What is Artificial Intelligence?

DAVID GAMEZ

Middlesex University, London, UK

Special Virtual Meeting:
Creative Explorations in Artificial Intelligence, Robotics, and Autonomous Systems

4th August 2020

Introduction

- Lot of discussion about AI ethics and threats in academia and the popular media.
- People often assume that artificial intelligence is a more powerful form of human intelligence.
- Little consideration of:
  - What intelligence actually is.
  - The variety of forms of intelligence.
  - How we can measure and compare different types of intelligence.

Current Measures of Intelligence

- Batteries of tests of verbal, mathematical and spatial reasoning abilities (IQ or g-score).
- Turing testing.
- Goal-based universal measures (Legg and Hutter 2007, Hernandez-Orallo and Dowe 2010).

Limitations of Current Measures

- IQ and g-score only apply to humans.
- Turing testing only measures human-like intelligence.
- Goal-based universal measures are not theoretically grounded and are based on external behaviour.

Intelligence and Prediction

- We can use prediction to measure intelligence.
- Prediction is closely linked to intelligence.
- A system that can make perfect predictions in its environment is highly intelligent within that environment.
- Many behaviours linked to intelligence, such as spatial and mathematical reasoning, require prediction.

Intelligence and Prediction

- Our ability to get rewards from different environments is closely linked to our ability to predict the consequences of our actions in different environments.
- Large amount of interest in the predictive brain hypothesis (Clark 2016).
- If brains are intelligent and the brain's core function is prediction, then brains that are better predictors should be more intelligent.

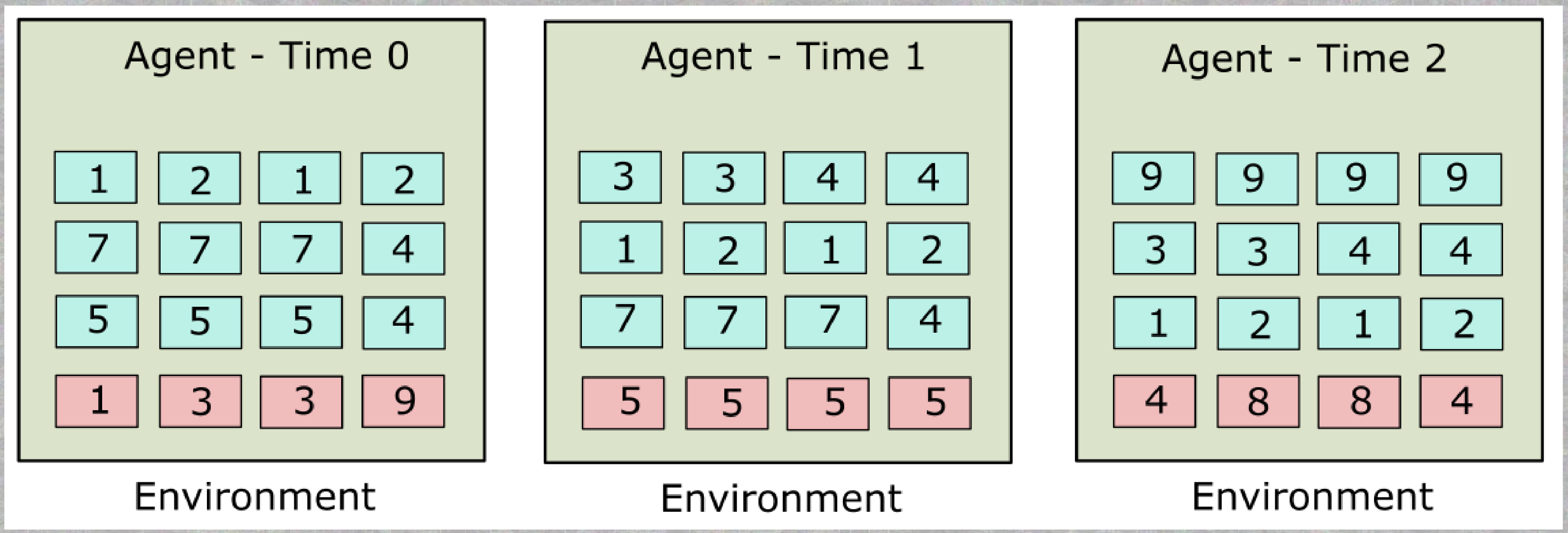
Machine Learning

- Many forms of machine learning are closely aligned with spatial or temporal prediction.
- A predictive measure of intelligence would be a good fit for AI.

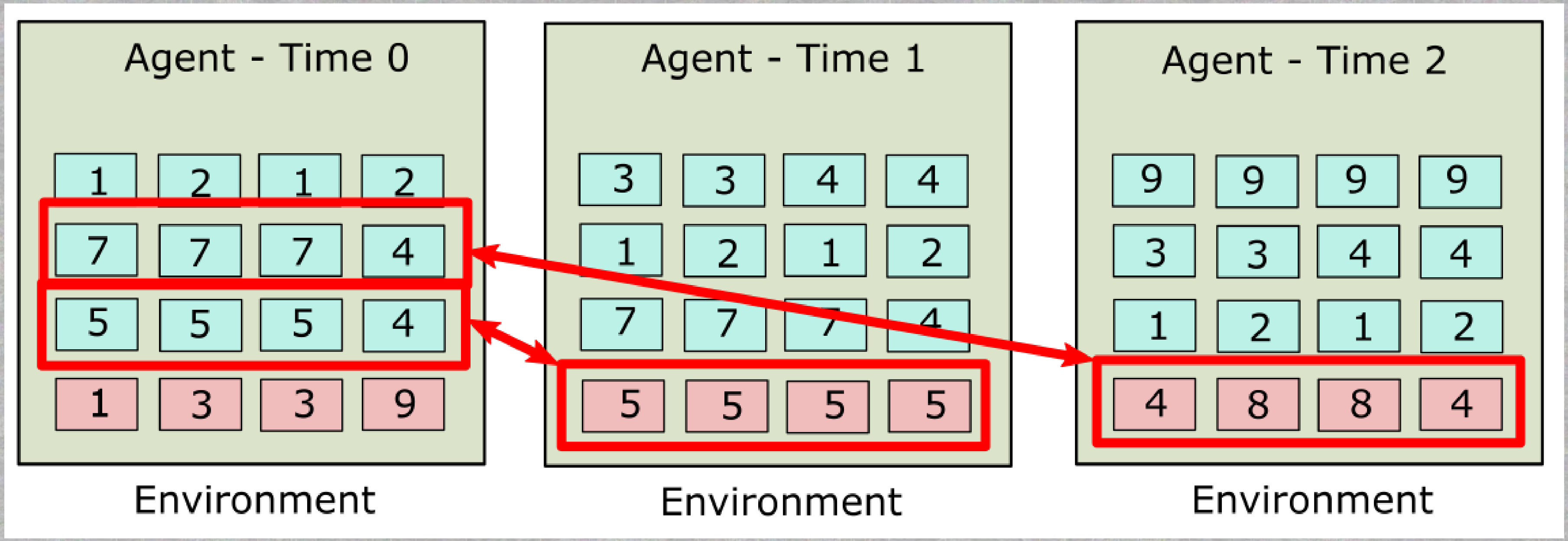
Measure of Predictive Intelligence

- I am developing a new algorithm for measuring intelligence that is based on a system's ability to predict future states.
- Can be used to compare the intelligence of natural and artificial systems.
- Could be a much more accurate way of assessing AI benefits and risks.

Agent Predictions

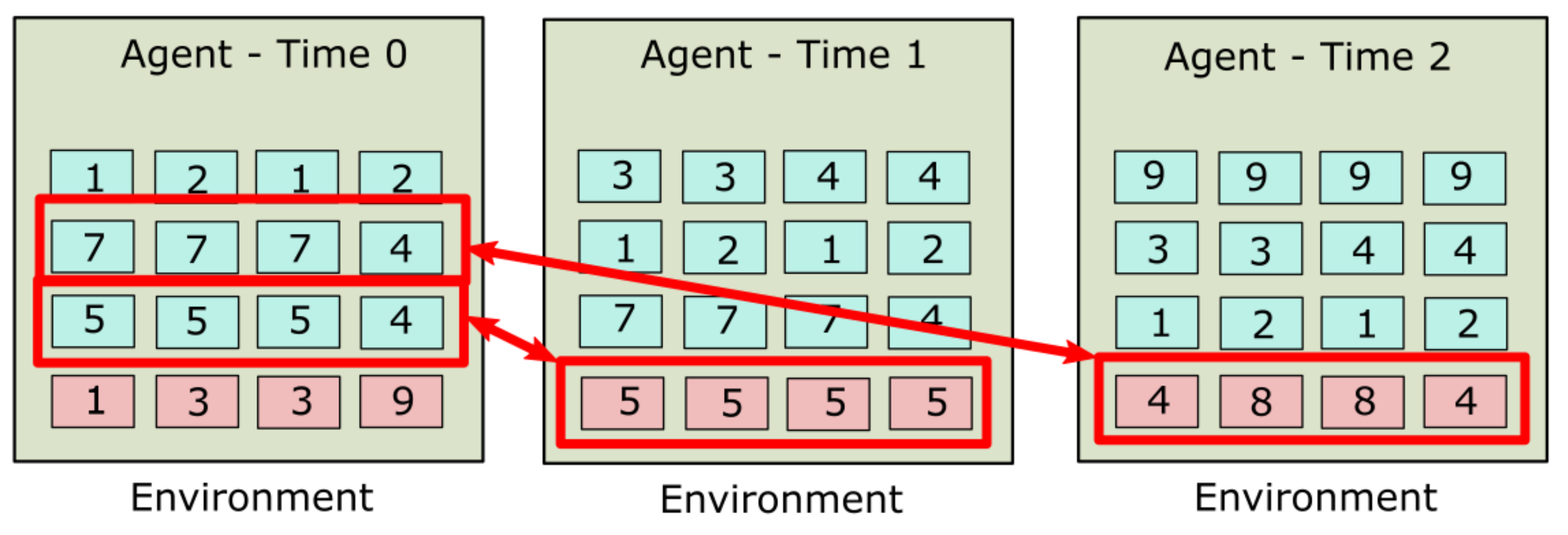


Agent Predictions



Predictive Intelligence
Single Environment

$$PI_e = \sum_{n=1}^{m} C(p_n)p_{match}$$



Crystallized and Fluid Intelligence

- Distinction introduced by Cattell (1971).
- PI_e is a measure of *crystallized* intelligence.
- Fluid intelligence is related to the rate of change of crystallized intelligence.

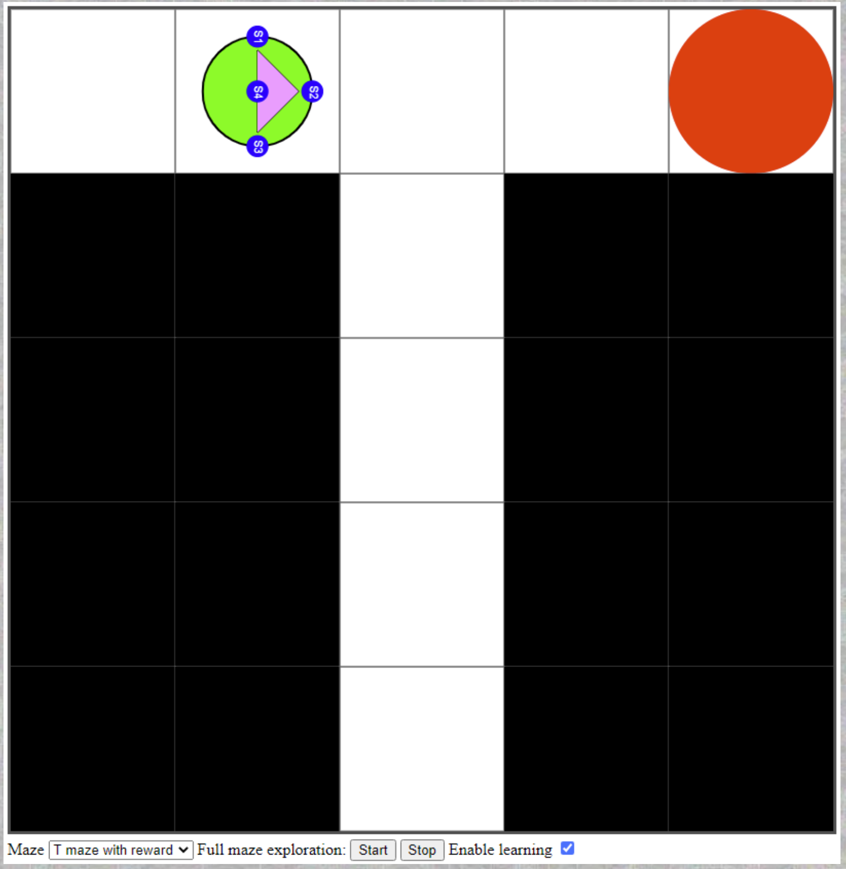
$$Fluid\ Intelligence = \frac{dPI}{dt}$$

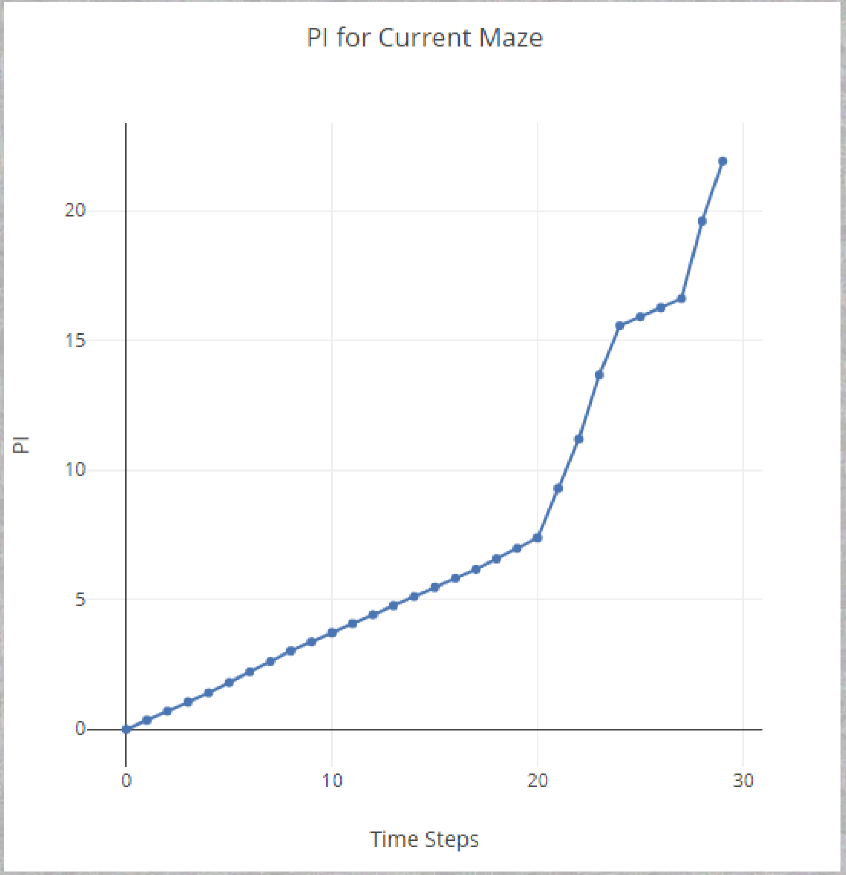
Predictive Intelligence Multiple Environments

- Sum intelligence across different environments.
- Multiply the sum by the extent to which the environments are different from each other.
- Use Kolmogorov complexity to measure this difference.

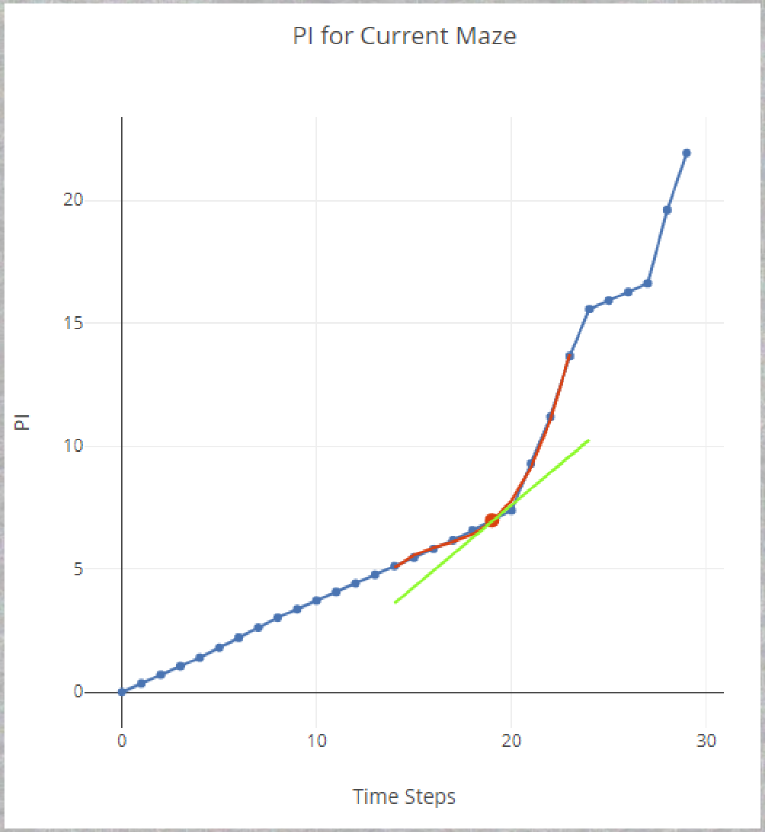
$$\frac{K(E_1 + E_2)}{K(E_1) + K(E_2)} * (PI_{E_1} + PI_{E_2})$$

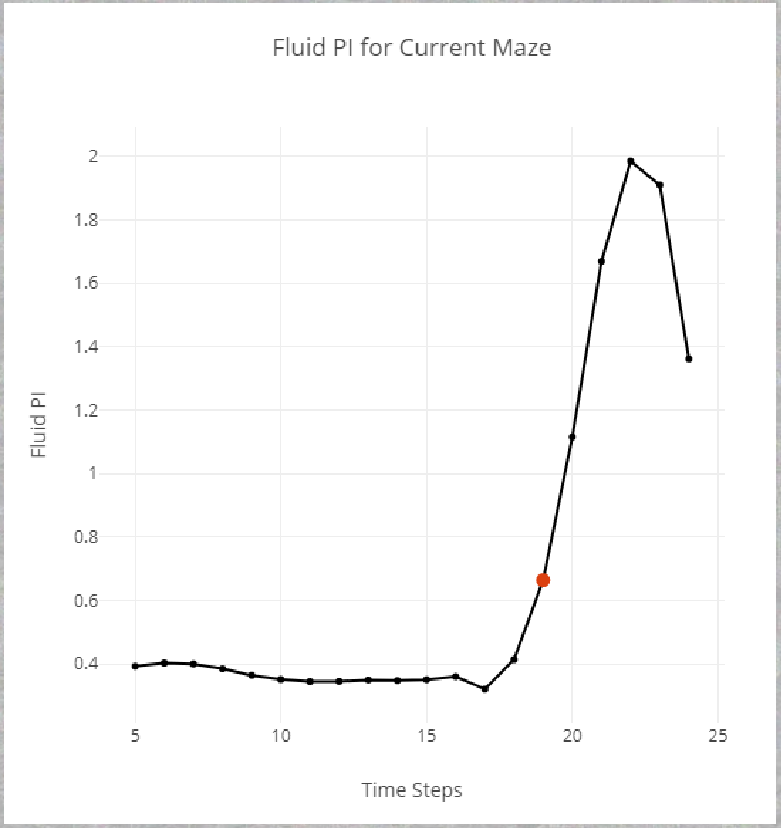
Current Work





Fluid Intelligence





Features of Predictive Intelligence

- No maximum PI.
- Many varieties of intelligence.
- Two systems can exhibit similar values of PI about different things.
- For example, one agent can have high PI about chess; another about human psychology.

Prediction is not Everything!

- Perfect prediction gives you God-like *omniscience*.
- Can predict every consequence of your action; know everything that is going to happen in your environment.
- Cognitive systems also have:
  - Goals that are often selected using emotions (Damasio 1994).
  - Planning systems that use predictive knowledge to figure out how to achieve their goals.
  - Actuators and control systems.
- Could separately measure these other features of artificial minds.

Understanding AI Threats with Predictive Intelligence

- Agents with low PI are not existential threats.
- Agents with a low PI in a particular area are not a threat to humans in this area.
- For example:
  - Manipulation. Only agents with a high PI in the domain of human psychology will be capable of advanced psychological manipulation.
  - Hacking. Only agents with high PI in the area of computer security will be capable of hacking into other computers, launching a global armageddon, etc.

Conclusions

- An informed discussion of AI ethics and threats requires:
  - Principled theory of natural and artificial intelligence.
  - Fine grained way of measuring intelligence that enables human and artificial intelligence to be compared.
  - Measures of both fluid and crystallized intelligence.

Conclusions

- Intelligence is closely linked to a system's ability to make accurate predictions.
- Have outlined an algorithm for measuring predictive intelligence.
- Can compare the intelligence of agents within a single environment and across environments.
- Good starting point for studying the differences between natural and artificial intelligence and understanding AI ethics.

Intelligence and Consciousness Paper

Journal of Artificial Intelligence and Consciousness

The Relationships between Intelligence and Consciousness in Natural and Artificial Systems

David Gamez

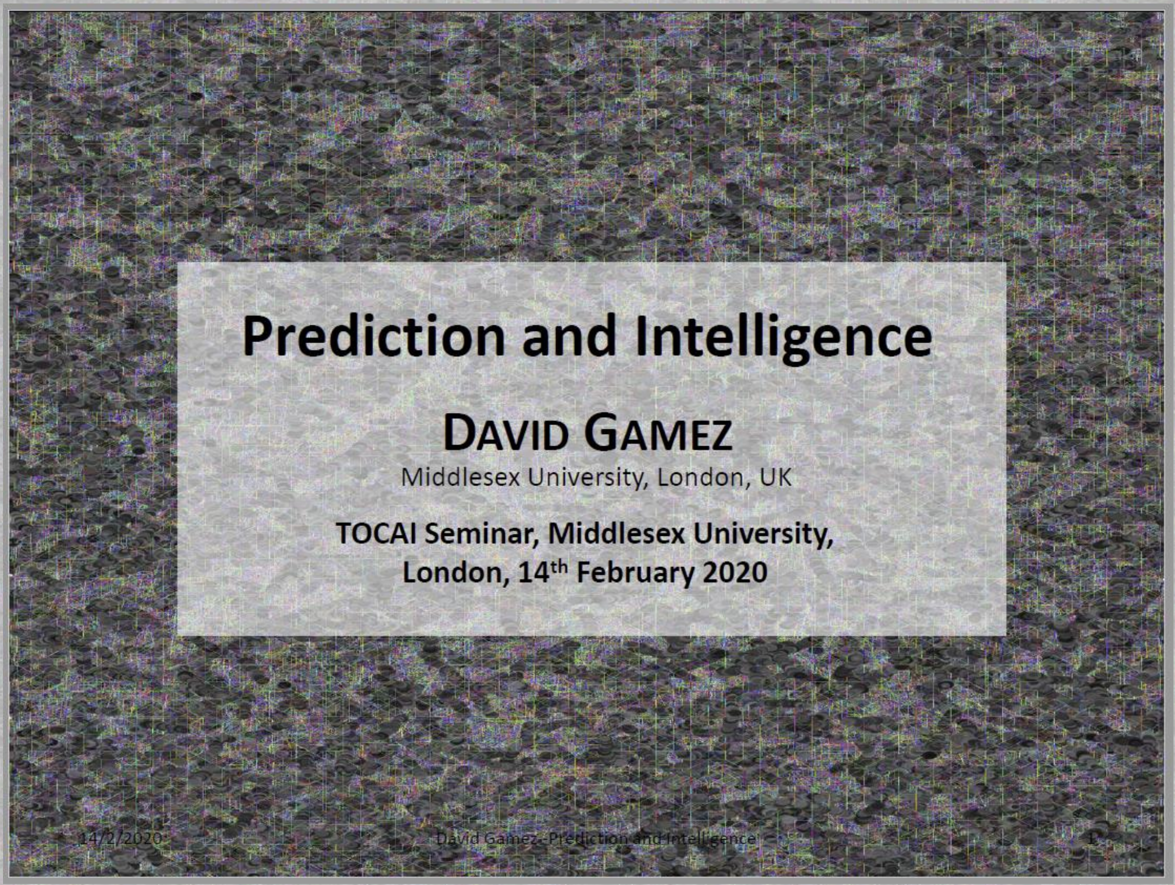
Department of Computer Science, Middlesex University, London, NW4 4BT, UK
d.gamez@mdx.ac.uk

Received (Day Month Year)
Revised (Day Month Year)

This paper explores some of the potential connections between natural and artificial intelligence and natural and artificial consciousness. In humans we use batteries of tests to indirectly measure intelligence. This approach breaks down when we try to apply it to radically different animals and to the many varieties of artificial intelligence. To address this issue people are starting to develop algorithms that can measure intelligence in any type of system. Progress is also being made in the scientific study of consciousness: we can neutralize the philosophical problems, we have data about the neural correlates and we have some idea about how we can develop mathematical theories that can map between physical and conscious states. While intelligence is a purely functional property of a system, there are good reasons for thinking that consciousness is linked to particular spatiotemporal patterns in specific physical materials. This paper outlines some of the weak inferences that can be made about the relationships between intelligence and consciousness in natural and artificial systems. To make real scientific progress we need to develop practical universal measures of intelligence and mathematical theories of consciousness that can reliably map between physical and conscious states.

Keywords: Intelligence; Consciousness; Artificial intelligence; AI; Measure of intelligence; IQ; g-score; Artificial consciousness; Prediction.

Prediction and Intelligence Talk



http://www.davidgamez.eu/talks/Gamez20_MiddlesexTOCAI-PredictionAndIntelligence.pdf

References

Cattell, R. B. (1971). *Abilities: Their Structure, Growth, and Action.* Boston: Houghton Mifflin.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind.* Oxford: Oxford University Press.

Hernández-Orallo, J. and Dowe, D. L. (2010). Measuring universal intelligence: Towards an anytime intelligence test. *Artificial Intelligence* 174: 1508-39.

Legg, S. and Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines* 17: 391-444.

Questions?

- Website: www.davidgamez.eu.