Magnus Johnsson

PERCEPTION, IMAGERY, MEMORY
AND CONSCIOUSNESS

ABSTRACT

I propose and discuss some principles that I believe are substantial for perception, various kinds of memory, expectations and the capacity for imagination in the mammal brain, as well as for the design of a biologically inspired artificial cognitive architecture. I also suggest why these same principles could explain our ability to represent novel concepts and imagine non-existing and perhaps impossible objects, while there are still limits to what we can imagine and 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, imagination, consciousness, self-organization, feature maps, associative learning, multimodal integration, cognitive architecture.

1. INTRODUCTION

The astonishing abilities of the mammalian brain 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 associatively, hierarchically, recurrently connected topology preserving feature representations that reflect the probability distribution of their input are important in the mammal 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 perception, imagery, and functional consciousness in the mammal

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 the sensory input is limited; multimodal integration. How various memory systems, imagery and perception fit together and 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 presumably been done through evolution by nature. By looking at how the mammal 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 functional consciousness in a biologically inspired cognitive architecture modeled on the mammal brain.

This systems level modeling approach means that, though not modeling crucial components and their interconnections in detail, general principles also adhered to by their biological counterparts should be identified and followed in the design of the cognitive architecture.

The components' functionality can be implemented with mechanisms that model the systems at a suitable level of accuracy. Later they can be re-implemented by other mechanisms for accuracy and performance reasons, or if more efficient implementations are found.

We could go about working on a bio-inspired systems-level cognitive architecture in various ways. At one extreme, we could work from a more holistic starting point by identifying crucial components and interactions found in the neural systems of biological organisms. Then we could implement maximally simplified versions of these and try to make them work together as well as possible. Examples of components in such an architecture inspired by a mammal brain could be a maximally simplified visual system and a maximally simplified mechanism, or set of mechanisms, corresponding to the Basal ganglia etc. Inspired by the work of Valentino Braitenberg (Braitenberg, 1984) and the robotics physicist Mark W. Tilden, I believe such simplified but complete cognitive architectures would still enact interesting behaviors.

At the other extreme, we could work on individual components while trying to optimize these to perform at a human level or beyond. Many artificial perception researchers work at this extreme, e.g. by creating computer vision systems that in some respects even exceed the abilities of humans.

 My approach is somewhere in the middle. I try to figure out general principles for not necessarily complete, but more composed architectures at an intermediary level. Hence my focus is not whether component imple-

mentations are optimized for performance. Following a systems level approach, individual components can be reimplemented iteratively at later stages for performance, accuracy or for other reasons, but this is not the focus here. Thus, the work on general principles can be isolated from the engineering questions of performance.

The perceptual parts of a cognitive architecture built according to these ideas employ to a large extent self-organizing topographical feature representations. Such feature representations are somewhat reminiscent of what has been found in mammal brains. These topographical representations are connected hierarchically, associatively and recurrently. Hierarchies of increasingly complex feature representations—and in the extension different architectural components, possibly distributed—self-organize while supervising each other's associative learning/adaptation over space (by associative connections) and over time (by recurrent associative connections). For example, such an architecture contains hierarchies of topographically ordered feature representations within sensory submodalities. To an extent, these hierarchies also cross the borders of different sensory submodalities, and even the borders of different sensory modalities. The topographically ordered feature representations connect associatively at various hierarchical levels within, but also across, sensory submodalities, modalities and to systems outside the perceptual parts, e.g. motor representations.

Below I discuss some principles that I believe are substantial to perception, various kinds of memory, expectations and imagery in the mammal brain and for the design of a bio-inspired artificial cognitive architecture. I also suggest why these principles could explain our ability to represent novel concepts and imagine non-existing and perhaps impossible objects, while there are still limits to what we can imagine and think about.

I will also present some ideas regarding how these principles could be relevant for an autonomous agent to become p-conscious (Block, 1995) in the sense defined by Bołtuć (2009), i.e. as referring to first-person functional awareness of phenomenal information. Whether such an autonomous agent would also be conscious in a non-functional first-person phenomenological sense, i.e. h-conscious, adopting again the terminology of Bołtuć (2009), and thus experience qualia of its own subjective first-person experiences of external objects and inner states, is another matter. The latter question belongs to the hard problem of consciousness (Chalmers, 2003). The difficulty with that problem is that a physical explanation in terms of brain processes is an explanation in terms of structure and function, which can explain how a system's behavior is produced, but it is harder to see why the brain processes are accompanied by subjective awareness of qualia. According to Chalmers (2003) all metaphysical views on phenomenal consciousness are either reductive or nonreductive, and he considers the latter to be more promising. Nonreductive views require a re-conception

of physical ontology. I suggest that the bio-inspired principles proposed in this paper have relevance for p-consciousness. Hence a cognitive architecture employing these ideas would probably become at least p-conscious. However, it is possible that h-consciousness is not a computational process, and I will not take a final position on the issue of phenomenal h-consciousness in this paper.

2. FEATURE REPRESENTATIONS

Topographically ordered maps are inherent parts of the human brain. There are continuously ordered representations of receptive surfaces across various sensory modalities, e.g. in the somatosensory and visual (van Essen, 1985) areas, in neuron nuclei and in the cerebellum.

The size of the representational area in such ordered representations depends on the behavioral importance and frequency of the represented input. For example, the representation of the fovea is much larger than the rest of the retina, and the representation of the fingertip is proportionally larger than the rest of the finger.

There are also more abstract topographically ordered representations in the brain, e.g. frequency preserving tonotopic maps (Tunturi, 1950; 1952; Reale, Imig, 1980) in primary auditory areas, and color maps in V4 (Zeki, 1980) in the visual areas.

In a model of such self-organized topographically ordered representations, essential relations among data should be made explicit. This could be achieved by forming spatial maps at an appropriate abstraction level depending on the purpose of the model. For a reasonable computational efficiency, the focus should be on main properties without any accurate replication of details. Reasonable candidates for a basic model corresponding to a topographically ordered representation in the brain satisfying these conditions are the Self-Organizing Map, SOM (Kohonen, 1988) and its variants. Such a basic model forms a fundamental building block—not to be confused with the crucial components discussed above—in the perceptual parts of a bio-inspired cognitive architecture.

Examples of suitable candidates, beside the SOM, are the Growing Grid (Fritzke, 1995) and the Growing Cell Structure (Fritzke, 1994). In addition to the adaptation of the neurons, these models also find suitable network structures and topologies through self-organizing processes. Other examples are the Tensor-Multiple Peak SOM, T-MPSOM (Johnsson et al., 2006) or the Associative Self-Organizing Map (Johnsson et al., 2009). The latter, or rather the principles it instantiates, are crucial for the principles of the perceptual parts of a cognitive architecture discussed in this paper and will be elaborated on below.

The SOM develops a representation that reflects the distance relations of the input, which is characteristic of lower levels of perception. If trained with a representative set of input, the SOM self-organizes into a dimensionality reduced and discretized topographically ordered feature representation also mirroring the probability distribution of received input. The latter means that frequent types of input will be represented with better resolution in the SOM. This corresponds to, for example, the development of a larger representational area of the fingertip than the rest of the finger in the brain, which was discussed above. Hence the SOM is reminiscent of the topographically ordered representations found in mammalian brains.

In a sense, the topographically ordered map generated by a SOM—and in the extension an interconnected system of SOMs—is a conceptual space (Gärdenfors, 2000) generated from the training data through a self-organizing process.

Due to the topology-preserving property of the SOM similar input elicit similar activity, which provides systems based on the SOM with an ability to generalize to novel input.

A SOM can be trained to represent various kinds of feature, including phenomenal ones. The latter would turn the SOM into a phenomenal content map (Damasio, 2010). For example, a SOM can be trained to represent directions of lines/contours (as in V1), colors (as in V4), or more complex features such as the postures and gesture movements of an observed agent (Buonamente et al., 2016), or the words of a text corpus ordered in a way that reflects their semantic relations (Ritter, Kohonen, 1989). Employing SOMs or other topographically ordered feature representations to represent phenomenal features together with the general design principles for a bio-inspired cognitive architecture suggested in this paper, would enable strong semantic computing (Bołtuć, 2018).

Other models of a self-organizing topology preserving feature representation are possible and might turn out to be more suitable for various reasons such as performance and accuracy. However, as also mentioned above, that is beyond the point of this paper, which aims at presenting higher level architectural principles where models of self-organizing topographically ordered representations are building blocks. Since I adhere to a systems-level modeling approach, subsystems of the cognitive architecture can be updated and substituted in an iterative fashion for improvement.

3. HIERARCHICAL FEATURE REPRESENTATIONS

How can self-organized topographically-ordered representations of a more abstract kind, e.g. a representation with semantically related symbols that occupy neighboring places be obtained in a cognitive architecture?

In the mammal brain there seems to be a principle of hierarchical ordering of representations, e.g. the executive and motor areas seem to be hierarchically ordered from more abstract to less abstract representations. Constraining the discussion to the perceptual parts of the mammal brain, we find that the different sensory modalities (visual, somatosensory, auditory, ...) adhere to a hierarchical organizational principle. For example, we find hierarchically-organized topology and probability-density preserving feature maps in the ventral visual stream of the visual system. These feature maps rely on the consecutive input from each other and tend to be hierarchically-ordered from representations of features of a lower complexity to representations of features of a higher complexity. Thus, we find ordered representations of contour directions in V1 in the Occipital lobe, of shapes in V2, of objects in V4, and of faces or complex facial features in the inferior temporal (IT) area of the Temporal lobe.

The hierarchical organization principle is employed artificially in Deep Neural Networks, i.e. in artificial neural networks with several hidden layers. A neural network that has been applied very successfully within the field of computer vision is the Deep Convolutional Neural Network (LeCun et al., 1998).

Here, when I discuss the hierarchical ordering principle for perceptual parts of a bio-inspired cognitive architecture, this principle is instantiated by hierarchical SOMs. The choice of SOMs is not based on performance, but on the fact that the hierarchical organization principle is also to be combined with other principles in the cognitive architecture elaborated on below. For the moment the SOM and its variants are considered good choices to explain and test principles.

Together with collaborators, the author has shown the validity of this hierarchical organizational principle repeatedly with hierarchical SOMs when applied to different sensory modalities. For example, in the case of the somatosensory modality, several experiments have been conducted to show how haptic features of an increasing complexity can be extracted in hierarchical self-organizing representations, e.g. from proprioceptive and tactile representations at the lower complexity end to self-organizing representations of shapes and sizes of the haptically explored objects (Johnsson et al., 2011a). Another example in the case of the visual domain where experiments have been done to show that hierarchies of ordered representations of postures at the lower complexity end to ordered representations of gesture movements of the observed agent can be self-organized (Buonamente et al., 2016).

4. SUPPLEMENTING SENSORY SIGNALS IN PERCEPTION

The brain supplements perceptions when the sensory input is not complete. This is evident from various visual illusions, e.g. the Kanizsa Triangle (Kanizsa, 1976) 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 semi-random eye movements known as saccades directing the movements toward particularly conspicuous and, in some sense interesting features. Supposedly we carry out similar semi-random movements with our hands and fingers to gain particularly interesting and useful tactile sensory input when we, for example, 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 input with expectations, from memory, of what is likely to be there.

A crucial aspect of biological cognition is an ability to simulate or influence perceptual activity in some brain areas due to the activity in other brain areas (Hesslow, 2002; Grush, 2004), e.g. the activity in areas of other sensory modalities. For example, when the visual perception of a lightning evokes an expectation of the sound of thunder, or when visual images/expectations of an object is evoked when it is felt in the pocket. Hence, one supplement to the afferent sensory signals in perception could be such simulated *Cross Modal Expectations*. These could even override actual input, which is evident from the McGurk-MacDonald effect (McGurk, MacDonald, 1976). If a person sees a video with someone making the sound /da/ on which the lips cannot be seen closing and the actual sound played is /ba/, the expectations evoked by the visual perception may have such an influence on the activity caused by the actual afferent auditory sensor signals that the person may still hear the sound /da/.

A variant of the SOM, the A-SOM (Johnsson et al., 2009), that adds adaptable associative connections between feature representations has been used to build artificial systems, e.g. (Johnsson, Balkenius, 2008), that demonstrate the supplementation of sensory input and the elicitation of cross-modal expectations.

5. NETWORKS OF FEATURE REPRESENTATIONS

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 input, are activated through the activity in other feature representations via associative connections. Hence the perceptions will be complete even with scarce sensory in-

put, because missing parts are filled in with likely guesses through internal simulations.

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 the feature maps into NFRs depend on how we look at it and how we choose to categorize the system into subsystems. The adaptive associative connections learn to associate simultaneous, or temporarily close, activity in various feature representations elicited by simultaneous, or temporarily close, but different ordinary input. This means that feature representations that later lack ordinary input will be activated by activity patterns associated with the ongoing activity in other feature representations in the NFR. For example, hearing the voice of a particular person would elicit activity patterns not only in the auditory hierarchies of feature representations that directly receives sensory input, but also in other, e.g. visual, feature representations in an intermodal NFR through associative activation. The total activity in the NFR will constitute episodic memories, imagination etc.

6. MEMORY

Although there are hierarchically-organized feature representations in the brain, it is questionable whether there are neurons—aka grandmother cells—that are the exclusive representatives of distinct individual objects. Though there is no total consensus regarding this, I consider it more likely that distinct individual objects are coded in a more distributed way as an ensemble of feature representations in the NFR, at various complexity levels, across several sensory (as well as non-sensory) modalities. Hence, the recognition of distinct individual objects consists in the simultaneous activation of a sufficiently large and unique subset of this ensemble of representations across various modalities.

Thus, the representation of a real or imagined concept or object is constituted by a set of associated activity patterns in various feature representations of the NFR distributed over multiple modalities. Such associated activations of topologically ordered feature representations preserve an internal ordering of activation and could be seen as forming a *Conceptual Space* (Gärdenfors, 2000).

The activation of some feature representations will tend to trigger expectations/imaginations of features of the distinct individual object in other representations across various modalities, presumably associated by associative connections in a way similar to the activation of more features of high-

er—or lower—complexity in hierarchically connected feature representations (which can as well be cross-modal).

I believe that such connectivity—associative and hierarchical—between feature representations of various modalities and at various complexity levels are what enables the filling in of missing parts of our perception by imagination, but also that they enable our various kinds of memory.

Different kinds of memory are, I believe, using the same kind of feature representations across modalities in our brains. What differs between different kinds of memory is rather how they are activated and precisely what ensemble of feature representations in the NFRs that are activated.

For example, one could speculate—in a simplified way—that 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 mammal brain, the frontal lobes, whereas perception as such, is the activation of an ensemble of feature representations due to afferent sensory signals, together with the filling-in of missing parts due to cross-modal as well as top-down expectations at various levels of hierarchies.

In episodic memory and imagination (i.e. internal simulation) the sets of associated networks of feature representations (which can also be non-sensory, such as motor representations) are activated internally (at least partly) in the cognitive architecture/brain. The associatively connected representations (actually associated activity patterns in the underlying wetware / hardware) are what lends memory and imaginations their associative characters.

The important point here is that it is reasonable to believe that the same (simple) principles are behind both the supplementation of perception and the associative character of memory and imagination; and the distributed, associative and hierarchical character of the intra- (and inter) modal representations they all (perception, imagination and memory) rely on. That different faculties go into each other also explains why we tend to both keep memories alive (by strengthening associative connections through reactivation) and sometimes change them over time (via imagination) when we recapitulate.

Semantic memory presumably corresponds to more persistent associations due to repeatedly overlapping activity from many various perceptual and episodic examples over time, thus forming prototypes in conceptual spaces. This also makes semantic memory more persistent, as well as more resistant in a deteriorating/aging system.

The point here is that all kinds of memory, perceptions, imaginations, and expectations are proposed using simultaneous and / or sequential activations of ensembles / subsets of the same NFRs across various modalities in the brain. In fact, I think that there is no reason that the representa-

tions should be constrained to the brain only, but that associated representations could also be various kinds of activity in / of the body, such as e.g. postural / breathing patterns, hormonal configurations etc. This would also explain why the change of posture / breathing patterns can change the state of the mind. In the extension, even "representations" that we interact with—and continuously reconfigure—in the environment outside the body—including the representations within other agents, such as humans, pets, machines etc.—are presumably included.

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.

This kind of feature ensemble coding also enables / explains the ability to represent completely novel categories / concepts in the brain / cognitive architecture, and the ability to create and imagine non-existing and perhaps impossible concepts, objects, etc. This is because representations are composed of sufficiently large ensembles of associated multi-modal features, and novel associated ensembles and sometimes associated ensembles corresponding to concepts and imaginations that do exist (but have not yet been seen or reached) or do not exist in our physical reality (e.g. unicorns) can emerge.

Of course, there are limits to what we can imagine and conceptualize, and perhaps even think about. For example, we are unable to visualize objects in spaces of a higher dimensionality than three. However, such limitations are just to be expected if all perceptions, memories and imaginations are made up of distributed (in space and time) activations of ensembles of associated features, and there are constraints on what kind of features can be represented in the brain (or cognitive architecture), which is likely. The constraints are probably set by biological limitations that exist due to a lack of evolutionary pressure, as well as determined by the development of the organism in its environment. An example of the latter is that cats raised in an environment consisting entirely of vertical lines during a critical developmental phase during infancy will be unable to see horizontal lines (Blakemore and Cooper, 1970). That there are constraints on what kind of features that can be represented also implies the possibility that all that we can think about regarding reality is not necessarily corresponding to all that there would have been to think about, had we been wired differently.

In accordance with the reasoning above, it is reasonable to assume that the need for associative connections—corresponding to axon bundles in the neural system of a biological organism—between feature maps at various complexity levels within as well as between different modalities are of significance in a cognitive architecture based on self-organizing topographically-ordered feature representations. Such associative connections need to be adaptive (by adjustable parameters corresponding to modifiable synapses in

the neural system of a biological organism) to enable the learning of associations between the activity in various feature representations.

7. IMAGINATION

In addition to an ability to automatically develop, and continuously readapt, sensory and other representations, and their interconnections that connect simultaneous activity within them spatially, a bio-inspired autonomous agent needs an ability to learn to associate activations of representations over time. This is desirable because it enables the autonomous agent to remember and re-enact sequences of perceptual—and other—activity across modalities and levels of hierarchy.

With such an ability an autonomous agent can remember sequences of perceptions, and if the ability is generalized, other things as well, e.g. motor activities. Such perceptual sequences could, for example, be visual landmarks. To the perceived visual landmarks, appropriate motor activity could be associated. With perceptual sequences simultaneously learned in other modalities together with cross-modal associations, the sequential memories are reinforced and thus diminish the influence of noise and limitations in sensory input. The perceptions (and preparatory responses etc.) corresponding to missing input in some modalities—sensory and other—will be imagined, i.e. elicited through cross-modal activation. If suddenly the agent would lack input to some, or all, sensory modalities, it would still be able to operate and to some extent carry out actions associated with imagined perceptions of the environment. With this kind of ability an agent would also be able to sit idle imagining various scenarios and the likely consequences of carrying out different kinds of actions. The latter is valuable for survival and will also accelerate the agent's learning.

The idea to internally elicit activity patterns in perceptual, motor and other circuits over time (activation sequences) and in space (in various feature maps across different modalities), corresponding to the patterns that would have been elicited had there been sensory input and had the actions been carried out, is closely related to the simulation hypothesis by Hesslow (2002). It could in the extension also be the foundation for providing agents with an ability to guess the intentions of other agents, either by directly simulating the likely perceptual continuations of the perceived behavior of an observed agent, or by internally simulating its own likely behavior in the same situation under the assumption that the other agent is similar in its assessments, experiences and values that drives it.

A mechanism that implements self-organizing topographically ordered feature representations that can be associatively and recurrently connected with an arbitrary number of other representations and with arbitrary time

delays is the Associative Self-Organizing Map (A-SOM). Hence the A-SOM would in some cases be a better choice, than the standard SOM, to use as one of the basic building blocks in the perceptual parts of a cognitive architecture. An A-SOM can learn to associate the activity in its self-organized representation of input data with arbitrarily many sets of parallel inputs and with arbitrarily long-time delays. For example, it can learn to associate its activity with the activity of other self-organization maps, or with its own activity at one or more earlier times. This allows for cross-modal expectations. For example, if a sensory modality, say the visual system in a cognitive architecture, produces a certain internal pattern of activity due to sensory input, then activity patterns are elicited in other sensory modalities corresponding to the patterns of activity that are often triggered in these other sensory modalities through sensory inputs that usually occur simultaneously, even when they do not. Due to the ability of the A-SOM to associate its activity with its own activity at one or more earlier times, a mechanism for sequence completion that can be used for internal simulation is made possible. This is consistent with those abilities necessary for an autonomous agent described above. The A-SOM has been successfully tested in many simulations (e.g., Johnsson et al., 2011b) in several different domains, as well as together with real sensors such as tactile sensors (Johnsson, Balkenius, 2008) and cameras (Buonamente et al., 2015), and when simulating likely continuations of sequences of strings of symbols and words (Gil et al., 2014). It has been used to simulate the sensory activity patterns likely to follow some initially perceived movements of actions/gestures (Buonamente et al., 2015). In the domain of music, a further developed and more mature and generalized version of the A-SOM has been used to simulate the sensory activity patterns likely to follow those elicited by the initial parts of perceived Bach chorale melodies (Buonamente et al., 2018).

Associative connections are in place between different representations at various levels of feature complexity. Simultaneously-activated feature representations develop stronger associative connectivity. The result is that we will find strongly interconnected sets of feature representations—and other kinds of circuits—in the brain/architecture. As humans, we label these and call them systems/components of one kind or another (depending on the particular discipline, the prevailing paradigm and zeitgeist), though we should keep in mind that these categorizations and demarcations are our inventions and thus somewhat arbitrary.

The inter-connectivity of the feature representations within a modality/submodality tend to be strong because it has been reinforced by simultaneous activations originating from the receptors of the modality specific sensory organs. Thus, connective configurations / subsystems in the brain / architecture develop through the repeated simultaneous activation of sets of self-organizing feature representations.

However, the feature representations within a modality also connect to feature representations in other modalities / systems, only to a lesser extent. This is due to the statistically fewer simultaneous activations of feature representations in other modalities. Various systems activate each other through these associative connections that have learned to associate activity that normally come together. Hence, if the activity within one system, perhaps triggered through afferent signals from sensory organs or from some other part of the brain/architecture, tend to correlate with the activity of other systems, perhaps triggered by the afferent signals from other sensory organs or other parts of the brain/architecture, then the inter-connectivity of the systems is reinforced. The foundation for these correlated activities in various systems is that sensory stimuli, and the consequences of an agent's actions, are related in a non-random way due to the statistical regularities of the properties of the world. These statistical regularities will be reflected in the associative connectivity between various systems.

In reality the 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, internal visual simulations of the person, corresponding to reality but also pure fantasies, etc. Internally simulated perceptual expectations in turn may trigger exploratory behavior and attention aiming at confirming the expectations by obtaining additional sensory input. All this founded on associatively connected networks of topologically ordered feature representations.

Taken together, all this means that NFRs containing topologically ordered feature representations with intra- and intermodal adaptable associative connections enable perception, various forms of memory and imagination. In addition, it provides a mechanism for representing the ongoing activity in one system/NFR with the activity of other systems/NFRs.

8. CONSCIOUSNESS

Consciousness is about experiencing perceptions, including the perceptions of our own actions; imagery; memories. But who is experiencing it? I am considering functional consciousness here, thus leave the problem of qualia out of the discussion.

In the discussion about cross-modal expectations and internal simulations above, I discussed how activity in some feature representations can elicit reasonable activity in other feature representations through associative connections. The elicited activity in the latter representations correspond to the activity that normally would or could occur simultaneously, or timed,

with the activity in the first representations even though the latter lack any afferent input ultimately originating from sensors.

I believe that the same mechanism with adaptive associative connections in the case of a bio-inspired cognitive architecture, or nerve bundles with synapses in the case of a neural system of a biological organism, between different subsets of feature representations, at various levels of abstraction, is significant for the realization of at least p-consciousness. From this perspective, the elicitations of activity in some feature representations by the activity in other feature representations via associative connections can be viewed as if the activity in the latter system (composed of the activity in connected, perhaps distributed, feature representations), in a sense, is represented by the activity in the former system (of associatively connected feature maps).

Various systems could perhaps also 'observe' each other simultaneously as well. The mechanisms and principles sketched above could be behind or be used for a kind of summarization of the observed subsystem's or subsystems' activity at a possibly different and more abstract level.

As also argued by Hesslow and Jirenhed (2007), perceptual simulation could explain the appearance of an inner world. A remaining question is 'who' is observing regardless of whether it is perceptions ultimately elicited from sensory organs, internal simulations originating from within the brain, or some combination thereof. My proposal is that they are observed by other connected configurations of systems whose activity summarizes/represents the observed internal simulations and perceptions, because their corresponding activity correlates due to the learning represented in the adaptive associated connections. The same systems could perhaps have multiple functions while also "observing" each other simultaneously as well.

Put differently, this could be seen as one system observing the various *Phenomenal Maps* of another system, whether these are activated due to sensory signals or through internal simulations (imagination, episodic memory, working memory etc.).

Still another way to put it is that some systems are aware of, i.e. p-conscious / functionally conscious of, other systems' perceptual activity.

The activity of associatively connected configurations of feature representations correlates because the adaptations of the associative connections between the representations, and the adaptions of the representations themselves happens simultaneously, continuously and dynamically. At a lower perceptual level this means that the activation of feature representations in some sensory modalities will elicit activity in feature representations in other sensory modalities and consequently sensory expectations / supplementations in those other modalities, as discussed above.

Thus, I believe that adaptive associative connections between and within various configurations of strongly connected feature representations at vari-

ous levels of complexity or abstraction are of significant importance for realizing p-consciousness, i.e. functional consciousness, in a cognitive architecture.

REFERENCES

C. Blakemore, G. F. Cooper, *Development of the Brain Depends on the Visual Environment*, Nature, 228, 1970, pp. 477–478.

N. Block, *On a Confusion about a Function of Consciousness*, Behavioral and Brain Sciences, 18, 1995, pp. 227–287.

P. Bołtuć, *Strong Semantic Computing*, Procedia Computer Science, 123, 2018, pp. 98–103.

____ , *The Philosophical Issue in Machine Consciousness*, International Journal of Machine Consciousness, 1 (1), 2009, pp. 155–176.

V. Braitenberg, *Vehicles: Experiments in Synthetic Psychology*, MIT Press, Cambridge MA 1984.

M. Buonamente, H. Dindo, A. Chella, M. Johnsson, *Simulating Music with Associative Self-Organizing Maps*, Journal of Biologically Inspired Cognitive Architectures, 25, 2018, pp. 135–140.

M. Buonamente, H. Dindo, M. Johnsson, *Discriminating and Simulating Actions with the Associative Self-Organizing Map*, Connection Science, 2 (27), 2015, pp. 118–136.

M. Buonamente, H. Dindo, M. Johnsson, *Hierarchies of Self-Organizing Maps for Action Recognition*, Cognitive Systems Research, (39), 2016, pp. 33–41.

D. J. Chalmers, *Consciousness and Its Place in Nature*, in: Blackwell Guide to the Philosophy of Mind, S. Stich, T. Warfield (eds.), Blackwell Publishing, Malden, MA 2003, pp. 102–142.

A. Damasio, *Self Comes to Mind: Constructing the Conscious Brain*, Pantheon, Cambridge 2010.

D. van Essen, *Functional Organization of Primate Visual Cortex*, Cerebral cortex, 3, 1985, 259–329.

B. Fritzke, *Growing Cell Structures—A self-organizing Network for Unsupervised and Supervised Learning*, Neural Networks, 9 (7), 1994, pp. 1441–1460.

____ , *Growing Grid—a Self-organizing Network with Constant Neighborhood Range and Adaptation Strength*, Neural Processing Letters, 5 (2), 1995, pp. 9–13.

P. Gärdenfors, *Conceptual Spaces—The Geometry of Thought*, MIT Press, Cambridge, MA 2000.

D. Gil, J. Garcia, M. Cazorla, M. Johnsson, *SARASOM – A Supervised Architecture based on the Recurrent Associative SOM*, Neural Computing and Applications, 5 (26), 2014, pp. 1103–1115.

R. Grush, *The Emulation Theory of Representation: Motor Control, Imagery and Perception*, Behav. Brain. Sci., 27, 2004, pp. 377–442.

G. Hesslow, *Consious Thought as Simulation of Behaviour and Perception*, Trends Cogn. Sci., 6, 2002, pp. 242–247.

G. Hesslow, D.-A. Jirenhed, *The Inner World of a Simple Robot*, J. Consc. Stud., 14, 2007, pp. 85–96.

M. Johnsson, C. Balkenius, *A Robot Hand with T-MPSOM Neural Networks in a Model of the Human Haptic System*, in: the Proceedings of Towards Autonomous Robotic Systems 2006, 2006, pp. 80–87.

____ , *Sense of Touch in Robots with Self-Organizing Maps*, IEEE Transactions on Robotics, 3 (27), 2011a, pp. 498–507.

____ , *Associating SOM Representations of Haptic Submodalities*, in: The Proceedings of Towards Autonomous Robotic Systems 2008, 2008, pp. 124–129.

M. Johnsson, C. Balkenius, G. Hesslow, *Associative Self-Organizing Map*, in: The Proceedings of the International Joint Conference on Computational Intelligence (IJCCI) 2009, 2009, pp. 363–370.

M. Johnsson, M. Martinsson, D. Gil, G. Hesslow, *Associative Self-Organizing Map*, in: Self-Organizing Maps—Applications and Novel Algorithm Design, MA: Intech, 2011b, pp. 603–626.

G. Kanizsa, *Subjective Contours*, Scien. Am., 234 (4), 1976, pp. 48–52.

T. Kohonen, *Self-Organization and Associative Memory*, Springer Verlag, Berlin–Heidelberg 1988.

Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, *Gradient-Based Learning Applied to Document Recognition*, in: Proceedings of the IEEE, 1998, pp. 2278–2324.

H. McGurk, J. MacDonald, *Hearing Lips and Seeing Voices*, Nature, 264, 1976, pp. 746–748.

R. A. Reale, T. H. Imig, *Tonotopic Organization in Auditory Cortex of the Cat*, J. Comp. Neurol., 192, 1980, pp. 265–291.

H. Ritter, T. Kohonen, *Self-Organizing Semantic Maps*, Biol. Cybern., 61, 1989, pp. 241–254.

A. R. Tunturi, *Physiological Determination of the Arrangement of the Afferent Connections to the Middle Ectosylvian Auditory Area in the Dog*, Am. J. Physiol., 162, 1950, pp. 489–502.

A. R. Tunturi, *The Auditory Cortex of the Dog*, Am. J. Physiol., 168, 1952, pp. 712–717.

S. Zeki, *The Representation of Colours in the Cerebral Cortex*, Nature, 284, 1980, pp. 412–418.

ABOUT THE AUTHOR — a cross-disciplinary researcher. He is currently Associate Professor in autonomous systems at Malmö University in Sweden and Researcher at Magnus Johnsson AI Research AB. More information at www.magnusjohnsson.se

Email: magnus@magnusjohnsson.se