

Proceeding Paper

Perceptions, Imagery, Memory, and Consciousness in Man and Machine [†]

Magnus Johnsson ^{1,2}

¹ Department of Computer Science and Media Technology, Malmö University, 205 06 Malmö, Sweden; magnus.johnsson@mau.se or magnus@magnusjohnsson.se

² Magnus Johnsson AI Research AB, 243 34 Höör, Sweden

[†] Presented at Philosophy and Computing Conference, IS4SI Summit 2021, online, 12–19 September 2021.

Abstract: I propose a number of principles that I believe are substantial for various faculties of the mammalian brain, such as perception, expectations, imagery, and memory. The same principles are also of interest when designing an artificial but biologically inspired cognitive architecture. Moreover, I discuss how the same principles may lie behind the ability to represent new concepts and to imagine fictitious and impossible objects, while also giving us reasons to believe that there are limits to our imagination and to what it is possible for us to think about. Some ideas regarding how these principles could be relevant for an autonomous agent to become functionally conscious are discussed as well.

Keywords: perception; memory; expectations; imagery; consciousness; self-organization; feature maps; associative learning; multimodal integration; cognitive architecture

1. Introduction

The mammalian brain's astonishing abilities raise the question of the principles by which it is organized. Since it is evolved rather than designed, such principles should be simple rather than complicated. This seems to be contradicted by the brain's remarkably advanced abilities. I believe that this contradiction is false and that the advanced capabilities of the brain are indeed based on fairly simple principles, but which are reused over and over again at different levels of complexity.

Below I explain and motivate why I think rather simple principles regarding self-organization; internal self-supervision; and laterally, hierarchically, and recurrently connected topology-preserving feature representations that reflect the probability distributions of their inputs are important in the mammalian brain, as well as how that insight can be used artificially.

Simple principles, I believe, employed over and over again by nature at various levels of complexity are behind astonishingly complex abilities, such as perceptions, imagery, and functional consciousness in the mammalian brain. The same principles can explain why we sometimes tend to perceive our expectations rather than what is really out there; how we construct and fill in the gaps in our perceptions within and between various sensory modalities when sensory inputs are limited; and multimodal integration. How various memory systems, imagery, and perception fit together can be explained by the same principles.

I will also discuss how the corresponding faculties could be implemented in an artificial biologically inspired cognitive architecture by employing the same principles in the same way as has been done through evolution by nature. By looking at how the mammalian brain is structured and by identifying its crucial components and how these are interconnected, knowledge can be obtained that, together with the identified principles, enables a systems-level approach to modeling perception as well as the integration of various sensory modalities, memory, imagery, the generation of an inner world, and



Citation: Johnsson, M. Perceptions, Imagery, Memory, and Consciousness in Man and Machine. *Proceedings* **2022**, *81*, 132. <https://doi.org/10.3390/proceedings2022081132>

Academic Editor: Peter Boltuc

Published: 29 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

functional consciousness in a biologically inspired cognitive architecture modeled on the mammalian brain.

2. Topologically Ordered Feature Representations as Building Blocks

Topology-preserving feature representations, i.e., self-organizing topographical representations of features that reflect the probability distributions of their inputs, are inherent parts of the sensory modalities as well as other systems in the mammalian brain. Such representations can be found in, for example, somatosensory, visual [1], and auditory areas, in neuron nuclei, and in the cerebellum. Due to the topology-preserving property of the representations, similar features are represented close to each other. That they also reflect the probability distributions means that the size of the representational areas depends on the importance and frequency of inputs. There are topology-preserving feature representations of various levels of complexity, from somatosensory areas containing somatotopic representations of the skin surface to frequency-preserving tonotopic maps [2–4] to color maps [5] in visual areas, as well as representations of more complex features.

Such topology-preserving feature representations could be modeled at an appropriate abstraction level and used as building blocks in an artificial cognitive architecture by self-organizing artificial neural networks, known as self-organizing maps (SOMs) [6], and variants, such as the associative SOM (A-SOM) [7]. These ANNs mimic crucial properties of the topology-preserving representations found in the mammalian brain, such as self-organization into topologically ordered feature representations that reflect the probability distributions of their inputs. SOMs can self-organize into, e.g., color maps, semantic maps [8], and posture maps [9].

3. Representational Hierarchies

Various sensory modalities, as well as other systems, in the mammalian brain seem to be ordered in hierarchies of increasingly more complex/abstract feature representations. More complex features rely on inputs from representations of less complex features, though there is probably also a flow of information in the opposite direction. Hence, in the visual modality we find the dorsal (where) and ventral (what) streams. Both streams originate in the primary visual areas in the occipital lobe, whereas the former ends up in the parietal lobe and the latter in the temporal lobe. For example, in the ventral stream there are ordered feature representations of contour directions in visual area one (V1), of shapes in visual area two (V2), of objects in visual area four (V4), and of complex facial features in the inferior temporal (IT) area.

In the auditory modality there seems to be similar (what, how, and where) streams. One could speculate whether the auditory and visual where streams, both ending up in the parietal areas, coincide to some extent.

This principle of a hierarchy of increasingly more complex feature representations has been exploited artificially in deep neural networks [10], but also in the use of hierarchical SOMs applied to various sensory modalities, e.g., [9,11].

4. Perception, Imagery, Memory, and Consciousness

The brain supplements perceptions when the sensory inputs are not complete. This is evident from various visual illusions, e.g., the Kanizsa Triangle [12], where the contours of a triangle can be perceived even though they are actually not there. Moreover, when our eyes scan the scenery before us, they are doing so by semirandom eye movements known as saccades, directing the movements toward particularly conspicuous and, in some sense, interesting features. Supposedly we carry out similar semirandom movement with our hands and fingers to gain particularly interesting and useful tactile sensory inputs when we, e.g., ransack our pockets for a particular key or grope about to find the doorknob in the dark. When we perceive, our brains seem to fill in the gaps of sensory inputs with expectations, from memory, of what is likely to be there.

The ability to influence and simulate perceptual activity in certain brain areas due to activity in other brain areas is crucial for biological cognition [13,14]. Hence, the supplementing of perceptions could also be due to cross-modal expectations. These could even override actual inputs, which is evident from the McGurk–MacDonald effect [15]. If you watch a video recording of a person making the sound /da/ (which means that the lips are not closing), but with the sound substituted with the sound /ba/, you may still perceive the sound you hear as being /da/.

A variant of SOMs, the A-SOM [7], which adds adaptable associative connections between feature representations has been used to build artificial systems, e.g., [16], which demonstrate cross-modal expectations.

Through adaptable associative connections between hierarchies of topographically ordered feature representations, self-organizing intra- and intermodal networks of feature representations (NFRs) are obtained. Some feature representations can be part of several NFRs, and the particular division of feature maps into NFRs depends on how we look at them and how we choose to categorize systems into subsystems. Adaptive associative connections learn to associate simultaneous, or temporarily close, activity in various feature representations elicited by simultaneous, or temporarily close, but different ordinary inputs. This means that feature representations that later lack ordinary inputs will be activated by activity patterns associated with ongoing activity in other feature representations in NFRs. For example, hearing the voice of a particular person would elicit activity patterns not only in the auditory hierarchies of feature representations that directly receive sensory inputs, but also in other, e.g., visual, feature representations in an intermodal NFR through associative activation. The total activity in NFRs will constitute episodic memories, imagery, etc.

The representation of a real or imagined concept or object is constituted by a set of associated activity patterns in various feature representations of NFRs distributed over multiple modalities. Hence, there is no need for any grandmother cells. Such associated activations of topologically ordered feature representations preserve an internal ordering of activation, and could be seen as forming a conceptual space [17].

To learn to represent novel concepts, objects, or possible objects, there is no need for new feature representations, because they are formed through associating activity patterns in existing feature maps in novel ways.

Hence, it is possible to create representations of various kinds of novel objects and concepts: existing; possible but not existing, e.g., unicorns; and impossible as well as nonexistent. What is possible to represent in this way by NFRs in the brain/artificial architecture is only constrained by the set of available feature representations that can be associated. In the mammalian brain, such feature representations are likely formed in early developmental phases and then to a large extent fixed and closed, see, e.g., [18]. It follows that there are likely constraints to what we can think.

NFRs containing topologically ordered feature representations with intra- and intermodal adaptable associative connections enable perception, various forms of memory, imagery, and functional consciousness.

In perception, sensory signals from receptors, together with information about involved exploratory actions, such as eye or hand movements, activate sets of feature maps. Those parts of the associated networks of feature representations that are not elicited directly by sensory inputs are activated through the activity in other feature representations via associative connections. Hence, the perceptions will be complete even with scarce sensory inputs, because missing parts are filled in with likely guesses through internal simulations. In episodic memory and imagery (i.e., internal simulation), the sets of associated networks of feature representations (which can also be nonsensory, such as motor representations) are activated internally in the cognitive architecture/brain. Semantic memory corresponds to more persistent associations due to repeatedly overlapping parts of activation from various episodic examples, thus forming prototypes in conceptual spaces. This also makes semantic memory more persistent. The working memory supposedly again employs networks of the same building blocks of feature representations obtained

during early developmental phases, but now activated in a more transient and temporary way, perhaps from, in the case of the mammalian brain, the frontal lobes.

In reality, these various cognitive functions are not separated from each other in a neat way. Rather, they blend and mix into each other. For example, the perception of hearing a familiar person's voice can trigger both episodic memories and internal visual simulations of the person, corresponding to reality but also pure fantasies, etc. Internally simulated perceptual expectations may, in turn, trigger exploratory behavior and attention aiming at confirming the expectations by obtaining additional sensory inputs. All of this is founded on associatively connected networks of topologically ordered feature representations.

The A-SOM has been successfully tested in many simulations [7] in several different domains regarding the ideas expressed in this paper. It has been tested with real sensors [16,17] as well as when simulating continuations of sequences [19–21].

Consciousness is about experiencing perceptions, including the perceptions of your actions; imagery; and memories. However, who is experiencing? I am considering functional consciousness here, thus leaving the problem of qualia out of the discussion.

When something is perceived, corresponding activity is elicited in a subset of associatively connected feature maps. This, in turn, elicits activity in other associatively connected subsystem(s), perhaps also composed of associatively connected feature maps. This means that the elicited activity patterns in the other connected subsystem(s), in a sense, represent the ongoing activity in the system (of associatively connected feature maps). This is equivalent to the second system observing the various phenomenal maps in the first system, whether these are activated due to sensory signals or through internal simulations (imagery, episodic memory, working memory, etc.).

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Van Essen, D. Functional organization of primate visual cortex. *Cereb. Cortex* **1985**, *3*, 259–329.
2. Tunturi, A.R. Physiological determination of the arrangement of the afferent connections to the middle ectosylvian auditory area in the dog. *Am. J. Physiol.* **1950**, *162*, 489–502. [[CrossRef](#)] [[PubMed](#)]
3. Tunturi, A.R. The auditory cortex of the dog. *Am. J. Physiol.* **1952**, *168*, 712–717. [[CrossRef](#)] [[PubMed](#)]
4. Reale, R.A.; Imig, T.H. Tonotopic organization in auditory cortex of the cat. *J. Comp. Neurol.* **1980**, *192*, 265–291. [[CrossRef](#)] [[PubMed](#)]
5. Zeki, S. The representation of colours in the cerebral cortex. *Nature* **1980**, *284*, 412–418. [[CrossRef](#)] [[PubMed](#)]
6. Kohonen, T. *Self-Organization and Associative Memory*; Springer: Berlin/Heidelberg, Germany, 1988.
7. Johnsson, M.; Balkenius, C.; Hesslow, G. Associative Self-Organizing Map. In Proceedings of the International Joint Conference on Computational Intelligence (IJCCI), Madeira, Portugal, 5–7 October 2009; pp. 363–370.
8. Ritter, H.; Kohonen, T. Self-Organizing Semantic Maps. *Biol. Cybern.* **1989**, *61*, 241–254. [[CrossRef](#)]
9. Buonamente, M.; Dindo, H.; Johnsson, M. Hierarchies of Self-Organizing Maps for Action Recognition. *Cogn. Syst. Res.* **2016**, *39*, 33–41. [[CrossRef](#)]
10. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
11. Johnsson, M.; Balkenius, C. Sense of Touch in Robots with Self-Organizing Maps. *IEEE Trans. Robot.* **2011**, *27*, 498–507. [[CrossRef](#)]
12. Kanizsa, G. Subjective Contours. *Scien. Am.* **1976**, *234*, 48–52. [[CrossRef](#)]
13. Hesslow, G. Conscious Thought as Simulation of Behaviour and Perception. *Cogn. Sci.* **2002**, *6*, 242–247. [[CrossRef](#)]
14. Grush, R. The Emulation Theory of Representation: Motor Control, Imagery and Perception. *Behav. Brain Sci.* **2004**, *27*, 377–442. [[CrossRef](#)]
15. McGurk, H.; MacDonald, J. Hearing Lips and Seeing Voices. *Nature* **1976**, *264*, 746–748. [[CrossRef](#)] [[PubMed](#)]
16. Johnsson, M.; Balkenius, C. Associating SOM Representations of Haptic Submodalities. *Towards Auton. Robot. Syst.* **2008**, *2008*, 124–129.
17. Gärdenfors, P. *Conceptual Spaces—The Geometry of Thought*; The MIT Press: Cambridge, MA, USA, 2000.

18. Blakemore, C.; Cooper, G.F. Development of the Brain Depends on the Visual Environment. *Nature* **1970**, *228*, 477–478. [[CrossRef](#)] [[PubMed](#)]
19. Buonamente, M.; Dindo, H.; Johnsson, M. Discriminating and Simulating Actions with the Associative Self-Organizing Map. *Connect. Sci.* **2015**, *27*, 118–136. [[CrossRef](#)]
20. Gil, D.; Garcia, J.; Cazorla, M.; Johnsson, M. SARASOM—A Supervised Architecture Based on the Recurrent Associative SOM. *Neural. Comp. Appl.* **2014**, *26*, 1103–1115. [[CrossRef](#)]
21. Buonamente, M.; Dindo, H.; Chella, A.; Johnsson, M. Simulating Music with Associative Self-Organizing Maps. *Biol. Inspired Cogn. Archit.* **2018**, *25*, 135–140. [[CrossRef](#)]