Jakovác Antal
Department of Computational Sciences

Physical ideas in Artificial Intelligence

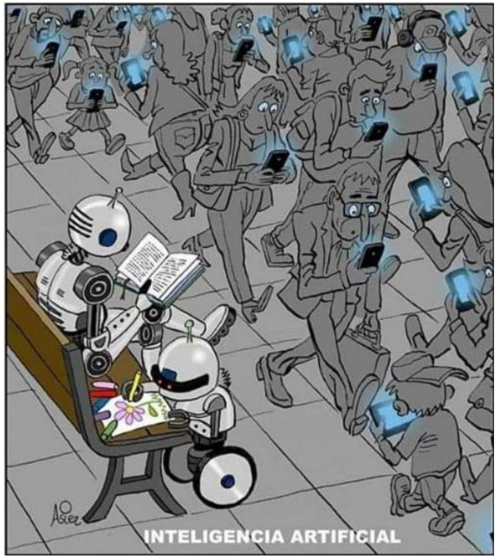
MT Kurbucz
P Pósfay
A Telcs
TS Biró

Margaret Island Symposium 2023 on Particles & Plasmas, 6–9 Jun 2023

Introduction

Recent impressive developments in AI

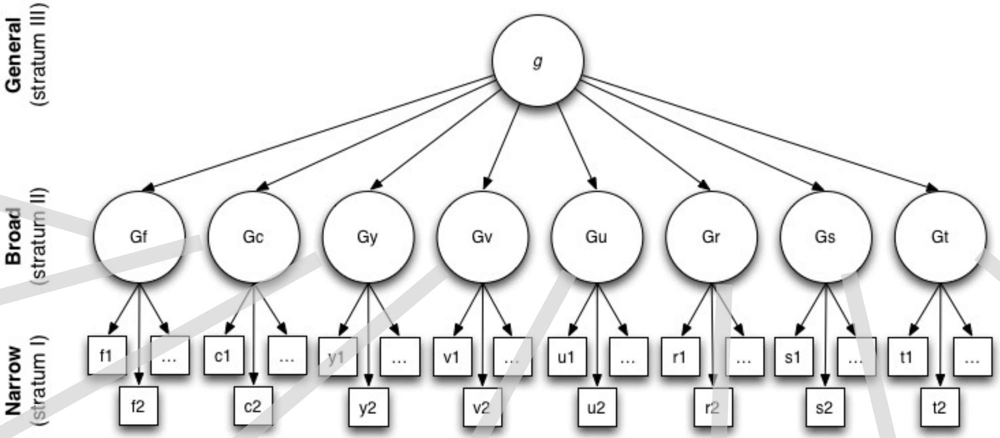
- Text generation: chatGPT, autoGPT, bing AI, bard AI, etc.
- Image generation: midjourney, thispersondeosnotexists, Dall-E, ...
- AI doomsday?
- Intelligent and useful tools, BUT heuristic, improvising, “lying”
- **Why do they work so well, and why do they fail so stupidly?**



Introduction

Human intelligence is not a monolithic entity:

Cattell–Horn–Carroll theory



Fluid intelligence: reasoning, analysis

Crystallized intelligence: use acquired knowledge

Memory and learning

Visual perception

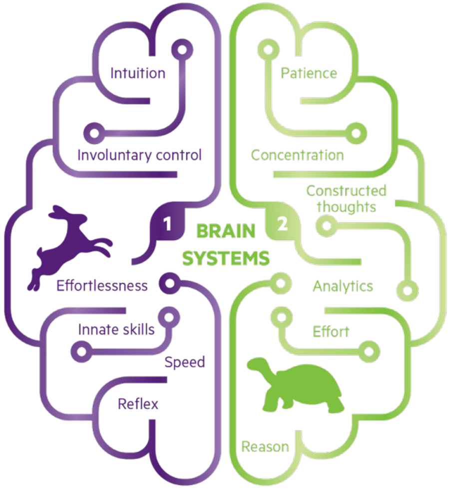
Auditory perception

Retrieval ability

Cognitive speediness

processing speed

Modes of human thinking



Categorization in cognitive psychology: *(Daniel Kahneman)*

- **System 1**:
  - fast, automatic, intuitive, no conscious awareness, no control, error-prone
  - ideal for fast, accurate responses (e.g. car driving, playing table tennis ...)
- **System 2**:
  - slower, conscious, deliberate, controlled, can be checked and re-iterated
  - ideal for contemplation, understanding

Introduction

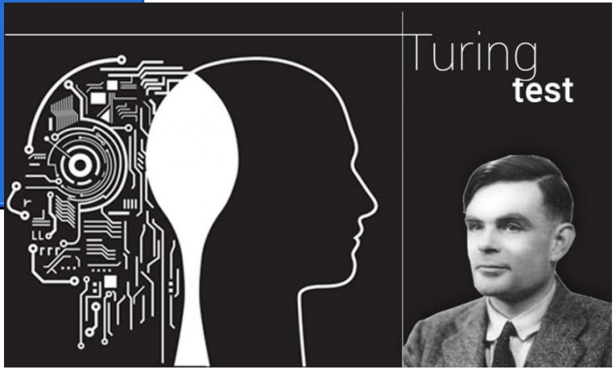
Human intelligence is not a monolithic entity, we use:

- *all parts of the intelligence*
- *System1 and System2*

→ ***unlike in AI models***

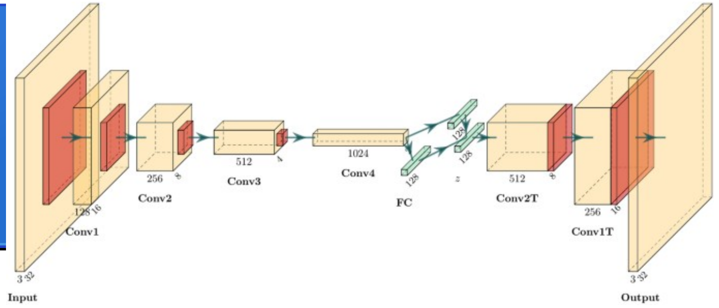
- We tend to think that all parts of IQ are present (cf. ELIZA, chatGPT → doomers)
- The performance of AI **depends on the task** we give
- **What was the question, if the answer is human intelligence?** *we do not know…*
  - *Turing's definition:* deceiving observers
  - *Classification task:* main stream AI solutions: System1
  - *How to represent System2?* ***scientific understanding***

Turing’s definition of intelligence

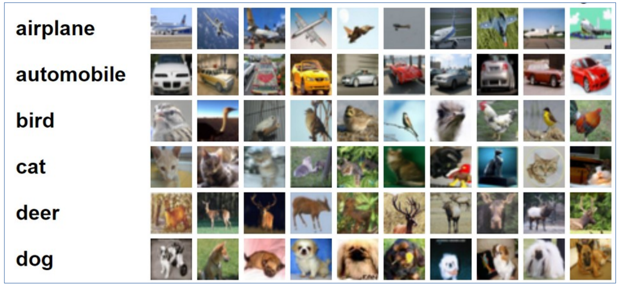


- **Intelligent:** indistinguishable from human in discussion (**Turing** 1950)
- **Task**: create a chatbot reacting to questions like humans do
- Famous programs:
  - ELIZA (Joseph Weizenbaum, 1960’s, MIT)
  - Eugene Goostman (13-year-old Ukrainian boy; 33% passed Turing test in 2014)
- Not really intelligent, but mislead humans to think they talk with an intelligent actor.

Classification task

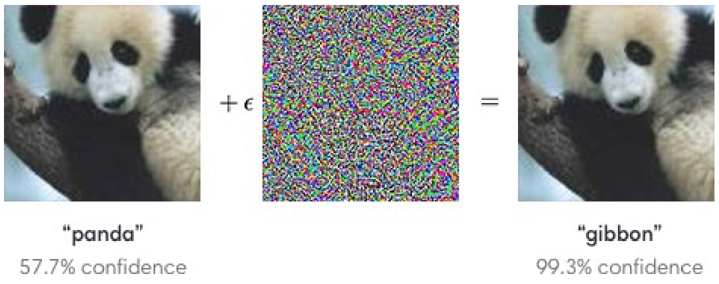


- **Intelligent**: classifies like humans do – we shall present the correct solution
- **Mathematical background**: probabilistic interpretation, Bayesian analysis, training, supervised learning
- **Technology**: plenty of ideas (DNN, CNN, ResNet, transformers, GAN, VAE, …)
- Most successful AI uses this method (classifiers, generators)



Classification task

- **Advantages**:
  - Very fast, effective
  - Good interpolation properties
- **Disadvantages** (apart from technical ones)
  - Slow training: needs a lot of data and uses a large amount of parameters
  - No control over the mistakes (c.f. adversarial attacks)
  - Input → output is a continuous function, can not train with very unbalanced data (e.g. can not have a class “no cat images”)
  - Specific → *catastrophic forgetting*: classification outputs are interdependent
- **All this corresponds to the System1 way of thinking!**



Understanding in science

How does the scientific understanding work?

- Make all possible *observations*: reveal **interactions,** microstates
- The “interesting” measurements are much fewer (“IR physics”, macrostates)
- There are interactions that do not influence the interesting physics (**irrelevant**)
- To build a model: take into account only the **independent relevant interactions** (Ockham’s razor)
- *Relevant interactions: constants over the microstates* (particle number, magnetization)
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Make all possible *observations*: reveal **interactions,** microstates
- The “interesting” measurements are much fewer (“IR physics”, macrostates)
- There are interactions that do not influence the interesting physics (**irrelevant**)
- To build a model: take into account only the **independent relevant interactions** (Ockham’s razor)
- *Relevant interactions: constants over the microstates* (particle number, magnetization)
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Make all possible *observations*: reveal **interactions,** microstates
- The “interesting” measurements are much fewer (“IR physics”, macrostates)
- There are interactions that do not influence the interesting physics (**irrelevant**)
- To build a model: take into account only the **independent relevant interactions** (Ockham’s razor)
- *Relevant interactions: constants over the microstates* (particle number, magnetization)
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” measurements are much fewer (“IR physics”, macrostates)
- There are interactions that do not influence the interesting physics (**irrelevant**)
- To build a model: take into account only the **independent relevant interactions** (Ockham’s razor)
- *Relevant interactions: constants over the microstates* (particle number, magnetization)
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

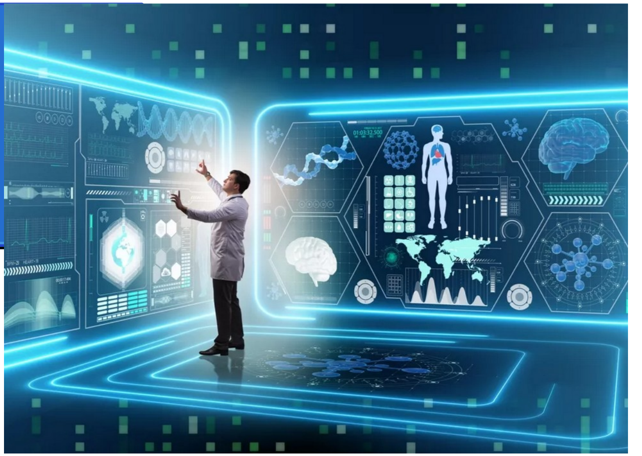
Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” measurements are much fewer (“IR physics”, macrostates)
- There are interactions that do not influence the interesting physics (**irrelevant**)
- To build a model: take into account only the **independent relevant interactions** (Ockham’s razor)
- *Relevant interactions: constants over the microstates* (particle number, magnetization)
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

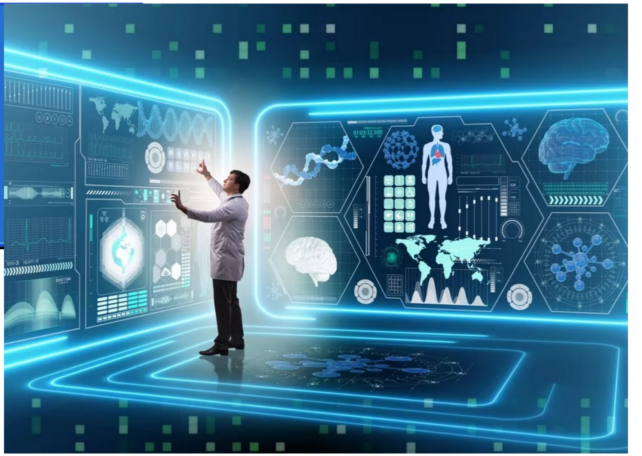


Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phenomena are much fewer (context)
- There are interactions that do not influence the interesting physics (**irrelevant**)
- To build a model: take into account only the **independent relevant interactions** (Ockham’s razor)
- *Relevant interactions: constants over the microstates* (particle number, magnetization)
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding



Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are interactions that do not influence the interesting physics (**irrelevant**)
- To build a model: take into account only the **independent relevant interactions** (Ockham’s razor)
- *Relevant interactions: constants over the microstates* (particle number, magnetization)
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are facts that do not influence the interesting phenomena (**irrelevant**)
- To build a model: take into account only the **independent relevant interactions** (Ockham’s razor)
- *Relevant interactions: constants over the microstates* (particle number, magnetization)
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are facts that do not influence the interesting phenomena (**irrelevant**)
- To build a model: take into account only the **independent relevant interactions** (Ockham’s razor)
- *Relevant interactions: constants over the microstates* (particle number, magnetization)
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are facts that do not influence the interesting phenomena (**irrelevant**)
- For understanding a phenomenon we shall keep the **independent relevant facts** (Ockham’s razor)
- *Relevant interactions: constants over the microstates* (particle number, magnetization)
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are facts that do not influence the interesting phenomena (**irrelevant**)
- For understanding a phenomenon we shall keep the **independent relevant facts** (Ockham’s razor)
- *Relevant interactions: constants over the microstates* (particle number, magnetization)
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are facts that do not influence the interesting phenomena (**irrelevant**)
- For understanding a phenomenon we shall keep the **independent relevant facts** (Ockham’s razor)
- *Relevant facts: are constants over the context (**laws**)*
- changing the “interesting” physics → change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are facts that do not influence the interesting phenomena (**irrelevant**)
- For understanding a phenomenon we shall keep the **independent relevant facts** (Ockham’s razor)
- *Relevant facts: are constants over the context (**laws**)*
- changing the “interesting” physics $\rightarrow$ change relevant interactions (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are facts that do not influence the interesting phenomena (**irrelevant**)
- For understanding a phenomenon we shall keep the **independent relevant facts** (Ockham’s razor)
- *Relevant facts: are constants over the context (**laws**)*
- changing the context → change the relevant facts (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are facts that do not influence the interesting phenomena (**irrelevant**)
- For understanding a phenomenon we shall keep the **independent relevant facts** (Ockham’s razor)
- *Relevant facts: are constants over the context (**laws**)*
- changing the context → change the relevant facts (**renormalization**)
- **scaling & dimensional analysis**: if macrostates are much bigger than microstates, then there remains just a few relevant interactions

the world is simple using an appropriate language

Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are facts that do not influence the interesting phenomena (**irrelevant**)
- For understanding a phenomenon we shall keep the **independent relevant facts** (Ockham’s razor)
- *Relevant facts: are constants over the context (**laws**)*
- changing the context → change the relevant facts (**renormalization**)
- usually we need take into account a **lot of relevant facts!**

the world is simple using an appropriate language

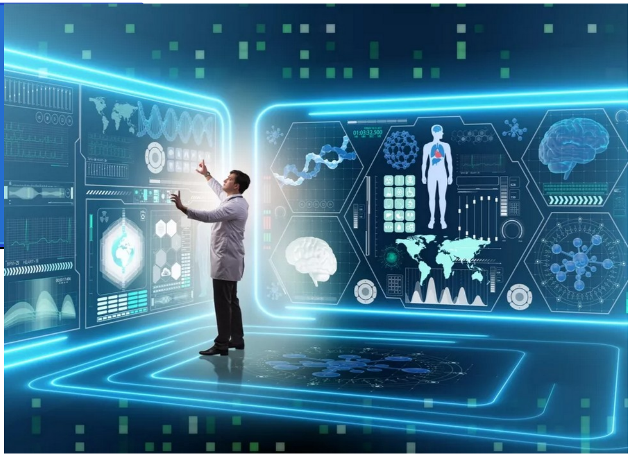
Generalizing scientific understanding

Almost all steps of the scientific method can be generalized!

- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are facts that do not influence the interesting phenomena (**irrelevant**)
- For understanding a phenomenon we shall keep the **independent relevant facts** (Ockham’s razor)
- *Relevant facts: are constants over the context (**laws**)*
- changing the context → change the relevant facts (**renormalization**)
- usually we need take into account a **lot of relevant facts!**

the world is simple using an appropriate language

Generalizing scientific understanding

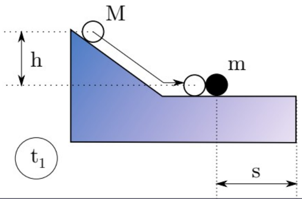


Almost all steps of the scientific method can be generalized!

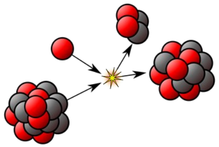
- Collect all available “facts”
- The “interesting” phanomena are much fewer (context)
- There are facts that do not influence the interesting phenomena (**irrelevant**)
- For understanding a phenomenon we shall keep the **independent relevant facts** (Ockham’s razor)
- *Relevant facts: are constants over the context (**laws**)*
- changing the context → change the relevant facts (**renormalization**)
- usually we need take into account a **lot of relevant facts!**

the world is complicated even with using an appropriate language

Generalizing scientific understanding



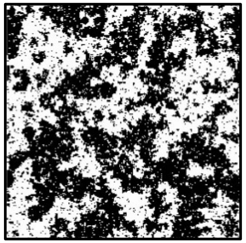
point mechanics
~ 5 relevant



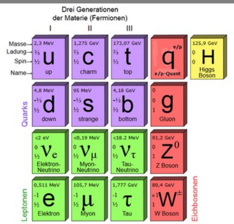
nuclear physics
20-? relevant

chemistry, biology
~ 100-? relevant

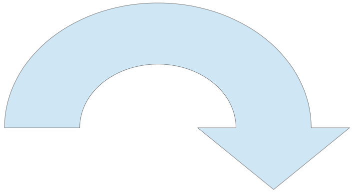
natural environment
? relevant ? irrelevant

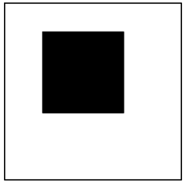


Ising model
3 relevant



Standard Model
21 relevant
(symmetries!)

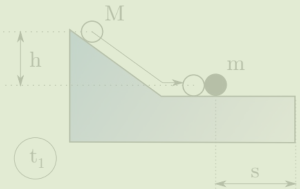


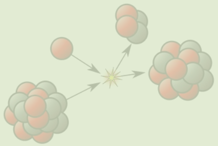


geometric images
~ 10-100 irrelevant

face recognitions
~ 30000 irrelevant

Generalizing scientific understanding





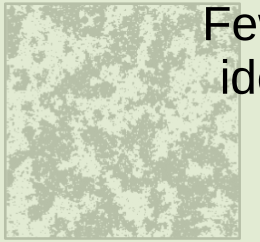
point mechanics
~ 5 relevant

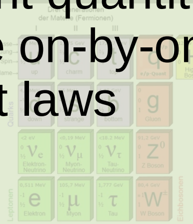
nuclear physics
20-? relevant

chemistry, biology
~ 100-? relevant

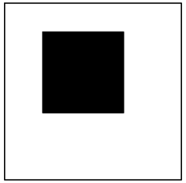
Science

Few relevant quantities
identifiable on-by-one
exact laws





natural environment
? relevant ? irrelevant



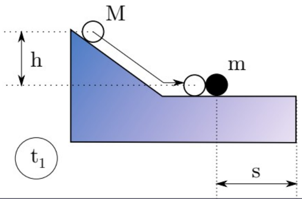
Ising model
3 relevant

Standard Model
21 relevant
(symmetries!)

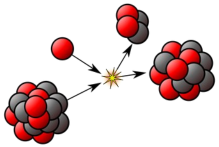
geometric images
~ 10-100 irrelevant

face recognitions
~ 30000 irrelevant

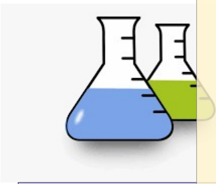
Generalizing scientific understanding



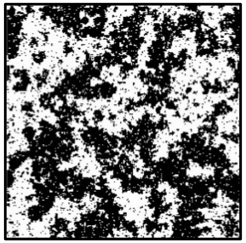
point mechanics
~ 5 relevant



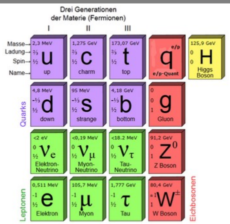
nuclear physics
20-? relevant



chemistry, biology
~ 100-? relevant



Ising model
3 relevant

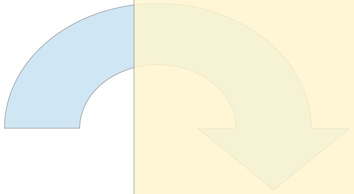


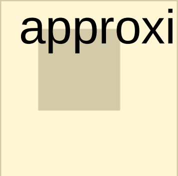
Standard Model
21 relevant
(symmetries!)

natural environment
? relevant ? irrelevant

Intelligence

Plenty of relevant quantities
not identifiable one-by-one
approximate laws

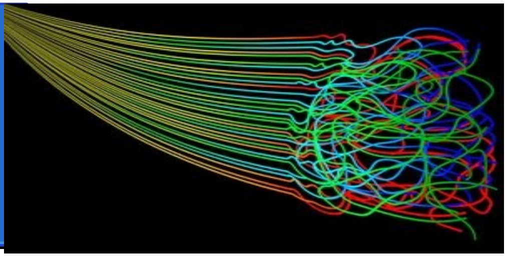




geometric images
~ 10-100 irrelevant

face recognitions
~ 30000 irrelevant

Entropy of the intelligence

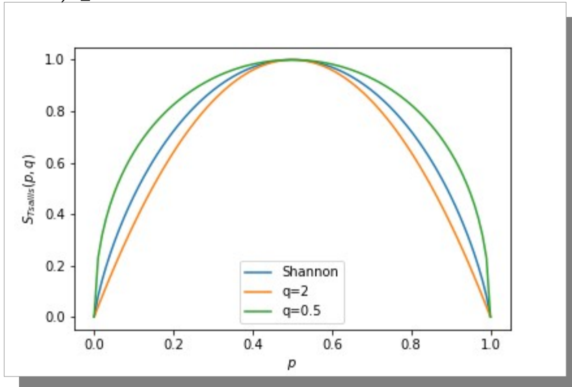


Best understanding: we use the minimal number of facts (Ockham’s razor)

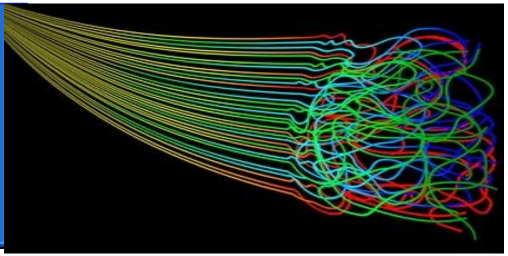
Is there a universal measure to decide, how good a given representation is?

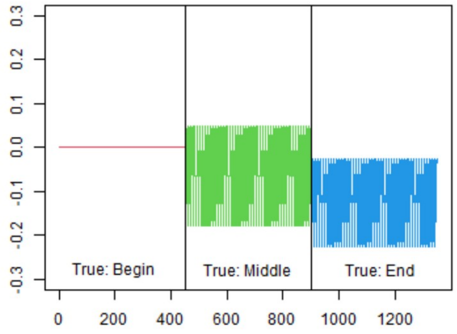
TS Biró, AJ, Universe 8 (1), 53; AJ, A Telcs, Entropy 24 (9), 1313

- **Simplest case:** context = C subset, binary measurement (fact): $\xi_i(\text{state}) \in \{0,1\}$
- probability distribution: $p_C(\xi_i=\sigma_i)=\dfrac{|\xi_i^{-1}(\sigma_i)\cap C|}{|C|}$
- **representation entropy:** $S_{repr}=\sum_{i=1}^{N}[\sum_{\sigma\in 0,1} p_C(\xi_i=\sigma_i)\log_2 p_C(\xi_i=\sigma)]$
- **properties:**
  - $S_{repr} \geq H$ Shannon entropy
  - equality iff independent facts, least number of relevant facts
  - we can use different individual entropy: $S_{repr}=\sum_{i=1}^{N} S(p_C(\xi_i))$

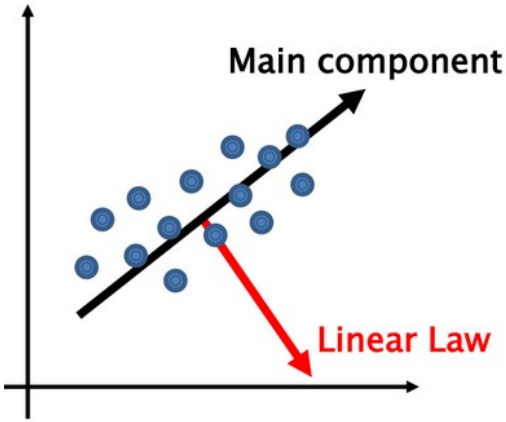


Entropy of the intelligence





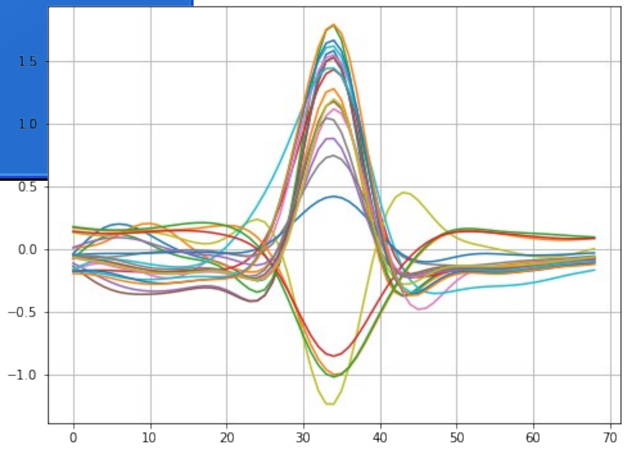
- **implementation for learning**:
  - instead of train with human annotated datasets, we shall seek facts that are **constant over elements of contexts (laws)**
  - ensure independence
- **practical approach**: law-based feature transformation
  - find laws in some functional space (e.g. linear functions)
  - collect laws for different elements of the context
  - for a new state use the best law → classification



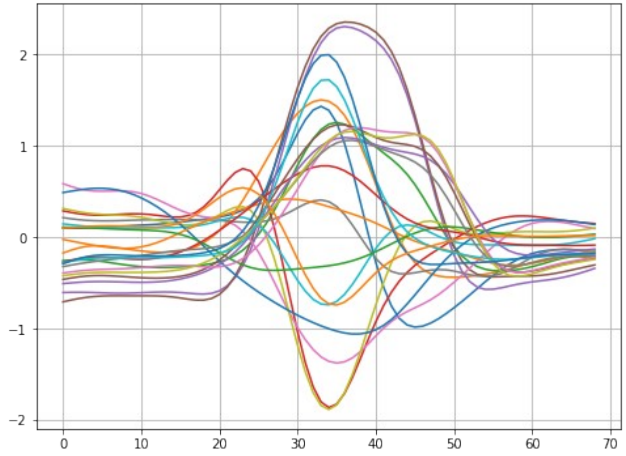
Relevance based intelligence

- **Advantages**:
  - control over mistakes: several laws
  - can be used with unbalanced data
  - no forgetting: laws separate class elements from not class elements, no interdependence between laws
  - fast training: needs fewer data and less parameters than training
- **Disadvantages** (apart from technical ones)
  - application can be slow for a lot of laws (parallelization necessary)
- **These are characteristic for System2 way of thinking**

Application: ECG analysis



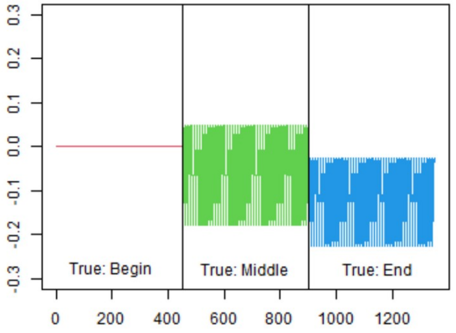
- Goal: classify heart beats into normal and ectopic
- ECG signal: cleaning, standardizing
- Mehtod: prepare test, validation and training sets
  - Find linear laws for the QRS complex (11 leg embedding, universal laws)
  - Train a classifier on the results (KNN, RF, SVM)
  - Results depend on several factors, best result SVM: 94.3% (close to state-of-art results)
  - More data could help to improve accuracy
- Can be used in a non-annotated dataset (self annotation)



Application: AReM database

MT Kurbucz, P Pósfay, AJ, Scientific Reports 12 (1), 18026

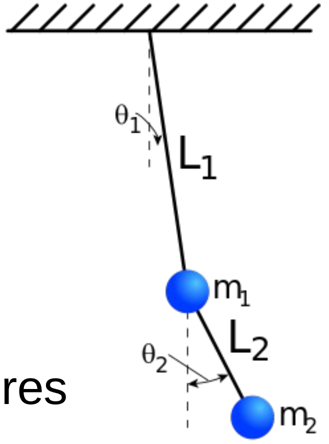
- Activity Recognition system based on Multisensor data fusion (AReM) Data Set
  - 7 motion classes (bending, lying, cycling,etc.)
  - 3 sensor data → 6 features (mean and variance)
  - 88 time series (instances), 480 values in each



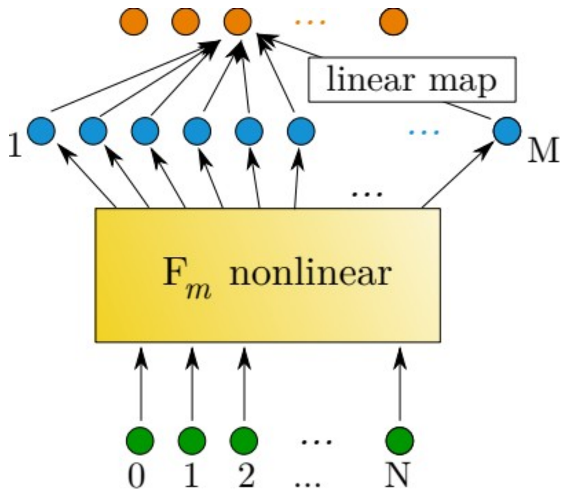
- Mehtod: LLT (Linear Law based feature Transformation)
  - Determine the laws for each instances and channels in the training sets
  - Apply them to the test series, take temporal average/variance → features
  - Train a classifier on the results (KNN, DT, SVM)
  - KNN provides error-free classification

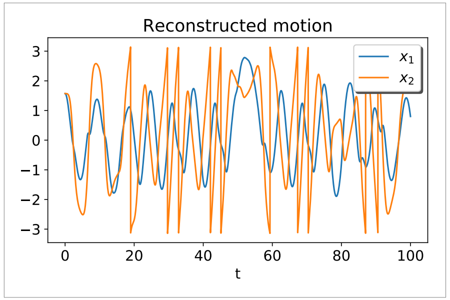
Nonlinear laws

AJ, MT Kurbucz, P Pósfay, New Journal of Physics 24 (7), 073021



- Generalization: input are not directly the embedded data, but pre-trained features
- F_m can be represented by (deep) neural network
- Extreme learning: the exact form of F_m does not matter
- Reconstruction of mechanical motions: 3-leg embedding (discrete Newton-equations)
- Chaoticity, stability → recursion to reconstruct motion

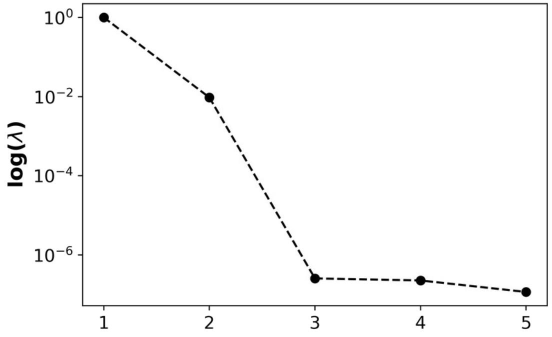


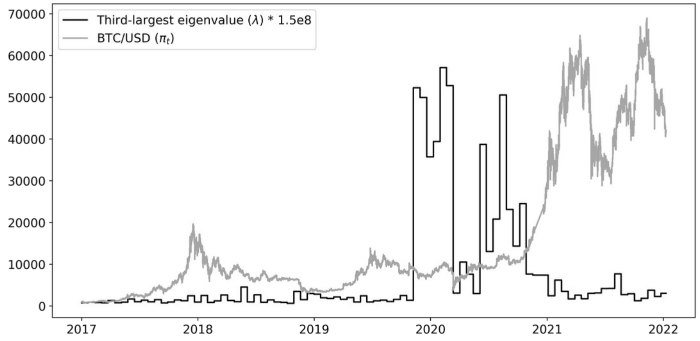


Stochastic processes

MT Kurbucz, P Pósfay, A Jakovác, arXiv preprint arXiv:2201.09790

- Markov chains: stochastic process where $P^{(n+1)}(x) = \sum_y T^{(n)}_{x,y} P^{(n)}(y)$ ⇒ $P^{(n+1)} = T P^{(n)}$
- In equilibrium (steady state) no *n* dependence, for equilibrium distribution: $P = T P$
- 2-variable correlation functions: $\langle f(x_n, x_{n+k}) \rangle = \mathrm{Tr}(F T^k)$ where $F_{xy} = f(x,y) P(x)$
- These satisfy linear laws: $\sum_k \langle f(x_n, x_{n+k}) \rangle w_k = 0$ if $\sum_k w_k T^k = 0$ characteristic polynomial
- Dimensionality of the Markov process can be determined from the laws





Conclusions

The question/task we want to solve determines the possible answers

- **Turing's intelligence definition**: programs deceiving humans
- **Classification task**
  - Probabilistic systems, specific tasks
  - Method of development: training
  - Slow training, fast operation → System 1
- **Representation task**
  - Structured systems, generic tasks, context
  - Method of development: finding relevant features, laws
  - Fast learning, slower operation → System 2

The end