

# In Search of Computational Correlates of Artificial Qualia

Antonio Chella, Salvatore Gaglio

Dipartimento di Ingegneria Informatica Università di Palermo  
Viale delle Scienze, building 6, 90128 Palermo, Italy  
{chella, gaglio}@unipa.it

## Abstract

In previous papers we presented a robot cognitive architecture organized in three computational *areas*. The *subconceptual* area is concerned with the processing of data coming from the sensors. In the *linguistic* area representation and processing are based on a logic-oriented formalism. The *conceptual* area is intermediate between the subconceptual and the linguistic areas and it is based on the notion of *conceptual spaces*. The robot, starting from the 3D information stored in the conceptual area and from the data coming from sensors and processed by the subconceptual area, is able to build a 2D viewer dependent reconstruction of the scene it is perceiving. This 2D model corresponds to what the robot is seeing at any given time. We suggest that the conceptual and the linguistic areas are at the basis of the robot artificial qualia.

## Introduction

It has been questioned if robots may have qualia, i.e., qualitative, phenomenal experiences in the sense discussed, among others, by Chalmers (1996).

We are not interested in the problem of establishing whether robots can have real phenomenal experiences or not. For our present concerns, speak of robot's "artificial qualia" in a sense similar to Aleksander (1996). We use this expression in a somewhat metaphorical sense: we call "artificial quale" a state that in some sense corresponds to the "phenomenal" experience of the robot, without making any hypothesis concerning the fact that the robot truly experiences it.

In previous papers (Chella et al. 1997, 2000) we presented a robot cognitive architecture organized in three computational *areas* - a term which is reminiscent of the cortical areas in the brain.

The *subconceptual* area is concerned with the processing of data coming from the sensors. Here information is not yet organized in terms of conceptual structures and categories. From the point of view of the artificial vision, this area includes all the processes that extract the 3D model of the perceived scene. In the *linguistic* area representation and processing are based on a logic-oriented formalism. We adopt the term "linguistic" instead of the overloaded term "symbolic", because we want to stress the

reference to formal languages in the knowledge representation tradition.

The *conceptual* area is intermediate between the subconceptual and the linguistic areas. Here, data is organized in conceptual "gestaltic" structures, that are still independent of any linguistic description. The symbolic formalism of the linguistic area is interpreted on aggregation of these structures.

We suggest that the conceptual and the linguistic areas are at the basis of the robot artificial qualia. In our model, the robot, starting from the 3D information stored in the conceptual area and from the data coming from sensors and processed by the subconceptual area, is able to build a 2D, viewer dependent reconstruction of the scene it is perceiving. This 2D model corresponds to what the robot is seeing at any given time. Its construction is an active process, driven by both the external flow of information and the inner model of the world.

## The Cognitive Architecture

The proposed architecture (Fig. 1) is organized in computational "areas". In our model, the areas are concurrent computational components working together on different commitments. There is no privileged direction in the flow of information among them: some computations are strictly bottom-up, with data flowing from the subconceptual up to the linguistic through the conceptual area; other computations combine top-down with bottom-up processing.

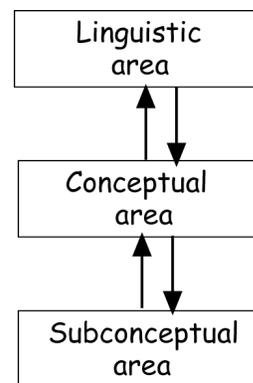


Figure 1: The cognitive architecture

## Conceptual Spaces

The conceptual area, as previously stated, is the area between the subconceptual and the linguistic area. This area is based on the theory of conceptual spaces (Gärdenfors 2000).

Conceptual spaces provide a principled way for relating high level, linguistic formalisms with low level, unstructured representation of data. A conceptual space CS is a metric space whose dimensions are the quantities generated as outputs of computational processes occurring in the subconceptual area, e.g., the outputs of the neural networks in the subconceptual area. Different cognitive tasks can presuppose different conceptual spaces, and different conceptual spaces can be characterized by different dimensions. Examples of possible dimensions, with reference to object perception tasks, are: color components, shape parameters, spatial coordinates, motion parameters, and so on. In general, dimensions are strictly related to the results of measurements obtained by sensors. In any case, dimensions do not depend on any specific linguistic description. In this sense, conceptual spaces come before any symbolic or propositional characterization of cognitive phenomena.

We use the term *knoxel* to denote a point in a conceptual space. The term *knoxel* (in analogy with the term *pixel*) stresses the fact that a point in CS is the knowledge primitive element at the considered level of analysis.

The conceptual space CS acts as a workspace in which low-level and high-level processes access and exchange information respectively from bottom to top and from top to bottom. However, the conceptual space is a workspace with a precise geometric structure of metric space and also the operations in CS are geometrics: this structure allow us to describe the functionalities of the robot awareness in terms of the language of geometry.

It has been debated if visual perception is based on a 3D representation, as presupposed by Marr (Marr 1982). In the present architecture, we maintain the Marrian approach, according to which our *knoxel* corresponds to a moving 3D shape.

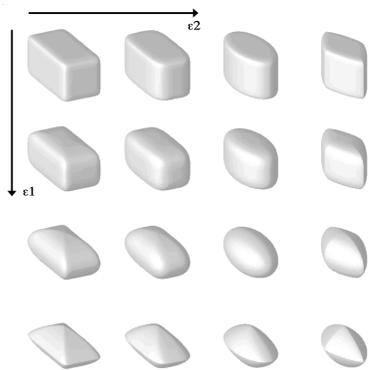


Figure 2: Superquadric shapes obtained by changing the form factors.

## Object and Scene Representation

In (Chella et al. 1997) we assumed that, in the case of static scenes, a *knoxel*  $k$  coincides with a 3D primitive shape, characterized according to Constructive Solid Geometry (CSG) schema. In particular, we adopted *superquadrics* (Jaklič et al. 2000) as the primitive of CSG. Superquadrics allow us to deal with a compact description of the objects in the perceived scene. This approach is an acceptable compromise between the compression of information in the scene and the necessary computational costs. Moreover, superquadrics provide good expressive power and representational adequacy.

Superquadrics are geometric shapes derived from the quadric parametric equation with the trigonometric functions raised to two real exponents. Fig. 2 shows the shape of a superquadric obtained by changing its form factors.

In order to represent composite objects that cannot be reduced to single *knoxels*, we assume that they correspond to groups of *knoxels* in CS. For example, a chair can be naturally described as the set of its constituents, i.e., its legs, its seat and so on.

Fig. 3 (left) shows a hammer composed by two superquadrics, corresponding to its handle and to its head. Fig. 3 (right) shows a picture of how hammers are represented in CS. The concept hammer consists of a set of pairs, each of them is made up of the two components of a specific hammer, i.e., its handle and its head.

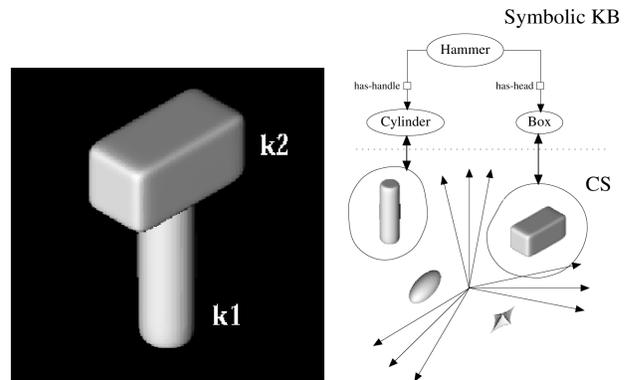


Figure 3: A hammer made up by two superquadrics and its representation in the conceptual space.

## Dynamic scenes

In order to account for the perception of dynamic scenes, we choose to adopt an intrinsically dynamic conceptual space. It has been hypothesized that simple motions are categorized in their wholeness, and not as sequences of static frames. In other words, we assume that simple motions of geometrically primitive shapes are our perceptual primitives for motion perception.

In our dynamic conceptual space, a knoxel now corresponds to a “generalized” simple motion of a superquadric. By “generalized” we mean that the motion can be decomposed in a set of components each of them associated with a degree of freedom of the moving superquadric.

A way of doing this, is suggested by the well known Discrete Fourier Transform (DFT). Given a parameter of the superquadric, e.g.,  $\mathbf{a}_x$ , consider the function of time  $\mathbf{a}_x(t)$ ; this function can be seen as the superimposition of a discrete number of trigonometric functions. This allows the representation of  $\mathbf{a}_x(t)$  in a discrete functional space, whose basis functions are trigonometric functions.

By a suitable composition of the time functions of all superquadric parameters, the overall function of time describing superquadrics parameters may be represented in its turn in a discrete functional space. We adopt the resulting functional space as our dynamic conceptual space. This new CS can be taught as an “explosion” of the space in which each main axis is split in a number of new axes, each one corresponding to a harmonic component. In this way, a point  $\mathbf{k}$  in the CS now represents a superquadric along with its own simple motion. This new CS is also consistent with the static space: a quiet superquadric will have its harmonic components equal to zero.

In Fig. 4 (left) a static CS is schematically depicted; Fig. 4 (right) shows the dynamic CS obtained from it. In the CS on the left, axes represent superquadric parameters; in the rightmost figure each of them is split in the group of axes, that represent the harmonics of the corresponding superquadric parameter.

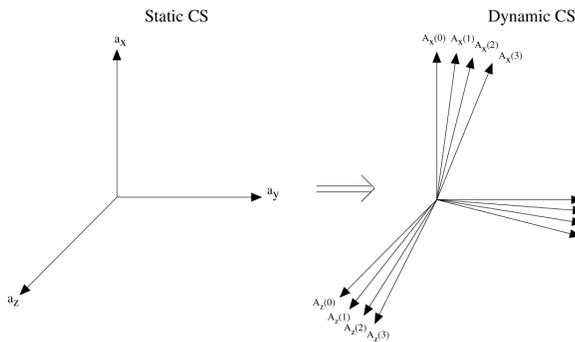


Figure 4: An evocative, pictorial representation of the static and dynamic conceptual spaces.

### Situations and Actions

Let us consider a scene made up by the robot itself along with other entities, like objects and persons. Entities may be approximated by one or more superquadrics. Consider the robot moving near an object. We call *situation* this kind of scene. It may be represented in CS by the set of the

knoxels corresponding to the simple motions of its components, as in Fig. 5 (left) where  $\mathbf{k}_a$  corresponds to an obstacle object, and  $\mathbf{k}_b$  corresponds to the moving robot.

A situation is therefore a configuration of knoxels that describe a state of affairs perceived by the robot. We can also generalize this concept, by considering that a configuration in CS may also correspond to a scene imagined or remembered by the robot.

For example, a suitable imagined situation may correspond to a goal, or to some dangerous state of affairs, that the robot must figure out in order to avoid it. We added a binary valuation that distinguish if the knoxel is effectively perceived, or it is imagined by the robot. In this way, the robot represents both its perceptions and its imaginations in conceptual space.

In a perceived or imagined situation, the motions in the scene occur simultaneously, i.e., they correspond to a single configuration of knoxels in the conceptual space.

To consider a composition of several motions arranged according to a temporal sequence, we introduce the notion of *action*: an action corresponds to a “scattering” from one situation to another situation in the conceptual space, as in Fig. 5 (right).

We assume that the situations within an action are separated by instantaneous events. In the transition between two subsequent configurations, a “scattering” of some knoxels occur. This corresponds to a discontinuity in time that is associated to an instantaneous event.

The robot may perceive an action passively when it sees some changes in the scene, e.g., a person in the robot environment changing her position. More important, the robot may be the actor of the action itself, when it moves or when it interacts with the environment, e.g., when it pushes an object. In both cases, an action corresponds to a transition from a situation to another.

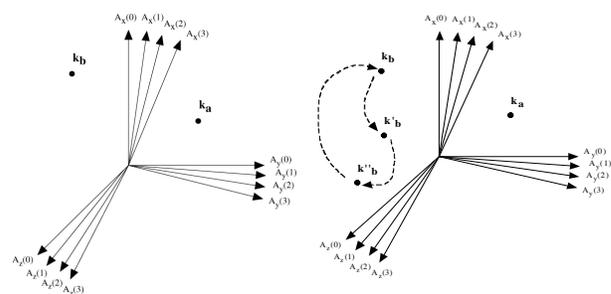


Figure 5: An example of situation and action in CS.

### Linguistic Area

The representation of situations and actions in the linguistic area is based on a high level, logic oriented formalism. The linguistic area acts as a sort of “long term memory”, in the sense that it is a semantic network of symbols and their relationships related with the robot

perceptions and actions. The linguistic area also performs inferences of symbolic nature.

In the current implementation, the linguistic area is based on a hybrid KB in the KL-ONE tradition (Brachman and Schmoltze 1985). A hybrid formalism in this sense is constituted by two different components: a terminological component for the description of concepts, and an assertional component, that stores information concerning a specific context.

In the domain of robot actions, the terminological component contains the description of relevant concepts such as *Situation*, *Action*, *Time\_instant*, and so on.

In general, we assume that the description of the concepts in the symbolic KB is not completely exhaustive. We symbolically represent only that information that is necessary for inferences.

The assertional component contains facts expressed as assertions in a predicative language, in which the concepts of the terminological components correspond to one argument predicates, and the roles (e.g. *precond*, *part\_of*) correspond to two argument relations.

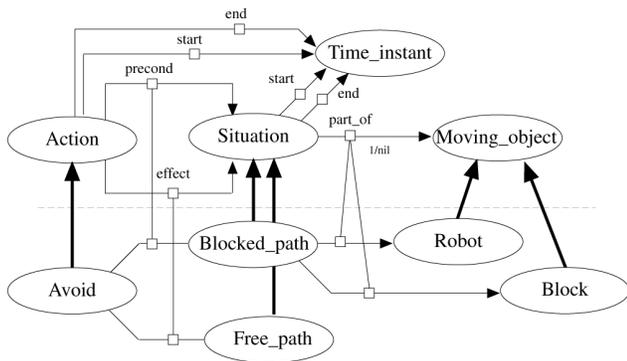


Figure 6. A fragment of the adopted KB.

Fig. 6 shows a fragment of the terminological knowledge base; in order to increase the readability, we adopted a graphical notation of the kind used by (Brachman and Schmoltze 1985). In the upper part of the figure some highly general concept is represented. In the lower part, the *Avoid* concept is shown, as an example of the description of an action in the terminological KB.

Every *Situation* has a starting and an ending instant. So, the concept *Situation* is related to *Time\_instant* by the roles *start* and *end*. A *Robot* is an example of a moving object. Also actions have a start instant and an end instant. An *Action* involves a temporal evolution (a scattering in CS). *Actions* have at least two parts, that are *Situations* not occurring simultaneously: the *precond* (i.e., the precondition) and the *effect* of the action itself.

An example of *Action* is *Avoid*. According to the KB reported in the figure, the precondition of *Avoid* is a *Blocked\_path* situation, to which participate the robot and

a blocking object. The effect of the *Avoid* action is a *Free\_path* situation.

In general, we assume that the description of the concepts in the symbolic KB (e.g. *Blocked\_path*) is not completely exhaustive. We symbolically represent only that information that is necessary for inferences.

The assertional component contains facts expressed as assertions in a predicative language, in which the concepts of the terminological components correspond to one argument predicates, and the roles (e.g. *precond*, *part\_of*) correspond to two argument relations. For example, the following predicates describe that the instance of the action *Avoid* has as a precondition the instance of the situation *Blocked\_path* and it has as an effect the situation *Free\_path*:

```
Avoid(av1)
precond(av1, bl1)
effect(av1, fr1)
Blocked_path(bl1)
Free_path(fr1)
```

The linguistic area is the area where the robot interacts with the user: the user may performs queries by using the symbolic language in order to orient the actions, for example, the user may ask the robot to search for an object. The user may performs queries by using the symbolic language in order to orient the actions, for example, the user may ask the robot to search for an object.

Moreover, the system may generate assertions describing the robot current state, its perceptions, its planned actions, and so on. However, the terms in our linguistic area are strictly “anchored” to knoxels in the conceptual area, in the sense that the meaning of the terms in the linguistic area is represented by means of the corresponding knoxels in the conceptual area. Therefore, in our architecture, symbolic terms are strictly related with the robot visual perceptions. The role of language is to “summarize” the dynamics of the knoxels at the conceptual area.

## Artificial Qualia

It has been questioned if robot may have “qualia”, i.e., qualitative, phenomenal experience. In our opinion it should be more correct to talk about the robot “artificial qualia” in the sense of (Aleksander 1996), so the problem is: during its mission tasks, has the robot some phenomenal experience?

In the proposed architecture, we have shown that the basis of the robot perception is the conceptual area, where the perceived scene is represented in terms of knoxels that describe the shape and the motion of the perceived entities, and the linguistic area where the scene is represented in terms of linguistic entities that summarize the dynamics of the knoxels in the conceptual space.

Now, we introduce an *iconic* area where a 2D reconstruction of the scene is built as a geometric projection of the knoxels in its conceptual space (where the

information about the scene is maintained in 3D) and from the data coming from sensors and processed by the subconceptual area.

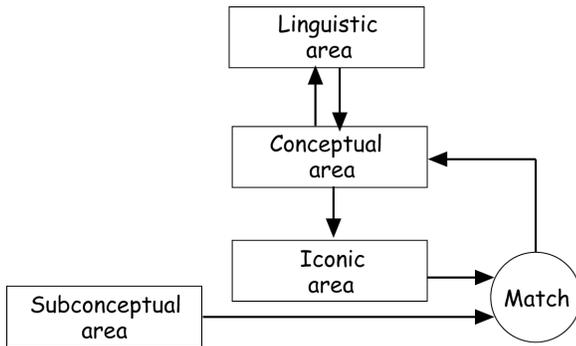


Figure 7. The revised architecture with the 2D iconic area and the perception loop.

Fig. 7 shows the robot architecture revisited to take into account the iconic area: in the revised architecture there is a perception loop between the conceptual area where the knoxels are represented, the iconic area and the subconceptual area. This perception loop has the role to adjust the match between the 2D iconic representation of the scene obtained from the knoxels in the conceptual area, and the external flow of perception data coming out from the subconceptual area.

We present the operation of the revised architecture with reference to the *CiceRobot* robotic project, an operating autonomous robot performing guided tours at the Archaeological Museum of Agrigento (Chella & Macaluso 2008).

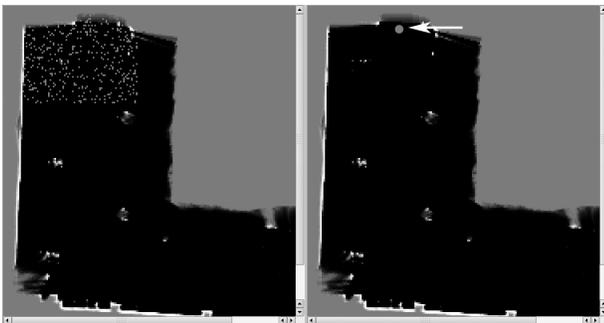


Figure 8. The initial distribution of expected robot positions (left), and the cluster of winning expected positions, highlighted by the arrow.

In order to compare the CS content with the external environment by the iconic area, the robot is equipped with a stochastic match algorithm based on particle filter (see, e.g., Thrun et al. 2005; details are reported in Chella &

Macaluso 2008). In brief, the algorithm generates a cloud of hypothesized possible positions of the robot (Fig. 8). For each position, the corresponding expected image scene is generated in the iconic area by means of geometric projection operations of the corresponding knoxels in CS. The generated image is then compared with the acquired image (Fig. 9).



Figure 9. The 2D image from the robot video camera (left) and the corresponding 2D image generated in the iconic area by the knoxels of CS (right).

An error measure  $\epsilon$  of the match is computed between the expected and the effective image scene. The error  $\epsilon$  weights the expected position under consideration by considering the distribution of the vertical edges in the generated and the acquired images (mathematical details in Chella & Macaluso 2008). In subsequent steps, only the winning expected positions that received the higher weight are taken, while the other ones are dropped.

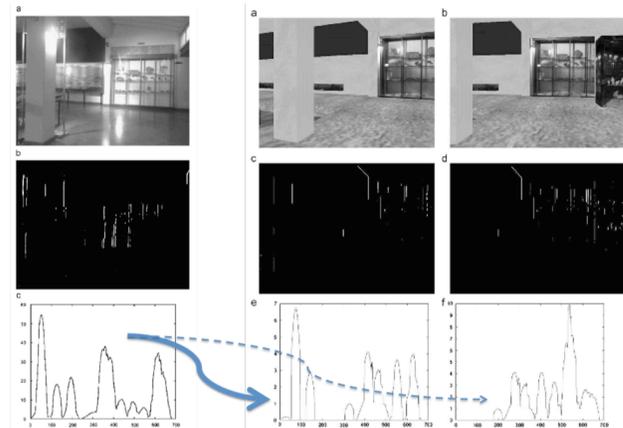


Figure 10. The image acquired by the robot camera along with the vertical edges and their distribution (left). The simulated image from CS corresponding to the hypothesis with the highest weight. (center) A simulated image corresponding to an hypothesis with a lesser weight. (right).

When the image generated in the iconic area matches with the image acquired from the robot camera, the knoxels in CS corresponding to the winning image are highlighted: they give rise to the description of the perceived scene by

means of the knowledge stored in the CS and the linguistic area, as described in previous Sects. As an example, in the situation described in Fig. 9, the assertional component of the KB generates the predicates stating that the robot is in a free path, and there is an armchair in front and on the right, and there is a column on the left:

```
Free_path(p1)
Armchair(a1)
Armchair(a2)
Column(c1)
Right_of(robot, a2)
Front_of(robot, a1)
Left_of(robot, c1)
```

We propose that the described reconstruction and match process constitutes the phenomenal experience of the robot, i.e., what the robot sees at a given instant. This kind of seeing is an active process, since it is a reconstruction of the inner percept in ego coordinates, but it is also driven by the external flow of information. It is the place in which a global consistency is checked between the internal model and the visual data coming from the sensors (Gaglio et al. 1984).

The synthesized pictures of the world so generated projects back in the external space the geometrical information contained in the knoxels in the conceptual space and, matched to incoming sensor data, it accounts for the understanding of the perceptive conscious experience. There is no need for a homunculus that observes it, since it is the ending result of an active reconstruction process, which is altogether conscious to the robot, which sees according to its own geometric (not yet linguistic) interpretation.

The phenomenal experience is therefore the stage in which the two flows of information, the internal and the external, compete for a consistent match by the particle filter algorithm. There a strong analogy with the phenomenology in human perception: when one perceives the objects of a scene he actually experiences only the surfaces that are in front of him, but at the same time he builds a geometric interpretation of the objects in their whole shape. In “gestaltian” terms, the robot in the described example perceives the whole armchairs and columns and not their visible sides only.

## Conclusions

According to the “quantum reality hypothesis” proposed by (Goertzel 2006), the described conceptual space has some similarities with the Goertzel *internal virtual multiverse*, in the sense that the CS is able to generate possible worlds and possible sequences and branches of events. Also the described match operation between the image acquired by the camera and the 2D reconstructed image from the iconic area may be seen as a sort of “collapse” of the several possible situations and actions in CS to a single perceived situation. Related ideas have been

proposed by Edelman (1989), Humphrey (1992) and Grush (2004).

The described model of robot perceptual phenomenology highlights open problems from the point of view of the computational requirements. The described architecture requires that the 3D reconstruction of the dynamic scenes and the match with the scene perceived by the robot during its tasks should be computed in real time. At the current state of the art in computer vision and computer graphics literature, this requirement may be satisfied only in case of simplified scenes with a few objects where all the motions are slow.

However, we maintain that our proposed architecture is a good starting point to investigate robot phenomenology. As described in the paper it should be remarked that a robot equipped with artificial phenomenology performs complex tasks better and more precisely than an “unconscious” reactive robot.

## References

- Aleksander, I. 1996. *Impossible Minds: My Neurons, My Consciousness*. London: Imperial College Press.
- Brachman, R.J. and Schmoltze, J.C. 1985. An overview of the KL-ONE knowledge representation system, *Cognitive Science*, 9, pp. 171-216.
- Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chella, A. and Macaluso, I. 2008. The Perception Loop in CiceRobot, a Museum Guide Robot, *Neurocomputing*, doi:10.1016/j.neucom.2008.07.011.
- Chella, A., Frixione, M. and Gaglio, S. 1997. A cognitive architecture for artificial vision, *Artificial Intelligence*, 89, pp. 73-111.
- Chella, A., Frixione, M. and Gaglio, S. 2000. Understanding dynamic scenes, *Artificial Intelligence*, 123, pp. 89-132.
- Edelman, G. 1989. *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.
- Gaglio, S., Spinelli, G. and Tagliascio, V. 1984. Visual Perception: an Outline of a Generative Theory of Information Flow Organization, *Theoretical Linguistics*, 11, Vol. 1-2, pp. 21-43.
- Goertzel, B. 2006. *The Hidden Pattern – A Patternist Philosophy of Mind*. Boca Raton: BrownWalker Press.
- Grush, R. 2004. The Emulator Theory of Representation: Motor Control, Imagery and Perception, *Behavioral Brain Sciences*, 27, pp. 377-442.
- Gärdenfors, P. 2000. *Conceptual Spaces*. Cambridge, MA: MIT Press.
- Humphrey, N. 1992. *A History of Mind*. New York: Springer-Verlag.
- Jaklič, A., Leonardis, A., Solina, F. 2000. *Segmentation and Recovery of Superquadrics*. Boston, MA: Kluwer Academic Publishers.
- Marr, D. 1982. *Vision*. New York: W.H. Freeman.
- Thrun, S., Burgard, W., Fox, D. *Probabilistic Robotics*. Cambridge, MA: MIT Press, 2005.