From Societies of Brains to Brains of Societies

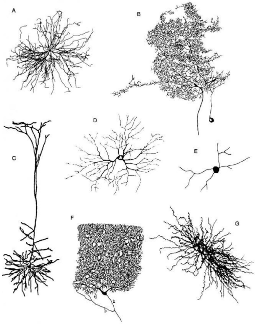
Reimer Kühn

Disordered Systems Group
Department of Mathematics
King's College London

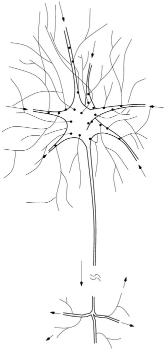
Cumberland Lodge, 19 Feb, 2022

Outline

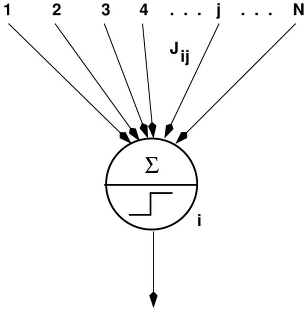
Neurons



diversity of neurons



schematic neuron



formal neuron

Human brain: $\sim 10^{11}$ neurons, each connected to $\sim 10^{4}$ (input & output),

$\lesssim 5-6$ handshakes between any pair

What Individual Neurons Do: Linear Separation

- Neurons: two state threshold elements, firing or non-firing

  $S_i \in \{1, 0\}$

  [W. S. McCulloch, W. Pitts (1943)]

- input output relation

  Post-synaptic potential:

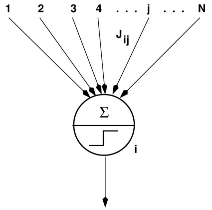
  $u_i = \sum_j J_{ij} S_j \equiv \boldsymbol{J}_i \cdot \boldsymbol{S}$

  $\Rightarrow$ output:

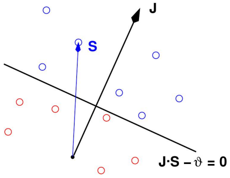
  $S_i = \Theta\left(u_i - \vartheta_i\right)$

- Synaptic couplings:

$$J_{ij} \begin{cases} > 0 & ; \text{excitatory} \\ < 0 & ; \text{inhibitory} \end{cases}$$



Neuron i receives N inputs (evidence!) S_j, gives weight J_{ij} to input S_j, sums all up, and compares to threshold ϑ_i. Fire if above, don't fire if below.

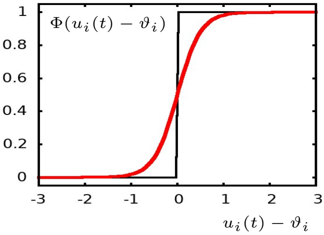


classification by linear separation

Neural Networks and Dynamics

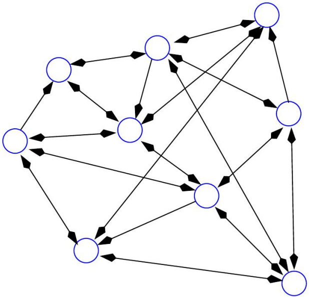
- Neurons interacting in networks: output of a neuron is used as input by others at later times (important when feedback loops exist).
- Dynamics (can be deterministic or probabilistic):

$$
\begin{aligned}
u_i(t) &= \sum_j J_{ij} S_j(t) \\
\mathrm{Prob}\big\{S_i(t+\Delta t) = 1\big\} &= \Phi\big(u_i(t) - \vartheta_i\big)
\end{aligned}
$$

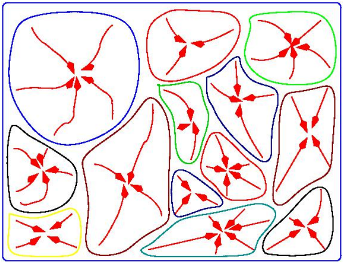


Recursive Networks – Attractors

- Deterministic dynamics: two types of global states: transient or persistent; the latter can be stable attractors (fixed points, limit cycles). ⇒ associative memory, motion control, ...
- Probabilistic dynamics: fluctuations about attractors of deterministic dynamics & occasional transitions between them (long-lived sets of states).



recursive Architecture

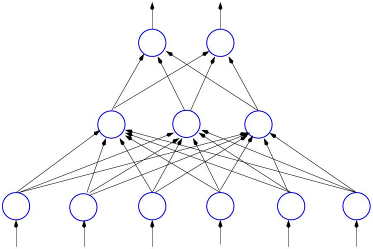


attractors, associative memory

attractors depend on $\{J_{ij}, \vartheta_i\}$: created by 'learning' or 'adaptation'

[D.O. Hebb (1949); J.J. Hopfield (1982)]

Feed-Forward Networks – Classification/Regression



feed-forward network
'question-answer machine'

- pattern recognition
- reading of written text
- triggers in high-energy physics
- control of autonomous vehicles
- playing the game of GO
- predicting protein structure ...
-
-
- trained on examples & able to generalize

Learning rules exist for systems with with smooth I/O relations.
Variants of these are the work-horses of the current AI revolution.

In human brains: mixture of modules with recursive and feed-forward architecture.

Outline

Information Processing & Learning

- neural firing states $\Longleftrightarrow$ information
- neural dynamics $\Longleftrightarrow$ information processing
- synaptic couplings $\Longleftrightarrow$ information processing capabilities
- processing capabilities evolve through learning/training

- Correspondences (high level)
  - neural firing states $\simeq$ brain states $\Longleftrightarrow$ cognitive states
  - neural dynamics $\Longleftrightarrow$ generates representations of world and acts on them; interprets, generates actions/reactions
  - synaptic couplings $\Longleftrightarrow$ cognitive repertoire
  - learning $\Longleftrightarrow$ adapt representations to improve adequacy, performance, success rate, survival probability, rewards, . . .

- Note: No CPU, slow but massively parallel hardware, highly fault tolerant, no separation between hardware and software

Learning — What is Involved?

- In order to establish and stabilize a desired set of (sequences of) neural activity patterns ($\simeq$ 'procedures') $\{\boldsymbol{S}^\mu\}$, $\mu = 1, 2, \ldots, p$,

$$S_i^\mu(t+\Delta t) = \Theta\Big(\sum_j J_{ij} S_j^\mu(t) - \vartheta_i\Big) \qquad \forall i, \mu, t \qquad (*)$$

$\Leftrightarrow$ **Require**: Parameters $\{J_{ij}, \vartheta_i\}$ must be found such that for each procedure μ and at all t, every neuron i should do what it is supposed to do in that situation, given the right input.

- Procedures could be realised as
  - fixed point attractors (associative memory)
  - firing sequences (motion control)
  - limit cycles (biological clocks)
  - input-output pairs (classifications of sensory data)
  - ...

Learning — When and Why Can it Fail?

- Recall: We want $\{J_{ij}\}$ and $\{\vartheta_i\}$ such that the network dynamics correctly implements all procedures $\boldsymbol{S}^\mu$, $\mu = 1, \ldots, p$.
- Analyse learning in terms of **version space**:

$$V_p = \Big\{ J_{ij}, \vartheta_i;\ \text{compatible with } \boldsymbol{S}^\mu \text{ for all } \mu = 1, \ldots, p \Big\}$$

- **Key point** to note:
  - Every new pattern/procedure puts new constraints on $\{J_{ij}, \vartheta_i\}$.
  - $\Rightarrow$ version space shrinks

$$V_0 \supseteq V_1 \supseteq V_2 \supseteq V_3 \supseteq \ldots$$

  - For given architecture/structure, there may exist

$$p_{\max} \quad \text{s.t.} \quad V_p = \emptyset \quad \text{for } p > p_{\max}$$

    independently of learning algorithm/strategy.

    [E. Gardner (1988)]

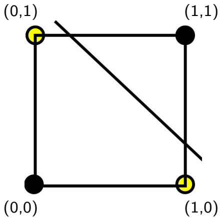
Learning — When and Why Can it Fail?

- $p_{\max}$ depends on architecture and (statistical) properties of $\{\boldsymbol{S}^\mu\}$. Typically of the order of the number N of input channels (adaptable parameters) per neuron.

- Architectures/structures have restrictions on the problems they can represent/solve. E.g.

$$S_0 = \mathrm{XOR}(S_1, S_2) \neq \Theta(J_{01}S_1 + J_{02}S_2 - \vartheta_0)$$

for any $J_{01}, J_{02}, \vartheta_0$.

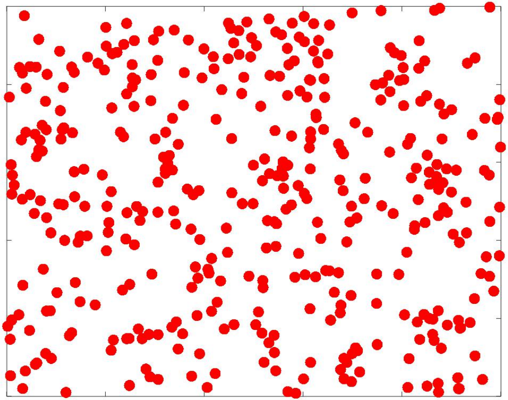


- As $p \to p_{\max}$ for given problem class, the version space can become fragmented (disconnected). Finding solutions that accommodate new patterns can become difficult (even impossible without violating some constraints 'on the way' to the new solution).

Learning — When and Why Can it Fail?

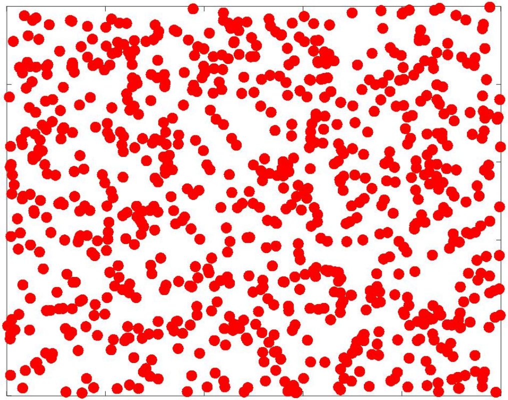
Version space: shrinking and fragmentation

Learning — When and Why Can it Fail?



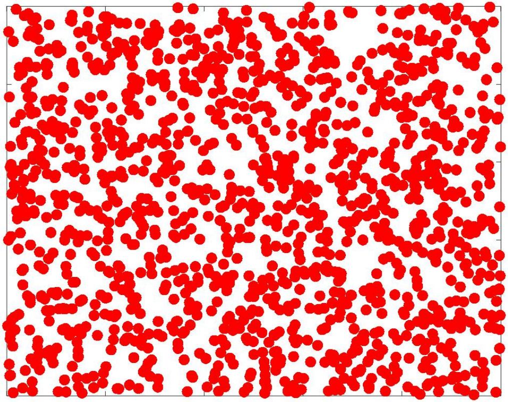
Version space: shrinking and fragmentation

Learning — When and Why Can it Fail?



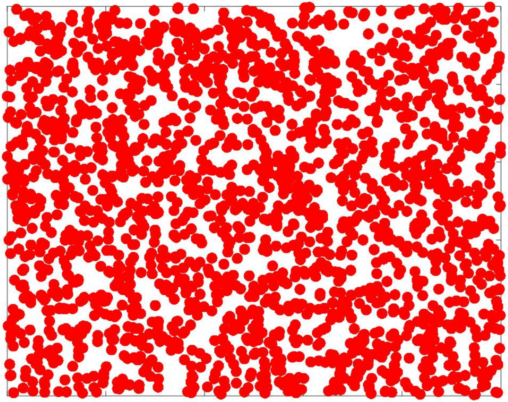
Version space: shrinking and fragmentation

Learning — When and Why Can it Fail?



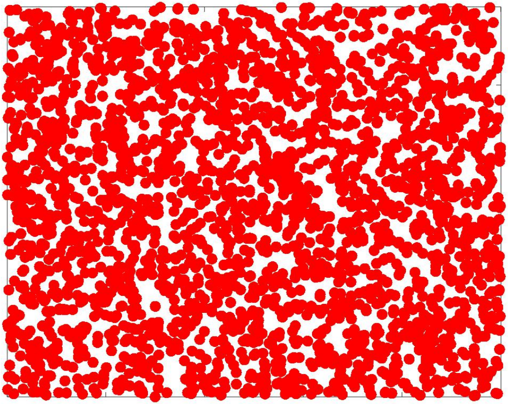
Version space: shrinking and fragmentation

Learning — When and Why Can it Fail?



Version space: shrinking and fragmentation

Learning — When and Why Can it Fail?



Version space: shrinking and fragmentation

Outline

Learning Systems — From Societies of Brains ...

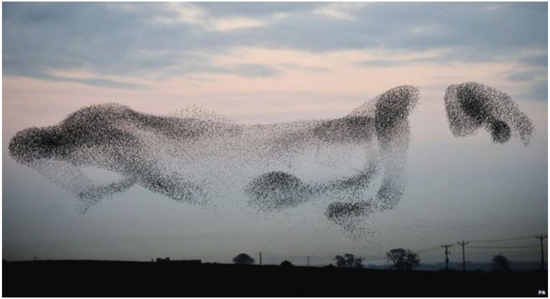
- Any complex system able to generate representations of the world is an information processing system.
- Need not be an individual agent, but could be a swarm, a team, an organization, or a society of complex agents.
- Information exchange in societies via language, actions, gestures, ...
- $\Rightarrow$ Collective information processing, if exchanged information influences/constrains (re-)action on receiving side.
- Any such system is capable of learning, if interactions (evaluations of exchanged information) are adapted, e.g. in order to improve performance, utility functions, ...

- Recall that this is essentially the mechanism underlying learning in brains!

. . . to Brains of Societies?

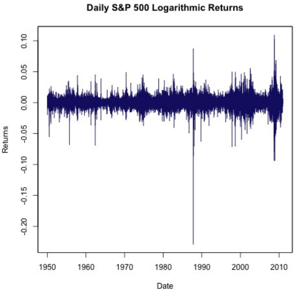
- Collective information processing in groups, communities, organizations, or societies will be guided by attractors.
    - subsets of system states or sequences of system state generated and reinforced by dynamics at the system level (autopoiesis)
    - pattern formation on short time-scales, (e.g. panics, emergence of rhythmic clapping after concert performances,. . .)
    - on longer time-scales: fashions, conventions , adoption of new technologies, . . . , trends in art and science, dominant scientific paradigms, moral value-systems).
- Recall: multiplicity/diversity of attractors in systems with given interactions; may observe occasional transitions between different attractors (spontaneous, or triggered by events)
- An attractor a system finds itself in may be contingent on past events. Degree of rigidity/fluidity of a society is a collective phenomenon (susceptibility).
- Important role of media of/technology for information exchange (reach, timescales . . .) [M McLuhan 1960s: Media as extensions of our senses . . .]

. . . to Brains of Societies?



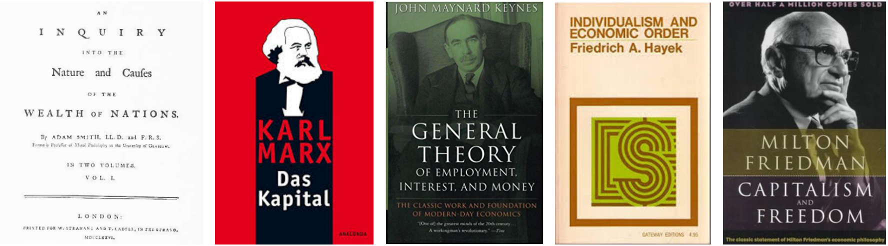
Murmuration of starlings

. . . to Brains of Societies?



Dynamics of Financial Markets:
(i) Fat tailed (leptokurtic) distributions of returns (student's t);
(ii) Fast decay of correlations of returns;
(iii) Very slow decay of correlations of volatilities (volatility clustering);
universal across virtually all markets.

. . . to Brains of Societies?



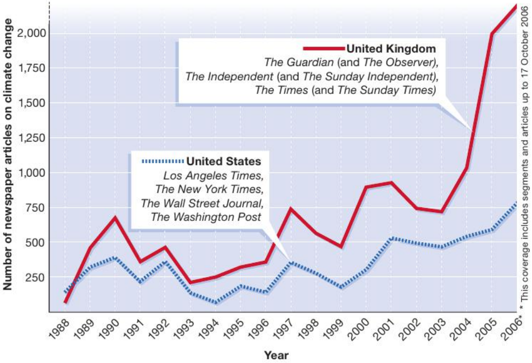
Evolution and Influence of Economic Theory

. . . to Brains of Societies?

MALLEVS
MALEFICARVM,
MALEFICAS ET EARVM
hæresim frameâ conterens,
EX VARIIS AVCTORIBVS COMPILATVS,
& in quatuor Tomos iustè distributus,
QVORVM DVO PRIORES VANAS DÆMONVM
versutias, præstigiosas eorum delusiones, superstitiosas Strigimagarum
ceremonias, horrendos etiam cum illis congressus; exactam denique
tam pestiferæ sectæ disquisitionem, & punitionem complectuntur.
Tertius praxim Exorcistarum ad Dæmonum, & Strigimagarum male-
ficia de Christi fidelibus pellenda; Quartus verò Artem Doctrinalem,
Benedictionalem, & Exorcismalem continent.
TOMVS PRIMVS.
Indices Auctorum, capitum, rerùmque non desunt.
Editio nouissima, infinitis penè mendis expurgata; cuique accessit Fuga
Dæmonum & Complementum artis exorcisticæ.
Vir siue mulier, in quibus Pythonicus, vel diuinationis fuerit spiritus, morte moriatur;
Leuitici cap. 20.
LVGDVNI,
Sumptibus CLAVDII BOVRGEAT, sub signo Mercurij Galli.
M. DC. LXIX.
CVM PRIVILEGIO REGIS.

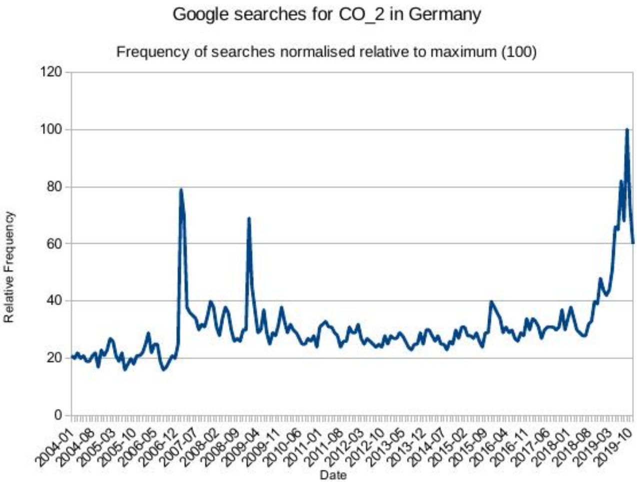
Witch trials in the early modern period
Estimated $40,000 - 60,000$ victims, 80% if which women
(Heinrich Kramer: Der Hexenhammer (1484))

Brains of Societies — Press Coverage & Public Debate



[T Boykoff & S R Rajan, EMBO Report **7** 207-211 (2007)]

Brains of Societies — Public Awareness



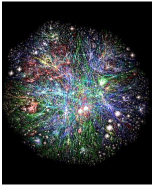
Public awareness: Google Searches for CO_2 in Germany; data from Google Trends. Peak in 2007 related to G8 Heiligendamm Summit (sustainability prominently on agenda for first time including (nonbinding) agreement aiming to at least halve CO_2 emissions by 2050); peak in 2009 combined effect G8 L'Aquila Summit and Copenhagen Summit on Climate Change

Brains of Societies — Collective Attention

- Results of Google search: barely looked at beyond page 5.
- Documents on page 1 "in the light of collective attention"
- Result of interaction between

  - page-rank: $\Leftrightarrow$ number (& importance!) of incoming links $\simeq$ votes
  - user behaviour: # of visits to a page ...
  - ...which are in turn influenced by page rank!

    $\Longrightarrow$ recursive dynamics with attractors!



Internet [opte.org (2007)]

- Some details at: `http://en.wikipedia.org/wiki/PageRank`
- Similar mechanisms at work in the domain of citation counts in science

Outline

What About Limitations?

- Explore whether the hypothesis about existence of fundamental limitations of representability in organizations survives scrutiny.
- Do this via series of approximations.
  - Start with society of two-state (yes/no) agents. Want:[1]

$$u_i^\mu = \sum_j J_{ij} S_j^\mu \,, \qquad S_i^\mu = \Theta\big(u_i^\mu - \vartheta_i\big) \qquad \forall i, \mu \qquad (*)$$

$\Leftrightarrow$ **Require**: Parameters $\{J_{ij}, \vartheta_i\}$ must be found such that an input $\boldsymbol{S} = (S_j)$ to agent i which represents 'context' μ, must generate the output of agent i that is required in context μ.

  - Standard theory for Mc Culloch-Pitts neurons applies. Typically $p_{\max} = \mathcal{O}(N)$. Precise values are known and depend on statistics; e.g. $p_{\max} = 2N$ for unbiased random patterns. [E. Gardner (1988)]

[1] Without loss of generality, choose procedures to consist of single I/O pairs

- First critique: states of agents are not binary
  - Answer: assume $S_i \in \mathbb{R}$ and graded response dynamics. Want

$$u_i^\mu = \sum_j J_{ij} S_j^\mu \ , \qquad S_i^\mu = g_i\big(u_i^\mu - \vartheta_i\big) \qquad \forall i, \mu \qquad (*)$$

    for some (continuous) response functions g_i.
  - Embedding routines (with tolerance ε):

> $\Leftrightarrow$ **Require**: Parameters $\{J_{ij}, \vartheta_i\}$ must be found such that an input $\boldsymbol{S} = (S_j)$ to agent i sufficiently close to 'context' μ, must generate an output of agent i sufficiently close to what is required in context μ.

  - Adapt standard theory to generalize to graded response neurons. Typically $p_{\max} = \mathcal{O}(N)$. Precise values are known and depend on statistics, input/output tolerance ε and the shape of the transfer-functions g_i. [D. Bollé, RK, J. van Mourik (1993)]

- Second critique: state variables communicated between agents are not simple scalars (language, gestures, actions, ...).

  - Answer: enlarge dimension. $\boldsymbol{u}_i = (u_i^a) \in \mathbb{R}^K$.

$$u_i^{\mu a} = \sum_{jb} J_{ij}^{ab} S_j^{\mu b} \quad , \quad S_i^{\mu a} = g_i^a \left(u_i^{\mu a} - \vartheta_i^a \right) \qquad \forall i, \mu \qquad (*)$$

  - Embedding patterns with tolerance ε then requires multidimensional generalization of previous argument.

$\Leftrightarrow$ **Require**: Parameters $\{J_{ij}^{ab}, \vartheta_i^a\}$ must be found such that an input $\boldsymbol{S} = (S_j^a)$ to agent i sufficiently close to 'context' μ, must generate an output of agent i sufficiently close *in all its dimensions* to what is required in context μ.

  - Same conclusions, although computations have not been done. Expect $p_{\max} = \mathcal{O}(KN)$

- Third critique: Agents are complex and have internal adaptable structure to 'compute' outputs.
  - Answer: could take internal adaptable structure into account

$$g_i^a\big(\boldsymbol{u}_i(t)\big) = g_{\boldsymbol{w}_i}^a\big(\boldsymbol{u}_i(t)\big)$$

    in which $\boldsymbol{w}_i$ stands for the collection of adaptable parameters within a function-class representable by agent i.

$\Leftrightarrow$ **Require**: Parameters $\{J_{ij}^{ab}, \boldsymbol{w}_i\}$ must be found such that an input $\boldsymbol{S} = (S_j^a)$ to agent i sufficiently close to 'context' μ, must generate an output of agent i sufficiently close *in all its dimensions* to what is required in context μ.

  - Conclusion about existence of fundamental limitations are not altered, if internal parameters are taken into account. Get further enlargement of expected $p_{\max}$.
  - Only partial results known. If g_i^a represents the computation of a feed-forward neural network with M hidden nodes, get an extra factor $\mathcal{O}(\ln M)$ for random binary patterns. [E. Baum, D. Haussler (1989)]

- Fourth critique: Individuals are not machines (but autonomous & unpredictable).
  - Answer: A difficult one. A tentative answer, as far as the question of existence of fundamental limitations is concerned:
    - Could model unpredictability and autonomy by replacing the g_i^a by random functions which produce a range of outputs with a statistics constrained by the inputs.

> $\Leftrightarrow$ **Require**: Parameters $\{J_{ij}^{ab}, \boldsymbol{w}_i\}$ must be found such that an input $\boldsymbol{S} = (S_j^a)$ to agent i close to 'context' μ, should *in all its dimensions and with sufficiently high probability* generate an output sufficiently close to what is required of agent i in context μ.

  - If this were not guaranteed for a procedure μ, there would be too many malfunctioning agents, which would prevent the reliable execution of that 'collective procedure'.
  - Conclusion about existence of fundamental limitations not altered. Details would depend of specification of probabilistic constraints.

Outline

Wrapping Up

- Note that collective information processing and learning is ubiquitous; it doesn't begin and it doesn't stop at the individual (neuron) level.
- Note hierarchy of levels: neurons $\rightarrow$ cortical columns $\rightarrow$ brain organization (visual, somatosensory, auditory, olfactory ... cortices, cerebellum ...) $\rightarrow$ brain $\rightarrow$ society of brains.
- Note that information exchange and processing between levels happens both ways: 'up and down'.
- Note that learning occurs at any level where evaluations of exchanged information are adapted, e.g. in order to improve performance.

Insights

- Collective information processing in groups, communities, organizations, or societies will be guided by attractors.
    - pattern formation on various time-scales, e.g.
    - swarming patterns, panics, emergence of rhythmic clapping
    - fashions, dynamics of economic cycles, adoption of new technologies
    - dominant paradigms or trends in art and science, moral value-systems
- Fundamental limitations are likely to exist for learning (i.e. the attainable cognitive repertoire in a given architecture) at all levels, including at the level of groups, organizations and societies.
    - slow-down of dynamics of change in complex organizations and highly evolved societies (fragmentation of version spaces)
    - some of the major transformations in history may be understood as finding 'extra dimensions' to accommodate solutions to problems previously unsolvable, given the structure (eg. hunter/gatherer to agricultural transition, division of labour & specialization, formation of states, invention of writing, printing with moveable letters, computers, the world wide web . . .)

Lessons?

- Note: Dialectics of innovation:

  (i) Version space expands, because existing procedures are made obsolete ($\Rightarrow$ constraints on parameters no longer required).
  (ii) Version space shrinks, because new procedures need to be implemented.

- Policy making or managing organizations should ideally ensure that balance is always on the right side!

  - For Innovations that don't produce additional problem solving capabilities, the effect of (i) must always be dominant.
  - Conversely, the only justification for dominance of effect (ii) is indeed an enrichment of the repertoire of information processing capabilities of an organization.

- Vehicles for enhancing collective cognitive repertoire: arts, sciences, collaboration, inclusion, participation, interdisciplinarity

Outline

Summary

- Neural information processing as paradigm for decentralized & collective information processing
- Hierarchy of levels of information processing & meanings
- Structure and limits of representability (version spaces)
- Society of brains as information processing systems
- Framework to rationalize pattern formation in societies on many time-scales
- Limits on learning adaptability and functionality for such systems
- Lessons for policy making

Thank you!