PHILOSOPHICAL TRANSACTIONS A

royalsocietypublishing.org/journal/rsta

Research

Article submitted to journal

Subject Areas:
AI and Information Processing, Dynamical Systems, Human Cognition

Keywords:
Neuromorphic Computation, Neuromorphic AI, Reasoning, Imagination, Consciousness and Cognition

Author for correspondence:
Peter Grindrod
e-mail: grindrod@maths.ox.ac.uk, pete@astut.co.uk

On The Next Generation For Neuromorphic Computing and Neuromorphic AI

Peter Grindrod[1]

[1]Astut Ltd and the Mathematical Institute, University of Oxford, UK

As we learn more about the cognitive inner workings of the brain's information processing and decision-making behaviours, this naturally leads us to consider alternative and additional ways to process information, from chips and architectures through to the generation of insights and decisions. In turn, this suggests some options that might respond to challenges that are not addressable by the present state-of-the-art. By their nature though, they may develop some common features and idiosyncrasies that are associated with human cognition; such as, individual expertise and blind spots, illusions, systematic errors, and transient mind-sets, as well as evolutionary advantageous fast-thinking facilities, applicable within novel and data-poor circumstances.

We will discuss some of the lessons learned by the reverse engineering of very large scale neuron-to-neuron simulations (1B neurons) within cortex-like, network-of-networks, architectures. We identify some elements of the dynamical behaviour of the inner sub-networks (neural columns) that are not exhibited by present-day neuromorphic chips, owing to conceptual and design limitations. We describe a novel mathematical framework that might encompass human cognitive processing alongside various future neuromorphic processing concepts.

We will also identify certain elements of human cognition, reasoning, and performance which present-day chips and present-day AI simply cannot fully emulate (match to a high standard, in some artificial way), or simulate (achieve in the same way). We will discuss how these aspects might catalyse some new fields of development for both processors and AI methodologies.

In short, we discuss how future research and development will respond in radical ways that are precluded by most present-day technologies.

THE ROYAL SOCIETY
PUBLISHING

1. Introduction

Neuromorphic concepts describe elements of engineering and processing that reflect the structures, features, and information processing observed within the human brain.

In this paper we will consider the human brain from two distinct perspectives: a physiological one, that examines some features of, and hard facts concerning, its information processing within a large-scale architecture; and a psychological (cognitive) reasoning one, that examines features and abilities of intelligent human reasoning.

Both perspectives are highly relevant to the development of next generation neuromorphic information processors and neuromorphic AI. As we shall see, some aspects are already driving novel research, while some aspects run counter to the present design of processors and AI algorithms.

Both perspectives tend to support the development of processing and AI that are more explainable, and therefore more trustworthy and justifiable. Yet this comes at the cost of being be more individual, possibly inconsistent, and somewhat subjective. As a result the outputs should be treated as trusted and reasoned advice, rather than as hard computational results and estimates. Neuromorphic ideas should go far beyond simple massively parallel arrays of processors and agent algorithms, driven by the power of modern computation. They should inspire new paradigms of processing and reasoning.

Our aim is to set out a number of clear research challenges that will overtly inspire, or are already covertly inspiring, novel processing and AI.

In particular, the physiological perspective calls for some lessons to be learned from the reverse engineering of cortex-like, large-scale, dynamical systems simulations, with a mathematical approach to the categorisation of the full range of the possible dynamics available to, and exhibited by, such systems. These are often precluded by chip design assumptions. They include the exploitation of multiple transient dynamical modes that each precondition and reduce the cognitive load (beyond the existing, and hugely impactful, *attention mechanisms*).

On the other hand, the reasoning perspective suggests some features of human intelligence that would be highly desirable (even if emulated rather than actually replicated).

2. What is going on inside a human brain?

The human brain's cortex is composed of about 1M neural columns, each of which contains about 10^4 densely interconnected neurons. The columns are arranged in a two-dimensional array (think of the cortex as being ironed out flat like a carpet, where each of the carpet piles is a single column). Some neurons are connected to neurons from near-neighbouring columns (not too far away). So the whole is a classic directed network-of-networks: the outer network, between the neural columns, is relatively sparse, whilst the inner networks, inside the neural columns, are relatively dense. Network scientists refer to such networks as having "high modularity" (with the inner, relatively dense, clusters called "modules").

The individual neurons themselves are excitable and refractory. This means that when a neuron is excited it *fires* in the cell centre (soma) and a spike of membrane depolarisation moves outwards along its branched axons, reaches some synapses; and there it jumps across onto a dendrite of a receiving neuron, with the spike moving inwards to the cell centre (soma) of the receiving neuron. Transmission of such signals takes place in a directed network-of-networks, with each neuron-to-neuron transmission taking some time to make it, if at all. When a neuron receives an incoming spike, it may itself respond with a spike, if it is ready: but once it has spiked it requires a little time, called the refractory time-period, for the various ions on either side of the cell membrane to re-equilibrate. If it receives a spike input during the refractory time-period, then it is not ready and no new spike will occur.

A single mass of densely directly connected neurons within a single neural column, once started-up, will just chatter to itself as signal spikes are passed around and around component cycles of its component directed network. This is all standard physiology.

(a) Large scale simulations of the cortex

In [1] the authors showed that the number of dynamical degrees of freedom exhibited by a single chattering neural column is roughly proportional to the logarithm of the number of neurons it contains (whilst the incoming and outgoing degrees of each neuron remain fairly similar). That phenomenon is not unrelated to conjectures about the number of independent (non-intersecting) cycles within random networks of many types. In turn, this predicts a *Goldilocks* effect where neural columns have to be large enough to exhibit robustly a good number of dynamical degrees of freedom, but should never be too large, as that would be wasteful. It would be more efficient, in terms of energy and volume, to have two smaller neural columns rather than have just one of twice the size. As a corollary we may deduce that neural columns should be broadly similar in size, as it would be inefficient to have range of sizes: you would get more total degrees of freedom (thus, better possible performance) with a uniform set of large enough columns. This is what is observed in the mammalian brain. Any humans who had brains with a wide range of differently sized neural columns would have been at a functional, and thus an evolutionary, disadvantage. So a dynamical systems view of neurons and neural columns implies a relatively uniform cortex architecture.

There is more to be implied regarding neuromorphic chips. Within such simulations [1] one must generate a suitable refractory time constant for each cell and a transmission time-delay (the time taken for a complete cell-to-cell spike transmission, successful or not). By direct demonstration one can show that if all the time-delays are set to the same constant (to one say), then the number of dynamical degrees of freedom quenches, dropping down radically. This is because there are many dead heats in walks (“walks" are successive, joined-up, neuron-to-neuron connections) of the same length from a certain sending neuron, via others, to a certain receiving neuron, much further away. So only one spike is generated at the receiving neuron (a dead heat). Whereas, with real valued time-delays, both these walks may have been independently viable (in causing receiver spikes) with spike arrivals separated by a difference in cumulative walk times that is greater than the corresponding receiving neuron’s refractory recovery time period.

Yet the present generation of neuromorphic chips, such as TrueNorth [2] (developed by IBM in realizing the vision of the DARPA *Systems of Neuromorphic Adaptive Plastic Scalable Electronics (SyNAPSE)* program) suffers under such an assumption. Similarly, Intel’s Loihi chip [3] possesses a digital architecture that “*approximates the continuous time dynamics using a fixed-size discrete time-step model*”.

Perhaps the dynamical neuromorphic baby has been thrown out with the bath water, as a result of the preference to only update the state of the chip (the neuronal array) at discrete (integer-based) times. In section (b) we discuss a much more wide and wise range of possible chip dynamics.

Moving up to whole cortex simulations, one may generate a full network-of-networks architecture and distribute cellular and connection-based properties appropriately (refractory times and transmission-based time-delays, respectively) as real values, not integers; and then make much larger-scale simulations.

This programme was carried out in [4], together with some reverse engineering of the output behaviours. When there are $N = 10^8$ up to 10^{10} individual neurons, one develops N spike trains in any experimental run. So the reverse engineering is rather *heavy lifting*. Ironically, it really requires a massive multi-processor computer (see [4]) in order to digitally simulate the natural, dynamics and processing observed within the biological system. The human brain’s 1.5kg of *wetware* instantiates everything robustly and at low energy. No wonder there is so much interest in neuromorphic processing: but is that ambition going far enough? Surely one should not be

royalsocietypublishing.org/journal/rsta Phil. Trans. R. Soc. A 0000000

content with massively parallel arrays of simple processes and claiming it to be “neuromorphic". How could and should mathematics help?

In separate experiments using a network of neural columns [4] one may force (stimulate) the full system by applying periodic spiking inputs at some chosen neuron or another and observing the response patterns (both in time and across the whole brain), at first locally, within the forced neuron’s own column, and then more widely, across the whole brain. One observes that the outputs taken from many, many, experiments can be represented by a hierarchical distribution of dynamical “modes". Each mode is a particular type of firing pattern over time and across the whole neuronal network. At any given level within the modal hierarchy they are mutually exclusive and competitive. Any neural architecture with real valued transmission delays and with excitable refractory units (cells) will exhibit this overall functionality.

The modes, both lower down (smaller) and higher up (larger) within the hierarchy, are candidates for qualia and other internal conscious phenomenological experiences. These too are arranged hierarchically from smaller experiences (the internal experience from the crunch of biting into an apple, or from the sight of the blueness of the sky, or from hearing the brassiness of a trumpet) up to higher phenomenological experiences (fear, anxiety, embarrassment, love). The “hard problem of consciousness", popularized by David Chalmers [5–7], is that of explaining how and why such phenomenological experiences might come about. Yet here, in the reverse engineering of a massive number of large-scale simulations, we have a possible answer.

In [8] it is argued that cognitive activity and the conscious phenomena are fully coupled. The acts of cognition (responding to incoming stimuli) result in the dynamical modes (the internal experiences), while the modes pre-condition the brain in advance of new incoming stimuli and the resulting cognition. Of course, there is an evolutionary fast thinking [9] advantage to any such pre-conditioning. In decision terms it reduces the possible decision set. Faster decision responses to stimuli may be of existential importance. However, it does not come for free: a brain’s immediate responses will depend on the internal experiences taking place (the present mode it is in) and there may well be a lack of consistency. Furthermore, any fast thinking, ruling out exhaustive logic, will become predictably irrational [10] and result in fast system heuristics [9].

In a later paper, [11] showed that the same internal hierarchical dynamical mode structures are still present if we replace the inner neural columns (each of 10,000 neurons) with medium dimensional clocks (each having, say, just ten or so phases) and allow these to be coupled together within an outer sparse network via phase-resetting directed connections. This is much easier to instantiate on a normal computer: but the internal phenomena - the entwinement of the present dynamical response (cognition) with the present preconditioning dynamical modes (the conscious experience) - is exactly the same. This also suggests a particular class of permissive unexplored neuromorphic chip dynamics.

Hence the design of neuromorphic processing needs to embrace real valued time delays between pairwise neurons (and pairwise groups of neurons, such as the neural columns), and novel AI algorithms might exploit the internal mode-based phenomena as pre-conditioners and accept the possibility of inconsistencies and irrationality within fast-thinking cognition.

In [7] the author contrasts the insights regarding cognition and consciousness made by direct simulations (and their consequent reverse engineering), with two other popular approaches to consciousness: integrated information theory (IIT) and quantum consciousness. In general though, the latter two approaches are not predictive: they do not produce testable outcomes that would be normal for the progression of (any) science [12,13]. On the other hand the identification of the dynamical modes with the internal feelings predicts that the latter should be competitive and mutually exclusive; and that they can be instantiated by assembling a suitable set of perceptions to cause yourself go to the happy place, or feel sad or embarrassed, by bringing the right kind of elements to mind [6]. Moreover, there is today simply no evidence that any (warm and wet) molecular biological systems have *lucked-in* to exploiting quantum effects for evolutionary advantage: it is merely hopeful and wishful thinking [7,14,15]. In general, quantum physics is far better to be deployed within the various national quantum technologies programmes (including quantum computing) that will capture quantum effects within labs and subsequently novel devices: they are not “neuromorphic".

The conceptual idea that (even transient) dynamical modes should precondition cognitive processing tasks is not new. It was the basis for the existing concept of an *attention mechanism*, which has been so successful within AI in reducing computational loads [16] (more recently discussed in [17] and the references therein). You simply do not need to boil the ocean for every sequential decision: an attentional focus is truely neuromorphic AI. Allowing models to focus on the most relevant parts of input data and decision requirements paved the way for breakthroughs such as the transformer architecture. Consequently such attention mechanisms have become essential within modern NLP and LLMs, enabling smarter, much more efficient, generative AI systems.

The motivation for this, as set out here, comes both from the existing advances in AI, and from the discoveries from reverse engineering large scale simulations (while the computational functionality required to carry out such simulations is readily available and deployable [4,11]).

Unfortunately not everything within large scale simulation has worked out so well. The tribulations of two very large science projects aiming to fully simulate human brains, in the US and the EU, have been well documented [18]. Those were caused by a variety of issues. These programmes have become focused on smaller, yet important, goals of brain mapping and building data processing facilities and new tools with which to study the brain. In [18] these big science projects were summarised, “*...instead of answering the question of consciousness, developing these methods has, if anything, only opened up more questions about the brain — and shown just how complex it is.*" Such a relative failure, and a bad experience, should not discourage us from learning the clear lessons from the reverse engineering of whole brain phenomena seen within simulations.

(b) Neural columns as dynamical systems over closed manifolds

Before we leave the processing issues within and across neural columns, it is useful to pause and re-consider the nature of the dynamics available.

Consider a single neural column: a directed network of excitable and refractory neurons, with suitable (real valued) time delays specified for each directed edge and refractory periods specified for each node. As in [1,4] the column will just chatter away to itself with every node producing a spike train over time. This situation is exactly what is modelled by (existing) neuromorphic chips, subject to simplifying assumptions (the integer or unitary time delays, for example). The reverse engineering of simulations in [1] examined the spiking output for independent periodicities. In general, after a transient, one observes a chaotic attractor, or a very long periodic attractor, containing many non-commensurate cycles. This observation led directly to the estimation of the number of dynamical degrees of freedom, discussed above, and hence to the possible simplification deployed in [11], by representing the column’s dynamics as a winding map over a $K-$dimensional torus, referred to above as a a $K-$dimensional clock, with a state described by K separate phases, and a corresponding flow. There are some other possibilities though. In [19] definitions and dynamical flows over other classes of k-dimensional closed manifolds are developed. These include non-orientable manifolds such as various generalisations of the Klein bottle or real projective spaces. The illustrations set out in [19] show how such spiking dynamics, starting out from a single nodal spike train, may be analysed so as to characterise the corresponding underlying attractor as such a suitable closed manifold. It is clear that the range and nature of dynamics available within even a single neural column is far more flexible and exotic than the highly restricted possibilities offered by present-day neuromorphic chips, such as in [2,3].

There is a need to set out the theoretical foundations, including a mathematical framework and vocabulary, for the development of next-generation neuromorphic information processing chips. At present, the neuromorphic computing sector is lacking in fundamental concepts concerning the parallelising and coupling of various unit processors in order to improve the embedding of certain AI paradigms. Moreover, the processors are often too simple (in terms of their state spaces and dynamics). To accelerate, this field needs to become both more permissive and yet more rigorous; and far closer to the nature of the information processing occuring within the human cortex is modelled.

The aim should be to enable more ambitious R&D concepts, and their direct inter-comparisons, by providing a well-defined and inclusive framework and theory, *The Mathematics of Human Cognition*, within which the dynamical *spiking* processing of the human cortex can sit along side those of particular chip designs.

This programme should address the following challenges.

a) Exploiting classes of compact closed manifolds, notably high dimensional (generalised) Klein bottles and tori [19], and suitable flows over them.
b) Exploiting novel methods such as topological data analysis methods that can analyse spike time-sequence data, from large spiking systems, and recognise, characterise, and calibrate suitable closed manifolds as attractors within the corresponding systems.
c) Establishing conditions upon large-scale spiking systems that imply the existence, or high likelihood, of various sub-classes of such attractors as models for neural columns.
d) Connecting-up of arrays of such dynamical systems within the outer network, as a means of creating novel neuromorphic chips.
e) Exploiting signal processing methods for the verification of behaviour within simulations, and any future novel physical instantiations, of such systems as candidates for neuromorphic chips.
f) Developing a common mathematical framework that subsumes the human brain's dynamical processing as well those of the various neuromorphic information processing concepts and platforms.
g) Engaging with the global neuromorphic computing sector, that is set to grow to 35T+ USD by 2035.

Although there are many variants between mere massively parallelised ("neuro-inspired") computational platforms and *true* neuromorphic chips and mechanisms, it is essential to define and set out the range of possibilities, and to provide a framework that encompasses all, and that will inspire and drive future neuromorphic information processing concept development.

In turn, world-class impacts for neuromorphic processing will be supported by appropriate mathematics: including topological data analysis, dynamical systems, large-scale simulation, and data scientific reverse engineering.

Novel neuromorphic paradigms are of present strategic interest to global corporates, such as Accenture, Arm, Brainchip, Cadence, HP, IBM, Intel, Knowm, Nvidia, Qualcomm, Samsung, Thales, and others, including national laboratories. Such processing will also underpin AI reasoning of the kind discussed in the next section.

3. Beyond simulation: further lessons from human reasoning

So far we have suggested some insights based on human brain simulation that might result in next generation developments of both neuromorphic chips and AI methods. In this section we will make some functional comparisons between human reasoning and what might be achievable within next generation neuromorphic AI and information processing. It is certainly a category error (and a common one) to present any neuromorphic chips or algorithms with tests that involve computations of the kind that our present (binary) chips and algorithms do so well; such as large sparse matrix decompositions and inversions, network-based calculations, any signal processing (as in past chip tests), and high bandwidth (data rich) calculations and decision making; rather than some tests associated with emulating some desirable attributes of human reasoning.

Each of these reasoning mechanisms might form the basis of novel neuromorphic AI paradigms and would extend the role and reach of AI, even as emulation. There is exciting work to do. Here we consider the following functional abilities, (a) through (e).

Of course, the cognitive capacities that we address below have each been studied within humans. For example, there is a substantial philosophical effort dealing with questions such as, "Where do ideas and hypotheses come from?". Rather than reflect on how such may occur within human brains, here we focus on how such abilities challenge next generation AI to emulate them (note, we do not expect that such AI will simulate them, working in the same way as a human brain).

There are open questions about how any of the (known, observed, information processing) dynamical properties within the human brain have anything to do with these higher cognitive capabilities? How do the dynamical properties give rise to abstraction, hypothesis generation, and so on? Indeed, though encouraging, our own massive dynamical simulations of brain dynamics are still to primitive to generate hypotheses, or imaginings, in response to any actual situational questions. On the other hand, even thinking aloud about these cognitive abilities (as here in this paper) is enough to catalyse, more radical, tactical ideas for next-generation neuromorphic AI functionality; especially in dealing with situations where the present-day, common, AI for data-rich decision-making cannot even get started [20]. This is driving recent interests in Explorative AI [21,22]

Encouragingly, there is a rich literature within computational neuroscience that explores how brains might disentangled representations to enable reliable generalization. However this is largely by way of explaining the human brain's performance, rather than setting out performance/emulation challenges for next generation AI.

(a) Abstraction

The human brain works at different levels. On the one hand it works deep down in the weeds, at the micro level of detailed stimuli, but in representing such knowledge it may implicitly or explicitly embed such data into metric spaces that have both local and global structures.

For example, language processing can represent a large set of words as a point cloud embedded within a suitable metric space, where word proximity is correlated with common associations with others sets of words. The more that two words share their near adjacent word sets within usage sentences and propositions in corpora of natural text, the closer they are embedded [23]. Such point clouds exhibit clustering of words commonly deployed in similar ways for similar purposes. For example, there might be a cluster of rude words about men, or a cluster of complementary words applicable to academics. Examining the clusters, we find some are highly related to such abstracted classification *labels* [24,25]. So such embeddings - while encoded at the micro word-to-word level - exhibit abstracted properties (labels) at the meso level (that of the clusters). The multi-scale manipulation and exploitation of such abstractions is potentially very useful. Suppose that a new word, of unknown meaning (to our brains), starts to be used. If the early usage appears to place it within a certain cluster, then we might at least apply the abstract property (the abstract label) and deduce that its meaning is some kind of complementary attribute that can be applied to an academic.

(b) Hypothesis generation

The generation of hypotheses, whether plausible or fallacious is an essential element of human reasoning. We can think of alternative hypotheses and decide we can accept one as a reason to act or take a decision. LLMs can already generate alternative hypotheticals, emulating a human.

Ask a question of ChatGPT [26] such as "*Why would I go to watch soccer?*" and ChatGPT produces a relevant list of alternative hypothetical and plausible reasons: The Atmosphere; The Drama (of the game) ; Community and Culture (of the fans). World-Class Talent (on display); Short and Sweet (no huge time commitment); and You Might Get Hooked.

Of course, there is little or no actual internal knowledge here of what the LLM is actually offering-up. The words and meanings and the various hypotheses make good sense though (to football fans like me), as if plucked from a super-large look-up table of all possible reasons to do anything. Whether generated on the fly or just looked-up, such generative AI emulates the human generation of hypotheses. It is not very imaginative though (see Section (e) below) . Yet it does a rather good job of organising previous knowledge (perhaps all of the old reasons that anybody might have ever given for going to watch soccer, captured within the conditioning corpora). But how might it be extended to imagine a new situation?

When you try *"Why and how might I teach a real dog to talk like a human?"* ChatGPT results in a good list of the benefits of (hypothetically) teaching a dog to talk, and then also some reasons why this is never going to be possible in practice.

How might such hypothesis generation be improved and extended in the future? If we ask a few mathematicians to prove a given conjecture, then they would likely each have a good go at assembling some useful concepts and definitions (some "mathematical furniture in the empty room") and some prior results that might be of help. They may attempt some separate lines of attack: perhaps assuming the converse and arguing for a contradiction. How might we emulate this? This type of task is and will become a large and active line of research. In this example it has the advantage that proofs are, by definition, explainable (the reasoning must be transparent). Methods such as trial and error (beloved in engineering) to construct logical arguments, or proof by induction, or proof by contradiction would all need to be in play. Should we allow access to Zorns's lemma? If necessary, yes.

(c) Refutation

Within many situations a human brain must respond urgently to one-off events (of the kind never before seen or experienced), there can be no previous knowledge of any similar decisions and their outcomes. However, hypothetical responses based on guesswork or anything else may be refuted by whatever is known in terms of situational constraints and expectations, or abstractions of standard and useful practices. This is one way that humans reason in such novel situations. There might be a whole battery of general constraints: infeasibility constraints, situational expectations and forecasts as constraints, abstracted experiences from other situations, and individual habits and heuristics as constraints. These may all refute some inappropriate (hypothesised) response options. Such AI reasoning, within some NLP logic-based setting, would be a form of novel *Explainable AI*, enabling decision support applications in situations where the present-day, classical, data-hungry, AI methods simply cannot get started.

This is a field of our present research, and will be addressed further in the near future, elsewhere.

(d) Counterfactual reasoning

Counterfactual reasoning refers to the human brain's ability to imagine the impacts and consequences of alternative possibilities, perhaps via "what if" scenarios, that differ from what actually happened or will happen. In doing so we learn from experiences (would things have turned out better under alternative actions or responses?), predict future outcomes (adopting alternative possibilities in any future similar or analogous situations), support hypothetical reasoning, develop policies for the future, and boost creativity. This a high level ability. It would likely require some abilities of abstraction, (alternative) hypothesis generation, imagination, and possibly refutation.

There is a growing interest in the use of counterfactuals within Explainable AI [27]; since explaining AI decisions may be regarded as a way to justify reliability and to establish trust. The recent review [28] provides some theoretical foundations and identifies some research objectives and challenges. It also presents some current methods based on *feature importance*, though those are really mechanistic post hoc tests and measures (of observations and consequences) rather than the adoption of fundamental reasoning abilities, as invoked above.

The use of counterfactual reasoning enabled by generative adversarial networks (GANs) is on the rise [29].

(e) Imagination

We have already discussed the generation of hypotheses and reasoned proofs (deductions), whether by guess, look-up table, chains of probability-based walks, or any other mechanism. An LLM such as ChatGPT [26] emulates the human ability to generate hypothetical responses (reasons) in response to prompts. This works whether the prompt is realistic, and possibly encountered within the conditioning corpora, or else is imaginary (and never encountered or even envisaged by any human in the past). On the other hand the LLM cannot as yet develop any novel and imaginative ideas (or situations) *ab initio*, for itself. It generates responses to prompts rather than creating any ideas and concepts for itself.

So how could we take this step without massively increasing the computational load? One way is via abstraction of the nature, concepts, or rules pertaining in similar or slightly parallel circumstances, followed by the reapplication of abstract experience within a novel specific circumstance. This is simply the use of abstraction followed by reapplication to focus a generative search. If we think of all of the *experience* that an AI has ever gained as an archive of specific propositions (its *facts of life*), then the archive might grow outwards just as in the manner of *novelty search* [30,31]. To do so via a focused development mechanism (in some manner) the propositions must themselves be embedded within a suitable metric space (that is the minimal structure able to furnish notions of *distance, novelty, and relative closeness*).

This type of idea needs to be addressed in future neuromorphic AI research, and there may be a number of issues to be faced down, not least the tendency of any search (via Lévy flights, for example [31]) in high dimensional spaces to leave many unexplored regions (gaps) behind the archive frontier in order to prioritise the shifting of that very frontier (the *leading edge* of experience). So some instances of such an imaginative AI may well leave gaps inside its archive within which other instances would forage, and vice versa. So be it. That is, of course, the same with different human brains. Once again, just as with any neuromorphic processing being mode-dependent (as discussed above), we aver that knowledge and performance of imaginative AIs would contain an individual, subjective, element.

4. Conclusion

We have set out a rather wide range of ideas for consideration by the wider research communities addressing next generation neuromorphic processing and AI.

A common feature of all such is the desire for more mathematical and logical (reasoning) precision to be deployed when comparing (much more directly) the performance of human brains with artificial processing and intelligence. It is expected that these ideas will inspire new paradigms and underpin neuromorphic and explainable artificial intelligence.

Acknowledgements. It is a pleasure to acknowledge the challenge, support, and advice received from Clive Bowman and Desmond Higham over many years; and that from Doug Saddy, Tamsin Lee, Zhivko Stoyanov, Ebrahim Patel, Chris Lester, Martin Brennan, and Ka Man (Ambrose) Yim – all collaborators within various stages of this research journey.
This research was partly funded by EPSRC grant number EP/Y007484/1 *Mathematical Foundations of Intelligence*.

References

1. Grindrod P, Lee TE. 2017 On Strongly Connected Networks with excitable-refractory dynamics and delayed coupling, Roy. Soc. Open Sci. (2017) 4(4):160912, doi:10.1098/rsos.160912.
2. DeBole MV, et al. 2019 TrueNorth: Accelerating from Zero to 64 Million Neurons in 10 Years. Computer 52(5):20–9.

3. Davies M, et al. 2018 Loihi, A Neuromorphic Manycore Processor with On-Chip Learning, in IEEE Micro, vol. 38, no. 1, pp. 82-99, Jan/Feb 2018, doi: 10.1109/MM.2018.112130359.
4. Grindrod P, Lester C. 2021 Cortex-like complex systems: what occurs within?. Front. Appl. Math. Stat. 7:627236. doi: 10.3389/fams.2021.627236.
5. Grindrod P. 2022 There is something that it is like to be me, preprint, ReserachGate, https://tinyurl.com/5n7s6h3a.
6. Grindrod P. 2028 On human consciousness: A mathematical perspective. Network Neuroscience 2018; 2 (1): 23-40. https://doi.org/10.1162/NETN_a_00030.
7. Grindrod P. 2024 Wishful thinking about consciousness. IgMin Res. May 02, 2024; 2(5): 302-308. https://www.igminresearch.com/articles/pdf/igmin180.pdf.
8. Grindrod P, Brennan M. 2023 Cognition and Consciousness Entwined, Brain Sci. 2023, 13(6), 872; https://doi.org/10.3390/brainsci13060872.
9. Kahneman D. 2011 Thinking, Fast and Slow. Farrar, Straus and Giroux.
10. Ariely, D. 2008 Predictably irrational: The Hidden Forces That Shape Our Decisions. Harper Collins.
11. Brennan M, Grindrod P. 2023 Generalised Kuramoto models with time-delayed phase-resetting for k-dimensional clocks, Brain Multiphysics, Vol 4, https://doi.org/10.1016/j.brain.2023.100070.
12. Maxwell N. 2017 Karl Popper, Science and Enlightenment, UCL Press, https://doi.org/10.2307/j.ctt1vxm8p6.
13. Mcleod S. 2023 Karl Popper: Theory Of Falsification, Simply Psychology Psychology Relationships, Research Methodology, https://www.simplypsychology.org/karl-popper.html Updated on July 31, 2023.
14. Lambert N, Chen YN, Cheng YC, Li CM, Chen GY, Nori F. 2013 Quantum biology. Nature Phys 9, 10–18, https://doi.org/10.1038/nphys2474.
15. McFadden J, Al-Khalili J. 2018 The origins of quantum biology, Proc. R. Soc. A.4742018067420180674, http://doi.org/10.1098/rspa.2018.0674.
16. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. 2017 Attention is all you need. Advances in neural information processing systems, 30.
17. Kang S, Kim J, Kim J, Hwang SJ. 2025 See What You Are Told: Visual Attention Sink in Large Multimodal Models, arXiv 2503.03321, https://arxiv.org/abs/2503.03321.
18. Mullin E. 2021 How big science failed to unlock the mysteries of the human brain, MIT Tech Review, https://www.technologyreview.com/2021/08/25/1032133/big-science-human-brain-failure/.
19. Grindrod P, Yim KM. 2025 Dynamical Systems on Generalised Klein Bottles. Entropy 27, 119, https://doi.org/10.3390/e27020119.
20. Bryant, MSFP. 2025, Tackling high-stakes problems with AI creativity, not data, November 06, 2025, https://preseednow.com/p/astut.
21. Suwandi, RC. 2025, The Science of Intelligent Exploration. Posterior Update. https://richardcsuwandi.github.io/blog/2025/exploration-in-ai/.
22. Hassabis, D. 2025, Lex Fridman Podcast No. 485 Transcript. https://lexfridman.com/demis-hassabis-2-transcript/#chapter7_ai_research.
23. Lenci A. 2018 Distributional models of word meaning, Annual Review of Linguistics, 4(1), pp. 151–171. doi:10.1146/annurev-linguistics-030514-125254.
24. Rogers A, Kovaleva O, Rumshisky A. 2020 A Primer in BERTology: What we know about how BERT works, arXiv 2002.12327, https://arxiv.org/abs/2002.12327.
25. Grindrod J. 2015 Word meanings in transformer language models, 2025, in preparation.
26. OpenAI. 2014 ChatGPT. Accessed [accessed 31/05/25].
27. Byrne RMJ. 2029 Counterfactuals in Explainable Artificial Intelligence (XAI): Evidence from Human Reasoning, Proc. Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, 6276–6282, https://doi.org/10.24963/ijcai.2019/876.
28. Ortigossa ES, Gonçalves T, Nonato LG. 2024 Explainable Artificial Intelligence (XAI) — From Theory to Methods and Applications, IEEE Access, 12, 80799-80846, doi: 10.1109/ACCESS.2024.3409843.
29. Del Ser J, Barredo-Arrieta A, Díaz-Rodríguez N, Herrera F, Saranti A, Holzinger A. 2024 On generating trustworthy counterfactual explanations, Information Sciences, 655, https://www.sciencedirect.com/science/article/pii/S0020025523014834.
30. J. Lehman J, Stanley KO. 2011 Abandoning objectives: evolution through the search for novelty alone, in Evolutionary Computation, 19, no. 2, pp. 189-223, https://ieeexplore.ieee.org/document/6793380

31. Bowman CE, Grindrod P. 2023 Desperately searching for something, Communications in Nonlinear Science and Numerical Simulation, 125, https://www.sciencedirect.com/science/article/pii/S1007570423002575/