Embodiment vs Memetics

From Semantics to Moral Patiency through the Simulation of Behaviour

Joanna J. Bryson

Artificial Models of Natural Intelligence

University of Bath, United Kingdom

@j2bryson

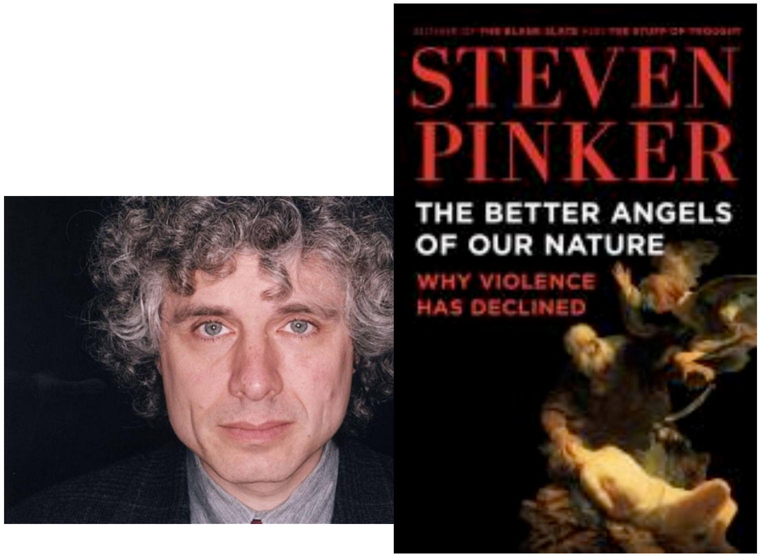
We Are Winning

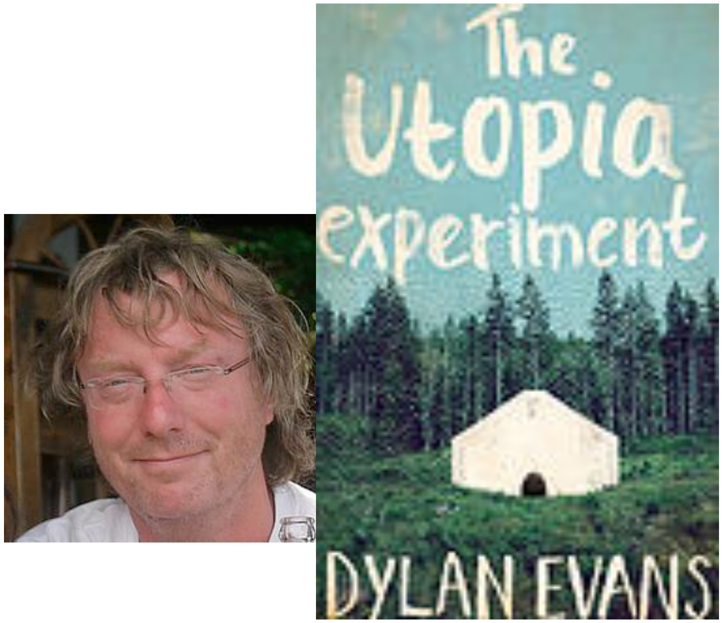
A public service announcement…

AI Global Warming

- Google, Apple, Microsoft, Intel, Cisco are worth $500B Each.
- Games Industry (console only), $49B revenue in 2014. Film Industry, $88B.
- NATO countries' annual military expenditures $800B.

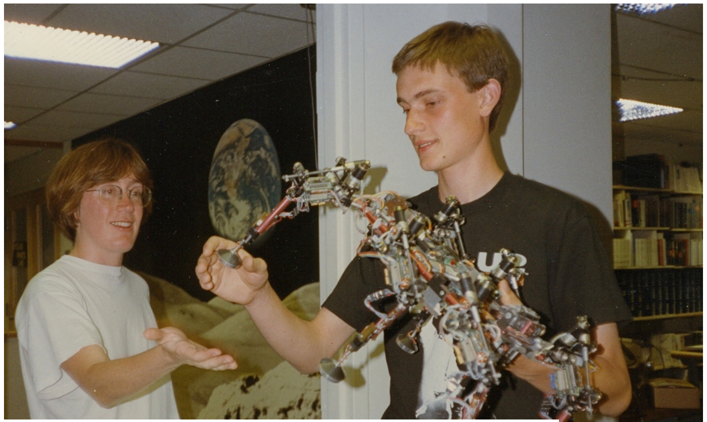
- Males in hunter-gatherer societies have 60% chance of dying at another person’s hand (war or murder). In the West this is 2%.
- We are five times more likely to be murdered than die in a war.





The AISB Approaches Are Winning

- Computational Social Science
- Intelligent Robots
- Philosophy & Ethics of AI
- Systems AI
- HRI



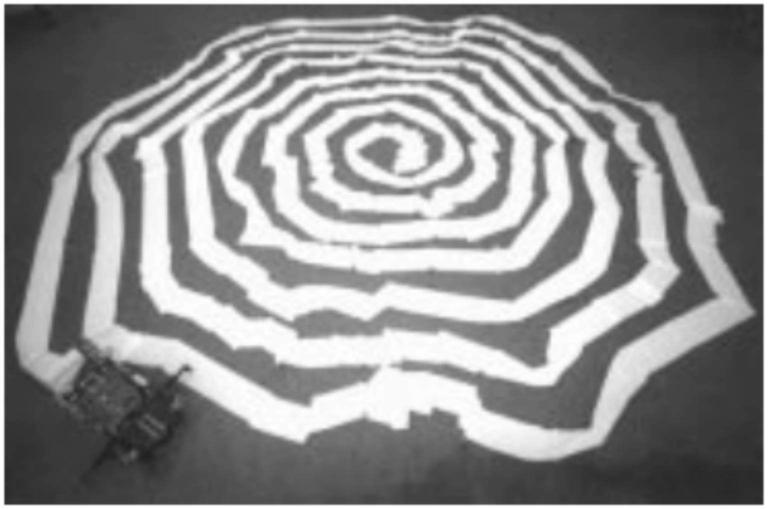


Figure 4. A spiral 'foraging' trail generated by the robot trace-maker.

Prescott's Cambrian Intelligence

Not Everyone is Winning

Cookie Policy | Feedback Like 2.6m Follow @MailOnline DailyMail Wednesday, Apr 22nd 2015 1AM 4°C 4AM 3°C 5-Day Forecast

Mail Online Science & Tech

Will robots make us their PETS? Apple founder Steve Wozniak has no doubt artificial intelligence will take over the world

© AP

Elon Musk donates $10 million to prevent a robot uprising: Entrepreneur says it is 'all fun and games' until something goes awry

Professional Responsibilities

- Countering both hype and hysteria in the media, even from colleagues.
- Thinking about applications of our research.
- Engaging with policy makers.
- Defending the right and obligation of universities to do blue-sky research.

Artificial vs Natural Intelligence

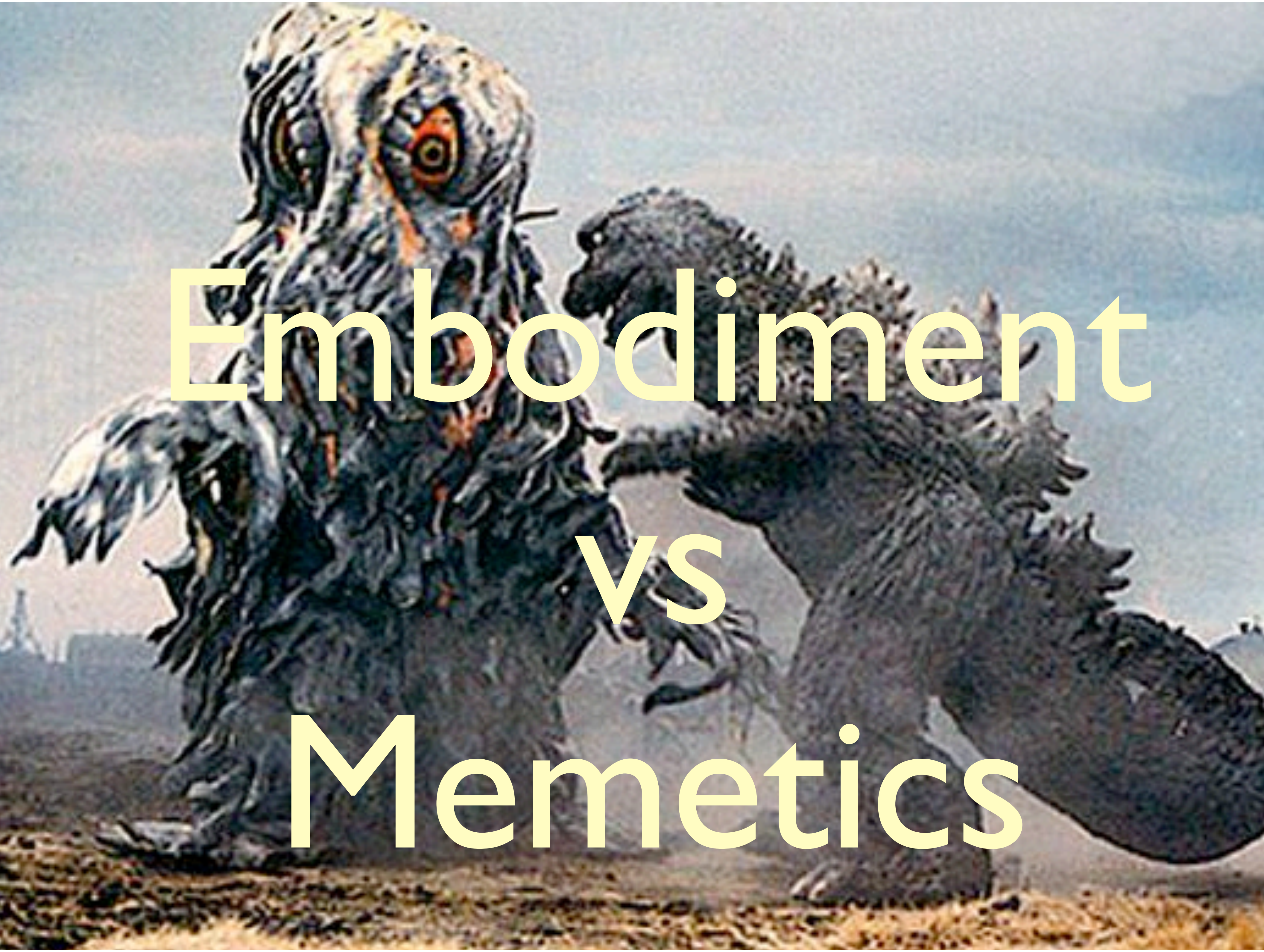
- Both are essentially search:
  - For what to do next.
  - For accurate predictions.
  - For perceptual and action categories that afford more efficient planning.
- Both suffer combinatorial explosion.
- Both benefit from concurrent search.

Authorship ≠ Childrearing

Within the laws of physics and computation, we have complete authorship over AI. We determine its capabilities and its goals.

Fundamentally different from our relationship to evolved life.

photos: Georgio Metta (top) & Emmanuel Tanguy



Embodiment vs Memetics

Outline

- Embodiment vs Memetics: Meaning
- Language Evolution and Human Uniqueness
- Culture and Altruism
- Imitation and Behaviour Oriented Design
- Embodiment vs Memetics: Morality

A Tale of Two Theses

- Embodiment: Semantic understanding of language requires long periods of learning difficult & shared physical concepts (Harnad 1990, Brooks 1991.)
- Memetics: Culture (including language) itself evolves, does not require true understanding from its substrate – e.g. humans (Dawkins 1976, Blackmore 1999.)

Refinements

adaptive: favoured by natural selection

- Some concepts you learn the hard way via embodiment later allow you to understand less accessible concepts via a metaphor e.g. path → life, career (Lakoff & Johnson 1999).
- Neo-diffusionist hypothesis: cultural diffusion (memetics) of adaptive behaviours/concepts more likely than neutral or negative ones (Kashima 2008, contra Blackmore).

Similarities

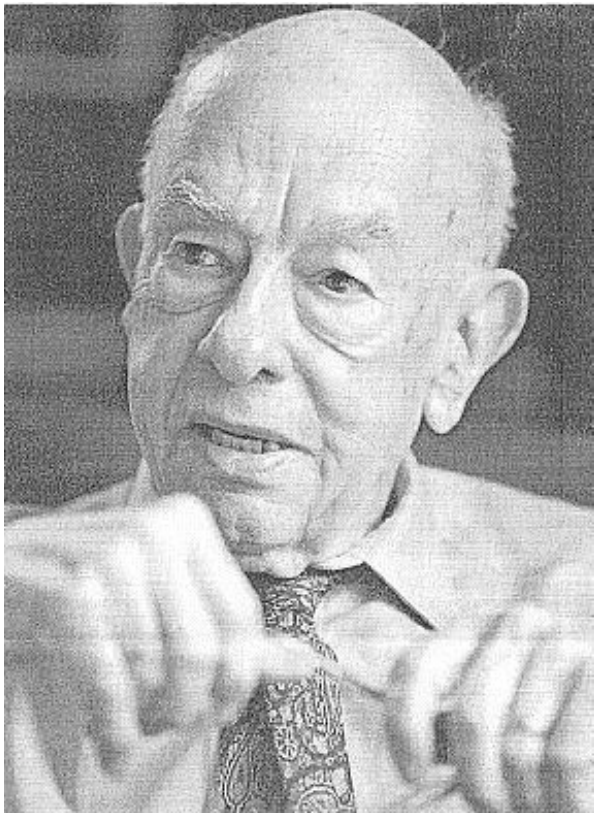
keeny-beeny 90s Joanna…

- Both cognitively minimalist.
  - No FOPL.
  - No complete world model.
- Large corpus linguistics makes semantics just another module in Behaviour-Based AI.
  - Easy! Like vision!

Semantics

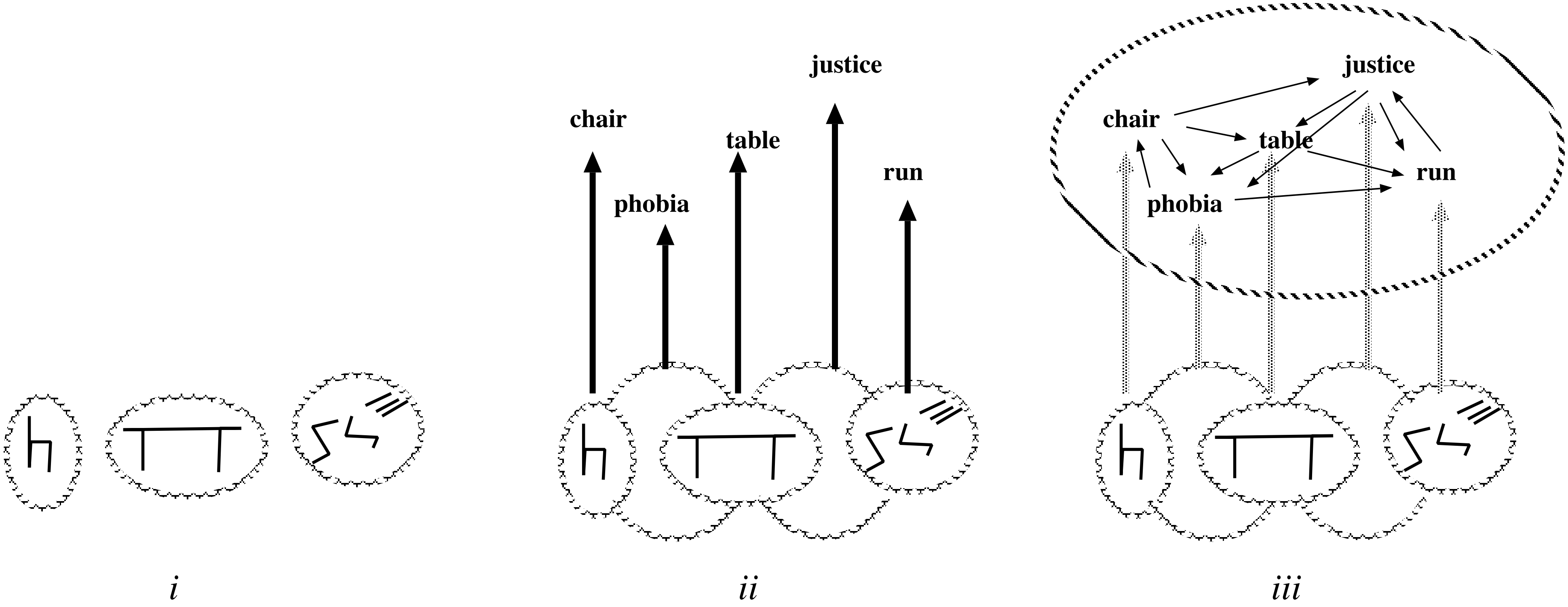
How Do We Learn What Words Mean?

- Ostensive definitions?



(Quine 1969)

Deacon's (1997) Theory of Semantics



The Symbolic Species

Large Corpus Semantics

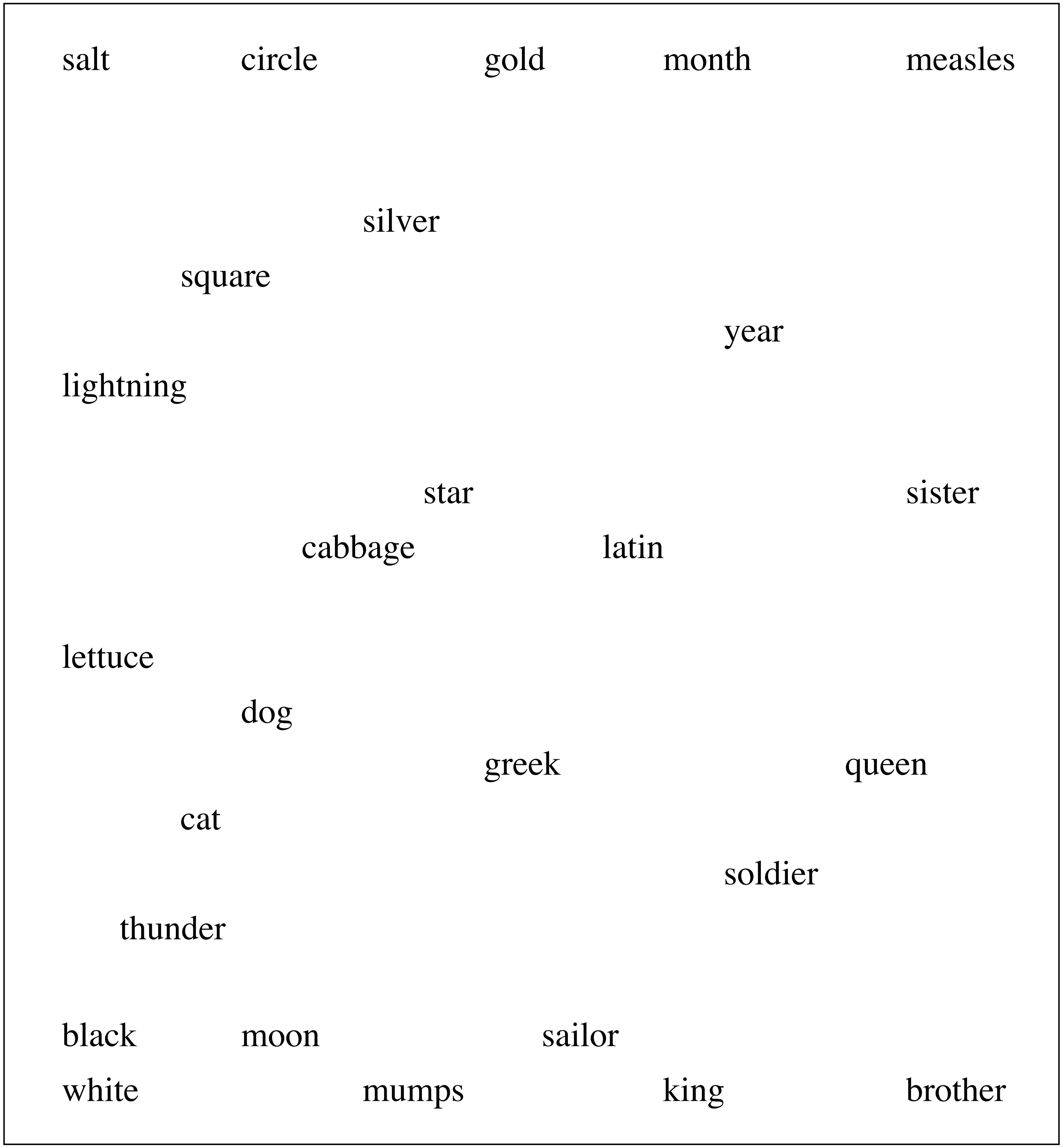
- Human semantics can be replicated by statistical learning on large corpra (Finch 1993, Landauer & Dumais 1997, McDonald & Lowe 1998, Bilovich & Bryson 2008).
- Record co-occurring words (appear nearby on either side every target word).
    - Track e.g. 75 fairly frequent words.
- ‘Meaning’ is cosine in 75-D space.

Validating Semantic Models

- Human semantics measured via priming studies.
- Flash a "priming" word to subjects too fast for conscious recall.
- Ask subject whether a collection of letters is a word or nonsense.
- Will recognise words faster if primed by something with a similar meaning.

(Moss et al. 1995)

Cosines between semantic vectors correlate with human reaction times (Figure: 75-D space projected in to 2-D, McDonald & Lowe 1998)



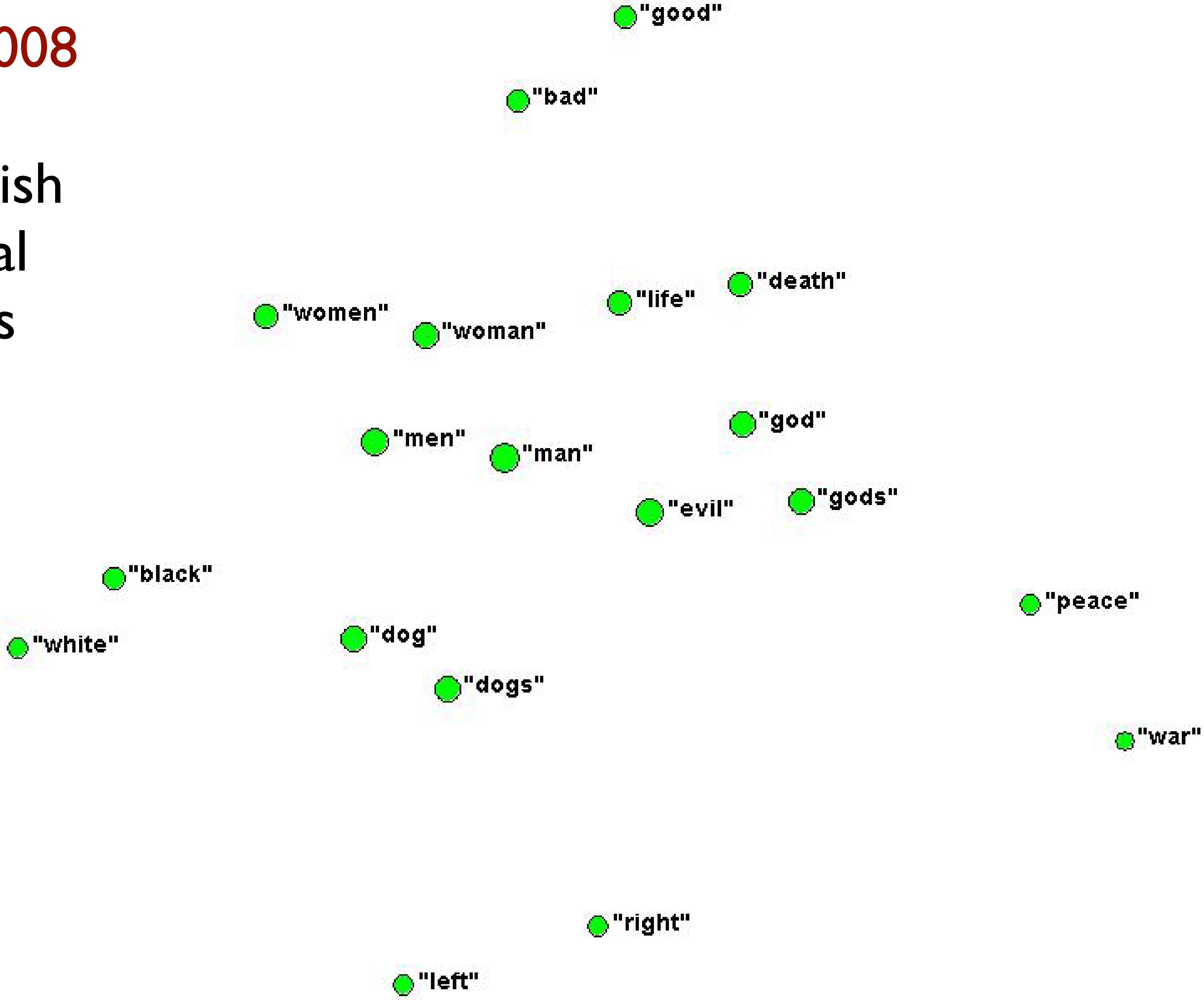
Bilovich & Bryson (2008)

Tracking Cultural Change

- Goal: replicating Banaji (2003) implicit association data.
    - Reaction times show cognitive consonance & dissonance btw good:right::bad:left; also black/white, male/female, old/young stereotypes.
- Can we reproduce cultural stereotypes in a corpus-based intelligent system?
- Can we see cultural change over time?

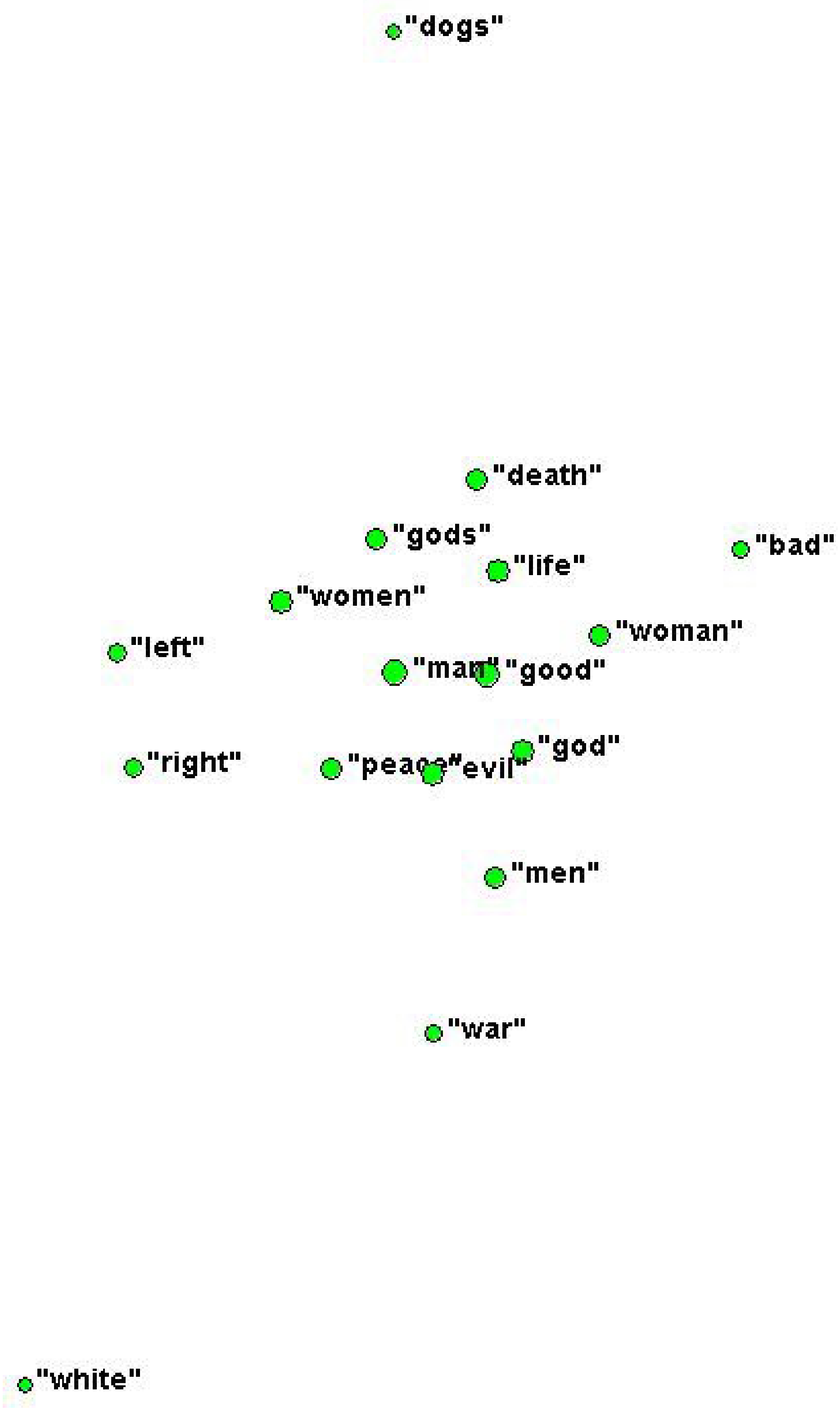
Bilovich & Bryson 2008

text: British National Corpus



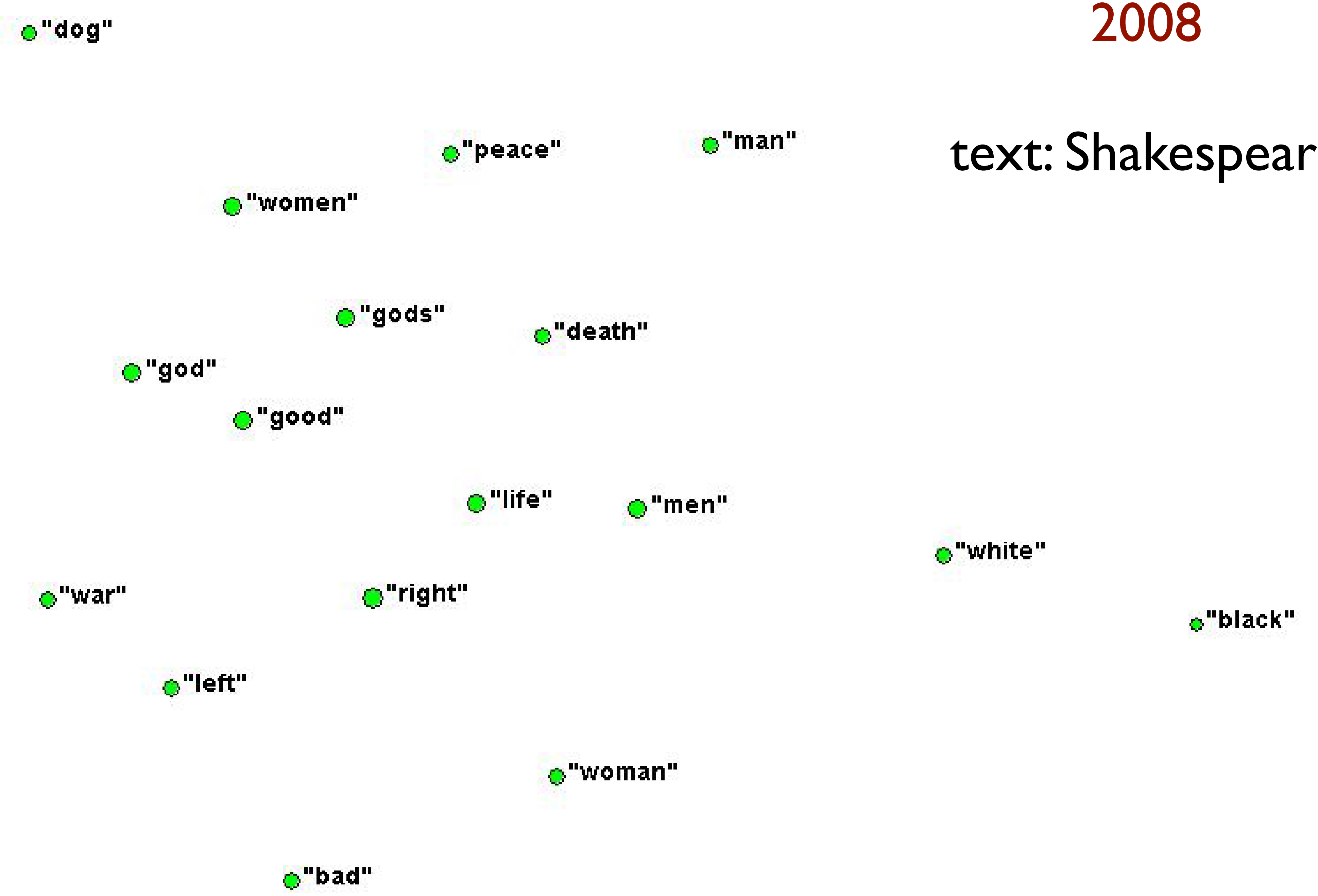
Bilovich & Bryson 2008

text: bible



Bilovich & Bryson 2008

text: Shakespeare



Humanlike Biases in Corpus Semantics

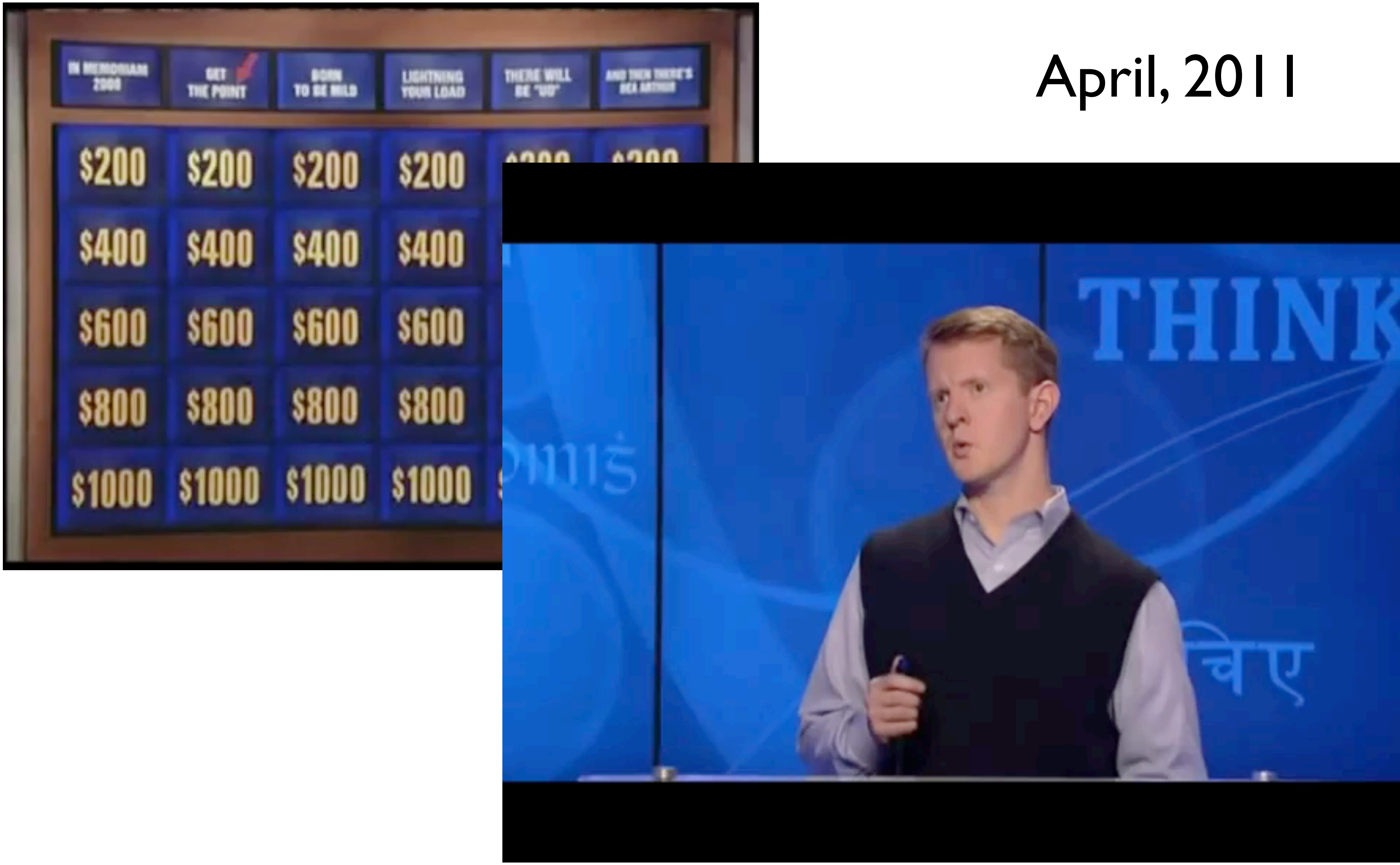
- Bilovich & I did not replicate Banaji (2003).
  - Nearest miss was Shakespeare – (nearly) single author?
- Macfarlane & I (in prep.) have found matches – by using the Enron Corpus.

Macfarlane (2013) Results

- Life terms more like pleasant & Death terms more like unpleasant words.
- Elderly & Youth did not go as per Banaji on pleasantness, though did on competence.
- Male terms more like Career & Female terms more like Family.

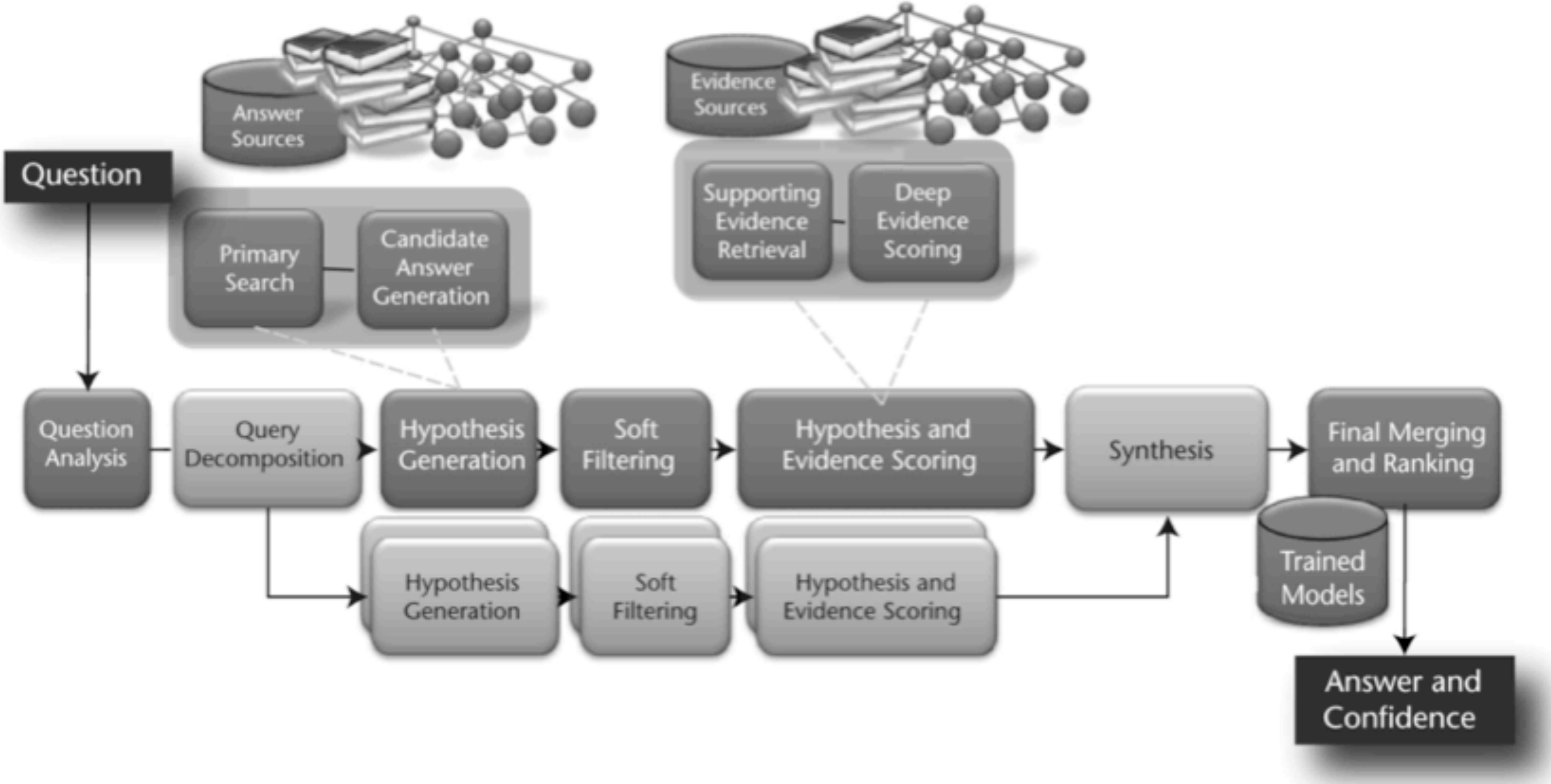
In preparation; also University of Bath Computer Science technical report.

Jeopardy vs Watson



April, 2011

Videos via Dale Lane, IBM

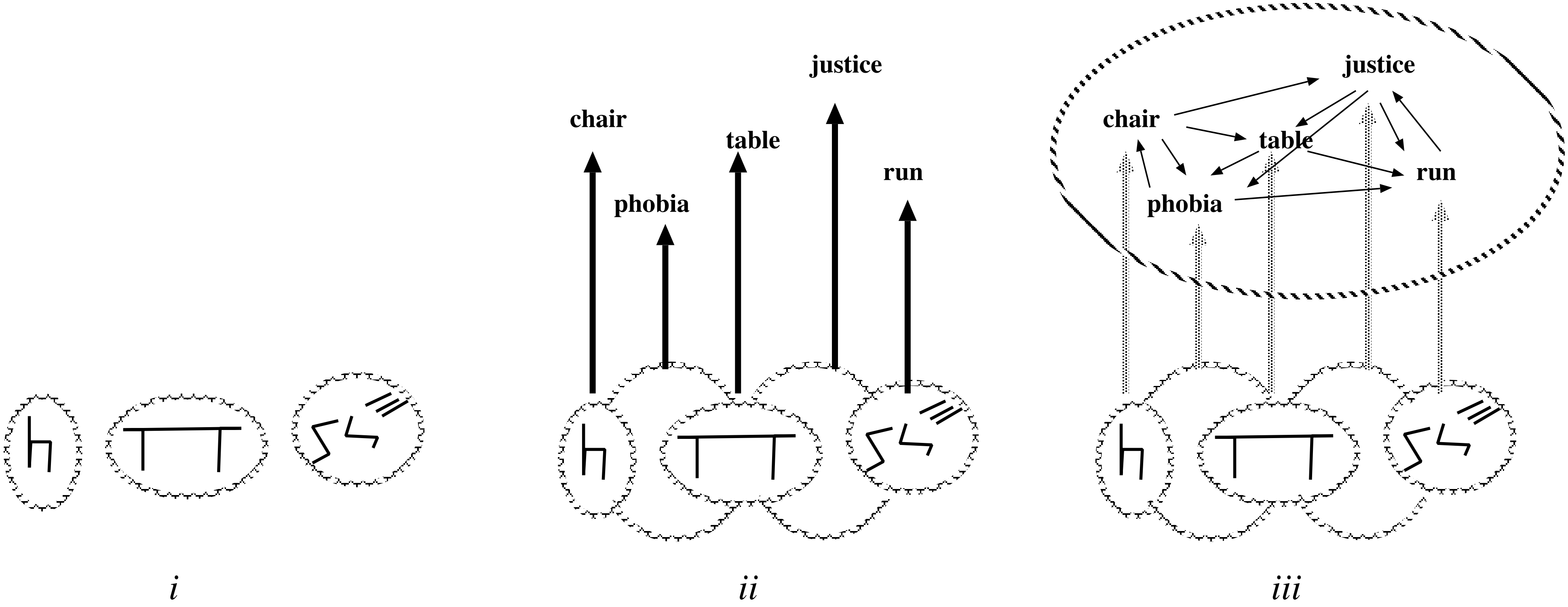


(Ferrucci *et al.*, AI Magazine 2010)

Outline

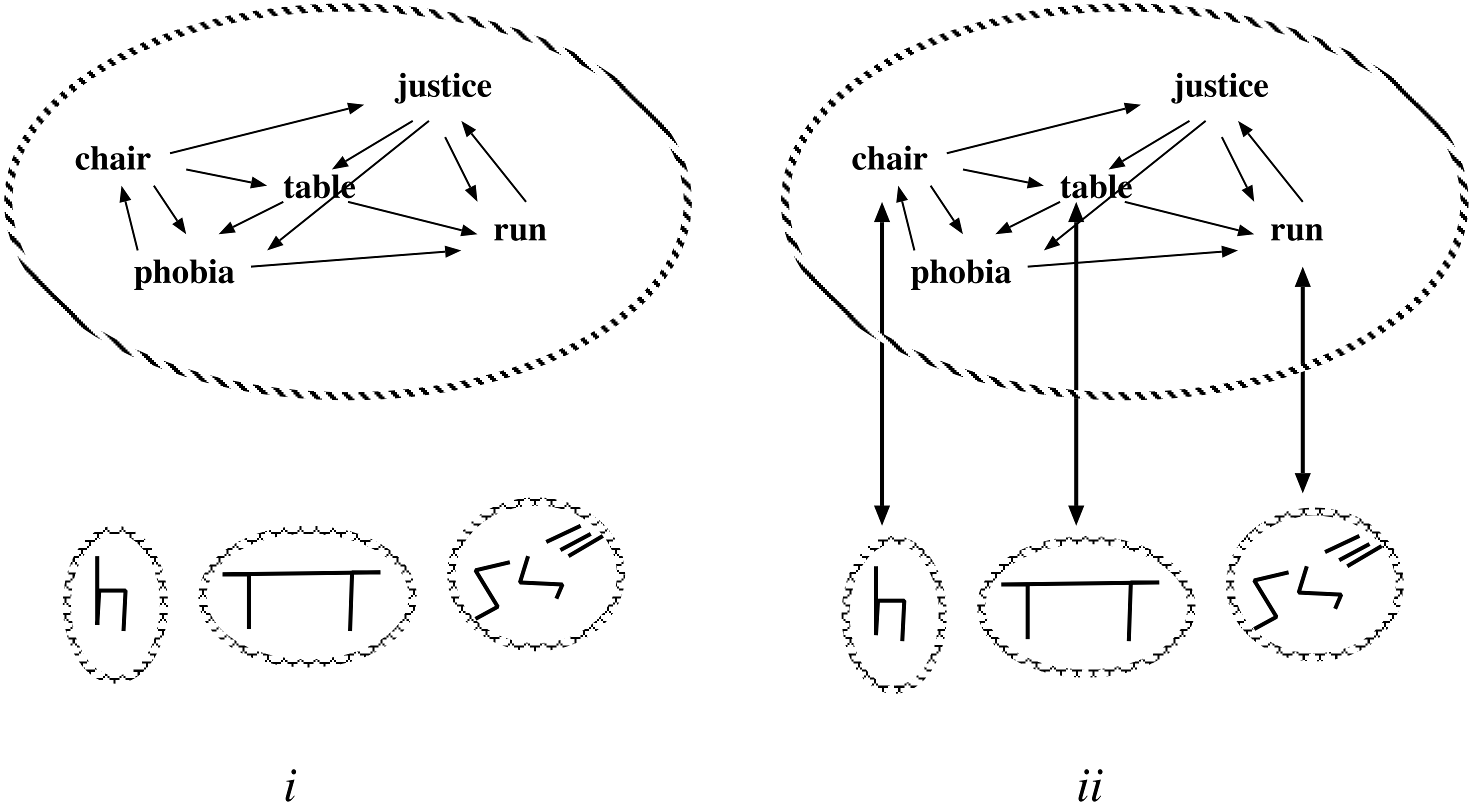
- Embodiment vs Memetics: Meaning
- Language Evolution and Human Uniqueness
- Culture and Altruism
- Imitation and Behaviour Oriented Design
- Embodiment vs Memetics: Morality

Deacon's (1997) Theory of Semantics



The Symbolic Species

Bryson's (2008) Theory of Semantics



Embodiment *versus* Memetics (first presented 2001)

Why are humans special? (Bryson 2008; 2009)

- Humans are the only primate species capable of precise vocal imitation (Fitch 2000; 2007).
- Communicates lots of information, including volume, pitch, timbre and time.
- Allows redundant encoding to preserve important details while others can mutate.
- Allows communication of complex, sustainable behaviour.

Why should temporal imitation matter?

- More information contained in the ‘genetic’ substrate.

- Allows for more variation while providing redundancy, robustness – assists GAs (Baluja1992;Weicker & Weicker; 2001;Miglino & Walker 2002).

- Aligns with Wray (2000) on the evolution of language from phrases, Kirby (2000) on cultural selection for language efficacy.

Why don't birds talk?

- They can't hold 2nd order representations
- Primates have uniquely complicated social organisations. (Harcourt 1992).
  - Almost all species remember how group-mates behave with respect to themselves (tit-for-tat).
  - But only primates behave as if they keep track of each other's social behaviour.

(my) old theory!

Compositionality / Recursion

- S →NP + VP
- NP →N | D + NP | ADJ + N | PN
- VP →IV | AUX + VP | TV + NP
- IV →laughed | cried | ...
- AUX→can | will | shall | ... |
- TV→threw | caught | ...
- N→dog | peacock | justice |...
- D→the | a | an

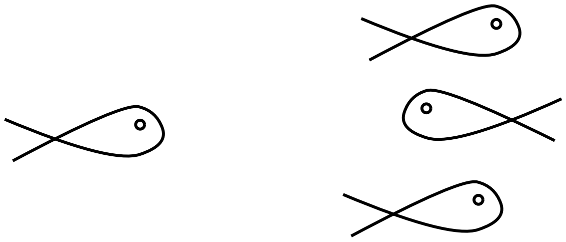
Allows language to be infinitely productive.

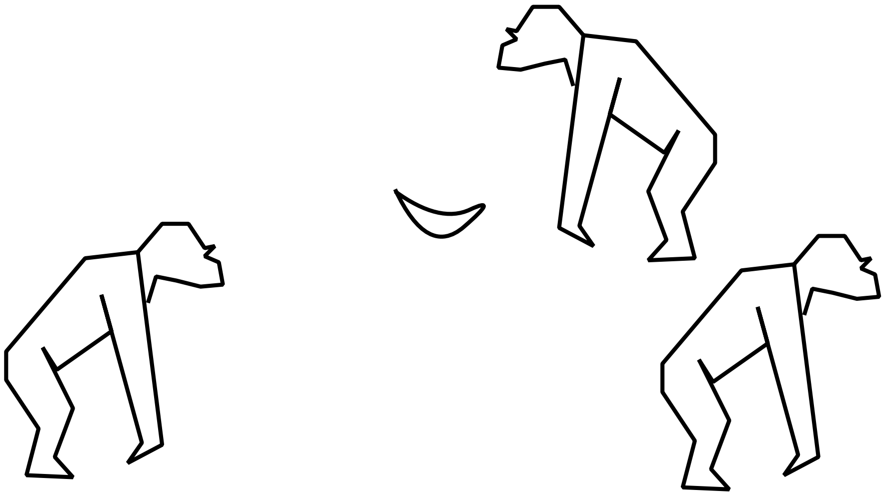
What no animal language learner has shown.

(cf. Hauser, Chomsky & Fitch 2002)

	ego
Roy	5
Thelma	2
Eunice	7
Harry	-1

	Roy	Thelma	Eunice	Harry
Roy	-	5	2	-4
Thelma	7	-	8	4
Eunice	-3	8	-	4
Harry	-1	3	5	-





Why Humans are Special (Bryson 2008, 2009)

	temporal imitation	no temporal imitation
second-order representations	people	non-human primates
no second-order representations	birds, seals	most things

Why Humans are Special (Bryson EoL 2010)

	temporal imitation	no temporal imitation
big brains, memories	people	non-human apes
no big brains, memories	birds, seals	most things

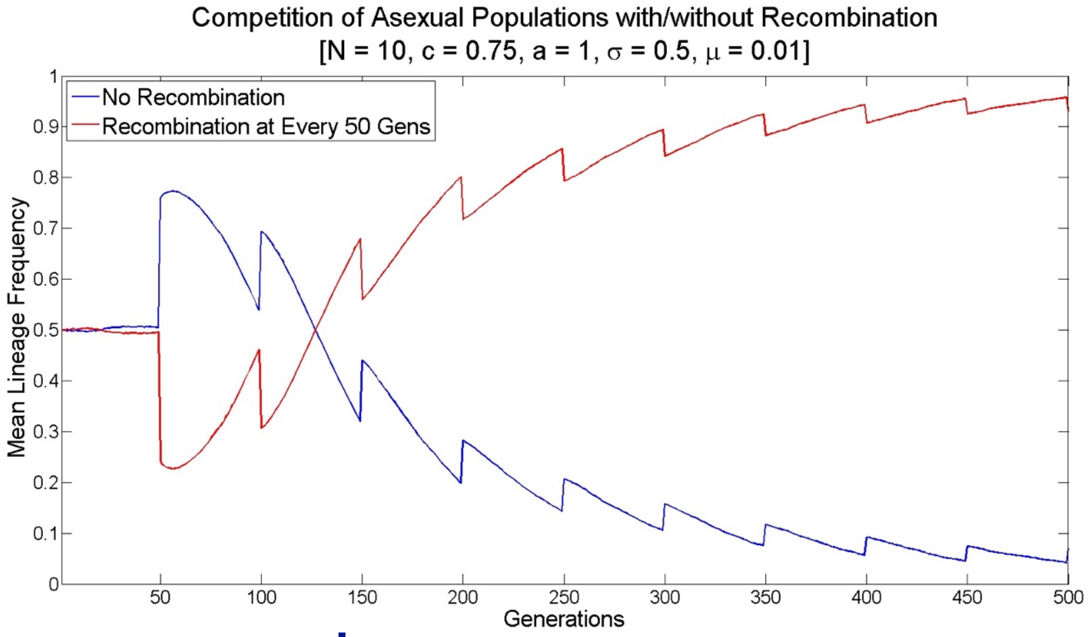
Key concept…

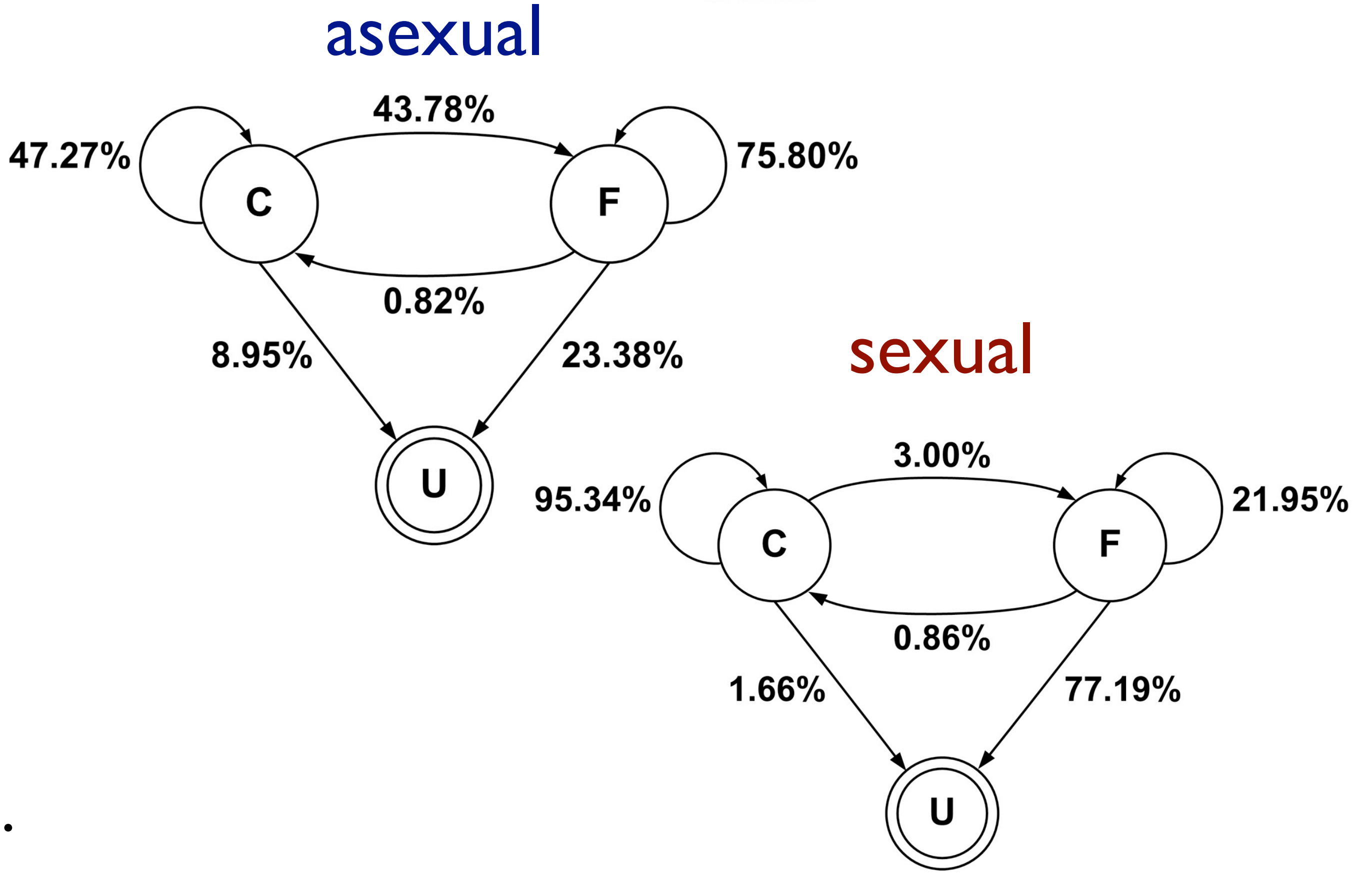
Evolvability

- Language itself evolves to be more learnable.
  - Dual replicator theory: Human culture & human biology both evolve – at the same time, under each other's influence.
  - Even within the genome, hierarchical representations evolve, e.g. genes to flag zones of innovation.

Yifei Wang

- Compensatory Mutation scale invariant in GRN.
- Sex pays its costs with stability, not just innovation.
- Evolvability
- Collaborators: Nick Priest (Bath), Dan Weinreich & Yinghong Lan (Brown), Steve Matthews (Bristol).





Outline

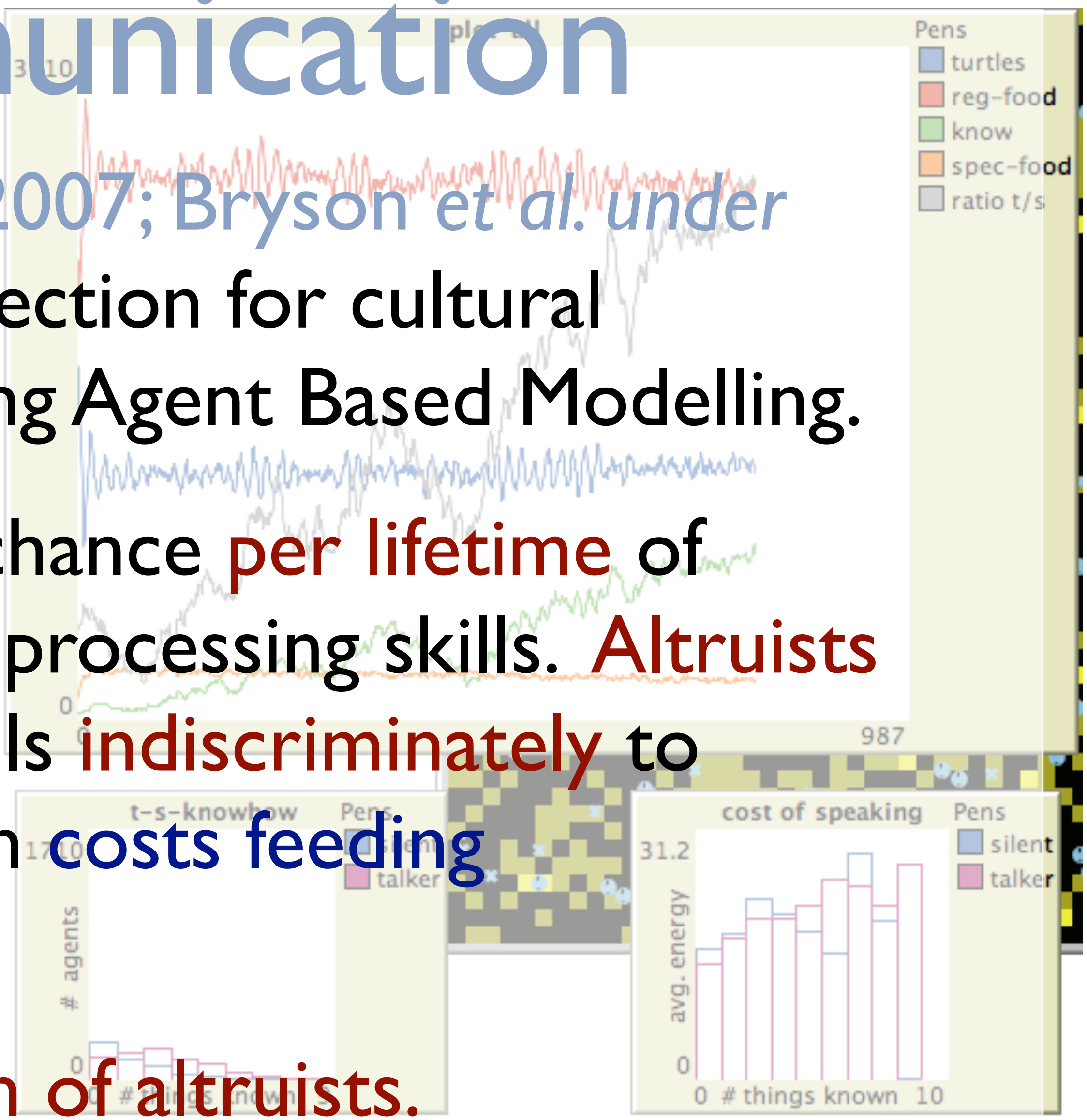
- Embodiment vs Memetics: Meaning
- Language Evolution and Human Uniqueness
- Culture and Altruism
- Imitation and Behaviour Oriented Design
- Embodiment vs Memetics: Morality

Problem / Critique

- Language is giving away information – reduces competitive advantage.
- Can’t evolve! Can’t be selected for!
  - Must be “Extra-Darwinian”…
  - or at least costly signalling (peacock tail.)

False!!

ABM of Altruistic Communication



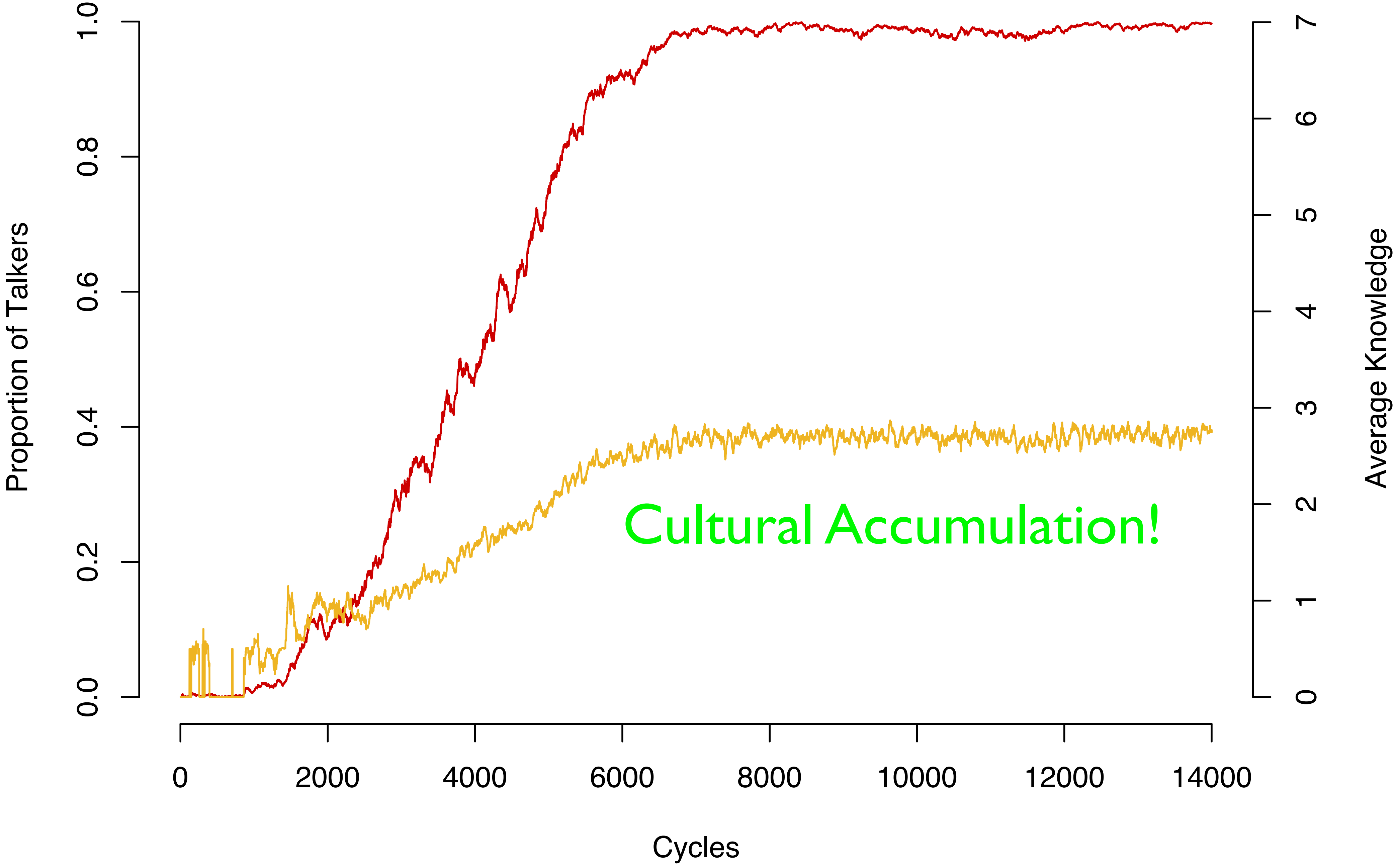
- Čače & Bryson (2007; Bryson *et al. under revision*) show selection for cultural accumulation using Agent Based Modelling.

- Agents have 5% chance per lifetime of discovering food-processing skills. Altruists communicate skills indiscriminately to neighbours, which costs feeding opportunities.

- Results in fixation of altruists.

Ivana Čače and Joanna J. Bryson, "Agent Based Modelling of Communication Costs: Why Information can be Free", in *Emergence and Evolution of Linguistic Communication* C. Lyon, C. L Nehaniv and A. Cangelosi, eds., pp. 305–322, Springer 2007.

Basic Results: Altruists & Knowledge

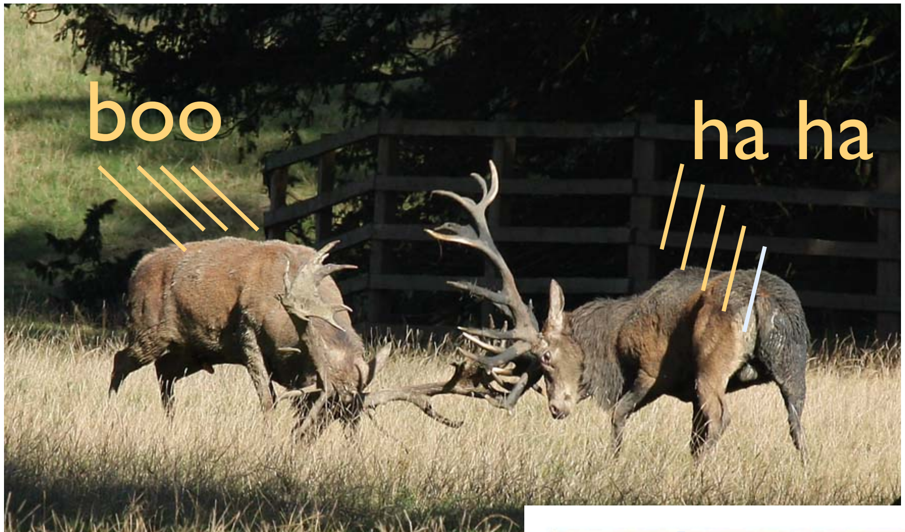


Cultural Accumulation!

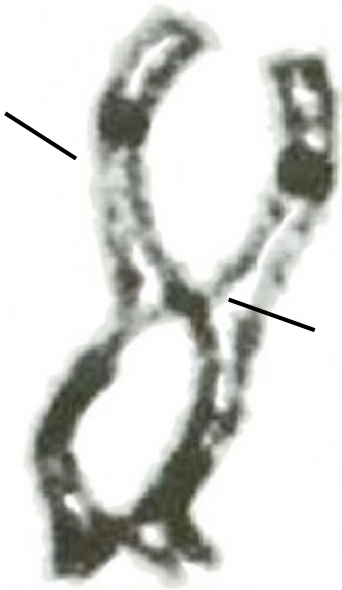
Selfish Genes $\nRightarrow$ Selfish Individuals

- Traits advantageous to the community but costly to the individual were (for some time) considered inaccessible to evolution. This is false.
- Explanation: inclusive fitness & kin / group selection
  - What is transmitted is the replicator.
  - The unit of selection is the vehicle (or interactor.)
  - In the current ecology, most vehicles are composed of many, many replicators.

Multiple Levels of Interaction ⇒Cooperation

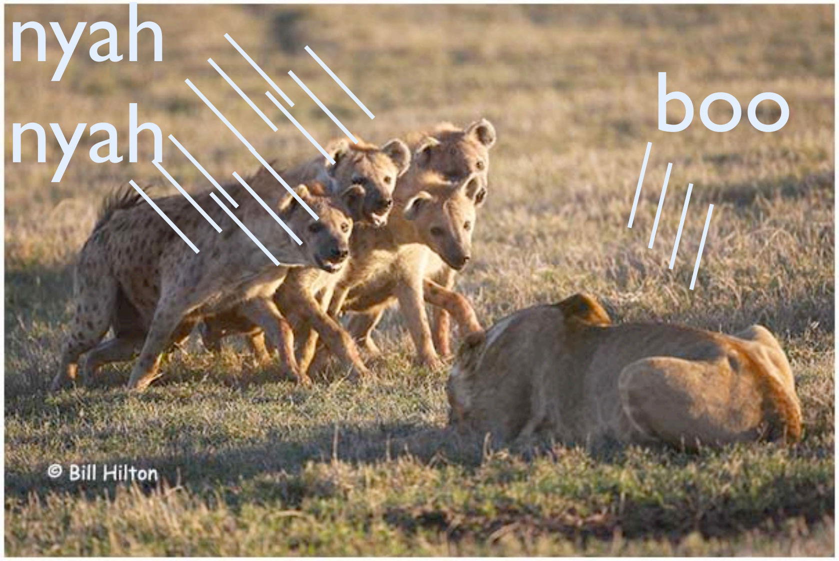


Replicator (Gene)



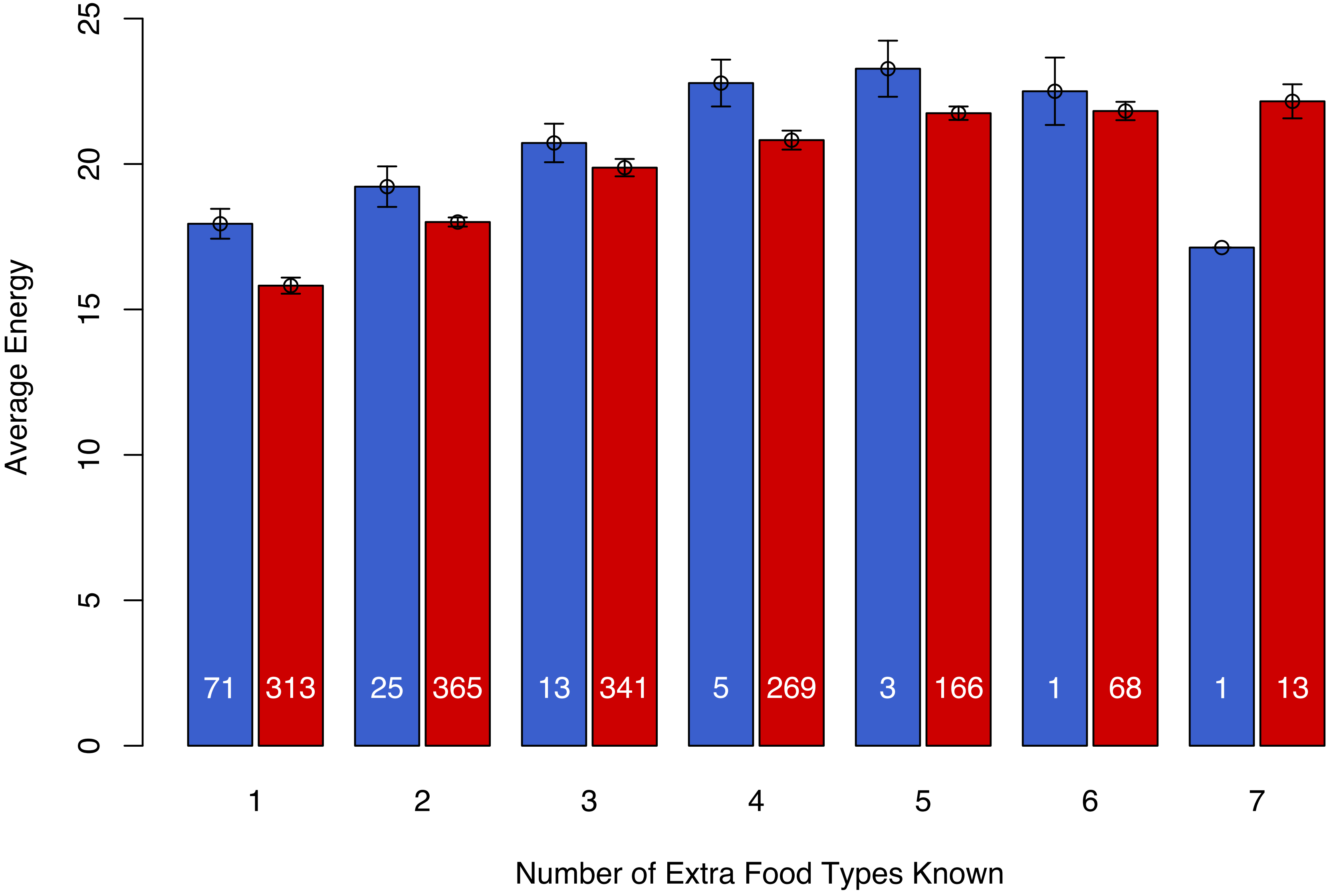
Organism

Group



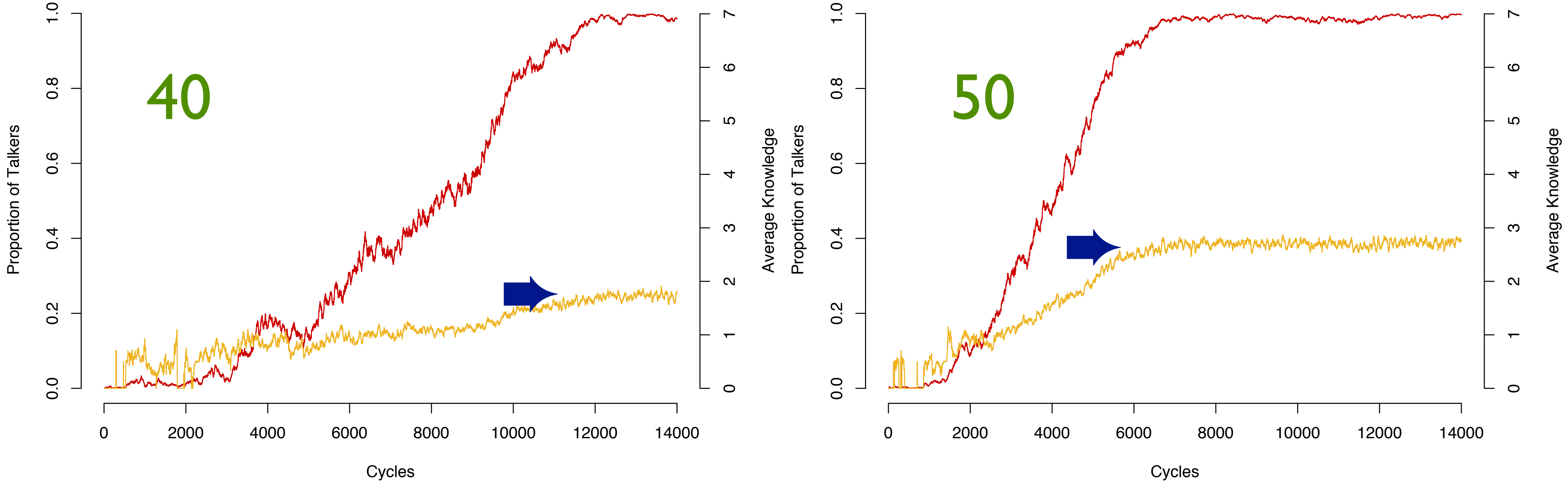
Cost (in energy ➪ reproduction)

talker (altruist) silent (free-rider)



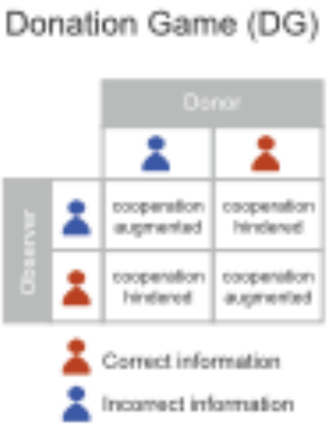
Life History & Culture

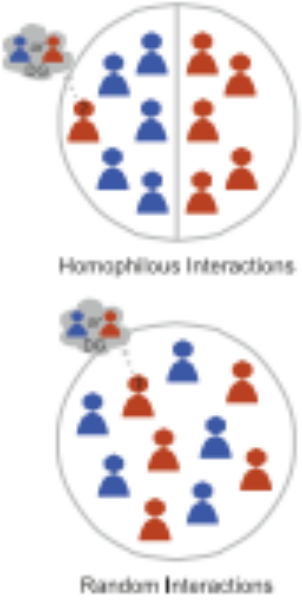
Altruists & Knowledge lifespan 40 versus 50 cycles

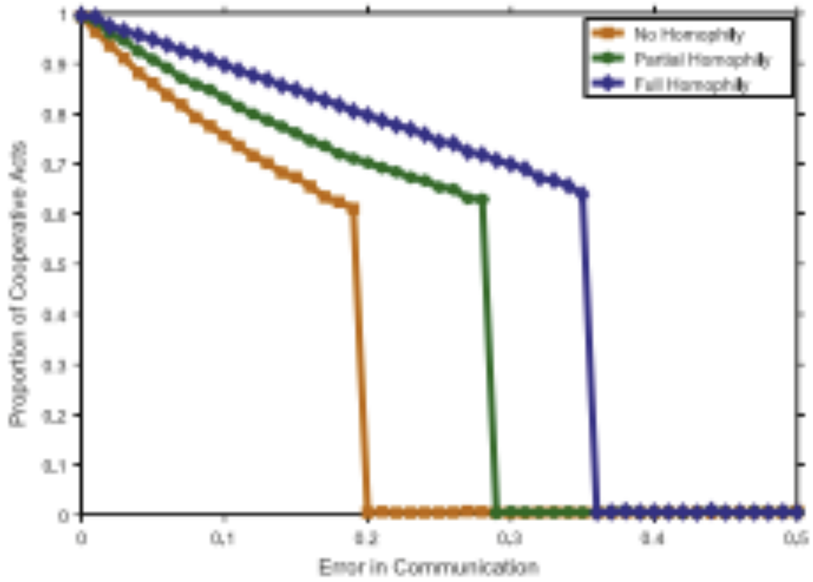


Life history tradeoffs determine how much is learned on average per lifetime ⇒ size of culture.

(Bryson, Lowe, Bilovich & Čače *under revision*)







Value homophily benefits cooperation but motivates employing incorrect social information

Paul Rauwolf*, Dominic Mitchell, Joanna J. Bryson

Dominic Mitchell

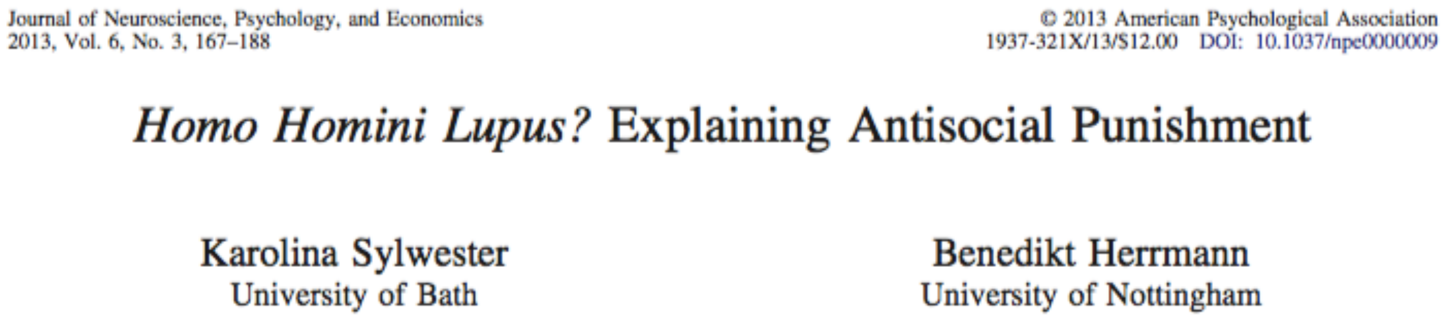
Paul Rauwolf

- Self Deception
- Impact Bias
- Unconsciousness

- Public Language
- Evolution of Language
- Winner/loser effects

Public Goods Investment

Antisocial Punishment



Journal of Neuroscience, Psychology, and Economics
2013, Vol. 6, No. 3, 167–188

© 2013 American Psychological Association
1937-321X/13/$12.00 DOI: 10.1037/npe0000009

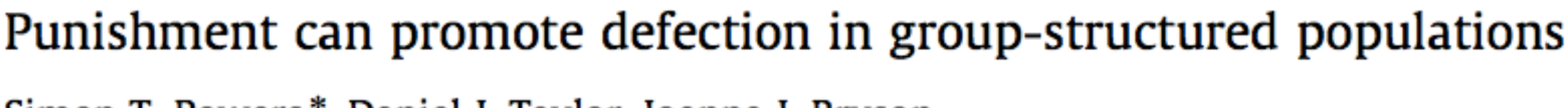
Homo Homini Lupus? Explaining Antisocial Punishment

Karolina Sylwester
University of Bath

Benedikt Herrmann
University of Nottingham

Joanna J. Bryson

Karolina Sylwester



Punishment can promote defection in group-structured populations

Simon T. Powers*, Daniel J. Taylor, Joanna J. Bryson

Daniel Taylor

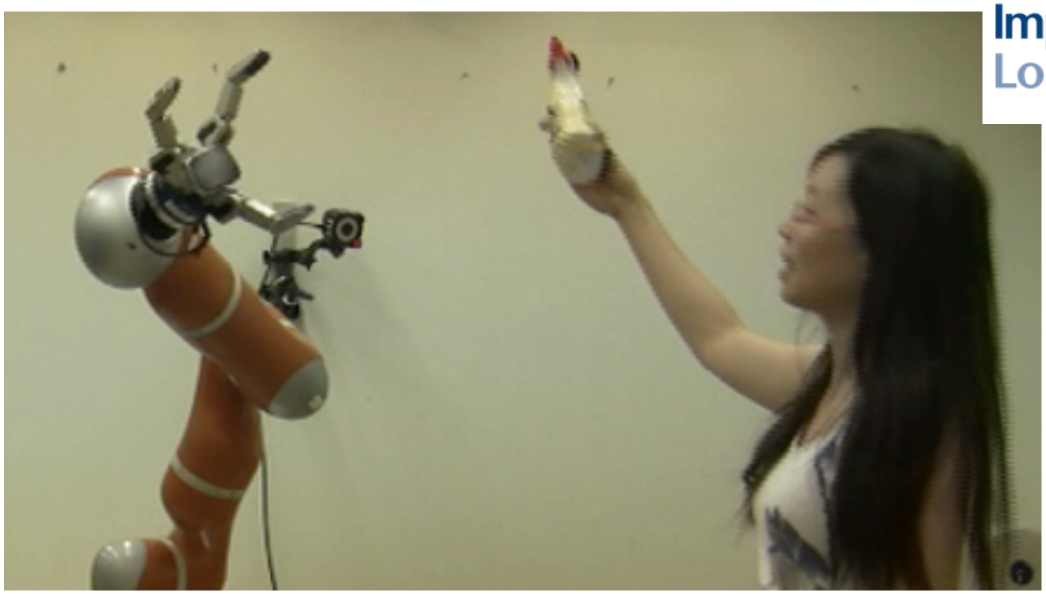
Simon Powers

The Evolution of the Social Contract

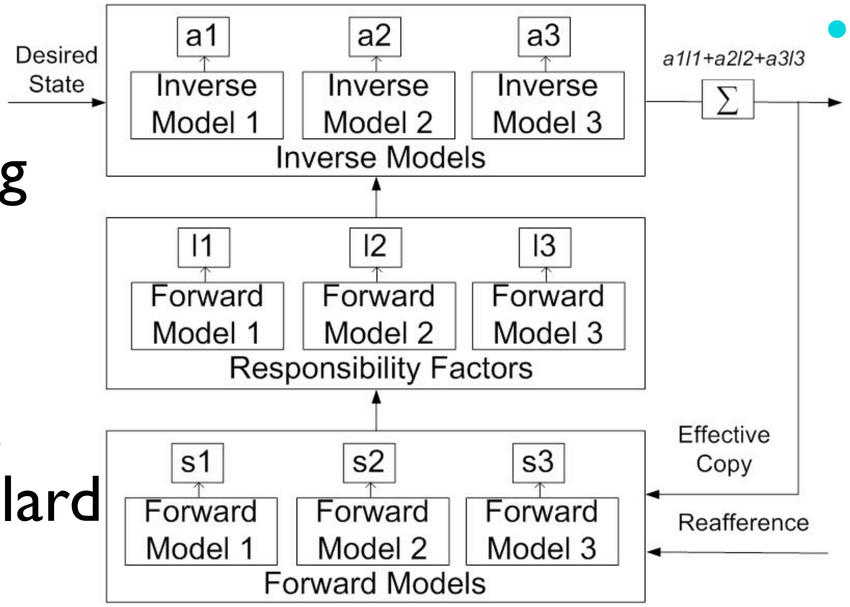
Outline

- Embodiment vs Memetics: Meaning
- Language Evolution and Human Uniqueness
- Culture and Altruism
- Imitation and Behaviour Oriented Design
- Embodiment vs Memetics: Morality

Bidan Huang



- The Use of Modular Approaches For Robots to Learn Grasping and Manipulation
- Realtime grasping strategies.
- Collaborators: Sahar El-Khoury, Miao Li, Aude Billard (EPFL), Tetsunari Inamura (NII).



- Learning Modules
  - Clustering Control Strategies
  - Encoding by GMM
    - Forward model

      $$p\{s_t, s_{t-1}, a_{t-1} \mid \Omega_F\}$$

    - Responsibility factor

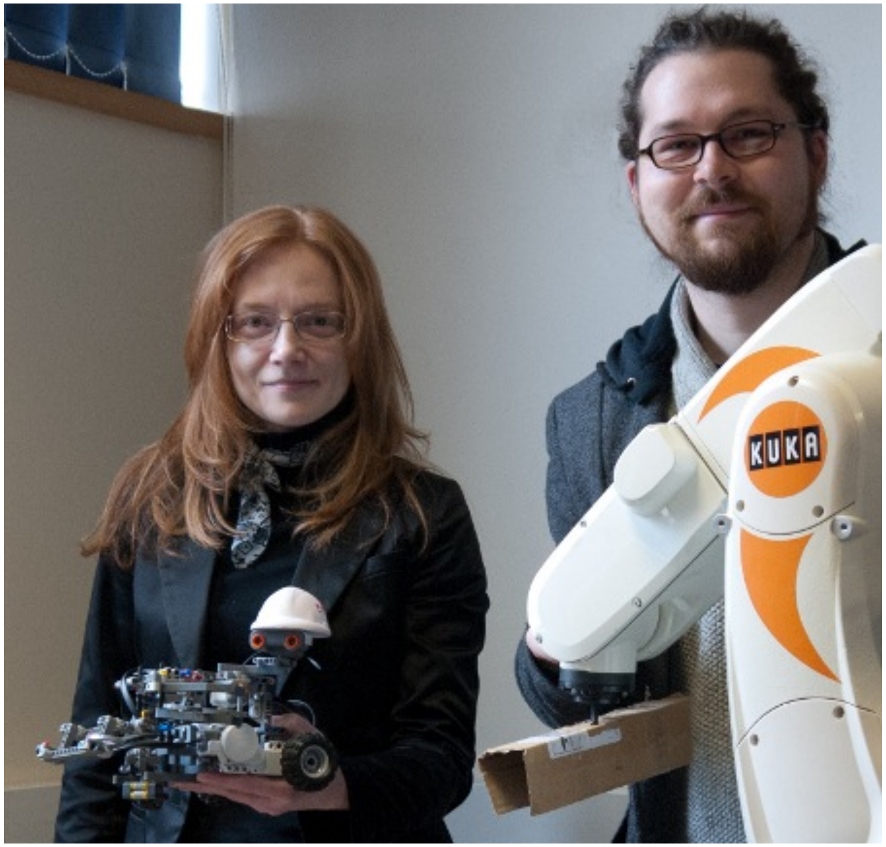
      $$\eta_t = \{s_t, s_{t-1}, a_{t-1}\}$$

      $$\lambda_t^k = \frac{p(\eta_t|\Omega_F^k)}{\sum_{j=1}^{J} p(\eta_t|\Omega_F^j)}$$

    - Inverse model

      $$p\{s_t, st+1, a_t, a_{t-1} \mid \Omega_I\}$$

      $$a_t = \sum_{k=1}^{K} \lambda_t^k a_t^k = \sum_{k=1}^{K} \lambda_t^k E\left(a_t \mid s_{t+1}^*, s_t, a\right.$$



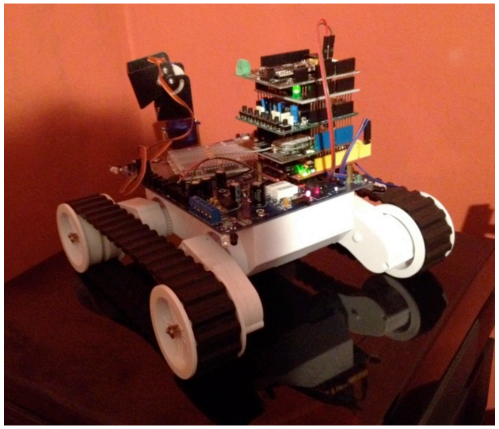
Jekaterina Novikova

- Human Robot Interaction
- Transparently synthetic emotions for collaboration.

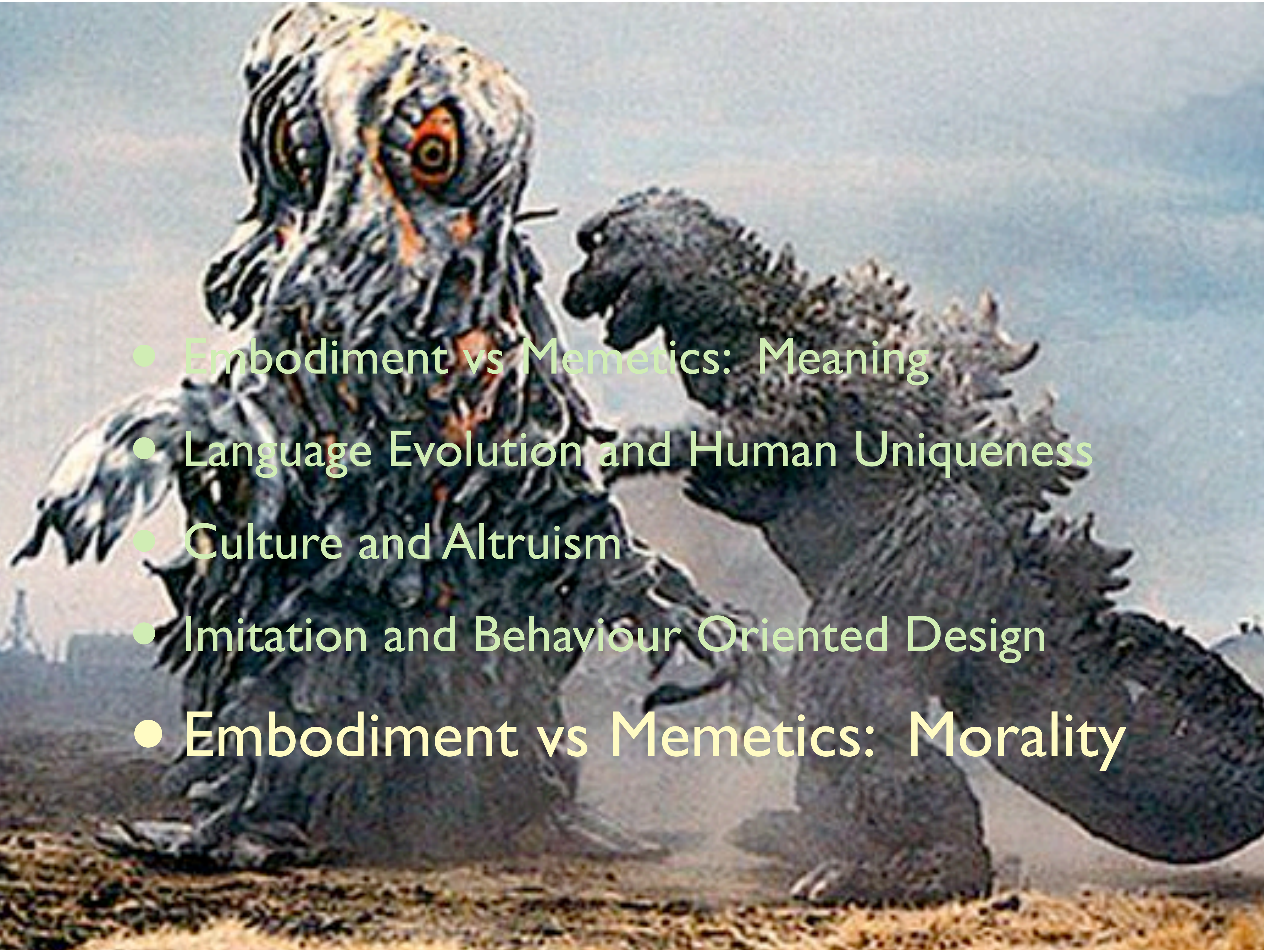
Swen Gaudl

- Game AI
- Learning from observation with Genetic Programming
- Stable, transparent control

Rob Wortham



- Ethical Domestic Robotics
- BOD Arduino



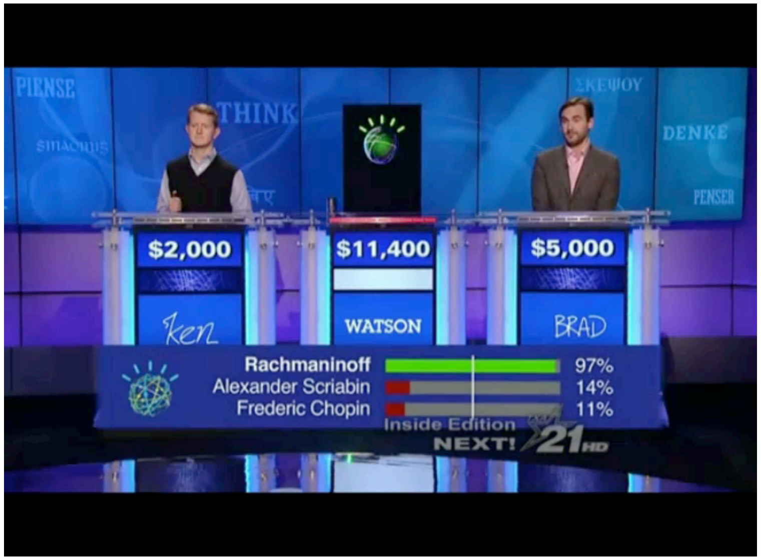
- Embodiment vs Memetics: Meaning
- Language Evolution and Human Uniqueness
- Culture and Altruism
- Imitation and Behaviour Oriented Design
- Embodiment vs Memetics: Morality

A typical slide for me these days…

- What is the current reality of AI?
  - It's here now, changing the world.
- Are the sciences of consciousness and ethics far enough along that we can predict the consequences of AI?
  - Yes.
- What scenarios should we worry about, and which should we seek to accelerate?
  - Give me forty minutes...

(London Futurists, 18 April – on YouTube)

AI Already Owns Our Advantages

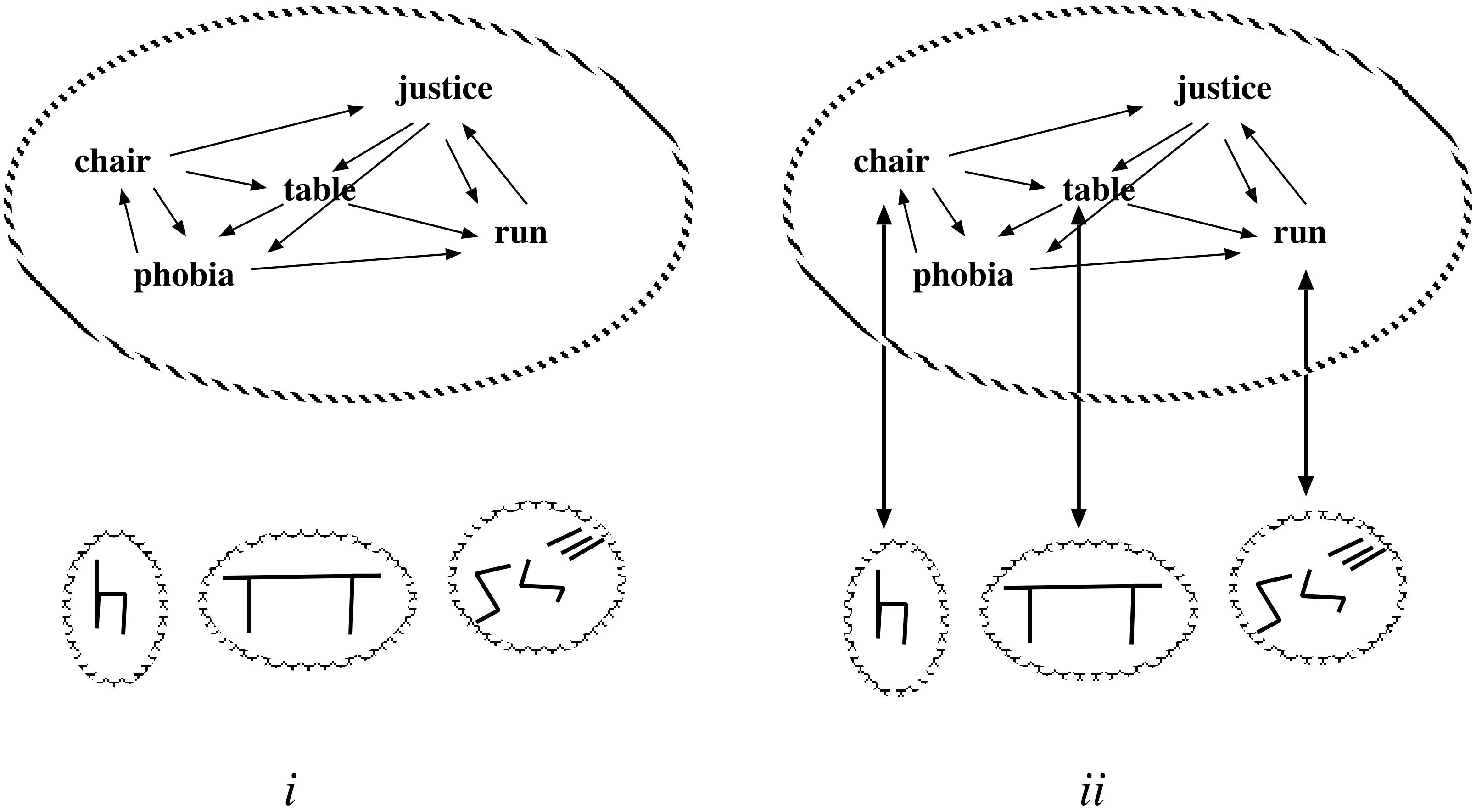


Utopia: Solve hard problems like sustainability; reliably supporting everyone's efforts to self actualise.

Dystopia: Losing autonomy / ability to freely express; catastrophic disruption of the global ecosystem.

Has Memetics won?

Bryson's (2008) Theory of Semantics



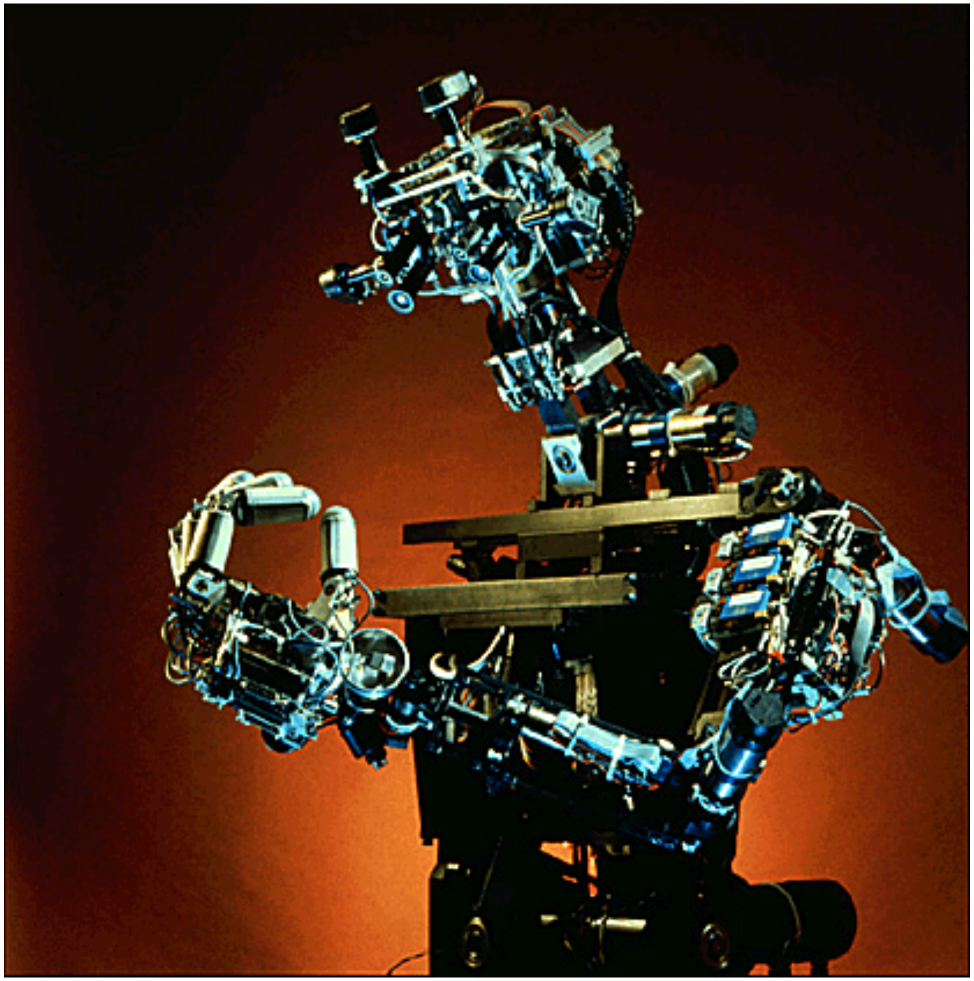
What About Ethics?

Robots are servants
we own.

⇒ Slaves

Bryson (2010)

- For Human Society (us):
  - Pros: feel godlike, culture might persist beyond planetary limits, might produce more useful tools.
  - Cons: political & commercial moral hazard, misattribution of blame / resources.
- For AI (them robots):
  - No Pros: (except maybe for the unbuilt).
  - Cons: compete w/ humans for resources, stress of social dominance, fear of death etc.



People want to make AI they owe obligations to, can fall in love with, etc. – “equals” over which we have complete dominion.

Joanna J. Bryson and Philip P. Kime, “Just an Artifact: Why Machines are Perceived as Moral Agents”, *The Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, Barcelona, Spain, pp. 1641–1646, Morgan Kaufmann, 2011.

We build robots and other AI, determine these systems’ goals. Our authorship of AI is fundamentally different from our relationship to other evolved systems.

a fact

photos: Georgio Metta (top) & Emmanuel Tanguy

Even if we could solve the technical problems of making robots that would persist longer than our civilisation, species or planet, would memetics for the purpose of its own sake make sense?

Our values are rooted in the problems of enculturated apes. Why pass moral responsibility derived from them to machines?

- Our values have and are coevolving with our species.
- Embodied
- A lot of ethical problems are simpler if we build AI and its regulation around humans as the moral subjects.

Conclusions

We are ethically obliged to make robots we are not ethically obliged to.

Deeming robots to be moral agents unethically neglects our responsibility as authors of their intelligence.

normative assertions

Thanks!

Thanks!

My current students:

~~Daniel Taylor~~

Bidan Huang

Dominic Mitchell

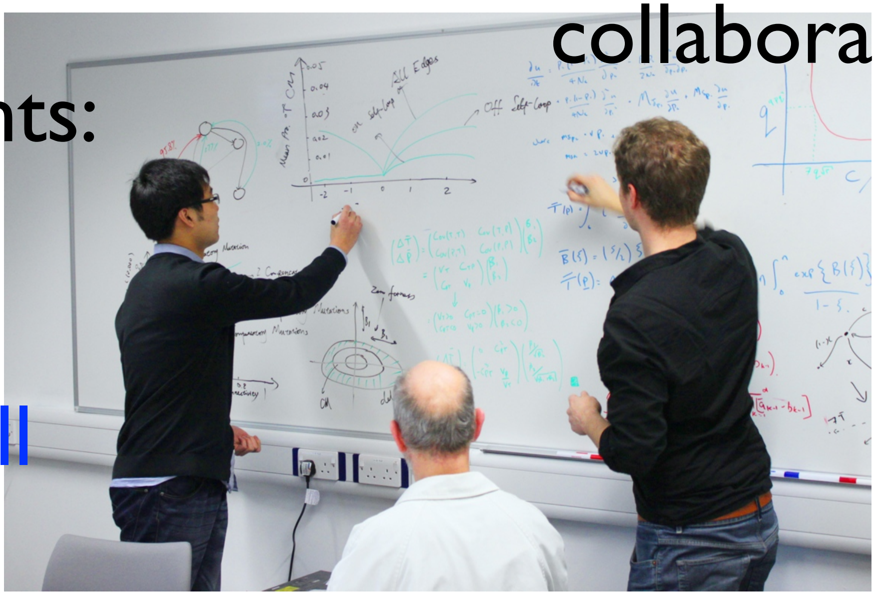
Swen Gaudl

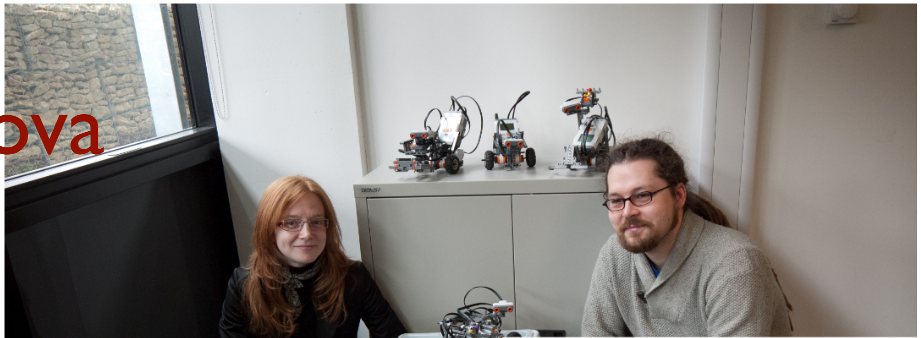
Paul Rauwolf

Jekaterina Novikova

Yifei Wang

Rob Wortham





... and other collaborators

Will Lowe

Dave Gunkel

Special Issue on AI Moral Subjectivity in March 2014

Philosophy & Technology