Conceptual similarity between attention heads and islets of inter-LINKed spine heads

Kunjumon Vadakkan

I will start explaining it as a story to a primary school student. Later, we will go in deeply. Imagine that we are walking through a forest full of mango trees. There are some rules here. 1) Each tree produces only one particular-colored mango on them. 2) Branches of one mango tree does not overlap with each other so that we will never see two mangoes of the same color together. 3) Each mango has a stalk (pedicel). There is only one mango on one stalk. 4) The distance between two mangoes on a tree branch is more than the size of a mango. 5) Branches of each mango tree go between the branches of many mango trees around them. So, we see different colored mangoes hanging down together. It is difficult to know which mango belongs to which tree. But when we trace each colored mango, we will reach only one tree. These are funny mango trees. Aren't they?

Also imagine another forest of mango trees hanging up in the air. Their roots are hanging in the air. Each one of the root tips comes down and touches one mango each that we are seeing.

Our brain is full of different layers of mango trees with mangoes on them and roots from the tree above them coming and touching on the mangoes of the lower layer. Now, we were thinking that memories are stored at the junction where the rootlet from the tree above touches mango. These junctions are called synapses and the change at this junction is called synaptic plasticity.

But I found a problem. To explain all the features of the brain, this type of connection alone is not sufficient. In addition to the above connection, there should be another connection. Remember, I told that since tree branches overlap heavily, we can see mangoes of different colors (from different trees) touching each other. When we learn something, there must be some changes occurring at the location where the skins of mangoes of different colors (that belong to different trees) touch each other.

When we do electron microscopy of the brain sections, we can see these mango-like structures (dendritic spines) from different tree-like structures (neurons) touching each other. But we have not yet proven that there will be some changes occurring between the skins of these mango-like structures when we learn. When artificial intelligence (AI) scientists tried to use this concept of mango-like structures (spines) interacting with each other in mathematical terms (using linear algebra), they started getting good AI. By keeping more mangoes to touch each other in a group, and by keeping them touching firmly for more time, we are getting better AI. Now that we are getting good AI, we can assume that the brain must also be working in the same manner – like mangoes from different tree touch and interact, spines of different neurons touch and interact.

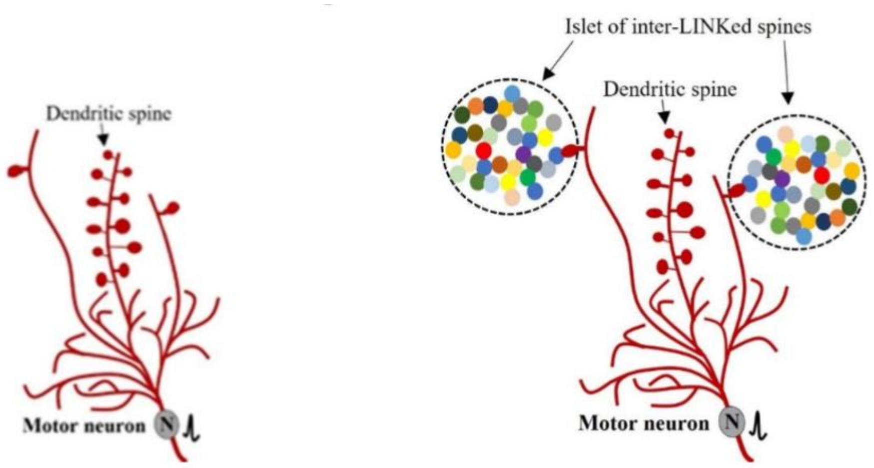
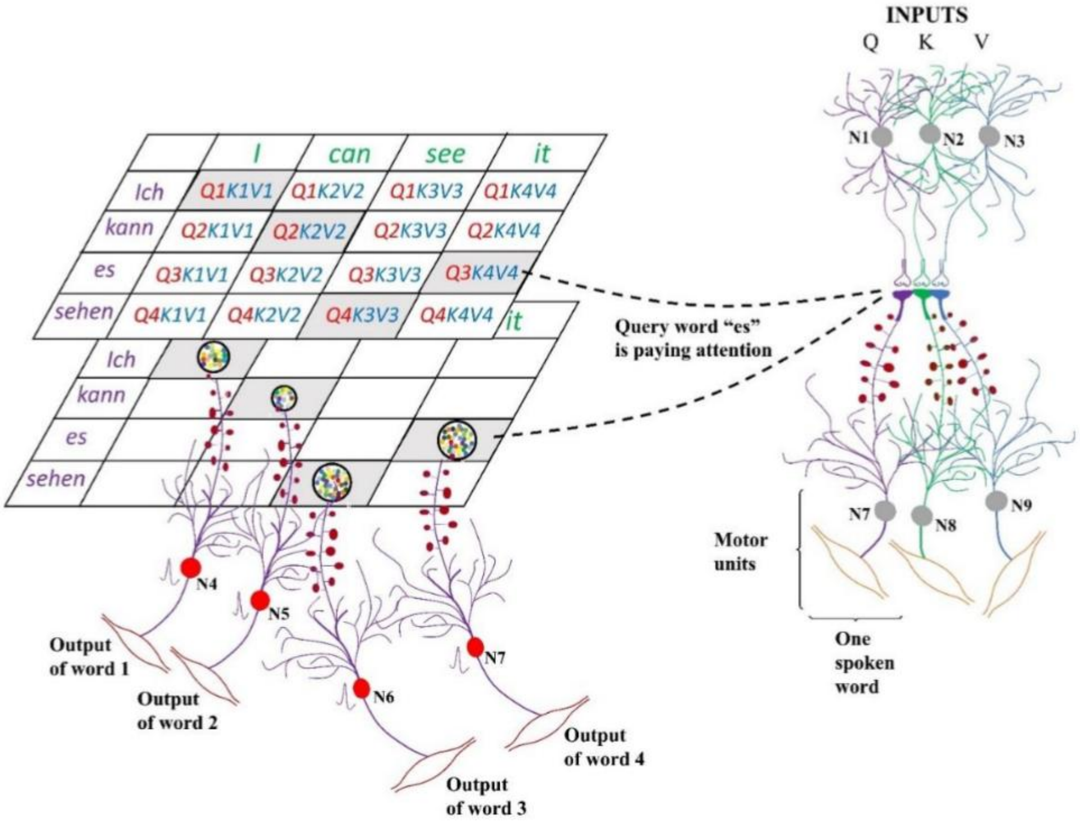


Figure 1. Left side: A mango tree (neuron) with red mangoes. Most branches are removed. 3 upgoing branches –2 with one mango each on either side and central branch with 11 spines on them. **Right side**: In the forest, the single mangoes were touching many mangoes of different colors from different trees. It is a photo we took before removing this tree.



	I	can	see	it
Ich	Q1K1V1	Q1K2V2	Q1K3V3	Q1K4V4
kann	Q2K1V1	Q2K2V2	Q2K3V3	Q2K4V4
es	Q3K1V1	Q3K2V2	Q3K3V3	Q3K4V4
sehen	Q4K1V1	Q4K2V2	Q4K3V3	Q4K4V4

Figure 2. Left side: A 4x4 matrix table is shown. The weight of the entries (weight of their relationships in a multi-domain space) determine the corresponding words between two languages during their translation. See 4 entries have corresponding 4 groups of mangoes. The entries determine the relationships between words = relation between touching mangoes. **Right side**: Proposed interactions between dendritic spines (postsynaptic terminals) of different neurons.

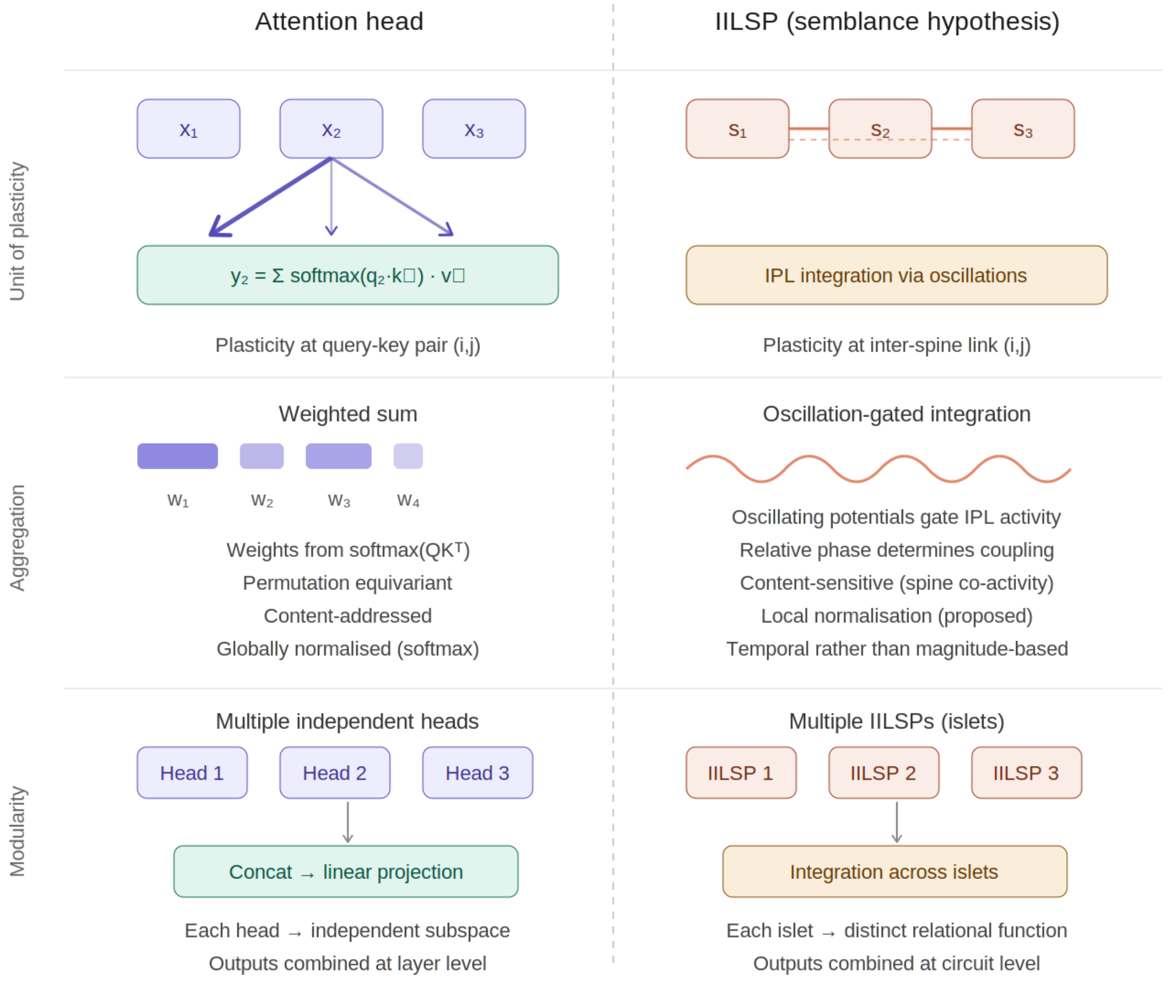


Figure 3. **The hidden logic between attention heads and islets of inter-LINKed spine heads.** This figure compares an AI mechanism called an attention head (used in language models) with a proposed brain mechanism called an IILSPs (Islet of Inter-Linked Spines). The both figure out which pieces of information are relevant to each other and combine them intelligently. Three rows highlight the key similarities. **Row 1**. Where does learning happen? Both systems learn at the relationship between pairs of elements, not at individual connections. In the attention head, a query token (x_2) weighs its relevance to every other token - thick lines mean strong relevance, thin lines mean weak. In the IILSP, learning is proposed to occur at special junctions (IPLs) between the branch-tips of different brain cells, shown as links between spines s_1, s_2, s_3. **Row 2**. How is information combined? The attention head blends information using a weighted sum - each input contributes in proportion to its relevance (shown as bars of different lengths), calculated from the content of the information itself. The IILSP is proposed to blend information through brain-wave synchrony - spines whose electrical rhythms align strengthen their connection. Both are flexible and content-sensitive; one uses mathematics, the other uses timing. **Row 3**. Do multiple units work in parallel? Yes, in both systems. A transformer runs many attention heads simultaneously, each capturing a different kind of relationship, then merges their

outputs. The brain is proposed to run many IILSPs in parallel, each capturing different relational patterns, with outputs pooled at the circuit level.

Overall message: This figure illustrates that a core computational idea, which is, **use pairwise relationships between elements, weight them by relevance, and run multiple parallel versions**. This has independently appeared in the mathematics of LLMs and the proposed IILSPs mechanism by the semblance hypothesis. This does not mean the brain is "doing the same thing as LLMs" but it does suggest that this particular style of computation may be a deep and powerful solution to the problem of integrating complex information, one that nature and engineers have arrived at from different starting points.

Note: Mangoes from different trees (easily distinguishable by their different colors) build relationships with each other even without their trees knowing about it. A tree comes to know about it only when there is a group action by several mangoes - particularly when many groups are in action and many of these groups contain mangoes belonging to that particular tree. In LLMs, words are dispersed within a multidimensional space. Similar words are located close to each other in this space. The relationships between words at different locations are denoted by lines connecting them. When too many lines (sequence, meaning, context, noun-verb [subject-predicate] usage) point to a specific word, that word is chosen as the next best word for speech output. I hope you can now see the conceptual similarities.

The above similarity suggests that these relationships are trainable and become trained through learning. The output is mostly predetermined by the extent of the training. What is unique to the brain is missing in the above model - namely, the first-person inner sensation or experience. The semblance hypothesis examined whether the connections to the mangoes of one layer of trees, to which the roots of trees from the layer above connect (synapses), possess any peculiar property that can endow mango-to-mango connections with the ability to generate a unitary mechanism for first-person properties. **Yes, there appears to be a surprising mechanism.** This forms the basis of the semblance hypothesis.