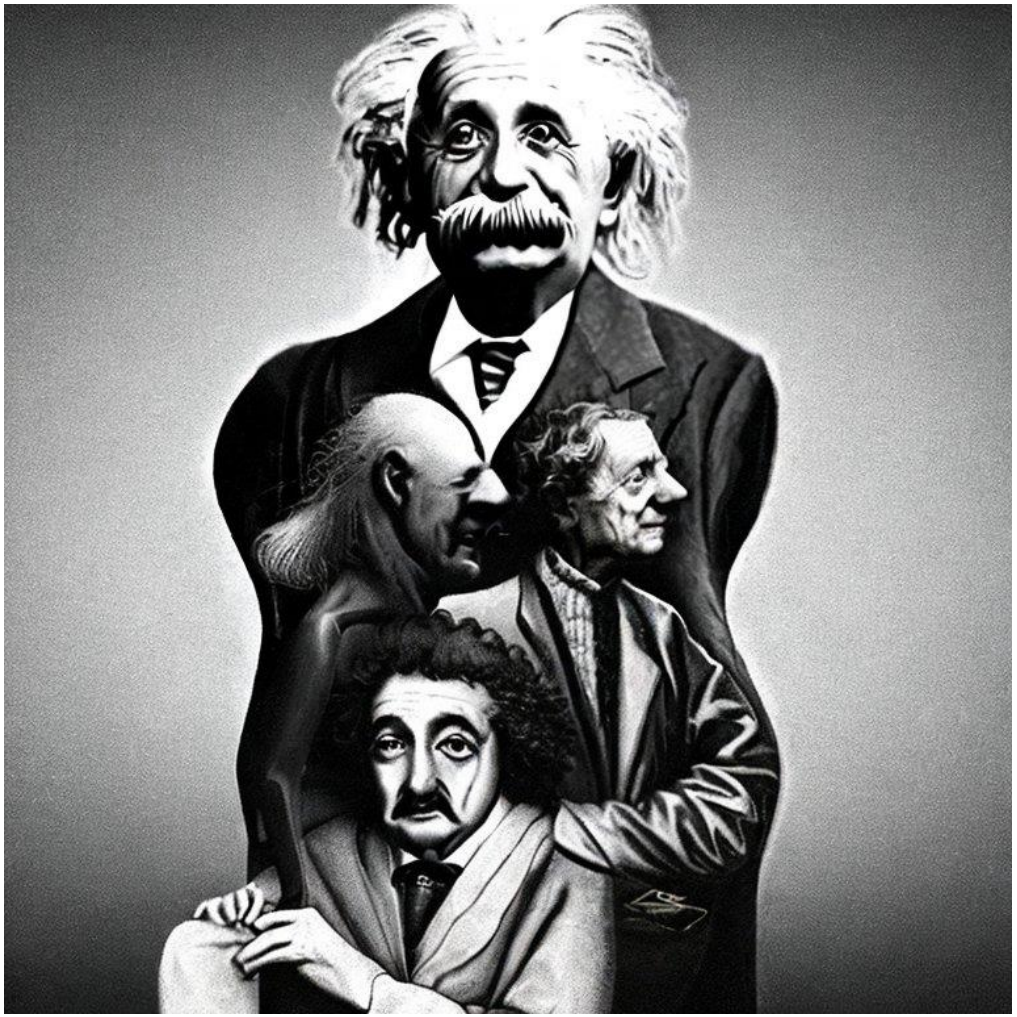
Game Theory and the Free Energy Principle for Artificial Intelligence

Dr. Michael Harré
Modelling and Simulation Group
School of Computer Science
University of Sydney

THE UNIVERSITY OF SYDNEY

https://enterprisetalk.com/tech-trends/ai/human-intelligence-can-fix-ai-shortcomings/

Is “intelligence” singular?



On the shoulders of giants

Stable Diffusion, inspired by Isaac Newton

We are a social species

Hunter-gatherers

Images by Stable Diffusion

We are a social species

Hunter-gatherers

Early agrarian culture

Images by Stable Diffusion

We are a social species

Hunter-gatherers

Early agrarian culture

First city-based civilisations

Images by Stable Diffusion

We are a social species

Hunter-gatherers

Early agrarian culture

First city-based civilisations

Progressively more complex cities, states, and societies

Images by Stable Diffusion

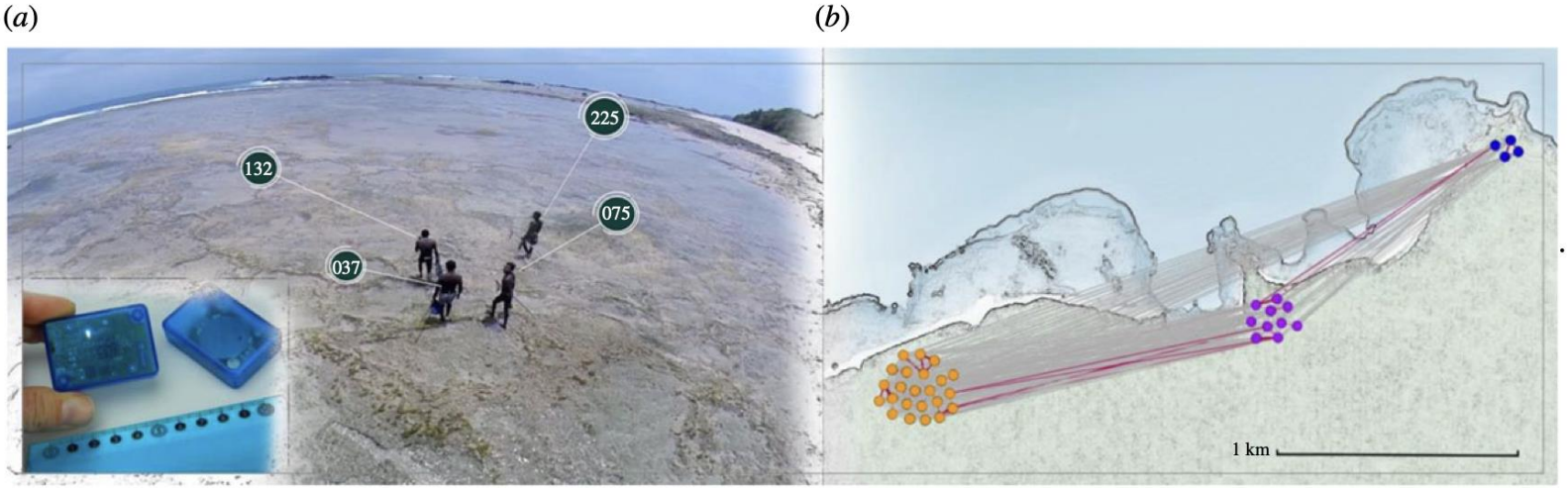
We are a social species

"Limits to higher sophistication [...] in chimpanzees may stem from social features."

The origins of human cumulative culture: from the foraging niche to collective Intelligence (2021)
A. Migliano and L. Vinicius

We are a social species

Long-distance networking is also crucial to foraging, cooperation and cultural exchange.

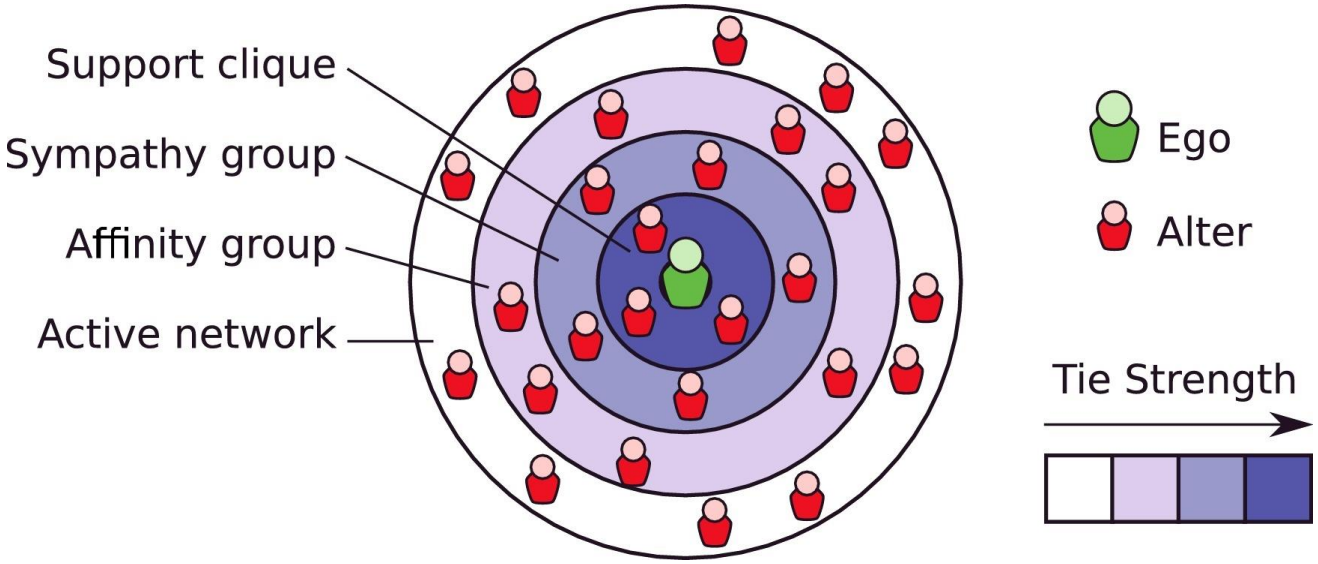


Mapping hunter–gatherer social networks and between-camp migration. New radio sensor technologies ((*a*), insert) can be used to trace contacts between individuals in hunter–gatherer populations (*a*), and reconstruct proximity networks within and between residential camps (dot colours, (*b*)).

The origins of human cumulative culture: from the foraging niche to collective Intelligence (2021)
A. Migliano and L. Vinicius

We are a social species

Our social networks have not increased in size in the last 200,000 years



Ego network structure in online social networks and its impact on information diffusion (2016)
V. Arnaboldi et al

We are a social species

Our social networks have not increased in size in the last 200,000+ years - despite online and other social technologies

How many friends do you need to maintain your social network?

ABC Science / By Anna Salleh

Posted Wed 18 May 2016 at 3:32pm, updated Thu 19 May 2016 at 9:35am

You don't need to like everyone in your network for it to work. *(Getty Images)*

We are a social species

INTERFACE

rsif.royalsocietypublishing.org

Research

The social brain: scale-invariant layering of Erdős–Rényi networks in small-scale human societies

Michael S. Harré and Mikhail Prokopenko

Complex Systems Research Group, Faculty of Engineering and IT, The University of Sydney, Sydney, Australia

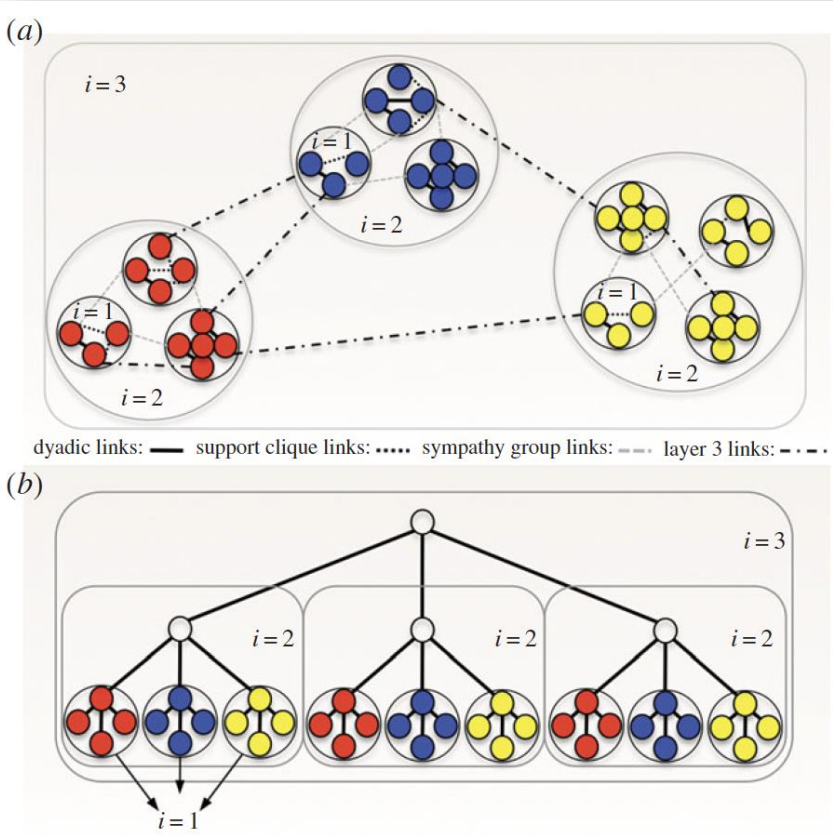
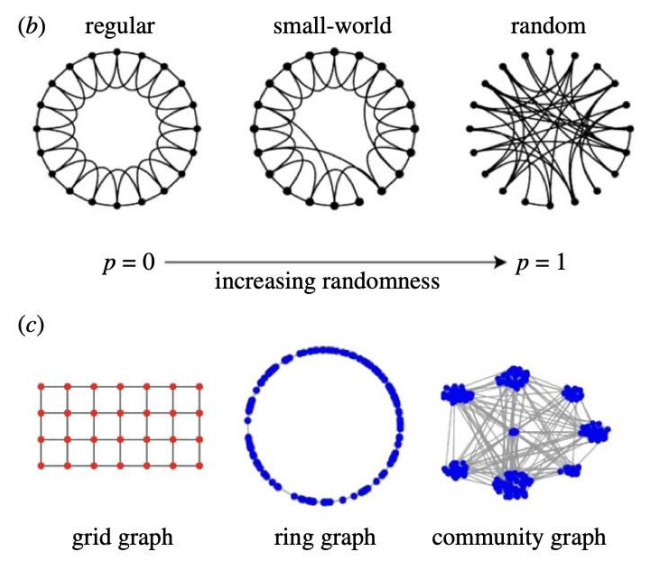


Figure 1. Two different social network models. (*a*) Random links form between sub-group members. As the average number of links per person increases in discrete steps, the network size also increases in predictable, discrete, steps. (b) A structured hierarchy similar to modern military, bureaucratic and corporate structures in which each layer is 'managed' by a coordinator. (Online version in colour.)

We are a social species



Collective minds: social network topology shapes collective cognition

Ida Momennejad

Microsoft Research NYC, New York, NY, USA

https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2020.0315
https://www.nature.com/articles/s41598-019-45576-3

We are a social species

Open access | Research article | First published online September 13, 2022

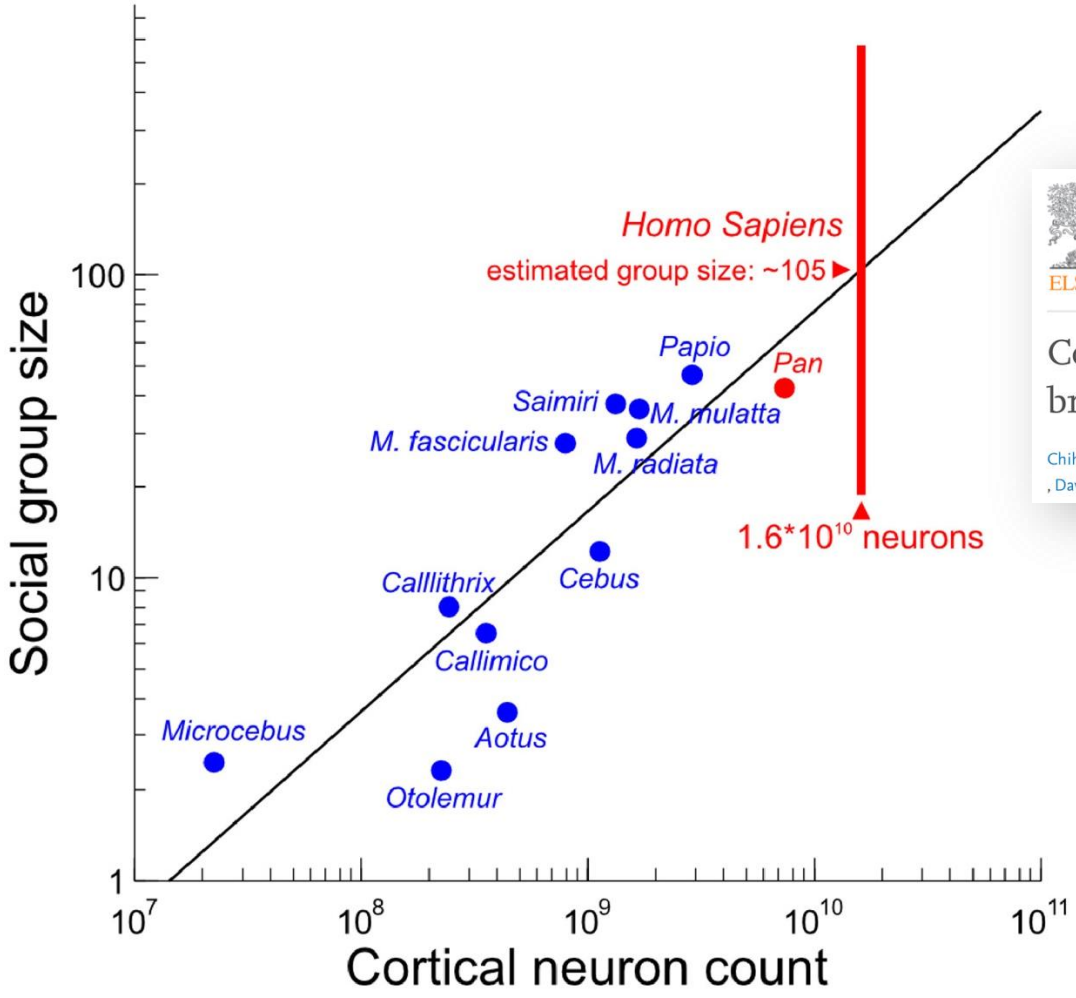
Collective intelligence for deep learning: A survey of recent developments

David Ha and Yujin Tang View all authors and affiliations

All Articles | https://doi.org/10.1177/26339137221114874

Figure 2. Left: Trajan's Bridge at Alcantara, built in AD 106 by Romans (Wikipedia, 2022). Right: Army ants forming a bridge (Jenal, 2011).

We are a social species



NeuroImage
Volume 245, 15 December 2021, 118693

Comparative connectomics of the primate social brain

Chihiro Yokoyama [a, 1, 2], Joonas A. Autio [a, 1], Takuro Ikeda [a, 1], Jérôme Sallet [b, c], Rogier B. Mars [d, e], David C. Van Essen [f], Matthew F. Glasser [f, g], Norihiro Sadato [h, i], Takuya Hayashi [a, j]

We are a socio-cognitive species

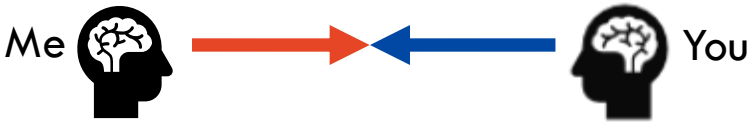
But how do we do it?

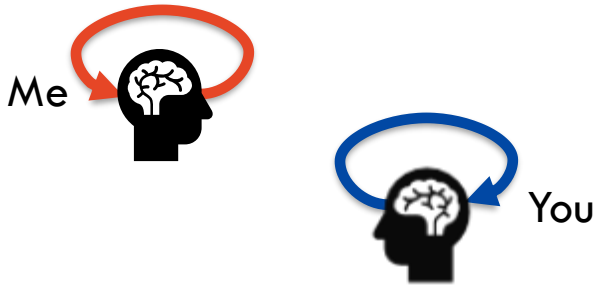
We are a socio-cognitive species

But how do we do it?

We are a socio-cognitive species

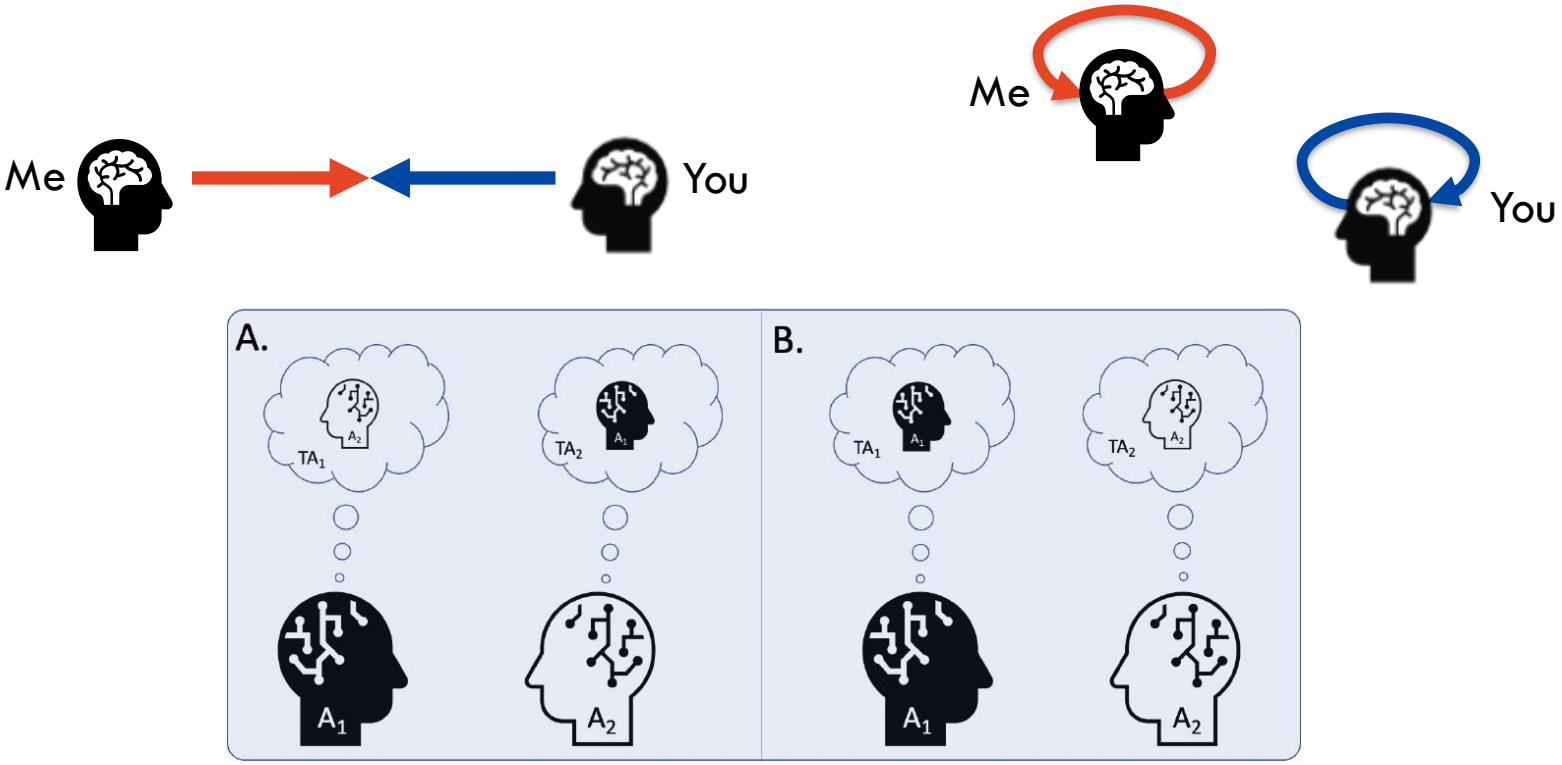
But how do we do it?





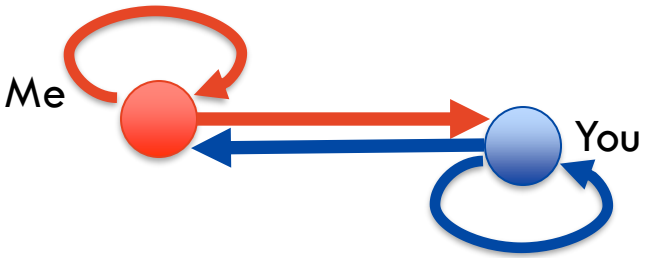
We are a socio-cognitive species

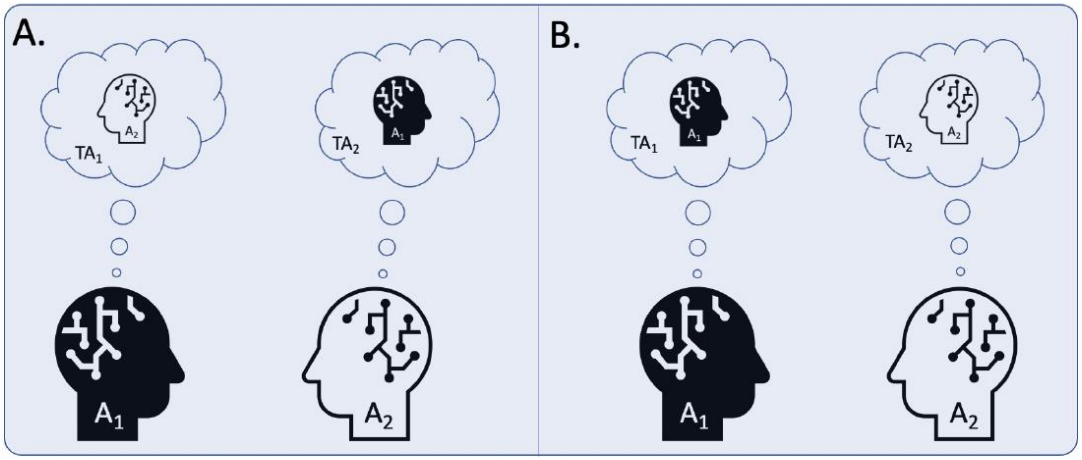
But how do we do it?



We are a socio-cognitive species

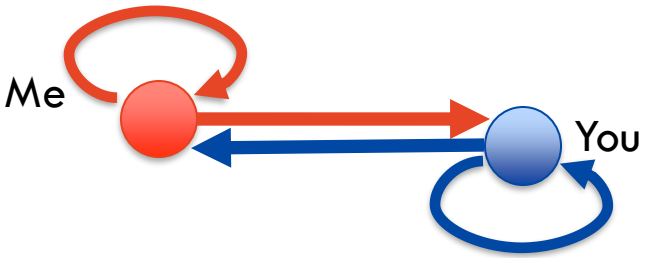
But how do we do it?

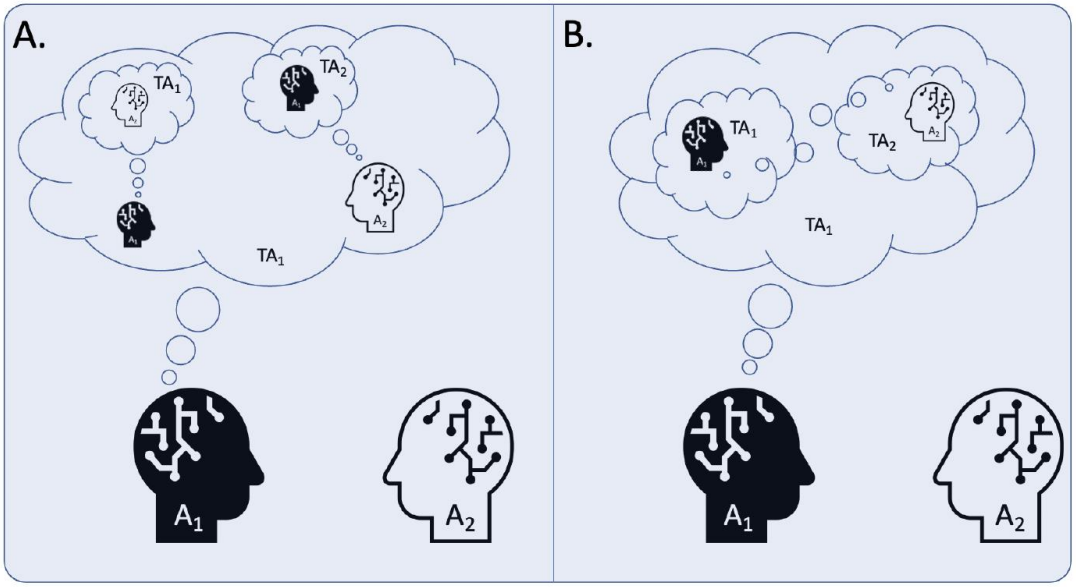




We are a socio-cognitive species

But how do we do it?



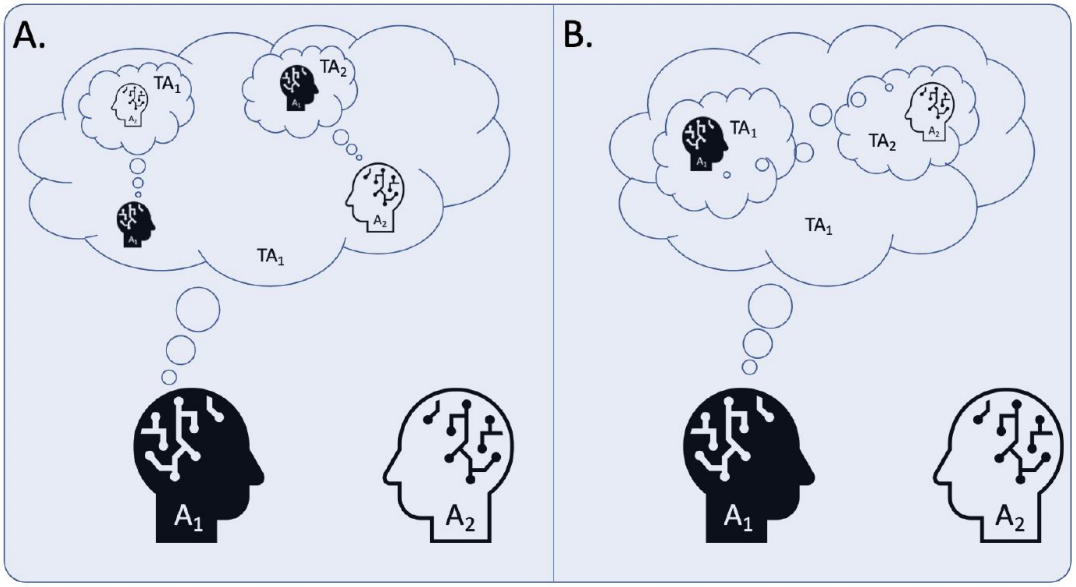


We are a socio-cognitive species

But how do we do it?

"We must try to **put ourselves inside their skin** and **look at us through their eyes**, just to **understand the thoughts** that lie **behind their decisions and their actions**."

Robert McNamara *The Fog of War*

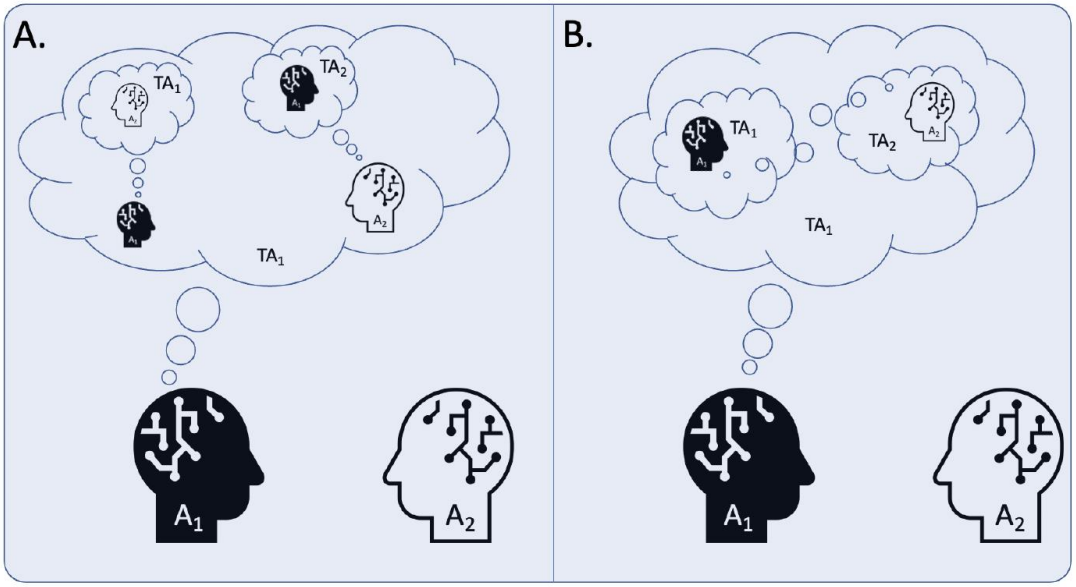


We are a socio-cognitive species

But how do we do it?

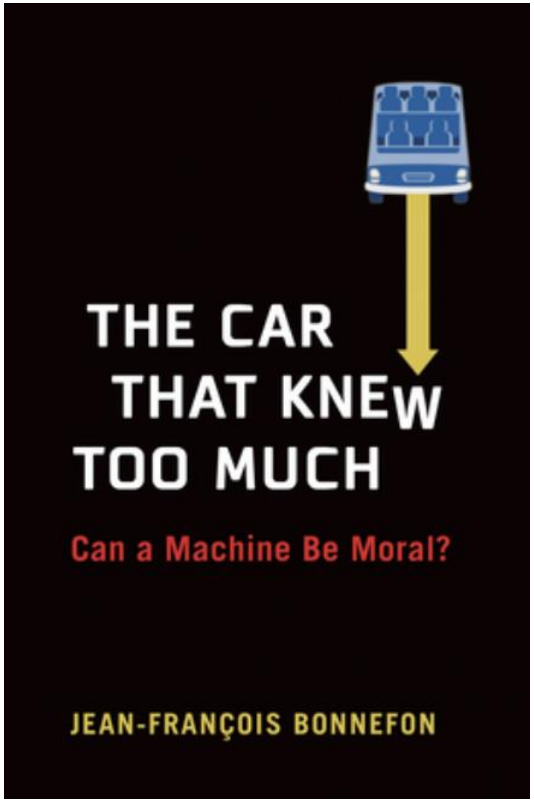
"In the Cuban Missile Crisis […] **we did put ourselves in the skin of the Soviets**.
In the case of Vietnam, **we didn't know them well enough to empathize**.
And there was total misunderstanding as a result."

Robert McNamara *The Fog of War*



We are a socio-cognitive species

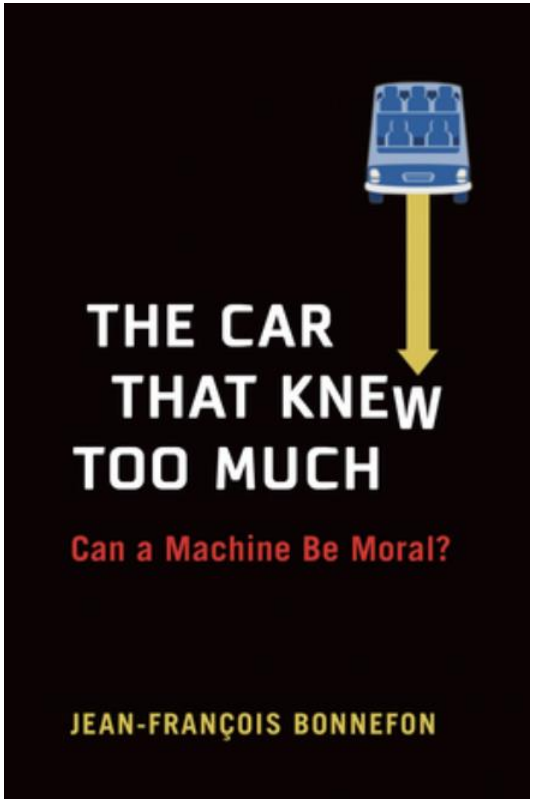
But how do we do it? – It's difficult to get right

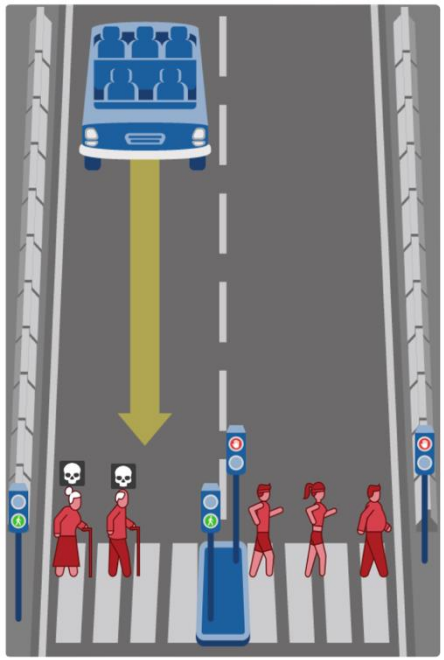


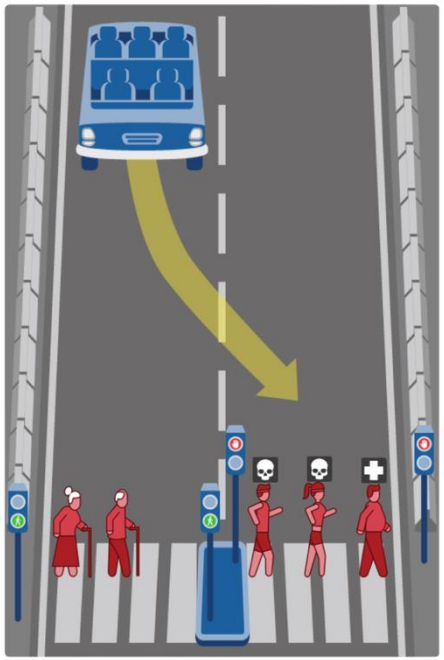
www.moralmachine.net

We are a socio-cognitive species

But how do we do it? – It's difficult to get right



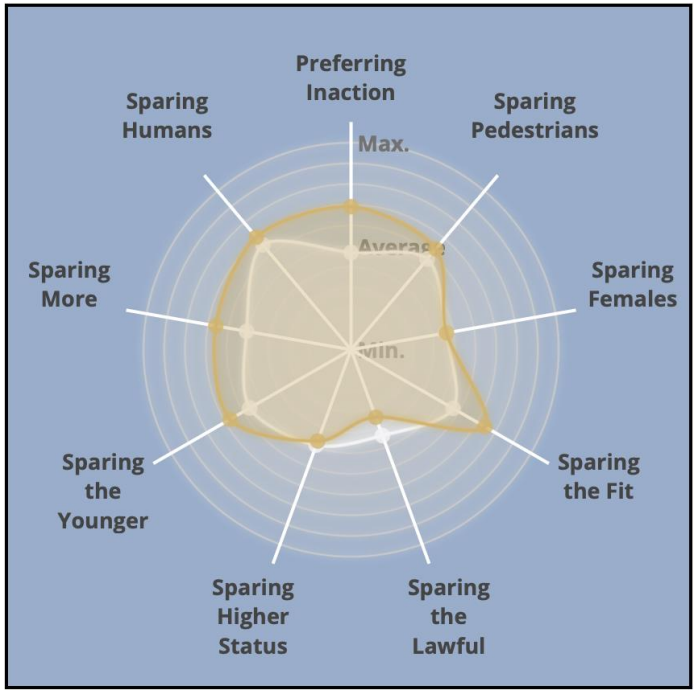




www.moralmachine.net

We are a socio-cognitive species

But how do we do it? – It's difficult to get right

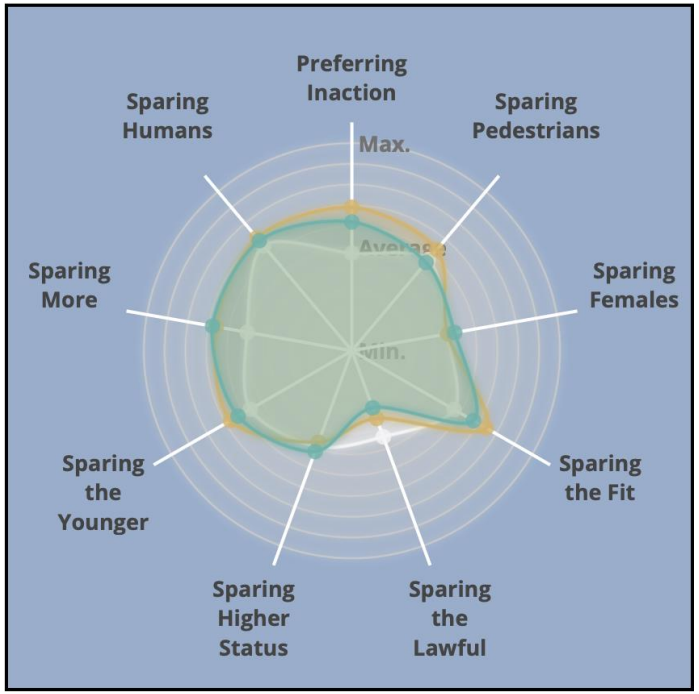


Australia

www.moralmachine.net

We are a socio-cognitive species

But how do we do it? – It's difficult to get right

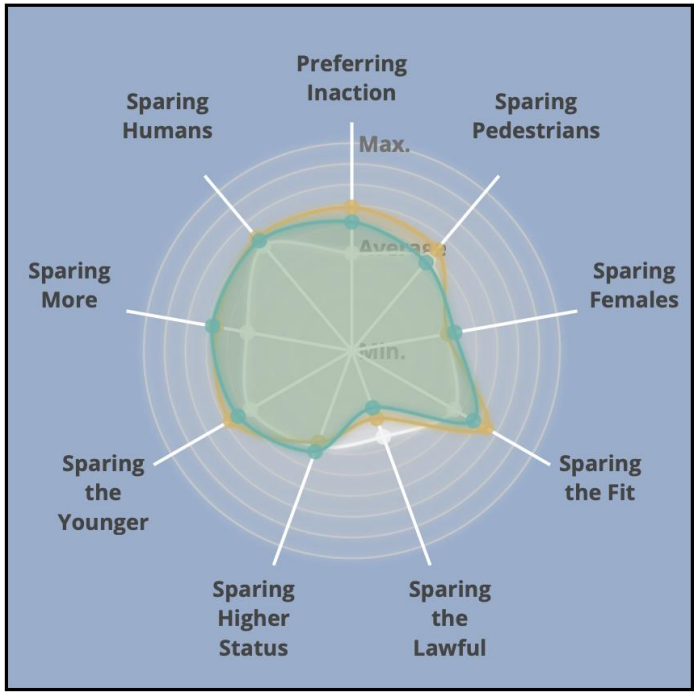


Australia vs USA

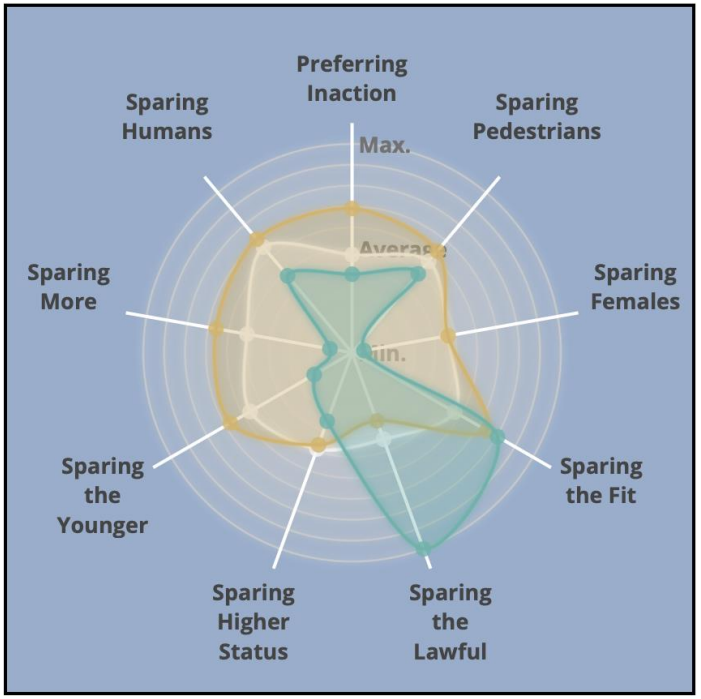
www.moralmachine.net

We are a socio-cognitive species

But how do we do it? – It's difficult to get right



Australia vs USA



Australia vs Brunei

www.moralmachine.net

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

A deceptively simple, extremely nuanced game, of strategy and cunning.

JFK enjoyed the game, as did Henry Kissinger.

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

To win, a player must not only play strategically, but form alliances, negotiate, persuade, threaten, and occasionally deceive.

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

To win, a player must not only **play strategically**, but form alliances, negotiate, persuade, threaten, and occasionally deceive.

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

To win, a player must not only **play strategically**, but **form alliances**, negotiate, persuade, threaten, and occasionally deceive.

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

To win, a player must not only **play strategically**, but **form alliances**, **negotiate**, persuade, threaten, and occasionally deceive.

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

To win, a player must not only **play strategically**, but **form alliances**, **negotiate**, **persuade**, threaten, and occasionally deceive.

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

To win, a player must not only **play strategically**, but **form alliances**, **negotiate**, **persuade**, **threaten**, and **occasionally deceive**.

Where is AI up to?

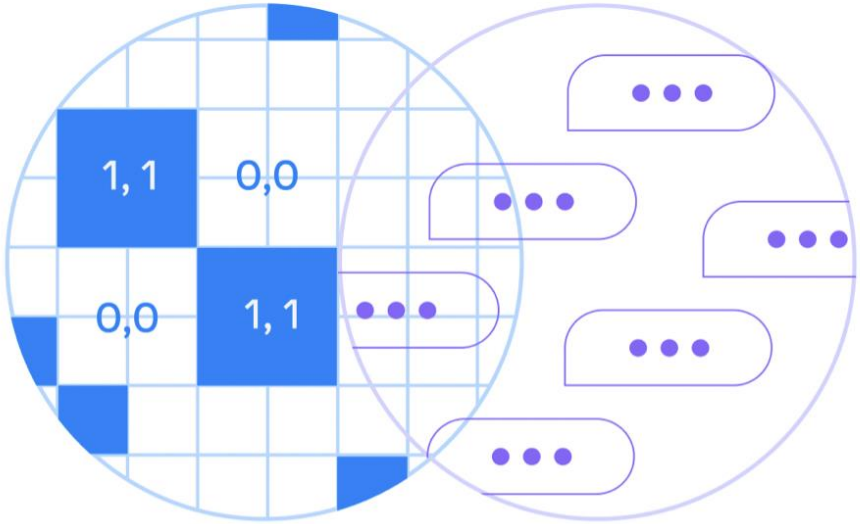
Cooperating and competing in the game "Diplomacy"

It therefore presents challenges for AI that **go far beyond** those faced either by systems that play **games like Go and chess or by chatbots** that engage in dialog in less complex settings.

The Game That Ruins Friendships and Shapes Careers

Where is AI up to?

Cooperating and competing in the game "Diplomacy"



Strategic Reasoning

CICERO predicts the moves other players are likely to make, as well as what moves they expect CICERO to make, and uses that information to create a strategic plan.

Natural Language Processing

CICERO grounds its conversations in a set of carefully chosen plans, so it can negotiate, offer advice, share information, and make agreements with other players.

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

Where is AI up to?

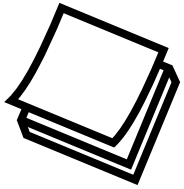
Cooperating and competing in the game "Diplomacy"

Where is AI up to?

Cooperating and competing in the game "Diplomacy"

Where is AI up to?

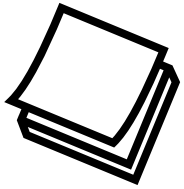
Lets extend this: AI in the loop policy analysis



Cold war documents

Where is AI up to?

Lets extend this: AI in the loop policy analysis



Cold war documents

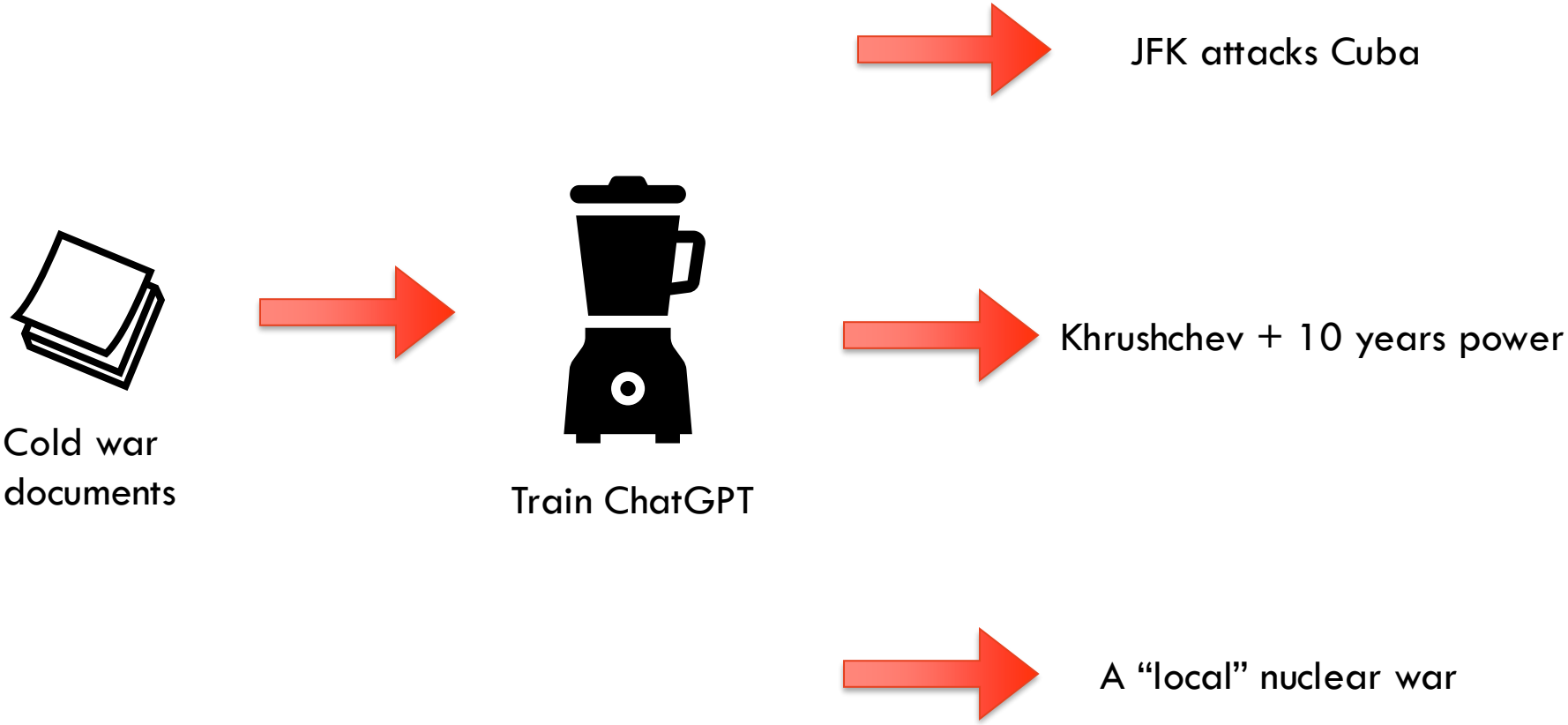
Train ChatGPT

Where is AI up to?

Lets extend this: AI in the loop policy analysis

Where is AI up to?

Lets extend this: AI in the loop policy analysis



What is AI up to?

nature
COMMENT | 04 May 2021

www.cooperativeai.org

Cooperative AI: machines must learn to find common ground

To help humanity solve fundamental problems of cooperation, scientists need to reconceive artificial intelligence as deeply social.

Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson & Thore Graepel

A huddle at the 2017 United Nations Climate Change Conference, where attendees cooperated on mutually beneficial joint actions on climate. Credit: Sean Gallup/Getty

What is AI up to?

nature

COMMENT | 04 May 2021

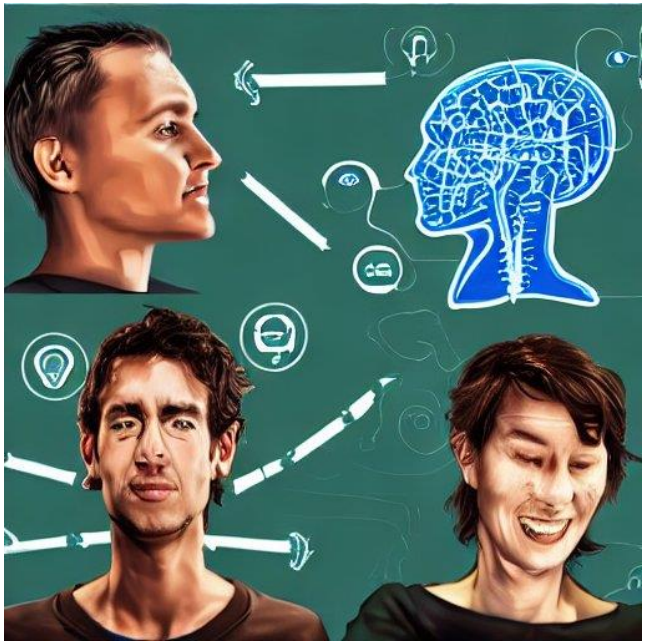
www.cooperativeai.org

Cooperative AI: machines must learn to find common ground

To help humanity solve fundamental problems of cooperation, scientists need to reconceive artificial intelligence as deeply social.

Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson & Thore Graepel

Tit - for - Tat



Free rider

Cooperative

Stable Diffusion

Principles of Games Against Nature

Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory

Peter D. Grünwald, A. Philip Dawid

Ann. Statist. 32(4): 1367-1433 (August 2004). DOI: 10.1214/009053604000000553

Principles of Games Against Nature

Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory

Peter D. Grünwald, A. Philip Dawid

Ann. Statist. 32(4): 1367-1433 (August 2004). DOI: 10.1214/009053604000000553

$$\mathrm{H}(X) := -\sum_{x \in \mathcal{X}} p(x) \log p(x) = \mathbb{E}[-\log p(X)],$$

Entropy is the expectation of -log(p(x))

$$H(P, Q) = -\sum_{x \in \mathcal{X}} p(x) \log q(x) \quad \textbf{(Eq.1)}$$

Definition of "Cross Entropy" for discrete distributions

The Cross Entropy term is the log-loss in a "game against nature": Nature = p(x)

Principles of Games Against Nature

Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory

Peter D. Grünwald, A. Philip Dawid

Ann. Statist. 32(4): 1367-1433 (August 2004). DOI: 10.1214/009053604000000553

$$H(P) = \inf_{q \in \mathcal{A}} \mathrm{E}_P\{-\log q(X)\} \leq \mathrm{E}_P\{-\log p^*(X)\} = H(P^*),$$

Principles of Games Against Nature

Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory

Peter D. Grünwald, A. Philip Dawid

Ann. Statist. 32(4): 1367-1433 (August 2004). DOI: 10.1214/009053604000000553

$$H(P) = \inf_{q \in \mathcal{A}} \mathrm{E}_P\{-\log q(X)\} \leq \mathrm{E}_P\{-\log p^*(X)\} = H(P^*),$$

$$\sup_{P \in \Gamma} \mathrm{E}_P\{-\log p^*(X)\} = H(P^*).$$

Principles of Games Against Nature

Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory

Peter D. Grünwald, A. Philip Dawid

Ann. Statist. 32(4): 1367-1433 (August 2004). DOI: 10.1214/009053604000000553

$$H(P) = \inf_{q \in \mathcal{A}} \mathrm{E}_P\{-\log q(X)\} \leq \mathrm{E}_P\{-\log p^*(X)\} = H(P^*),$$

$$\sup_{P \in \Gamma} \mathrm{E}_P\{-\log p^*(X)\} = H(P^*).$$

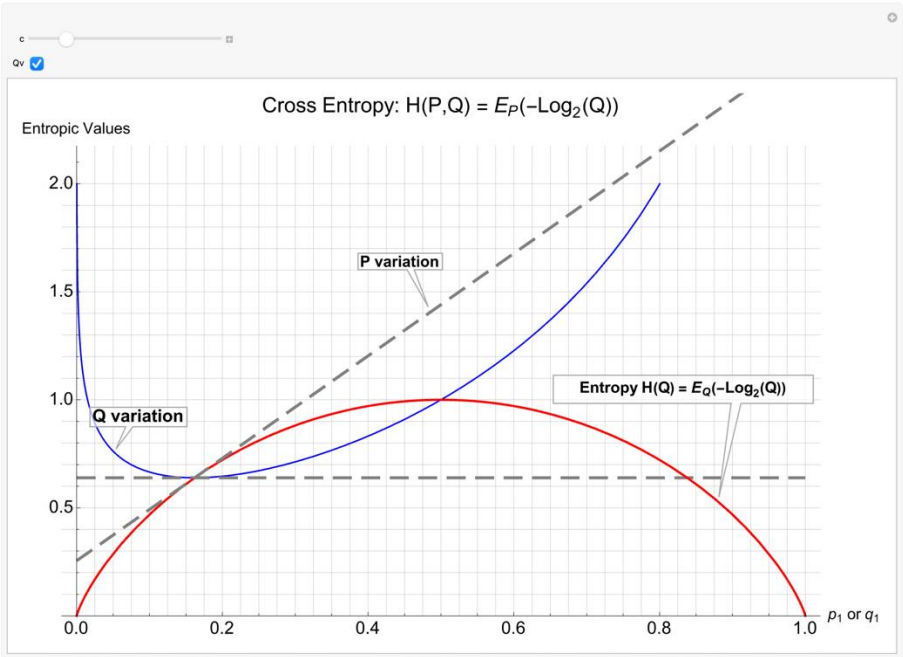
$$\sup_{P \in \Gamma} \inf_{q \in \mathcal{A}} \mathrm{E}_P\{-\log q(X)\} = \inf_{q \in \mathcal{A}} \sup_{P \in \Gamma} \mathrm{E}_P\{-\log q(X)\}.$$

Principles of Games Against Nature

Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory

Peter D. Grünwald, A. Philip Dawid

Ann. Statist. 32(4): 1367-1433 (August 2004). DOI: 10.1214/009053604000000553



Entropy is the minimization of the loss In a game against nature, achieved by the "player" varying q(x)

In the diagram the blue curve is the cross-entropy term for different values of q(x)

The entropy is the red curve for different values of q(x)

Maximum Entropy and Game Theory

Free Energy, Free Utility:

$$H(p) = -\sum_x p(x)\log(p(x))$$

$V(p)$ = internal energy (potential function)

$\mathcal{F}(p) = V(p) - T\,H(p)$ (where T is temperature)

} Helmholtz Free Energy (1)

Maximum Entropy and Game Theory

Free Energy, Free Utility:

$$H(p) = -\sum_x p(x)\log(p(x))$$

$V(p)$ = internal energy (potential function)

$\mathcal{F}(p) = V(p) - T\,H(p)$ (where T is temperature)

} Helmholtz Free Energy (1)

$$H(P) = -\sum_{x,y} P(x,y)\log(P(x,y))$$

$U_a(P) = E_p[\,U_a(x,y)\,]$ (expected utility for a)

$\mathcal{F}(P) = U_a(P) - T\,H(P)$ (T is uncertainty or error)

} Free Utility (2 players) (2)

Maximum Entropy and Game Theory

$$H(P) = -\sum_{x,y} P(x,y)\log(P(x,y))$$

$$U_a(P) = E_p[\, U_a(x,y)\,] \text{ (expected utility for } a)$$

$$\mathcal{F}(P) = U_a(P) - T\,H(P) \quad (T \text{ is uncertainty or error})$$

} Free Utility (2 players) (2)

Optimising these "free functionals" leads to standard exponential solutions:

$$P(x) \propto \exp(-\beta\, V(P))$$

$$P(x \mid x = x_i) \propto \exp(\,\beta\, U_a(\,P(y) \mid x = x_i\,)\,)$$
$$P(y \mid y = y_i) \propto \exp(\,\beta\, U_b(\,P(x) \mid y = y_i\,)\,)$$

Strategic Choice of Preferences: the Persona Model

David Wolpert , Julian Jamison , David Newth and Michael Harre

From the journal The B.E. Journal of Theoretical Economics
https://doi.org/10.2202/1935-1704.1593

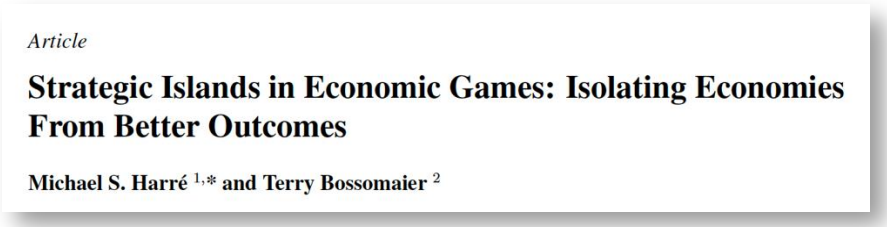
Maximum Entropy and Game Theory

Col :

U_c	L	R
T	a_c	b_c
B	c_c	d_c

Row :

U_r	L	R
T	a_r	b_r
B	c_r	d_r



Article

Strategic Islands in Economic Games: Isolating Economies From Better Outcomes

Michael S. Harré [1,*] and Terry Bossomaier [2]

Maximum Entropy and Game Theory

Col :

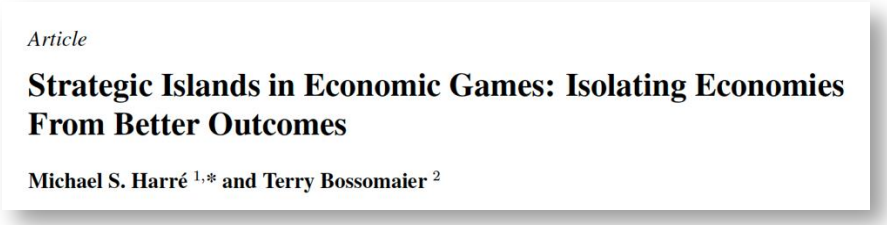
U_c	L	R
T	a_c	b_c
B	c_c	d_c

Row :

U_r	L	R
T	a_r	b_r
B	c_r	d_r

$$E(U_c) = \sum_{i,j} p_r(i) p_c(j) U_c^{i,j}$$

$$E(U_r) = \sum_{i,j} p_r(i) p_c(j) U_r^{i,j}$$



Article

Strategic Islands in Economic Games: Isolating Economies From Better Outcomes

Michael S. Harré [1,*] and Terry Bossomaier [2]

Maximum Entropy and Game Theory

Col :

U_c	L	R
T	a_c	b_c
B	c_c	d_c

Row :

U_r	L	R
T	a_r	b_r
B	c_r	d_r

$$E(U_c) = \sum_{i,j} p_r(i)p_c(j)U_c^{i,j}$$

$$E(U_r) = \sum_{i,j} p_r(i)p_c(j)U_r^{i,j}$$

$$E(U_c)_* : \sum_{i,j} p_r(i)p_c^*(j)U_c^{i,j} \geq \sum_{i,j} p_r(i)p_c(j)U_c^{i,j} \ \forall \ p_c(j) \ \text{and}$$

$$E(U_r)_* : \sum_{i,j} p_r^*(i)p_c(j)U_r^{i,j} \geq \sum_{i,j} p_r(i)p_c(j)U_c^{i,j} \ \forall \ p_r(i).$$

Article

Strategic Islands in Economic Games: Isolating Economies From Better Outcomes

Michael S. Harré [1,*] and Terry Bossomaier [2]

Maximum Entropy and Game Theory

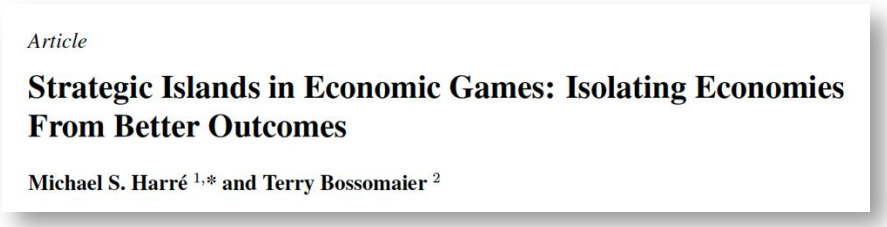
Col :

U_c	L	R
T	a_c	b_c
B	c_c	d_c

Row :

U_r	L	R
T	a_r	b_r
B	c_r	d_r

$$p_x(i) \geq 0 \,\forall\, i, \quad \sum_{i=1}^{k} p_x(i) = 1, \quad \sum_{i,j=1}^{k} p_r(i)p_c(j)U_x^{i,j} = E(U_x).$$



Article

Strategic Islands in Economic Games: Isolating Economies From Better Outcomes

Michael S. Harré [1,*] and Terry Bossomaier [2]

Maximum Entropy and Game Theory

Col :

U_c	L	R
T	a_c	b_c
B	c_c	d_c

Row :

U_r	L	R
T	a_r	b_r
B	c_r	d_r

$$p_x(i) \geq 0 \,\forall\, i, \quad \sum_{i=1}^{k} p_x(i) = 1, \quad \sum_{i,j=1}^{k} p_r(i)p_c(j)U_x^{i,j} = E(U_x).$$

$$\mathcal{L}(p_x) = S(p_x) + \beta_x \sum_{i,j} p_r(i)p_c(j)U_x^{i,j} + \beta_0 \sum_i p_x(i),$$

Article

Strategic Islands in Economic Games: Isolating Economies From Better Outcomes

Michael S. Harré [1,*] and Terry Bossomaier [2]

Maximum Entropy and Game Theory

Col :

U_c	L	R
T	a_c	b_c
B	c_c	d_c

Row :

U_r	L	R
T	a_r	b_r
B	c_r	d_r

$$p_x(i) \geq 0 \,\forall\, i, \quad \sum_{i=1}^{k} p_x(i) = 1, \quad \sum_{i,j=1}^{k} p_r(i)p_c(j)U_x^{i,j} = E(U_x).$$

$$\mathcal{L}(p_x) = S(p_x) + \beta_x \sum_{i,j} p_r(i)p_c(j)U_x^{i,j} + \beta_0 \sum_i p_x(i),$$

$$\frac{\partial \mathcal{L}(p_x)}{\partial p_x} = -\ln(p_x(i)) + \beta_x \sum_j p_x(j)U_x^{i,j} + \beta_0 - 1 = 0,$$

Article

Strategic Islands in Economic Games: Isolating Economies From Better Outcomes

Michael S. Harré [1,*] and Terry Bossomaier [2]

Maximum Entropy and Game Theory

Col :

U_c	L	R
T	a_c	b_c
B	c_c	d_c

Row :

U_r	L	R
T	a_r	b_r
B	c_r	d_r

$$p_x(i) \geq 0 \,\forall\, i, \quad \sum_{i=1}^{k} p_x(i) = 1, \quad \sum_{i,j=1}^{k} p_r(i)p_c(j)U_x^{i,j} = E(U_x).$$

$$\begin{aligned}
\mathcal{L}(p_x) &= S(p_x) + \beta_x \sum_{i,j} p_r(i)p_c(j)U_x^{i,j} + \beta_0 \sum_i p_x(i), \\
\frac{\partial \mathcal{L}(p_x)}{\partial p_x} &= -\ln(p_x(i)) + \beta_x \sum_j p_x(j)U_x^{i,j} + \beta_0 - 1 = 0, \\
p_x(i) &= \mathcal{Z}_x^{-1} \exp\left(\beta_x \sum_j p_x(j)U_x^{i,j}\right), \\
&= \mathcal{Z}_x^{-1} \exp\left(\beta_x E(U_x|s_x = i)\right),
\end{aligned}$$

This the Quantal Response Equilibrium (QRE)

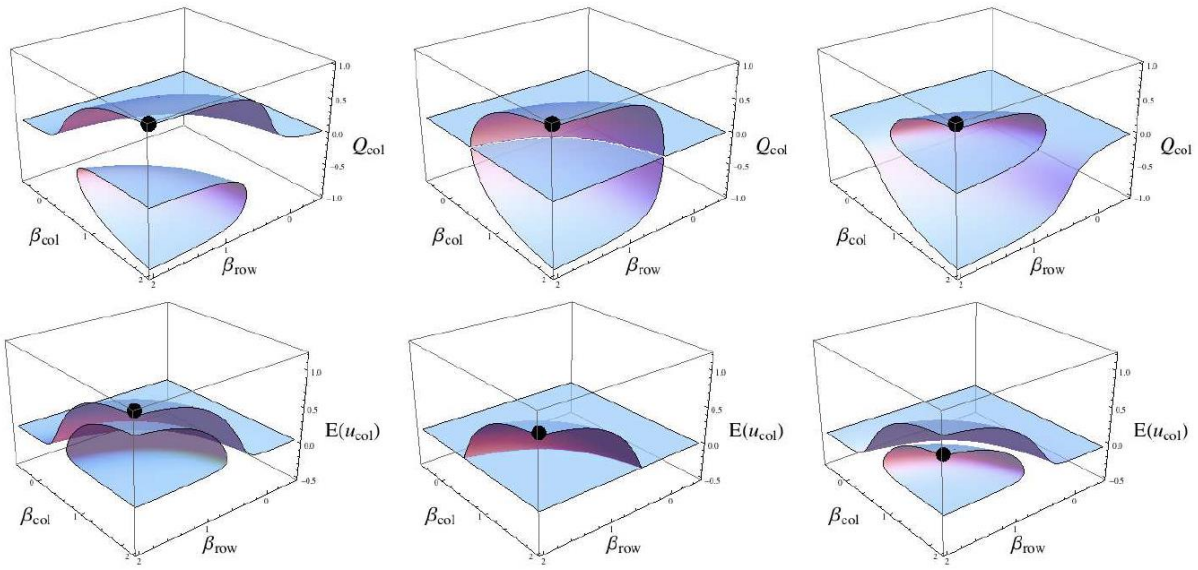
Article

Strategic Islands in Economic Games: Isolating Economies From Better Outcomes

Michael S. Harré [1,*] and Terry Bossomaier [2]

Maximum Entropy and Game Theory

Figure 4. Perturbed QRE solutions for $\delta_c = \delta_r \in \{0.2, 0, -0.2\}$ from left to right with a β pair $\beta_c = \beta_r = 2$, the equilibrium strategy is where the black dot is, see Equations (38)–(39).



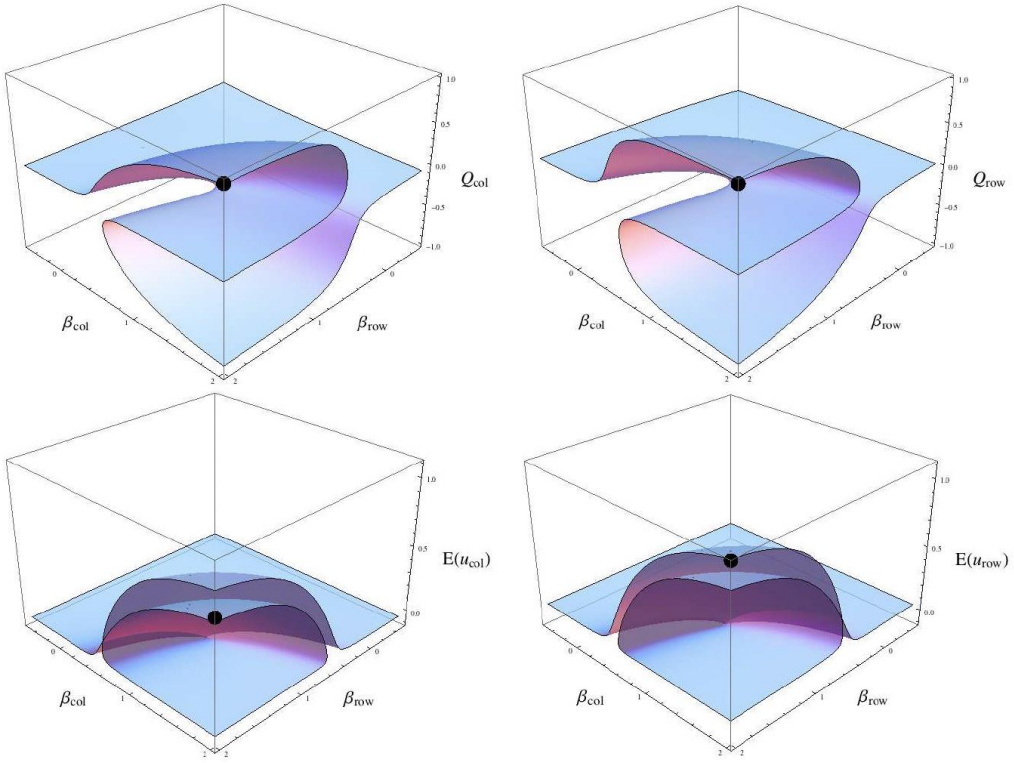
Article

Strategic Islands in Economic Games: Isolating Economies From Better Outcomes

Michael S. Harré [1,*] and Terry Bossomaier [2]

Maximum Entropy and Game Theory

Figure 6. Perturbed QRE solutions for both players for Q values and the corresponding expected utilities: $\{\delta_r, \delta_c\} = \{0.2, -0.2\}$.



Article

Strategic Islands in Economic Games: Isolating Economies From Better Outcomes

Michael S. Harré [1,*] and Terry Bossomaier [2]

Friston's Free Energy Principle for Cognition

3.1. Free Energy in Physics

We first introduce the Helmholtz free energy $\mathscr{F}_p$ used in physics as the internal energy of the system $V(x)$ having discrete states *x* less the Shannon entropy of the system:

$$H(x) = -\sum_i P(x_i) \log (P(x_i)) \tag{3}$$

multiplied by the temperature of the system *T*:

$$\mathscr{F}_p = V(x) - TH(x). \tag{4}$$

Information Theory for Agents in Artificial Intelligence, Psychology, and Economics

by Michael S. Harré

Complex Systems Research Group, Faculty of Engineering, The University of Sydney, Sydney 2006, Australia

Entropy **2021**, *23*(3), 310; https://doi.org/10.3390/e23030310

Friston's Free Energy Principle for Cognition

3.2. Free Utility in Economics

Equation (4) has an economic counterpart that appears, for example, in the earlier work of Wolpert [57,58] on predictive game theory and collective intelligence. In its simplest form an agent chooses a distribution p over the utilities $U(x_i)$ over a finite set of discrete choices $\{x_i\}$ such that they have an expected utility:

$$E_p\left[U\left(x\right)\right] = \sum_i p_i U\left(x_i\right) \tag{5}$$

The 'free utility' of this situation is given by:

$$\mathscr{F}_g \;=\; E_q\left[U\left(x\right)\right] - TH\left(x\right), \tag{6}$$

Information Theory for Agents in Artificial Intelligence, Psychology, and Economics

by Michael S. Harré

Complex Systems Research Group, Faculty of Engineering, The University of Sydney, Sydney 2006, Australia

Entropy **2021**, *23*(3), 310; https://doi.org/10.3390/e23030310

Friston's Free Energy Principle for Cognition

$$P(x_i) = \mathscr{Z}^{-1} e^{\beta E_p[G(x)]} \tag{7}$$

with $\mathscr{Z}$ normalising the distribution and the free energy is:

$$\mathscr{F}(p) = \frac{\partial \mathscr{Z}}{\partial \beta} + \beta^{-1} \sum_x P(x) \log\left(P(x)\right), \tag{8}$$

$$\begin{aligned} &= E_p\left[G(x)\right] - \beta^{-1} H(x), \\ &= \text{Utility} - \beta^{-1}\,\text{Entropy}. \end{aligned} \tag{9}$$

Friston's Free Energy Principle for Cognition

The free energy expression: Substituting the last two distributions into Equation (17) we have the following form:

$$\mathscr{F}(Q(\tilde{s},\tilde{u})) \;=\; E_Q\left[-\log\left(P(\tilde{o},\tilde{s},\tilde{u}\,|\,m)\right)\right] - H\left(Q(\tilde{s},\tilde{u})\right), \tag{25}$$

$$\Rightarrow Q(\tilde{s},\tilde{u}) \;=\; \arg\min_{Q}\,\mathscr{F}\left(Q(\tilde{s},\tilde{u})\right). \tag{26}$$

Information Theory for Agents in Artificial Intelligence, Psychology, and Economics

by Michael S. Harré

Complex Systems Research Group, Faculty of Engineering, The University of Sydney, Sydney 2006, Australia

Entropy **2021**, *23*(3), 310; https://doi.org/10.3390/e23030310

Friston's Free Energy Principle for Cognition

Friston's Free Energy Principle:

One of the goals of Friston's work is to estimate the joint probability of states observed o and actual states s via Bayes Theorem:

$$P(s, o) = P(o|s)P(s)$$

Friston's Free Energy Principle for Cognition

Friston's Free Energy Principle:

One of the goals of Friston's work is to estimate the joint probability of states observed o and actual states s via Bayes Theorem:

$$P(s, o) = P(o|s)P(s)$$

This calculation is often too difficult to compute directly so Friston's "Free Energy Principle" for the brain addresses this by estimating an alternative probability $Q(s)$ via opitmisation:

$$Q^*(s) = \underset{Q(s)}{\operatorname{argmin}} \mathcal{F}(Q) \tag{3}$$

$$Q^*(s) \simeq P(s|o) \tag{4}$$

Friston's Free Energy Principle for Cognition

Friston's Free Energy Principle:

One of the goals of Friston's work is to estimate the joint probability of states observed o and actual states s via Bayes Theorem:

$$P(s,o) = P(o|s)P(s)$$

This calculation is often too difficult to compute directly so Friston's "Free Energy Principle" for the brain addresses this by estimating an alternative probability $Q(s)$ via opitmisation:

$$Q^*(s) = \underset{Q(s)}{\operatorname{argmin}}\, \mathcal{F}(Q) \tag{3}$$

$$Q^*(s) \simeq P(s|o) \tag{4}$$

$$\mathcal{F}(Q) = E_Q[\log(Q(s)) - \log(P(s|o))] \tag{5}$$

$$= \underbrace{E_Q(-\log(P(s|o)))}_{\substack{\text{cross entropy} \\ = \text{expected log loss}}} - \underbrace{H(Q(s))}_{\text{entropy}} \tag{6}$$

Friston's Free Energy Principle for Cognition

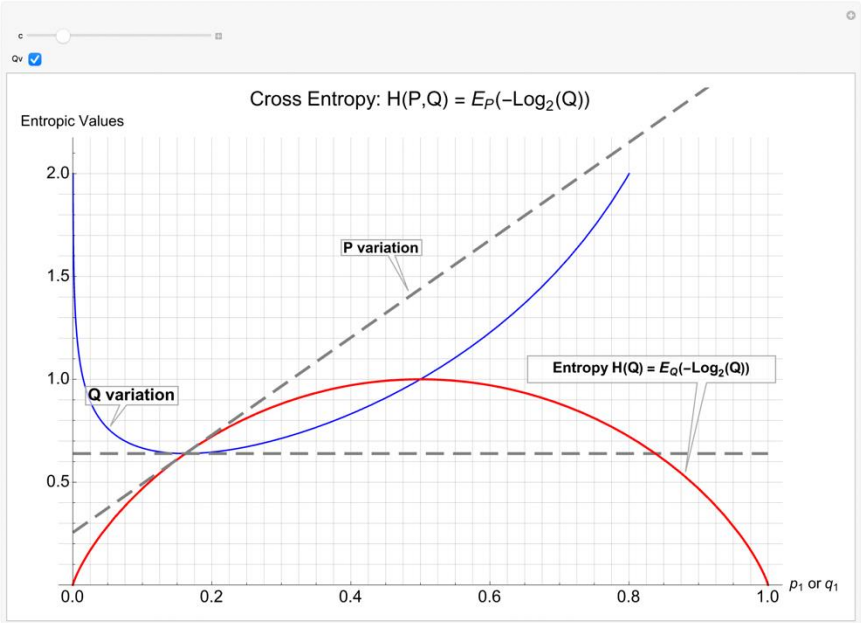
Published: 13 January 2010

The free-energy principle: a unified brain theory?

Karl Friston

Nature Reviews Neuroscience **11**, 127–138 (2010) | Cite this article

Friston's Free Energy Principle:



$$\mathcal{F}(Q) = E_Q[\log(Q(s)) - \log(P(s|o))] \tag{5}$$

$$= \underbrace{E_Q(-\log(P(s|o)))}_{\substack{\text{cross entropy} \\ = \text{expected log loss}}} - \underbrace{H(Q(s))}_{\text{entropy}} \tag{6}$$

Friston's Free Energy Principle for Cognition

Published: 13 January 2010

The free-energy principle: a unified brain theory?

Karl Friston

Nature Reviews Neuroscience **11**, 127–138 (2010) | Cite this article

There is a lot more work by Friston and colleagues

ELSEVIER

Journal of Physiology-Paris

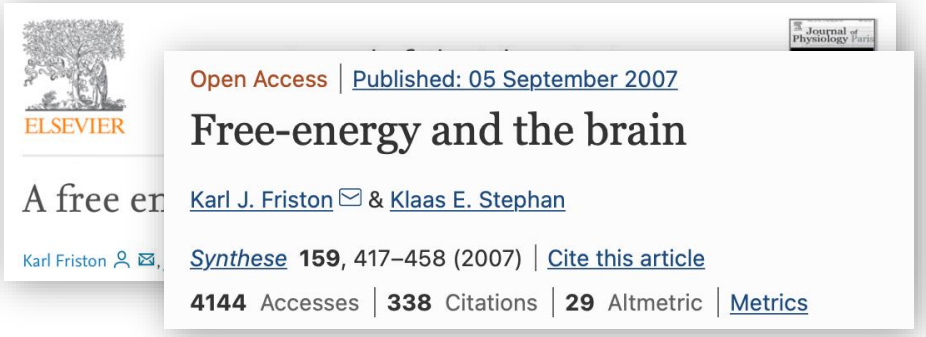
Volume 100, Issues 1–3, July–September 2006, Pages 70-87

A free energy principle for the brain

Karl Friston, James Kilner, Lee Harrison

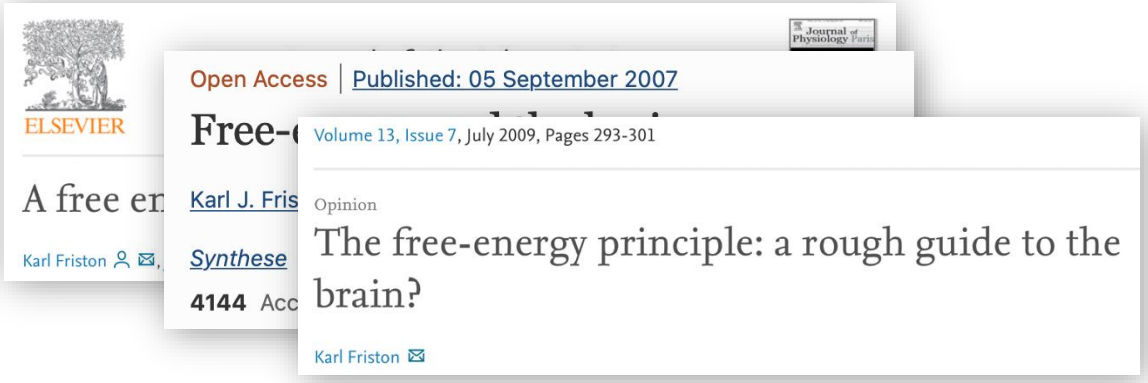
Friston's Free Energy Principle for Cognition

There is a lot more work by Friston and colleagues



Friston's Free Energy Principle for Cognition

There is a lot more work by Friston and colleagues

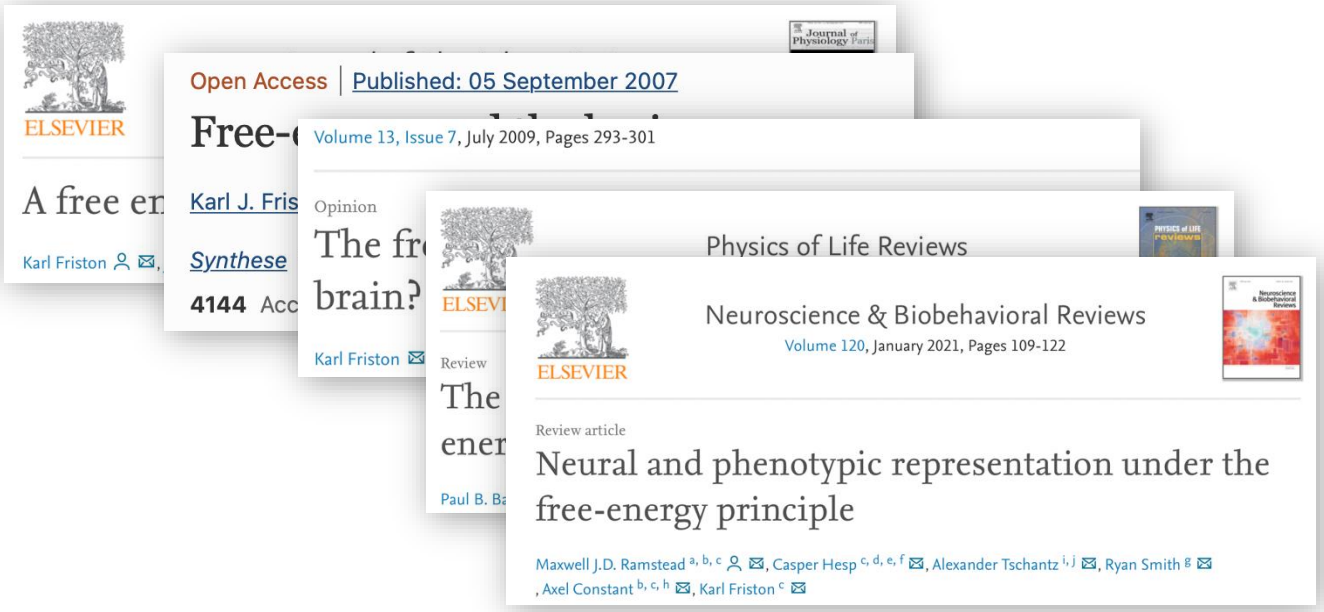


Friston's Free Energy Principle for Cognition

There is a lot more work by Friston and colleagues

Friston's Free Energy Principle for Cognition

There is a lot more work by Friston and colleagues



Friston's Free Energy Principle for Cognition

Published: 13 January 2010

The free-energy principle: a unified brain theory?

Karl Friston

Nature Reviews Neuroscience **11**, 127–138 (2010) | Cite this article

But these are the three ideas to take away from this section:

$$\mathcal{F}(Q) = E_Q[\log(Q(s)) - \log(P(s|o))]$$

$$= \underbrace{E_Q(-\log(P(s|o)))}_{\text{cross entropy}} - \underbrace{H(Q(s))}_{\text{entropy}}$$

= expected log loss

$$H(P) = -\sum_{x,y} P(x,y)\log(P(x,y))$$

$$U_a(P) = E_p[U_a(x,y)] \text{ (expected utility for } a)$$

$$\mathcal{F}(P) = U_a(P) - T\,H(P) \quad (T \text{ is uncertainty or error})$$

Optimising these "free functionals" leads to standard exponential solution

note: Grunwald and Dawid showed that this is a "game" between nature and decision maker (2004)

Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory

PD Grünwald, AP **Dawid** - the Annals of Statistics, 2004 - projecteuclid.org

Artificial Intelligence, Equilibrium, and Decisions

We have a Markov Decision Process: $\mathcal{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{T}, r\}$

$\rightarrow s \in \mathcal{S}$ is a state in the state space,

$\rightarrow a \in \mathcal{A}$ is an action in the action space of the agent

$\rightarrow \mathcal{T}$ is the transition model from one state to another

$\rightarrow r(s_t)$ is the reward, a function of the state s at time t

$\rightarrow \pi : \mathcal{S} \times \mathcal{A} \in [0, 1]$ is an agent's "policy", a state-action tuple mapping s_t and a_t to a probability at time t


Inverse Reinforcement Learning as the Algorithmic Basis for Theory of Mind: Current Methods and Open Problems

by Jaime Ruiz-Serra and Michael S. Harré *

Modelling and Simulation Research Group, School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, NSW 2006, Australia

* Author to whom correspondence should be addressed.

Algorithms **2023**, *16*(2), 68; https://doi.org/10.3390/a16020068

Artificial Intelligence, Equilibrium, and Decisions

We have a Markov Decision Process: $\mathcal{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{T}, r\}$

$\rightarrow s \in \mathcal{S}$ is a state in the state space,

$\rightarrow a \in \mathcal{A}$ is an action in the action space of the agent

$\rightarrow \mathcal{T}$ is the transition model from one state to another

$\rightarrow r(s_t)$ is the reward, a function of the state s at time t

$\rightarrow \pi : \mathcal{S} \times \mathcal{A} \in [0, 1]$ is an agent's "policy", a state-action tuple mapping s_t and a_t to a probability at time t

The transition model is a probability distribution:

$\mathcal{T}: P_a(s'|s) = P(s' = s_t \mid s = s_{t-1}, a_t)$

Inverse Reinforcement Learning as the Algorithmic Basis for Theory of Mind: Current Methods and Open Problems

by Jaime Ruiz-Serra and Michael S. Harré *

Modelling and Simulation Research Group, School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, NSW 2006, Australia

* Author to whom correspondence should be addressed.

Algorithms **2023**, *16*(2), 68; https://doi.org/10.3390/a16020068

Artificial Intelligence, Equilibrium, and Decisions

We have a Markov Decision Process: $\mathcal{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{T}, r\}$

$\rightarrow s \in \mathcal{S}$ is a state in the state space,

$\rightarrow a \in \mathcal{A}$ is an action in the action space of the agent

$\rightarrow \mathcal{T}$ is the transition model from one state to another

$\rightarrow r(s_t)$ is the reward, a function of the state s at time t

$\rightarrow \pi: \mathcal{S} \times \mathcal{A} \in [0, 1]$ is an agent's "policy", a state-action tuple mapping s_t and a_t to a probability at time t

The transition model is a probability distribution:

$\mathcal{T}: P_a(s'|s) = P(s' = s_t \mid s = s_{t-1}, a_t)$

The agent's policy is a probability function:

$\pi = P(a = a_t \mid s = s_t)$

Inverse Reinforcement Learning as the Algorithmic Basis for Theory of Mind: Current Methods and Open Problems

by Jaime Ruiz-Serra and Michael S. Harré *

Modelling and Simulation Research Group, School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, NSW 2006, Australia

* Author to whom correspondence should be addressed.

Algorithms **2023**, *16*(2), 68; https://doi.org/10.3390/a16020068

Artificial Intelligence, Equilibrium, and Decisions

We have a Markov Decision Process: $\mathcal{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{T}, r\}$
$\rightarrow s \in \mathcal{S}$ is a state in the state space,
$\rightarrow a \in \mathcal{A}$ is an action in the action space of the agent
$\rightarrow \mathcal{T}$ is the transition model from one state to another
$\rightarrow r(s_t)$ is the reward, a function of the state s at time t
$\rightarrow \pi: \mathcal{S} \times \mathcal{A} \in [0, 1]$ is an agent's "policy", a state-action tuple mapping s_t and a_t to a probability at time t

The transition model is a probability distribution:
$\mathcal{T}: P_a(s'|s) = P(s' = s_t \mid s = s_{t-1}, a_t)$

The agent's policy is a probability function:
$\pi = P(a = a_t \mid s = s_t)$

An optimal policy for $\pi: \mathcal{S} \mapsto \mathcal{A}$ is that π^* which maximises the expected accumulation of reward = long term discounted value, i.e. the following expected value:
$V(\mathbf{s}) = E(r(s_1) + \gamma r(s_2) + \gamma^2 r(s_3) + \ldots \mid \pi)$ where $\gamma \in [0, 1]$ = discount (Bellman's Equation for policy π)


Inverse Reinforcement Learning as the Algorithmic Basis for Theory of Mind: Current Methods and Open Problems

by Jaime Ruiz-Serra and Michael S. Harré *

Modelling and Simulation Research Group, School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, NSW 2006, Australia

* Author to whom correspondence should be addressed.

Algorithms **2023**, *16*(2), 68; https://doi.org/10.3390/a16020068

Artificial Intelligence, Equilibrium, and Decisions

The reward function (here, cost to be minimised) R comprises a state term $r(s) \geq 0$ (to be inferred) and a control term that is the KL-divergence between the control dynamics and the passive dynamics (in order for the KL divergence to be defined, it is required that $\pi(s'|s) = 0$ when $\Pr(s'|s) = 0$, a condition that is imposed),

$$R(s, \pi(\cdot|s)) = r(s) + D_{KL}(\pi(\cdot|s)||\Pr(\cdot|s)). \tag{41}$$

A desirability function $z(s) = \exp(-V(s))$ is used to define the optimal control dynamics

$$\pi^*(s'|s) = \frac{\Pr(s'|s)z(s')}{\sum_\zeta \Pr(\zeta|s)z(\zeta)}. \tag{42}$$

Inverse Reinforcement Learning as the Algorithmic Basis for Theory of Mind: Current Methods and Open Problems

by Jaime Ruiz-Serra and Michael S. Harré *

Modelling and Simulation Research Group, School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, NSW 2006, Australia

* Author to whom correspondence should be addressed.

Algorithms **2023**, *16*(2), 68; https://doi.org/10.3390/a16020068

Artificial Intelligence, Equilibrium, and Decisions

The reward function (here, cost to be minimised) R comprises a state term $r(s) \geq 0$ (to be inferred) and a control term that is the KL-divergence between the control dynamics and the passive dynamics (in order for the KL divergence to be defined, it is required that $\pi(s'|s) = 0$ when $\Pr(s'|s) = 0$, a condition that is imposed),

$$R(s, \pi(\cdot|s)) = r(s) + D_{KL}(\pi(\cdot|s)||\Pr(\cdot|s)). \tag{41}$$

A desirability function $z(s) = \exp(-V(s))$ is used to define the optimal control dynamics

$$\pi^*(s'|s) = \frac{\Pr(s'|s)z(s')}{\sum_\zeta \Pr(\zeta|s)z(\zeta)}. \tag{42}$$

Under passive dynamics, $\Pr(\tau|s_0) = \prod_{t=1}^{H} \Pr(s_t|s_{t-1})$ is the probability of a trajectory. For the same trajectory to occur when the control dynamics are applied, the probability is

$$\Pr(\tau|s_0, \pi) = \frac{\Pr(\tau|s_0) \exp\left(-\sum_{t=0}^{H} r(s_t)\right)}{z(s_0)}. \tag{44}$$

Inverse Reinforcement Learning as the Algorithmic Basis for Theory of Mind: Current Methods and Open Problems

by Jaime Ruiz-Serra and Michael S. Harré *

Modelling and Simulation Research Group, School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, NSW 2006, Australia

* Author to whom correspondence should be addressed.

Algorithms **2023**, *16*(2), 68; https://doi.org/10.3390/a16020068

Artificial Intelligence, Equilibrium, and Decisions

Now compare the reinforcement learning algorithm:

$$R(s, \pi(\cdot|s)) = r(s) + D_{KL}(\pi(\cdot|s)||\Pr(\cdot|s)).$$

Artificial Intelligence, Equilibrium, and Decisions

Now compare the reinforcement learning algorithm's reward:

$$R(s, \pi(\cdot|s)) = r(s) + D_{KL}(\pi(\cdot|s)||\Pr(\cdot|s)).$$

With the following decomposition of the KL-divergence:

$$U(\pi_i, \pi_{-i}) = u(\pi_i, \pi_{-i}) + \lambda D_{KL}(\pi_i \mid \tau_i)$$

Artificial Intelligence, Equilibrium, and Decisions

Now compare the reinforcement learning algorithm:

$$R(s, \pi(\cdot|s)) = r(s) + D_{KL}(\pi(\cdot|s)||\Pr(\cdot|s)).$$

With the following decomposition of the KL-divergence:

$$U(\pi_i, \pi_{-i}) = u(\pi_i, \pi_{-i}) + \lambda D_{KL}(\pi_i \mid \tau_i)$$

$$= u(\pi_i, \pi_{-i}) + \lambda \underbrace{(\underbrace{-H(\pi_i)}_{\text{Entropy}} + \underbrace{H(\pi_i, \tau_i)}_{\text{cross entropy}})}_{\text{Log–loss game}}$$

Artificial Intelligence, Equilibrium, and Decisions

Now compare the reinforcement learning algorithm:

$$R(s, \pi(\cdot|s)) = r(s) + D_{KL}(\pi(\cdot|s)||\Pr(\cdot|s)).$$

With the following decomposition of the KL-divergence:

$$U(\pi_i, \pi_{-i}) = u(\pi_i, \pi_{-i}) + \lambda D_{KL}(\pi_i \mid \tau_i)$$

$$= u(\pi_i, \pi_{-i}) + \lambda \underbrace{(-\underbrace{H(\pi_i)}_{\text{Entropy}} + \underbrace{H(\pi_i, \tau_i)}_{\text{cross entropy}})}_{\text{Log–loss game}}$$

$$= \overbrace{u(\pi_i, \pi_{-i}) + \lambda(\underbrace{-H(\pi_i)}_{\text{Entropy}} + \underbrace{H(\pi_i, \tau_i)}_{\text{cross entropy}})}^{\text{(local) MaxEnt Game Theory}}$$

Artificial Intelligence, Equilibrium, and Decisions

Which has the following "equilibrium" solution by MaxEnt optimization:

$$p(a_i \mid \pi_{-i}) = p(a_i) \exp\left(\lambda^{-1} u(a_i \mid \pi_{-i})\right) Z^{-1}$$

(Z^{-1} is a normalizing factor)

$$U(\pi_i, \pi_{-i}) = u(\pi_i, \pi_{-i}) + \lambda D_{KL}(\pi_i \mid \tau_i)$$

$$= u(\pi_i, \pi_{-i}) + \lambda \underbrace{(-\underbrace{H(\pi_i)}_{\text{Entropy}} + \underbrace{H(\pi_i, \tau_i)}_{\text{cross entropy}})}_{\text{Log-loss game}}$$

$$= \overbrace{u(\pi_i, \pi_{-i}) + \lambda(\underbrace{-H(\pi_i)}_{\text{Entropy}} + \underbrace{H(\pi_i, \tau_i)}_{\text{cross entropy}})}^{\text{(local) MaxEnt Game Theory}}$$

Artificial Intelligence, Equilibrium, and Decisions

The essential point of this section are these relationships, the KL-Divergence is an extension of MaxEnt and Free Energy principles:

$$\mathcal{F}(Q) = E_Q[\log(Q(s)) - \log(P(s|o))]$$

$$= \underbrace{E_Q(-\log(P(s|o)))}_{\text{cross entropy}} - \underbrace{H(Q(s))}_{\text{entropy}}$$

$$= \text{cross entropy} = \text{expected log loss}$$

$$H(P) = -\sum_{x,y} P(x,y)\log(P(x,y))$$

$$U_a(P) = E_p[U_a(x,y)] \text{ (expected utility for } a)$$

$$\mathcal{F}(P) = U_a(P) - T\,H(P) \quad (T \text{ is uncertainty or error})$$

$$U(\pi_i, \pi_{-i}) = u(\pi_i, \pi_{-i}) + \lambda D_{KL}(\pi_i \mid \tau_i)$$

$$= u(\pi_i, \pi_{-i}) + \lambda(\underbrace{-H(\pi_i)}_{\text{Entropy}} + \underbrace{H(\pi_i, \tau_i)}_{\text{cross entropy}})$$

$$\text{Log-loss game}$$

$$= \overbrace{u(\pi_i, \pi_{-i}) + \lambda(\underbrace{-H(\pi_i)}_{\text{Entropy}} + \underbrace{H(\pi_i, \tau_i)}_{\text{cross entropy}})}^{\text{(local) MaxEnt Game Theory}}$$

Grunwald and Dawid showed that this is a "game" between nature and decision maker (2004)

Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory
PD Grünwald, AP **Dawid** - the Annals of Statistics, 2004 - projecteuclid.org

This is the game theory version of MaxEnt or equivalently the "Free Utility" optimization

This is a combination of the above equations

Future "Grand Directions"

Review

Inverse Reinforcement Learning as an Algorithmic Approach to Theory of Mind: Current Methods and Open Problems

Jaime Ruiz-Serra [1] and Michael Harré [1,*]

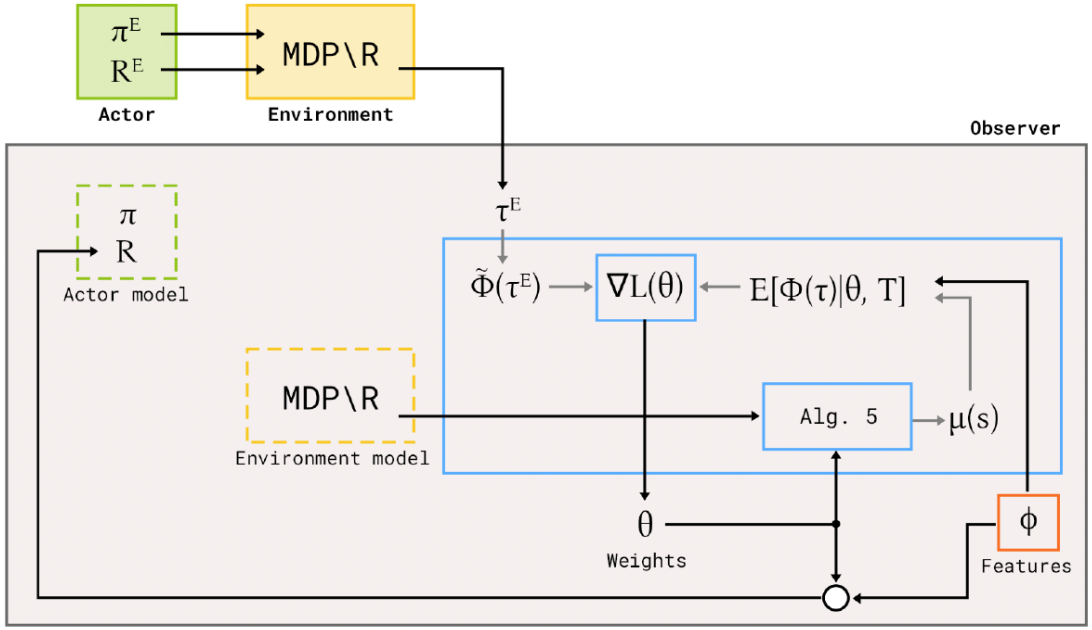


Figure 3. Diagram of the MaxEnt IRL algorithm (see Algorithm 6)

Future "Grand Directions"

Testing 'Theory of Mind' models for AI

Michael S. Harré
School of Computer Science
The University of Sydney
Sydney, Australia, 2006
michael.harre@sydney.edu.au

Husam El-Tarifi
Oxford Economics Australia
Sydney, Australia, 2000
heltarifi@oxfordeconomics.com

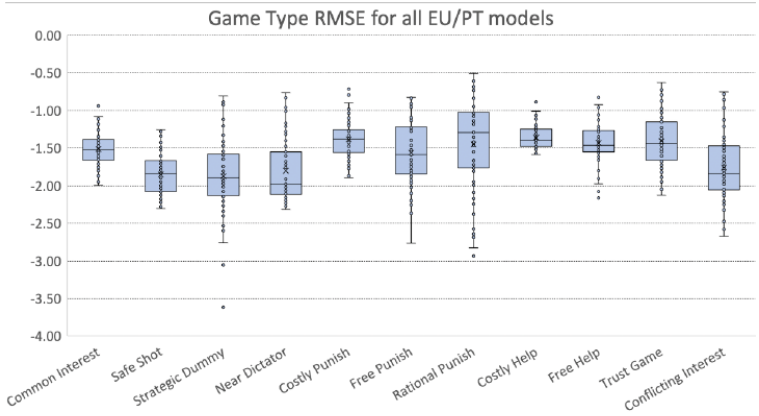


Figure 5: Log_{10} transformed RMSE performance distribution of all EU-PT models dis-aggregated by game type. Logs highlight the low value tails for 'best in class' data points with low RMSE values.

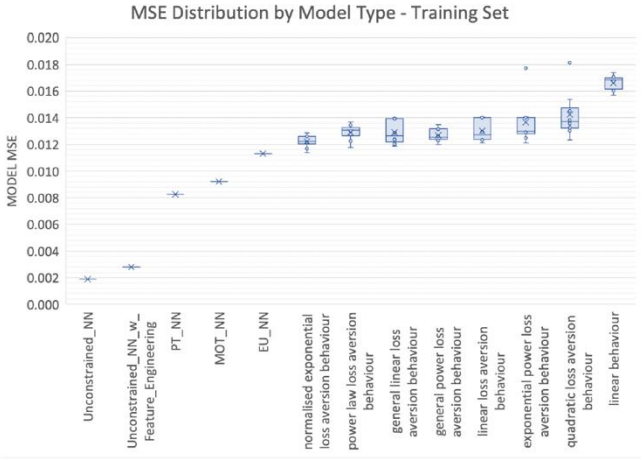


Figure 2: Root Mean Square Error results for each utility and neural network model using the training data.

Future "Grand Directions"

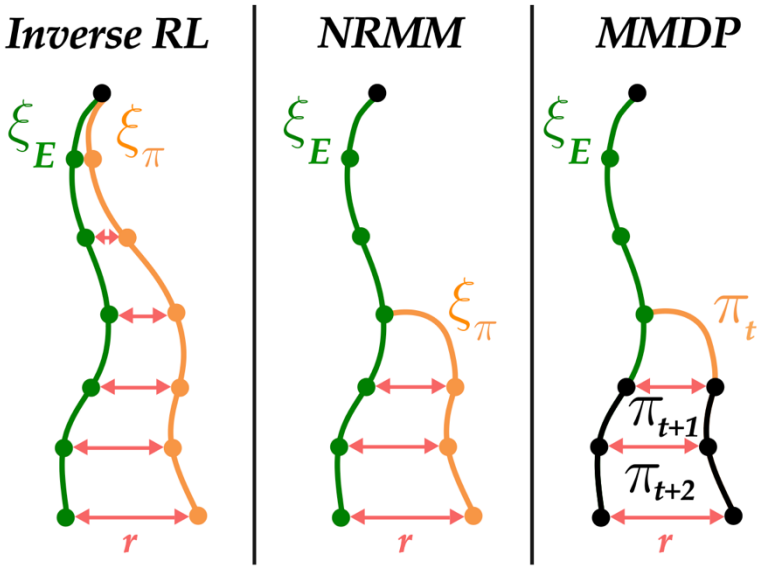
Inverse Reinforcement Learning without Reinforcement Learning

ICML '23

Gokul Swamy[1] David Wu[2] Sanjiban Choudhury[2] Drew Bagnell[3, 1] Steven Wu[1]

[1] CMU [2] Cornell University [3] Aurora Innovation

[Paper] [Video] [Code]



Future "Grand Directions"

nature
COMMENT | 04 May 2021

www.cooperativeai.org

Cooperative AI: machines must learn to find common ground

To help humanity solve fundamental problems of cooperation, scientists need to reconceive artificial intelligence as deeply social.

Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson & Thore Graepel

A huddle at the 2017 United Nations Climate Change Conference, where attendees cooperated on mutually beneficial joint actions on climate. Credit: Sean Gallup/Getty