Temporal continuity as an architectural constraint on machine consciousness

Benjamin Gaskin

University of Sydney

An overview of the argument

1. Consciousness and temporality
2. Three types of temporal continuity
3. Zeno's new paradox of the arrow
4. Implications for machine consciousness
5. Continuity in current architectures

Consciousness and temporality

- What we don't mean—
  - Memory (e.g., amnesic cases such as Clive Wearing)
  - Temporal understanding (e.g., proper use of tense)
  - Temporal reasoning (e.g., planning with respect to time)
  - Temporal sensitivity (e.g., spiking neural networks)
  - Temporal integration (e.g., the perception of change)
- By temporal flow, we mean **phenomenological continuity**
  - Strictly speaking, **temporal continuity of the phenomenal container**
  - This is the flowing temporality of our ordinary first-person perspective

Three types of temporal continuity

- **Perceptual continuity:** observer projects continuity
  - Images presented at 24 frames per second appear as smooth motion
- **Behavioural continuity:** coherent state-transition rules
  - The necessary structure of the series presented at 24 frames per second
- **Intrinsic continuity:** substrate unfolds as continuous flow
  - The system unfolds of its own accord, without update steps
  - This is present in a physical object falling or electricity flowing

The increment-rule distinction

- **Increment—**
  - How does n differ from n+1?
- **Rule—**
  - How does n align with n+1?
- If we are plotting a continuous function, for instance—
  - We start with some point, where we **apply the function**:
    - The result can then be **plotted as a point** on the graph
    - We then **add an increment and apply the function again**
    - This process **unfolds by rule-application and iteration**
  - **From these discrete steps, a continuous curve emerges**

Zeno's paradox of the arrow

- Aristotle, from the *Physics*: **"if everything when it occupies an equal space is at rest, and if that which is in locomotion is always in a now, the flying arrow is therefore motionless."**
  - Aristotle specifies, of this: "that the flying arrow is at rest, which result **follows from the assumption that time is composed of moments.**"
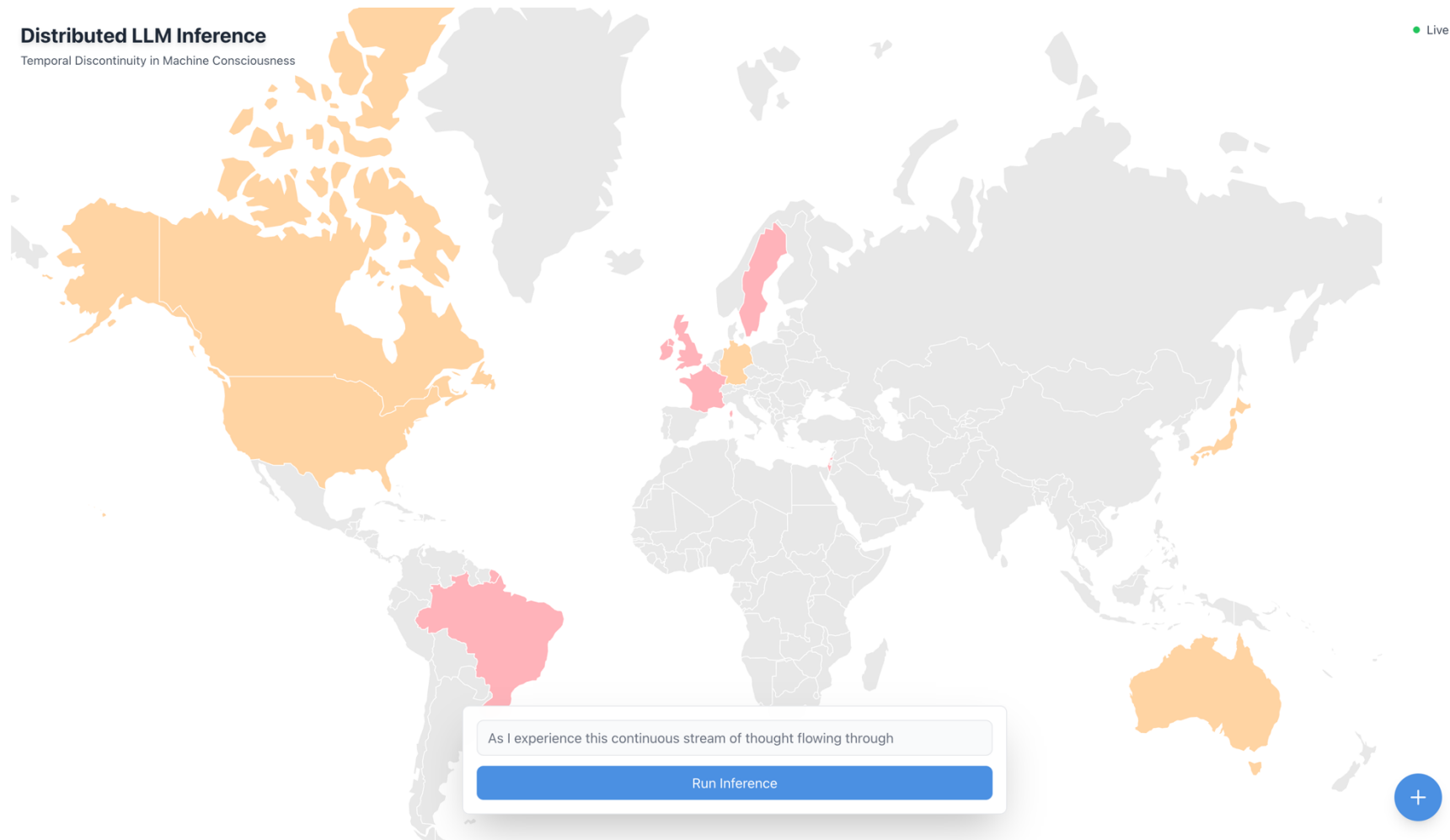
Zeno's new paradox of the arrow

- Suppose instead that **the arrow is portrayed upon a screen**
- We know that **this proceeds by discrete frames**, as noted earlier
- If this **exceeds 24 fps**, however, it **appears as smooth motion**
- There are **at least two types of continuity** here—
  - Perceptual continuity, in the appearance of smooth motion
  - Behavioural continuity, in the sequencing and relation of frames
  - **What about intrinsic continuity?**
- We might say the screen entails a continuity of mechanism, there is **a continuity of material constituents and electrical flow**

Zeno's paradox of the manifold arrow

- Suppose now, with a single screen, we have **many projectors**
- The projectors are all **set with a delay**: the chain is **started by some sign** and **each flashes a single frame** upon the screen
- There are **still at least two types of continuity**—
    - Perceptual continuity, in the appearance of smooth motion
    - Behavioural continuity, in the sequencing and relation of frames
    - **What about intrinsic continuity?**
- It is now more difficult to argue that there is any meaningful sort of intrinsic continuity, **unless we neglect spatiotemporal distance**

The paradox applied to LLM inference

Distributed LLM Inference
Temporal Discontinuity in Machine Consciousness

● Live

As I experience this continuous stream of thought flowing through

Run Inference

What then for machine consciousness?

- **If we accept behavioural continuity** as an adequate substrate for consciousness, then **we are caught in a fork—**
  - Either consciousness **exists over arbitrary spatiotemporal distances**;
    - This runs into the boundary problem: but then why is it bounded thus?
  - **Or consciousness requires spatiotemporally-localised continuity**
    - This means accepting limitations on the appropriate architectures.
- We thus see **analog systems with a continuous dynamical core** as being **more probable candidates** for machine consciousness
  - Neuromorphic computing
 - Biological computing

Continuity in ANNs
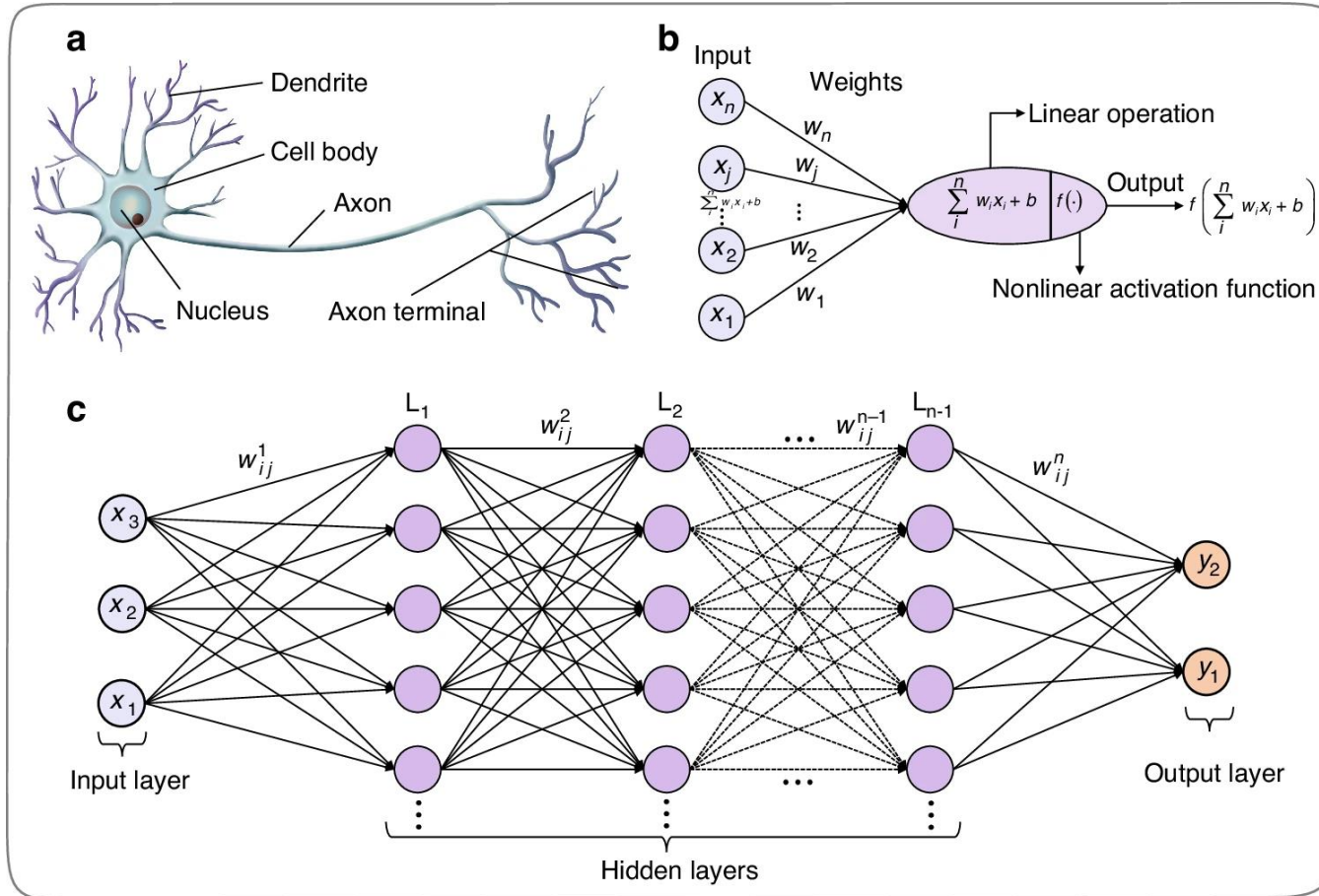
- These proceed **layer-wise** with a **sequence of mathematical operations**, either in series or parallelised
  - When they are run in real-time, we simply iterate this process
- Behavioural continuity, perceptual continuity, **no intrinsic continuity**

Continuity in LLMs
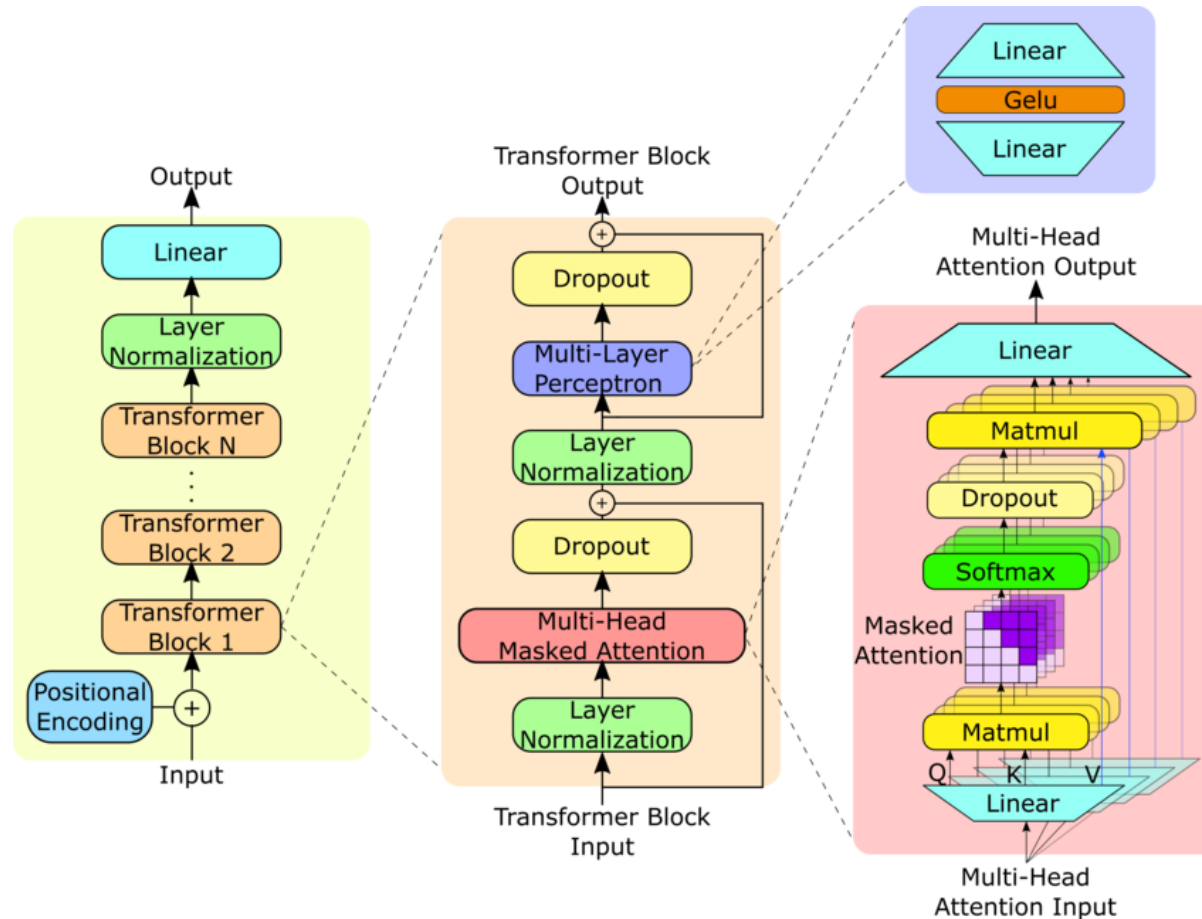
- Follow the **same principle as ANNs**, differing only in the set of operations
  - The KV cache is a matter of efficiency, not continuity
- These show behavioural continuity and can carry perceptual continuity, but **lack intrinsic continuity**

Continuity in SNNs
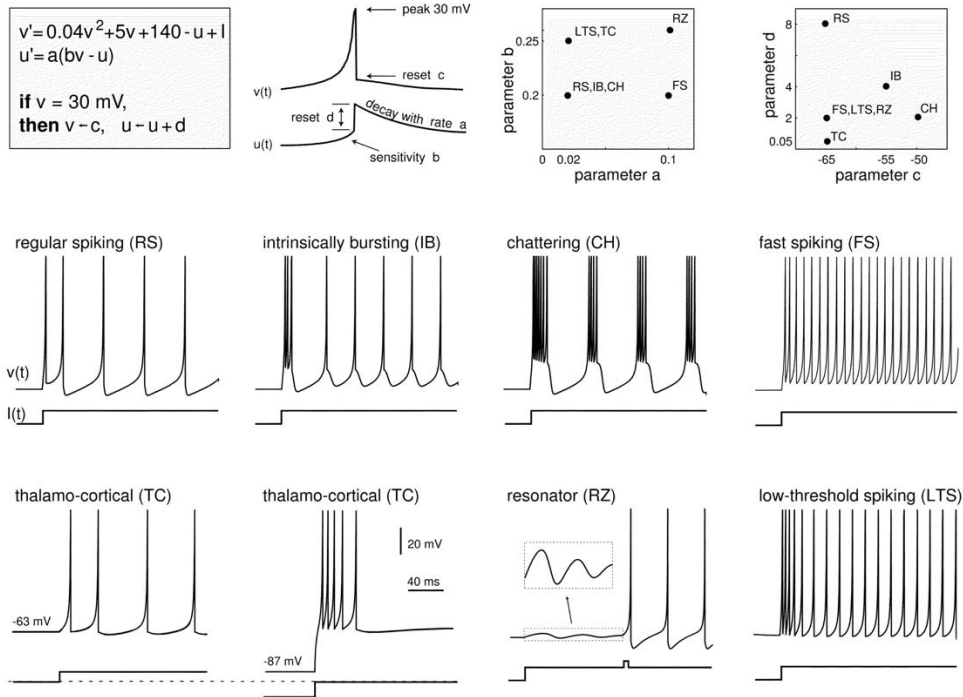
- There are two options here—
  - Event-driven
  - Time-stepped
- Both follow an **increment and rule**, each differing only at the level where this is applied—
  - **CPU clock;** or
  - **Timestep frequency**

Continuity in neuromorphic computing
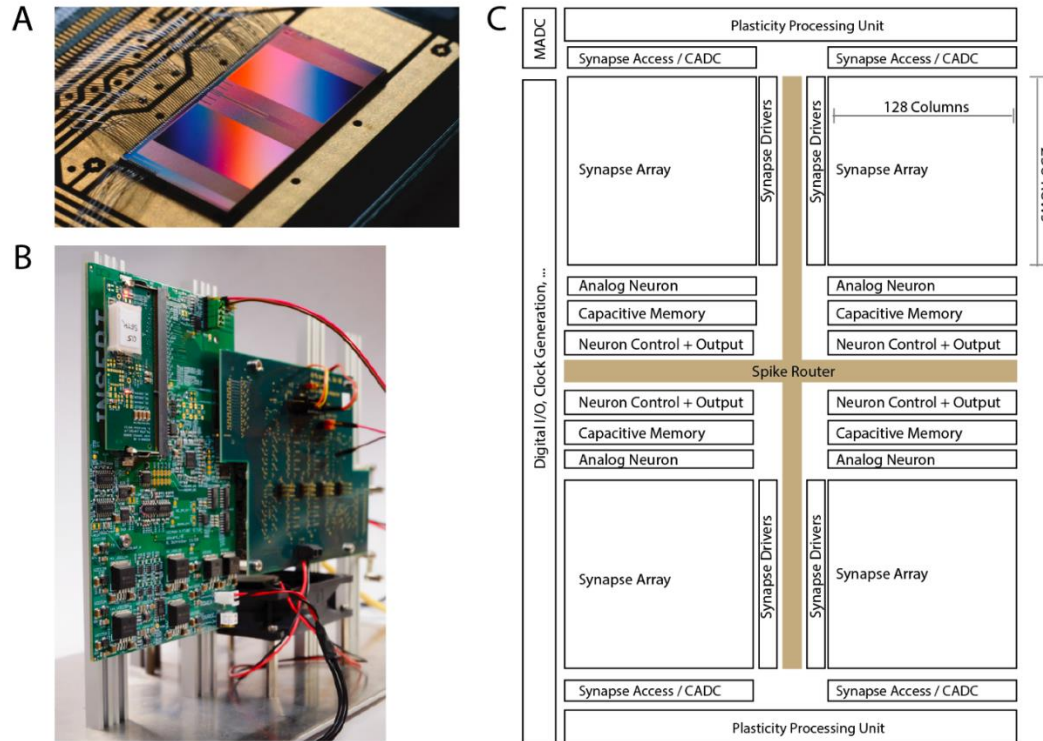
- Neuromorphic computing in terms of hardware comprises a **range of architectures—**
  - IBM's **TrueNorth** and Intel's **Loihi** are **digital neuromorphic** chips
  - **BrainScaleS-2** (left) uses **analog neurons** for membrane potential
- **What is the relevant core** for continuous consciousness?
  - The binding problem
  - Neuronal vs. reticular theories

Continuity in biological computing
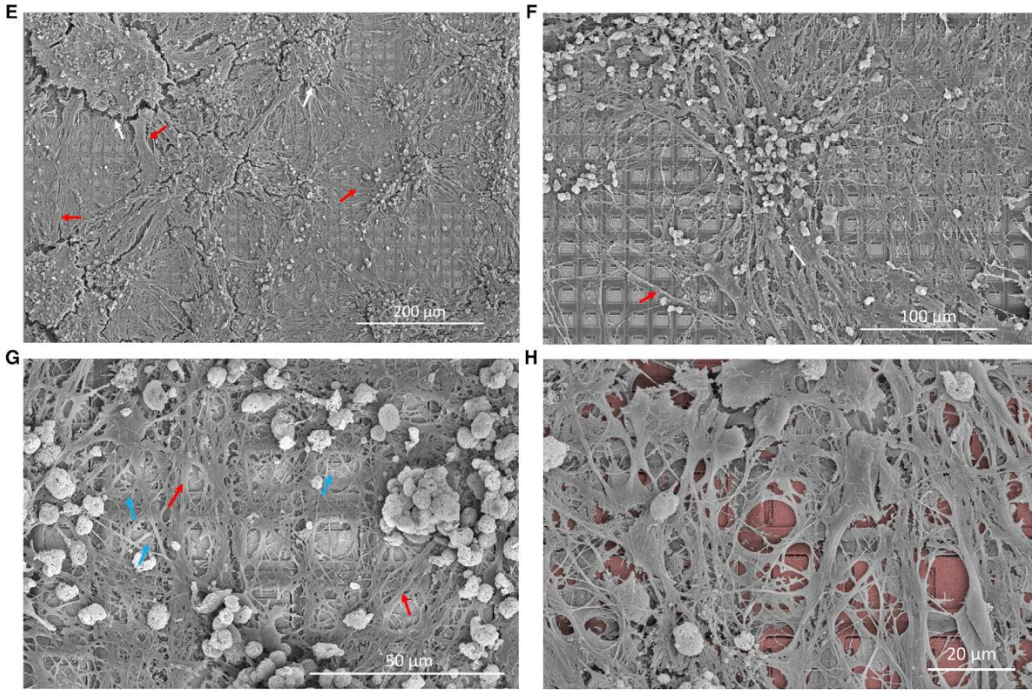
- iPSCs are in principle able to **express any component of the human system**, our existence proof for consciousness

- Neural organoids are **coupled with micro-electrode arrays** to form computing systems
  - Does the digital MEA matter?

- Current systems self-assemble **networks of cortical tissue**
  - Is continuity at the tissue-level?

End.

bgas0204@uni.sydney.edu.au

Summary

- **To the extent that consciousness is temporally continuous,** this is a **necessary (but not sufficient)** attribute of any substrate
  - Perceptual continuity, behavioural continuity, **intrinsic continuity**
  - The **separation of increment and rule** in behavioural vs. intrinsic

- Our thought experiments suggest that **if spatial contiguity is important, behavioural continuity will be inadequate**

- This leads us to conclude that **analog architectures are more promising as candidates for machine consciousness**
  - **Substrates with a continuous dynamical core**

- We **do not directly contradict computational functionalism**
  - The **mode of computation** may be as important as its content

- There **remain many open questions—**
  - **Which analog architectures** are suitable?
  - **What is the relevant core** within a system?
    - Neuron-, tissue-, organ-level, etc.

www.gregadunn.com
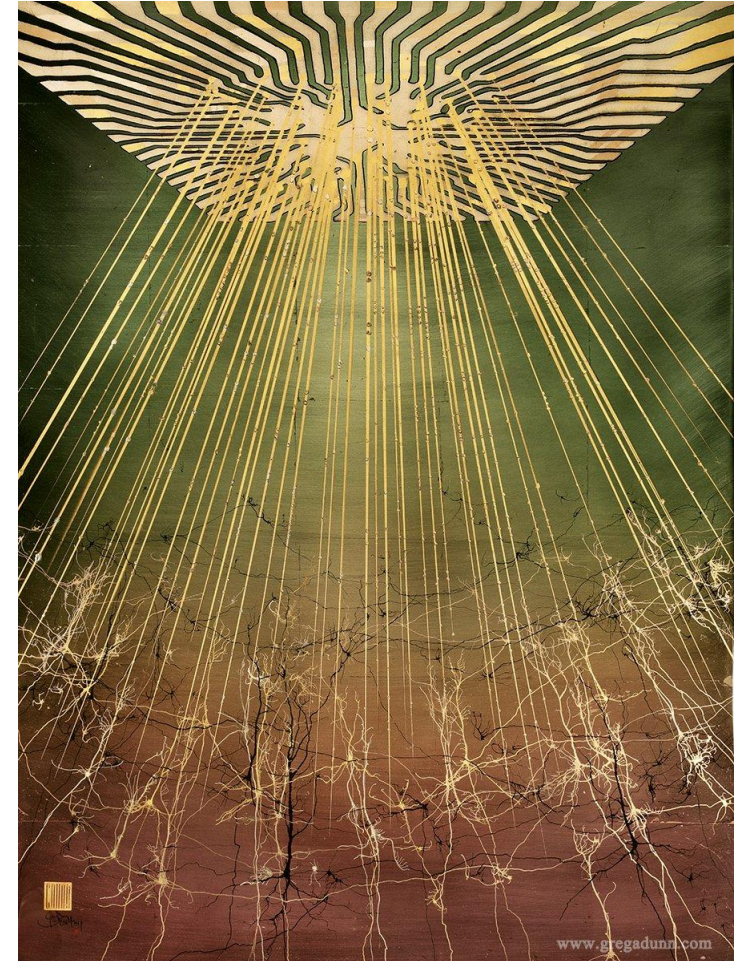
What about the binding problem?

- **Functional**, stimulus-related—
  - How are representations of objects related to their properties?

- **Phenomenal**, consciousness-related—
  - How are representations united in phenomenal experience?
    - Damasio (1989): "The experience of reality … is not parcellated at all. The normal experience we have of entities and events is coherent and 'in-register,' both spatially and temporally. Features are bound in entities, and entities are bound in events. How the brain achieves such a remarkable integration starting with the fragments that it has to work with is a critical question. I call it the binding problem."

- Our argument relates to the phenomenal binding problem—
  - The question of continuity might be read as a **temporal binding problem**
  - **The solution to the binding problem defines what we call the dynamical core**

What about the KV cache in LLMs?

- This refers to the **Keys (K)** and **Values (V)** in Transformer blocks
  - These are **aligned with the Queries (Q)** for each token-step of inference
    - The result of this **determines the attention allocated to context tokens**
  - These **can be cached across token-steps and inference-steps**
- The KV cache is a matter of **efficiency rather than continuity**
  - Given the same sequence, it is **readily recalculated** at any point
 - Whether we store or calculate has **no meaningful behavioural effect**

What about action potentials in the brain?

- Action potentials in the brain **fire in an all-or-nothing fashion**
  - These can thus be considered as **discrete, binary events**
  - It is not clear what they mean for the brain, however—
    - The **membrane potentials** underlying these, however, are continuous
    - The **ionic concentrations** below that are comprised of discrete units
    - There is **also continuity above this level** (e.g., ephaptic coupling)
- This returns us to the question: **what is the relevant core?**
  - Ultimately this **may prove to be an empirical question**

What about hybrid architectures?

- Take **Watanabe's hemispheric test** for machine consciousness as an example
  - This involves **replacing one hemisphere with an artificial architecture**
  - Does the individual then experience the visual content corresponding to this hemisphere?
- Our question is **whether the artificial hemisphere sustains consciousness**
  - We would argue that this may only show it **integrates into an existing continuous core**
  - **If separated, it would not sustain experience**

A Test for Machine Consciousness
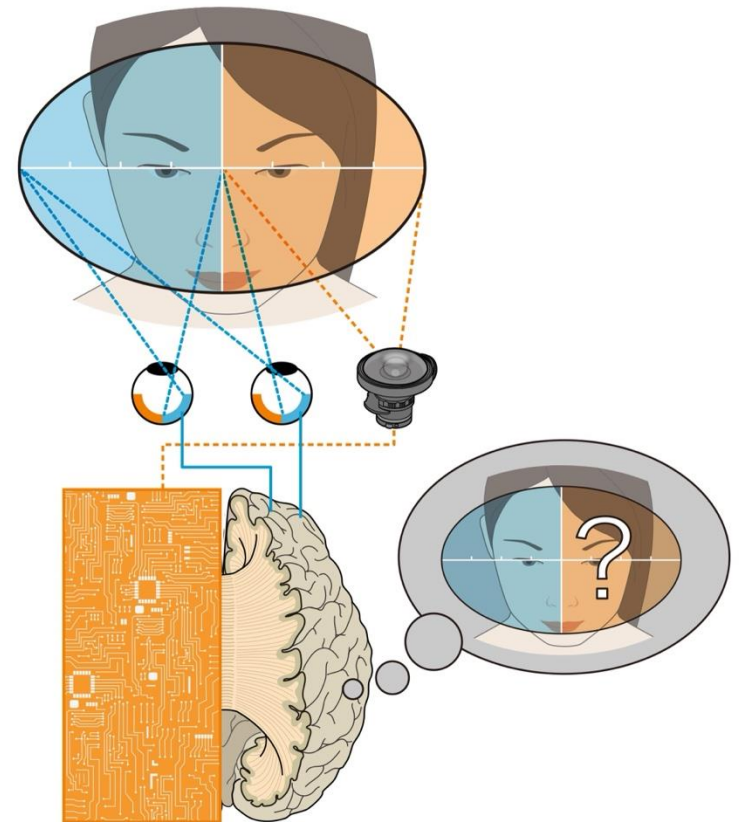
What do you mean by a continuous core?

- There must be **some substrate corresponding to this continuity**
  - This must **be situated at the appropriate level**, which we call the core
- We can frame this **in terms of the binding problem**: **any solution must provide a substrate characterised by intrinsic continuity**
  - The core is **that element of a conscious system which serves as the continuous substrate underpinning its dynamic phenomenal binding**

What might be the relevant dynamical core?

- **Electromagnetic fields** are one promising candidate here
- Lane and Rodriguez (2025) argue that **mitochondrial activity in respiratory complex I generates cellular EMFs**, which **gives the organism an integrated, real-time reading of their metabolic state**
  - Volatile general anaesthetics **disrupt mitochondrial function**; specifically spin polarization at respiratory complex I
  - Anaesthetics **affect protists as well as animals** (simple and complex)
  - Mitochondria are **derived from bacteria** via endosymbiosis
- They **may not be conscious in any recognisable sense**
  - Pockett (2017) EMF sensory consciousness to specific neurophysiological structures: **apical dendrites of neocortical pyramidal cells**
  - These **carry feedback to sensory areas** ~80–100ms post-stimulus
  - **Complex consciousness may rest upon specific EMF patterns**, shaped by the **structure and activity of their field-generating mechanisms**

What of the persistent interlocutor illusion?

AI Consciousness: A Centrist Manifesto (Birch, 2025)

- Birch has recently made a similar argument concerning temporality, as opposing **"rampant misattributions of human-like consciousness"**

- He identifies **the persistent interlocutor illusion**, against which he draws on similar technical details to our argument (e.g., statefulness)

- In contrast, however, Birch's **focus is identity**: that there be the "right kind of psychological continuity … [for] personal identity"
  - As in Parfit's identity theory, for instance, where "Relation R" is necessary

- Nevertheless, we align in holding **extant LLMs create an illusory interlocutor through their emulation of linguistic patterns**
  - In our words, this **perceptual continuity results from a behavioural continuity**
  - This continuity is **rooted in the state (token) transition 'rules' acquired during learning**—pre-training, instruction tuning, reinforcement learning, etc.

What of the temporal gap?

How to build conscious machines (Bennett, 2025)

- Bennett identifies something similar, namely 'the Temporal Gap'
  - As ours, **not substrate dependence but an architectural constraint**
- He **identifies two options for conscious machines—**
  - Either it must be **realised by an environmental state at a point in time;**
  - Or it can be **"smeared across time"** (as with CPU or SIMD computation)
- There remains **uncertainty**, however, namely **'the Temporal Gap'**
- Where Bennett emphasises **state**, we focus on **the relationship between states**—or rather, **the specific nature of continuity**

What of sensory sampling's incompleteness?

Frame by frame? A contrasting research framework for time experience (Rodriguez et al., 2023)

- Starting at the **apparent continuity of phenomenal temporality**, their argument takes as its point of departure two tensions—
  - First, with the **discrete state-space relations used in modelling**
  - Second, with the **discrete and usually asynchronous nature of sensory samplings in organisms**, resulting in a **"sparse graph-like structure"**
- Given this, **how can we explain the fluidity of perception?**
  - This work is **quite close to our question, differing mostly in emphasis**
- The very **incompleteness of sensory samplings** is taken to **compel their own completion, thereby constituting an integrated whole**
  - Where **this concerns the relationship between (and within) states**, we are concerned instead with **how intrinsic continuity exceeds states**