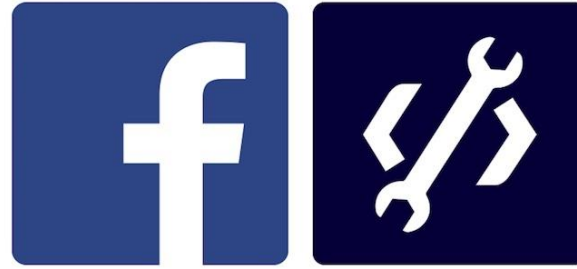


Artificial Neural
Network/Probabilistic
Learning Algorithm



WORLD

Non Robot AI



Facebook DeepText AI

FOSSBYTES

DeepText, a deep learning-based text understanding engine that can understand with near-human accuracy DeepText leverages several **deep neural network architectures**, including convolutional and recurrent neural nets...

*"As part of our bonkers ideas initiative, we are looking at various different ways to **incorporate artificial intelligence and bots**", Elizabeth Eastaugh, Director of Technology, **Expedia**, reported in The Guardian, May 2016*

In Summary – Why should you care?

1. **AI is everywhere. Today.** It decides what you see, and **WYSIATI**⁵
2. **Robots** are becoming **commonplace** for civil and military applications.
3. Humans are not evolved (genetically or culturally) to deal with non human autonomy (agency).
4. We need robots, AI and all complex autonomous systems to be **transparent, so that we can understand** the internal thinking that results in **their actions**.
5. Recent advances in AI are mainly in computational intelligence i.e. Artificial Neural Networks (ANN) and Probabilistic Learning Algorithms.
6. Transparency for ANN/Probabilistic approaches is a hard open research question (rule extraction).
7. **We are building AI that we don't understand. We must fix that.**

5. WYSIATI = What You See Is All There Is, from Daniel Kahneman, 2012, *Thinking, Fast and Slow*.



UNIVERSITY OF
BATH

Rob Wortham

Department of Computer Science
University of Bath

@RobWortham

r.h.wortham@bath.ac.uk

www.robwortham.com