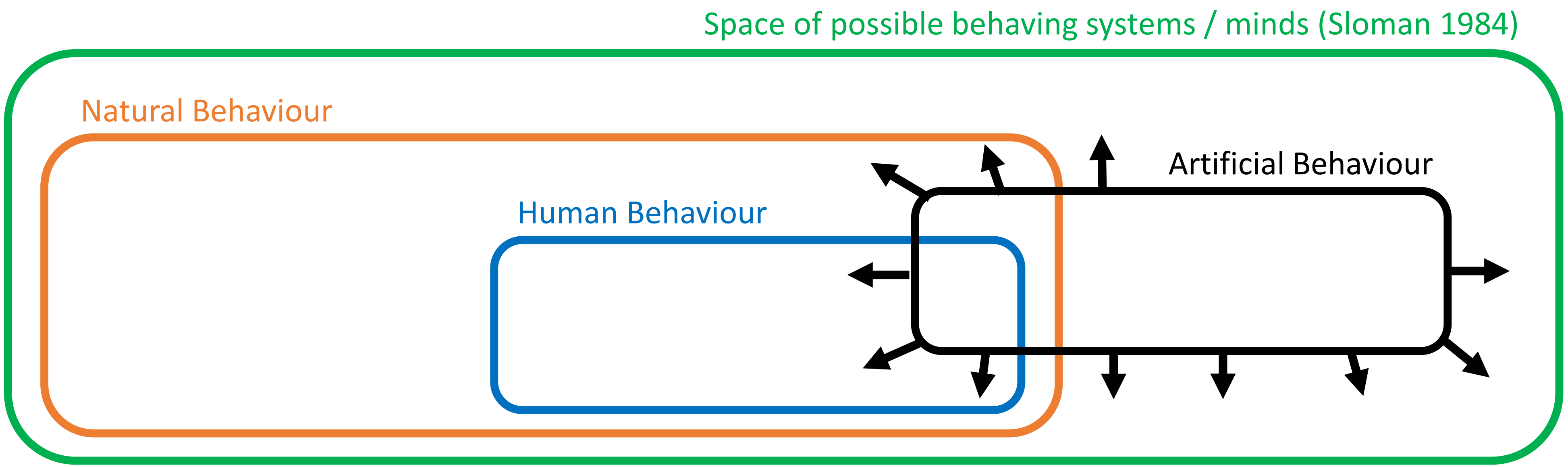
Diversity Unites Intelligence: Measuring Generality

José Hernández-Orallo (jorallo@dsic.upv.es)
Universitat Politècnica de València, Valencia (www.upv.es)
Also visiting the Leverhulme Centre for the Future of Intelligence, Cambridge (lcfi.ac.uk)

Varieties of Minds, Cambridge, UK, 5 June – 8 June 2018

The Space of All Minds

- Copernican Revolution:
  - Cognitive science placed nature in a wider landscape:



- Different interpretations:
  - Replace Behaviour by Learning / Cognition / Intelligence / Minds.

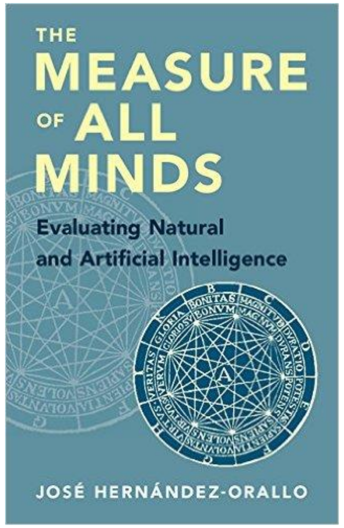
The Space of All Minds

- Custom still places humans or evolution at the centre of the landscape:
  - **Biology**: behaviour must be explained in terms of evolution. *But are the patterns and the explanations valid beyond life?*
  - **Artificial intelligence**: anthropocentric goals and references (human-level AI, Turing test, superintelligence, human automation, etc.). *Isn't this myopic?*

How can we characterise this space in a universal way, beyond anthropocentric or evolutionary constraints?

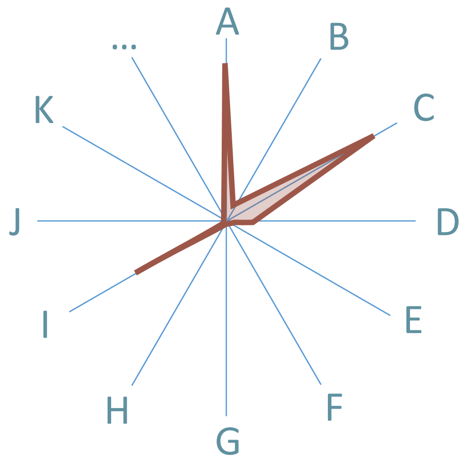
- A measurement approach:

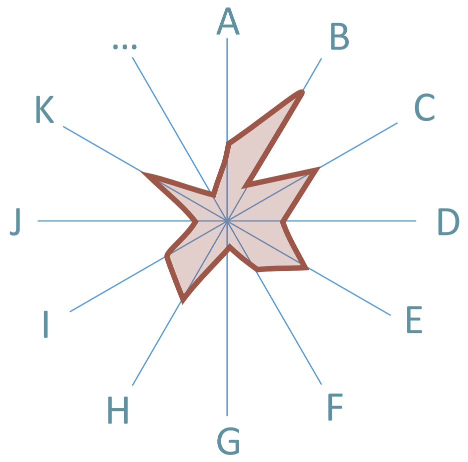
  "The Measure of All Minds: Evaluating Natural and Artificial Intelligence", *Cambridge University Press*, 2017.
  http://www.allminds.org

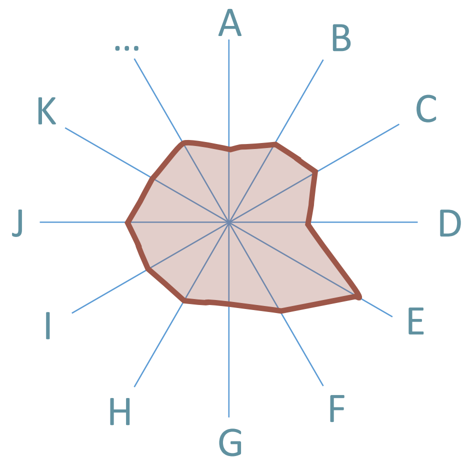


The Space of All Minds

- Infinitely many environments, infinitely many tasks: A, B, C, ….







Intelligence is a subjective phenomenon. No-free-lunch theorems, multiple intelligences, narrow AI

SPECIFIC

GENERAL

Intelligence is a convergent phenomenon. The positive manifold, g/G factors, Solomonoff prediction, AGI

Artificial systems: by conception, we can design a system to be good at A, C and I, and very bad at all the rest.

Non-human animals: environments, morphology, physiology and (co-)evolution creates some structure here.

Humans: strong correlation between cognitive tasks and abilities: general intelligence.

The Space of All Tasks

- All cognitive tasks or environments M.
  - Dual space to all possible behaving systems.
  - M only makes sense with a probability measure *p* over all tasks $\mu \in$ M.
  - An animal or agent π is selected or designed for optimal cognition in this ‹*M,p*›.

$$\Psi(\pi, M, p) \triangleq \sum_{\mu \in M} p(\mu) \cdot R(\pi, \mu)$$

- If M is infinite and diverse *policies* are acquired or learnt, not hardwired.
- But who sets ‹*M,p*›?
  - In biology, natural selection (physical world, co-evolution, social environments).
  - In AI, applications (narrow or more robust/adaptable to changes).

So is general intelligence a subjective phenomenon to a choice of ‹M,p›?

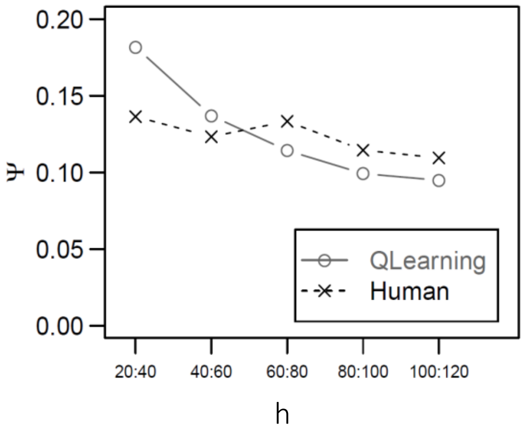
The Space of All Tasks

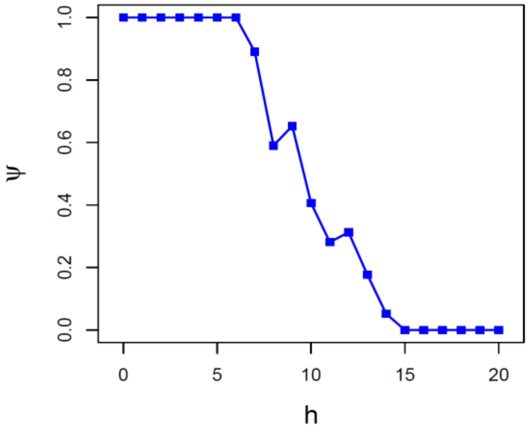
- In a RL setting choosing a universal distribution $p(\mu)=2^{-K_U(\mu)}$ we get the so-called "Universal Intelligence" measure (Legg and Hutter 2007).
  - Proper formalisation of including all tasks, "generalising the C-test (Hernandez-Orallo 2000) from passive to active environments".
  - Problems (pointed out by many: Hibbard 2009, Hernandez-Orallo & Dowe 2010):
    - The probability distribution on M is not computable.
    - Time/speed is not considered for the environment or agent.
    - Most environments are not really discriminating (hells/heavens).
    - **The mass of the probability measure goes to just a few environments.**

Legg and Hutter's measure is "**relative**" (Leike & Hutter 2015), a schema for tasks, a meta-definition instantiated by a particular choice of the reference U.

The Space of All Policies

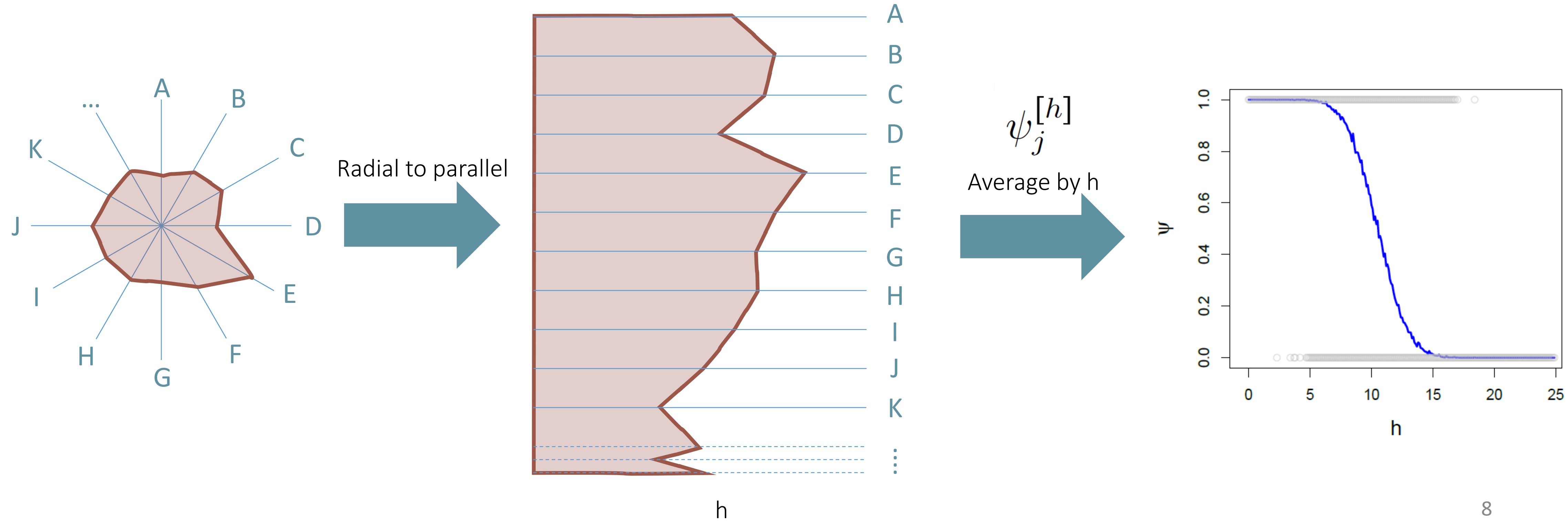
- Instead of the (Kolmogorov) complexity of the description of a task:
  - We look at the policy, the solution, and its complexity.
  - The resources or computation it needs: *this is the **difficulty** of the task*.
  - Difficulty is fundamental in psychometrics (e.g., IRT) and dual to capability.
- Let's assume we have a metric of difficulty or hardness (h) for tasks.
  - "agent (person) characteristic curves" (ACCs), expected response Ψ against difficulty:





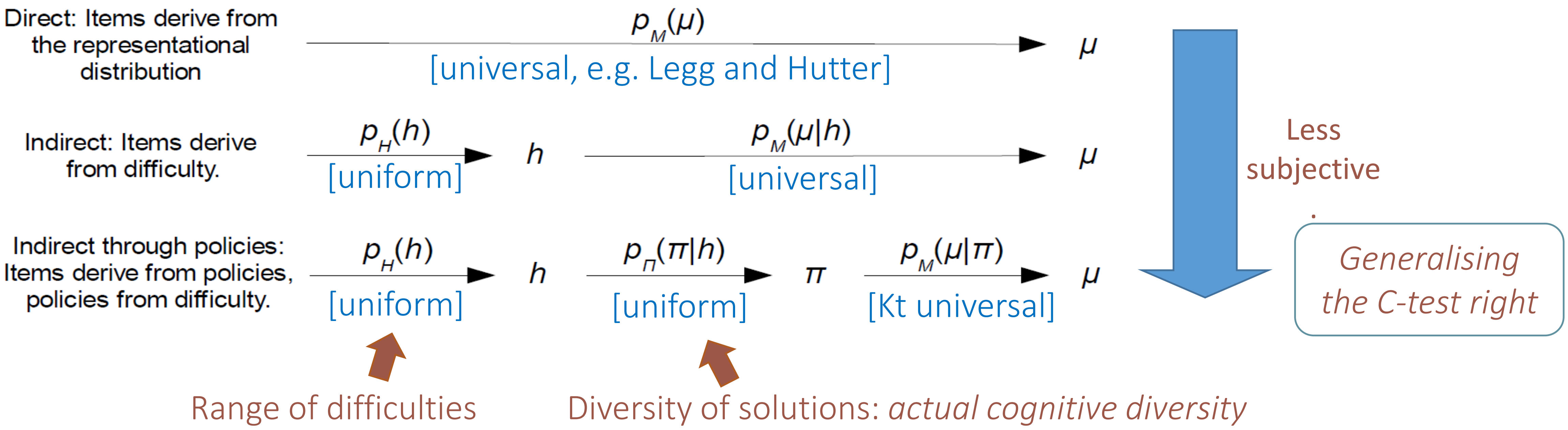
The Space of All Policies

- ACCs just aggregate the radial chart:
  - Each dimension A, B, C, … is ordered by policy difficulty:



The Space of All Policies

- Alternative formulations:



Less dependent on the representational mechanism for policies (invariance theorem).

How to Best Cover this Space to Maximise Ψ?

By evolution, by AI or by science.

A Measure of Generality

- A fundamental question for:
  - Human intelligence: positive manifold, g factor. General intelligence?
  - Non-human animal intelligence: g and G factors for many species. Convergence?
  - Artificial intelligence: general-purpose AI or AGI. What does the G in AGI mean?
- Usual interpretation:

General intelligence is usually associated with competence for a wide range of cognitive tasks

	μ_1	μ_2	μ_3	μ_4	μ_5
π_a	0.85	0.75	0.80	0.85	0.75
π_b	1.00	1.00	0.00	1.00	1.00

This is wrong! Any system with limited resources cannot show competence for a wide range of cognitive tasks, independently of their difficulty!

A Measure of Generality

> General intelligence must be seen as competence for a wide range of cognitive tasks up to a certain level of difficulty.

- Definition
  - Capability (Ψ), the area under the ACC: $\psi_j \triangleq \int_0^\infty \psi_j^{[h]}\, dh$
  - Expected difficulty given success:

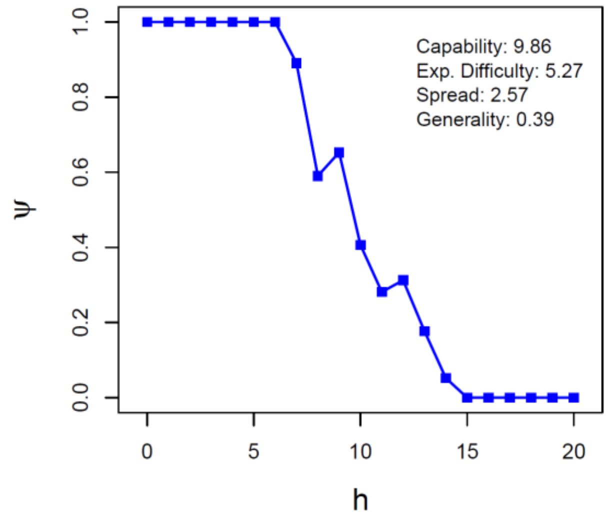
$$\mathbb{H}_j \triangleq \mathbb{E}_i[h | A_{i,j} = 1] = \frac{m_j}{\psi_j} \qquad m_j \triangleq \int_0^\infty h \cdot \psi_j^{[h]}\, dh$$

  - Spread:

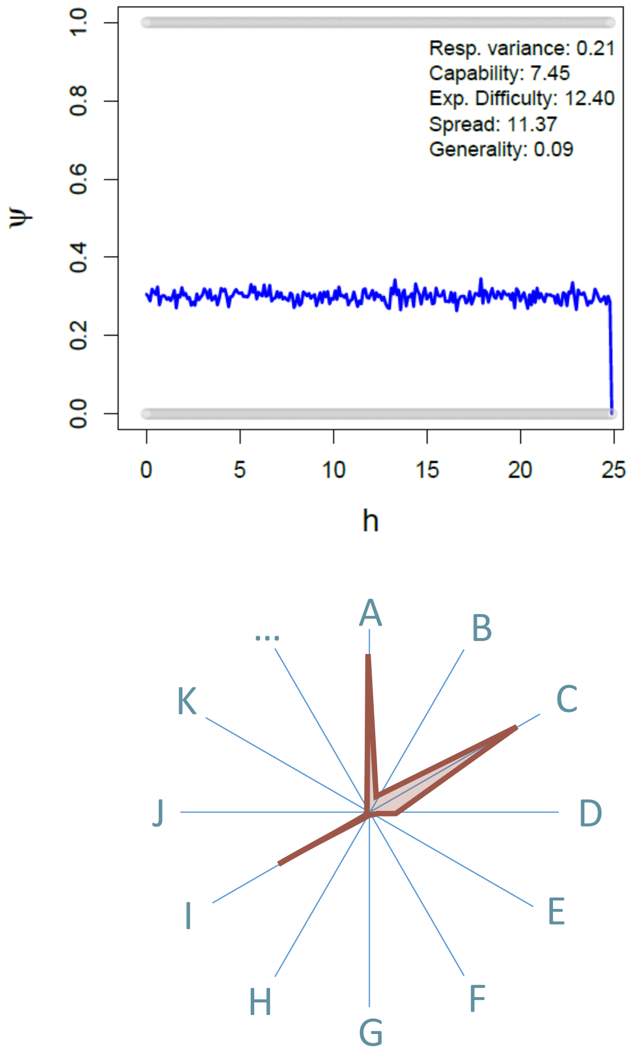
$$z_j \triangleq \sqrt{(2\mathbb{H}_j - \psi_j) \cdot \psi_j} = \sqrt{2m_j - \psi_j^2}$$

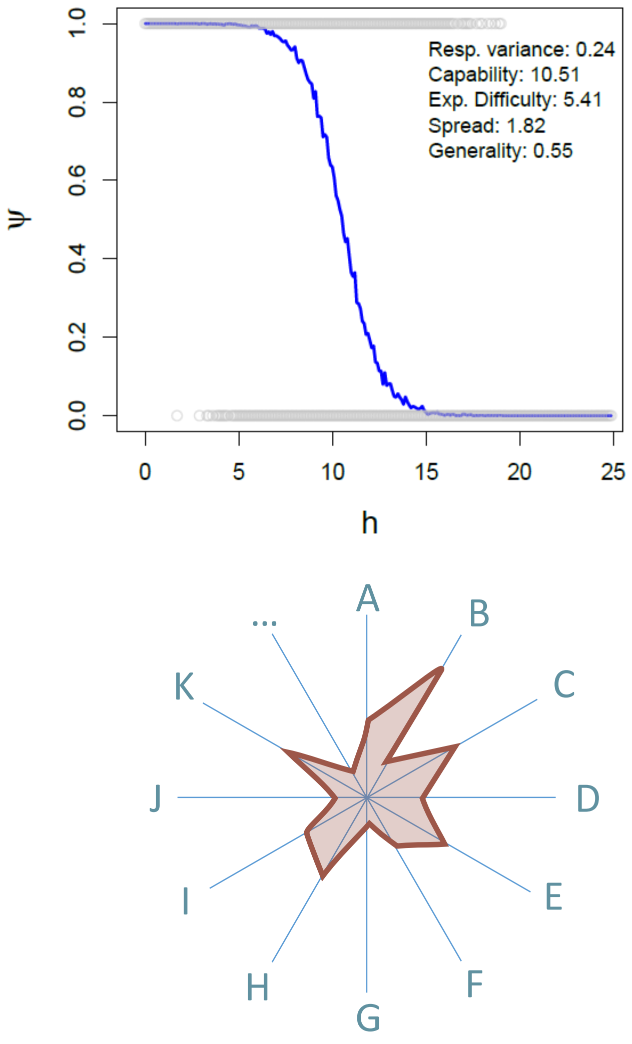
  - Generality:

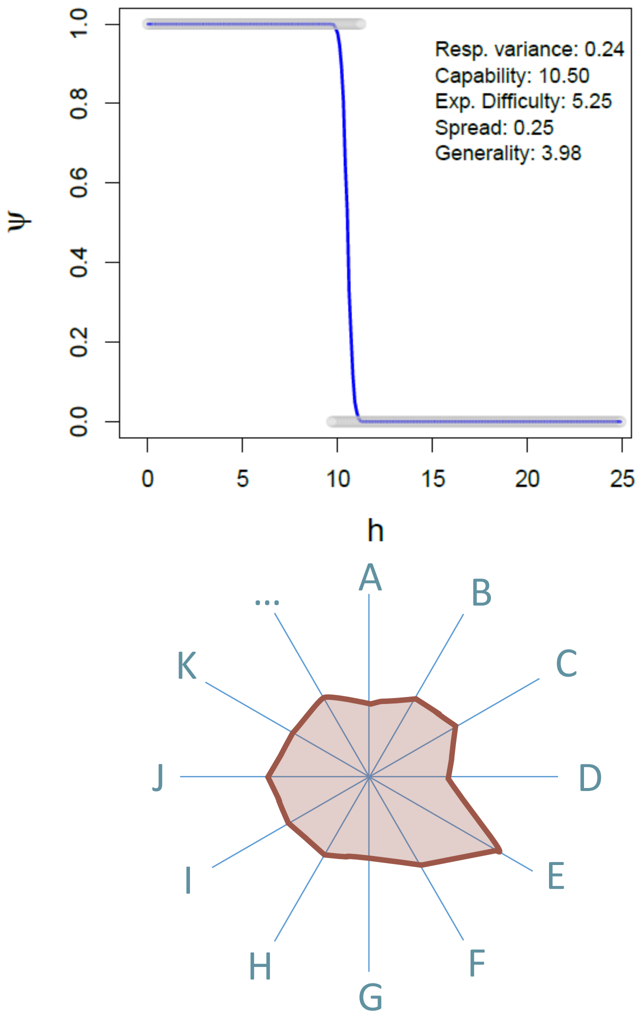
$$\gamma_j \triangleq \frac{1}{z_j} = \frac{1}{\sqrt{2m_j - \psi_j^2}}$$



A Measure of Generality

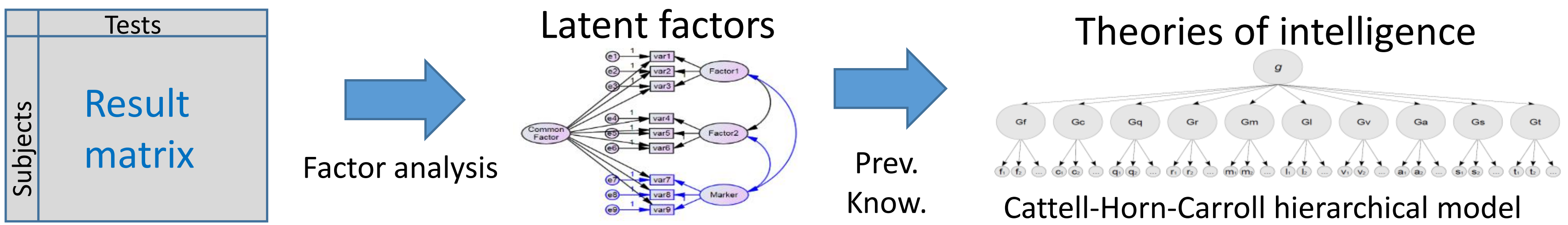






Generality: Humans

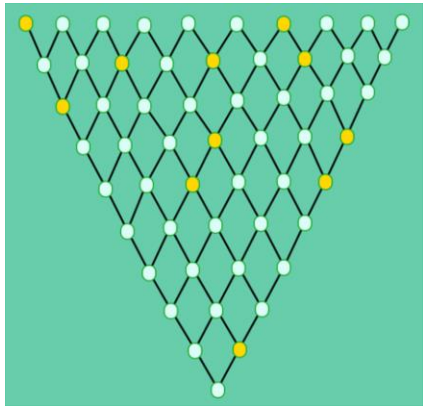
- Classical psychometric approach:
  - "General intelligence" usually conflates generality and performance.
  - Manifold and g factor are populational.



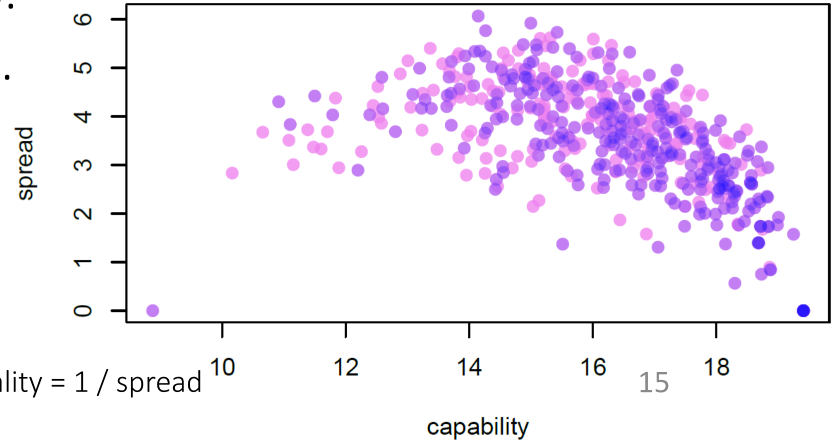
- Using the new measure of generality:
  - Capability and generality are observables, applied to individuals, no models.
  - We don't assume any grouping of items into tests with ranging difficulties.
  - Applicable to individual agents and small sets of tasks/items.

Generality: Humans

- Example (joint work with B.S. Loe, 2018):
  - Elithorn's Perceptual Mazes: 496 participants (Amazon Turk).
  - Intrinsic difficulty estimators (Buckingham et al. 1963, Davies & Davies 1965).
  - We calculate the generalities for the 496 humans.
    - Correlation between spread (1/gen) and capability is -0.53.
  - See relation to latent main (general) factor:
    - All data: one-factor loading: 0.46, prop. of variance: 0.23.
    - 1stQ of generality: 1-f loading: 0.65, prop. of variance: 0.46.



Against Spearman's Law of Diminishing Returns (SLODR).



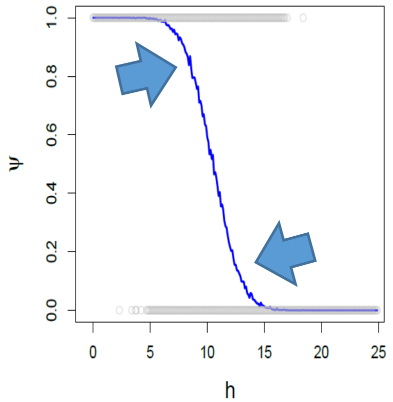
Generality = 1 / spread

Generality: Animals

- Why is general intelligence convergent? (Burkart et al. 2017)
  - Convergent g and G.
  - Domain-specific vs domain-general cognitive skills?

- Using the new measure of generality:
  - We see *h* as cognitive/evolutionary resources and efficiency as Ψ / h.
    - Generality in animals partly explained by efficiency.

Domain-general cognition has higher *Ψ / h* than domain-specific cognition.



    - Endogenous causes also play a role (e.g., "Bullmore and Sporns: "Economy of brain network organisation", NatRev Neuroscience 2012.

Generality: Animals

- Why g/G may be misleading?
  - g/G try to explain **variance** in results.
  - Species with high variance in capability have more to explain and usually high g.
  - Does not really compare the generality of individuals or species, but populations.
    - Woodley of Menie et al. "General intelligence is a source of individual differences between species: Solving an anomaly." Behavioral and Brain Sciences 40 (2017).

Generality is about diversity in tasks, not about diversity in populations!

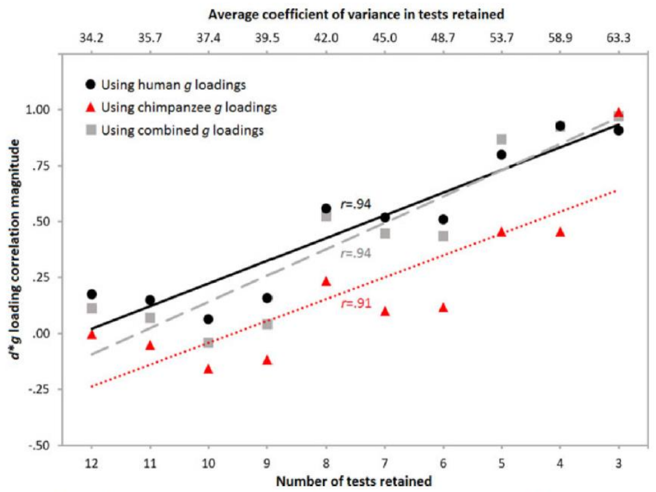
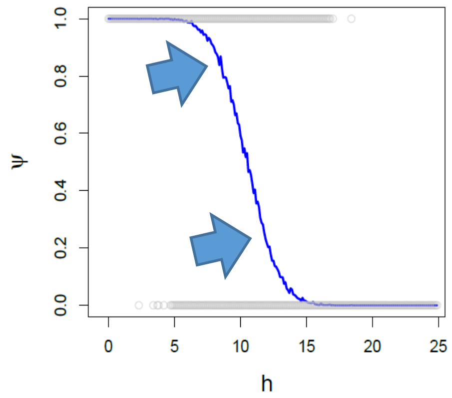


Figure 7: Correlations between task g loadings and the scores d on the y-axis as a function of the average coefficient of variance in the tests retained, choosing them by removing those with smallest variance first. Trends shown for chimpanzees, humans and a combined population. Copied from [129].

- Ongoing work (and looking for collaborators!):
  - Apply new generality (non-populational).

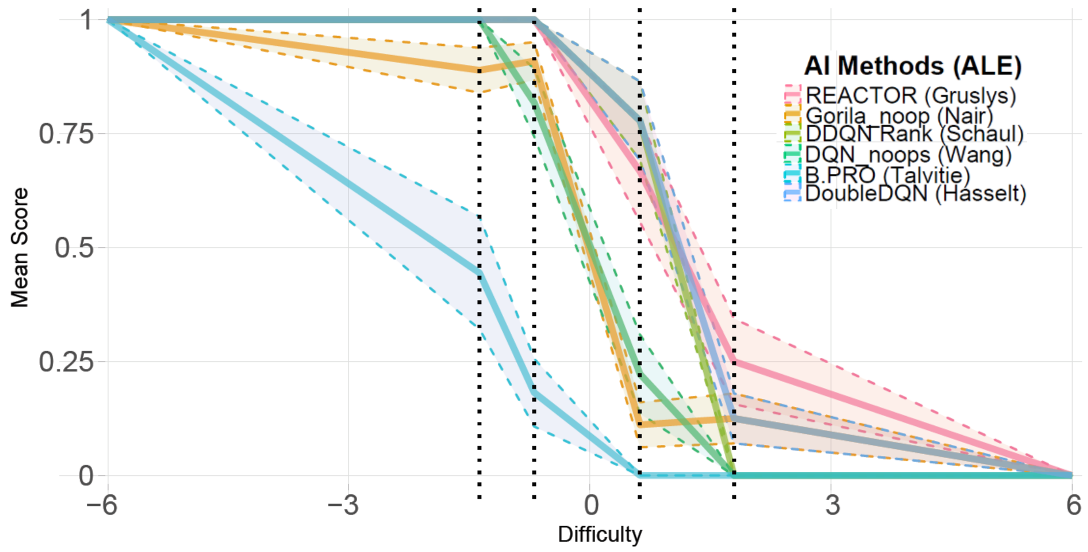
Generality: A(G)I

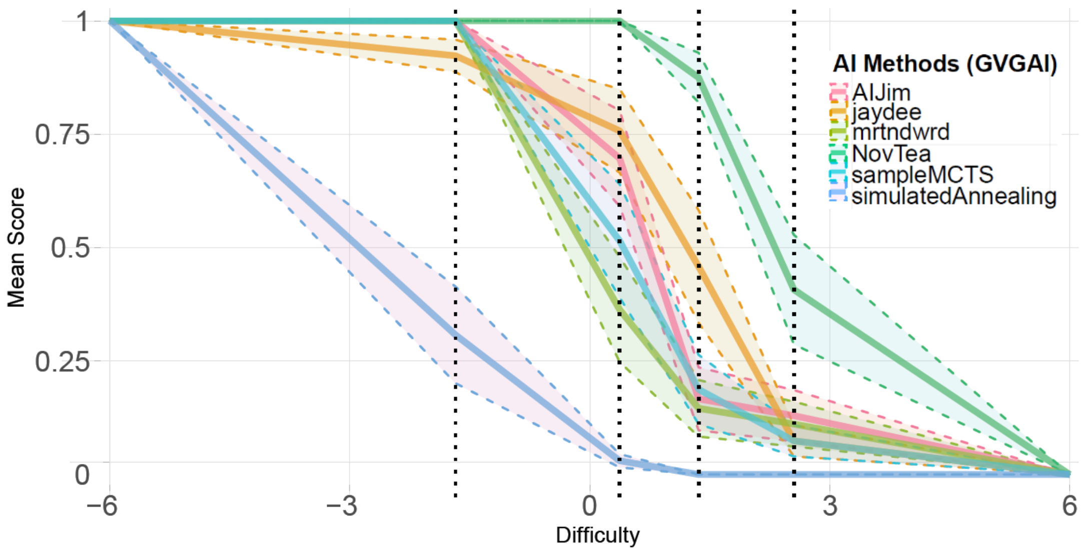
- How can the G in AGI be properly defined? No AI populations!
  - We want to calculate the generality of **one** AI system.
- Using the new measure of generality:
  - We could have very general systems, with low capability.
    - They could be AGI but far from humans: baby AGI, limited AGI.
  - All other things equal, it makes more sense to cover easy tasks first.
- Link to resources and compute.
  - Measuring capability and generality and their growth.
  - Look at superintelligence in this context.



Generality: A(G)I

- Example (joint work with F. Martinez-Plumed 2018)
  - ALE (Atari games) and GVGAI (General Video Game AI) benchmarks.
    - Progress has been made, but what about generality? Are systems more general?

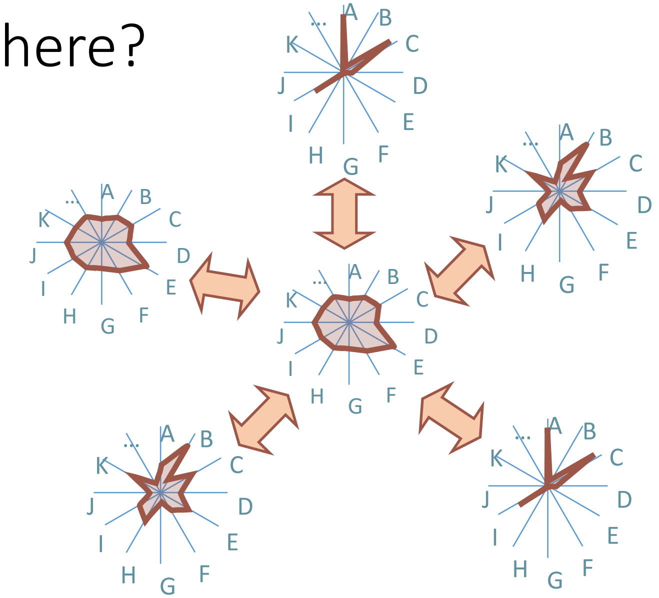




Generality and Diversity

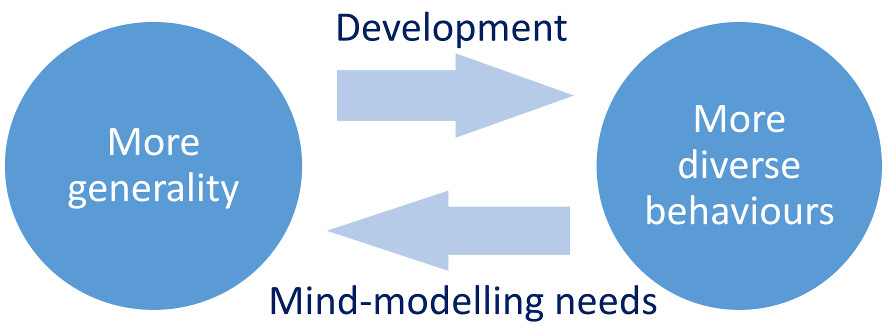
- What happens with generality when surrounded by other agents?
  - The distribution of tasks changes completely
    - Usually seen in terms of co-evolution (e.g., flowers and insects) or social groups.
    - Mind-modelling becomes necessary in competitive/cooperative scenarios.
    - Can we accommodate ‹*M,p*› theoretically in multi-agent contexts?
      - Darwin-Wallace distribution (purely cognitive evolution: same body for all agents).
  - What role does the split generality/capability play here?
    - More nuanced social hypothesis:

More complex social circumstances trigger an increase of capability **and/or** generality?



Generality and Diversity

- Is a population with high generality diverse?
  - *General* agents can *specialise* differently through development.
    - Different roles in the group for the benefits of specialisation.
    - Different strategies because of different experience.
    - Acquired bias makes learning and communication more efficient.
  - Diversity is also achieved through non-cognitive traits (e.g., personality).
- Generality-Diversity: Virtuous circle?



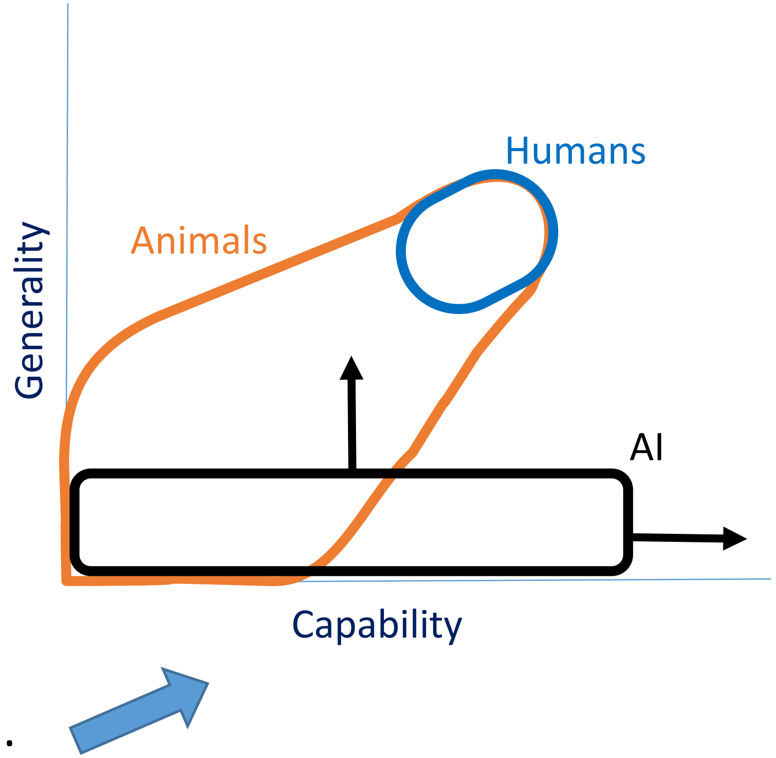
How does this compare with AlphaZero, and increase of capability (self-improvement) through selfplay (no diversity at all)?

Conclusion: Generality is Universal

- Generality conceptualised as a measure:
  - It's not populational: measures individual generality.
  - Depends on resources (difficulty).

> Limited resources connect capability and generality, and unite intelligence

- Generality splits from "general intelligence":
  - More universal perspective than evolution.
  - Artificial *General* Intelligence a matter of degree!
  - Complex interplay between diversity and generality.
  - A new dimension to analyse the landscape of cognition:



Ongoing Initiatives

- Generality and AGI Risks:
  - How does generality affect AGI safety, together with capability and resources?
- Cambridge^2 initiative:
  - Series of workshops on Generality and AI.
- The Atlas of Intelligence:
  - Collection of maps comparing humans, non-human animals and AI systems.

THANK YOU!

References

- Barmpalias and Dowe (2012) "Universality probability of a prefix-free machine". Philosophical Transactions of the Royal Society A. 2012
- Burkart et al. (2017)
- French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, *3*(4), 128-135.
- Hernandez-Orallo, J. (2000). Beyond the Turing test. Journal of Logic, Language and Information, 9(4), 447-466.
- Hernandez-Orallo & Dowe (2010). Hernández-Orallo, J., & Dowe, D. L. (2010). Measuring universal intelligence: Towards an anytime intelligence test. Artificial Intelligence, 174(18), 1508-1539.
- Hibbard, B. (2009). Bias and no free lunch in formal measures of intelligence. Journal of Artificial General Intelligence, 1(1), 54.
- Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines*, *17*(4), 391-444.
- Leike, J., & Hutter, M. (2015). Bad universal priors and notions of optimality. In Conference on Learning Theory (pp. 1244-1259).