Philosophical Problems in Science (2025), **79**, 3–19
(Zagadnienia Filozoficzne w Nauce)
doi:10.59203/zfn.79.732

ARTICLE

The problem of scene understanding by artificial autonomous agents

Marzena Bielecka* and Andrzej Bielecki

AGH University of Krakow
*Corresponding author. Email: bielecka@agh.edu.pl

Abstract
In this paper, the problem of scene understanding in the context of autonomous robotics is considered. The implemented solutions, based on various types of artificial intelligence systems, are described and the perspectives of their development are discussed. In particular, the problem of cognition that reflects the structure and nature of the environment in which the artificial autonomous agent operates, is studied. The specifics of scene models and algorithms for their construction and verification of their adequacy are discussed. The presentation of the existing partial solutions, including the results obtained by the authors, is the starting point of the analysis. Possibilities of new approaches are analysed, especially those that are biologically inspired. The potential fundamental limitations are studied. These problems are discussed in the context of epistemology and in the frame of cybernetics. The analysis of the relationship between philosophy and artificial intelligence in the aspect of cognitive vision systems of autonomous robots is the aim of this paper.

Keywords: epistemology, autonomous agent, scene understanding, cognition, robotic autonomy

1. Introduction

Until recently, the human being was the only cognitive subject analysed on philosophical grounds in the context of epistemology. However, the rapid development of artificial intelligence implies, among others, its wide application in embodied systems such as artificial autonomous agents, for instance, autonomous robots and autonomous unmanned aerial vehicles (Bielecki 2024b; Floreano and Wood 2015). Currently, artificial autonomous agents are characterized by only limited autonomy. A prerequisite for achieving full autonomy is adequate modelling of cognitive functions that are associated with sensory, creating, for example, the cognitive vision system (Tadeusiewicz 1992). The full understanding of the environment in which the agent operates is one of the aspects of the said functions. Such understanding consists of recognizing individual objects, their mutual spatial relationships, linking them with the appropriate semantics and cognitive models the agent is equipped with, for instance, by comparing the perceived scene with an incorporated terrain map. It also includes the integration of signals from various types of sensors and the creation of environment models that will be a reference point in the process of understanding the scene (Katsev et al. 2011; Li et al. 2017; Wang et al. 2017). The experience gained so far allows us to assume that drawing ideas from biological prototypes may be an effective approach to developing systems for understanding the environment (Hu et al. 2017; Siagian and Itti 2009; Tejera et al. 2018) which, among others, can be a starting point for human-robot interactions (Kang and Han 2023; Quintas 2023; Umbrico et al. 2023).

The development, implementation, and verification of efficiency of robotic cognitive systems can shed new light on classic philosophical issues. These include the concept of truth and its verification, the reference of the cognitive subject to physical reality, information processing within the system in the context of the nature of information itself, and testing the adequacy of the model from an ontological and functionalist perspective. These problems are situated in an intensively explored stream of studies that concerns the relationship between philosophy and technology (Polak 2023).

While contemporary advancements in artificial intelligence have led to the emergence of sophisticated architectures, such as vision—language models (VLM) and affordance—learning agents, these systems do not necessarily circumvent the fundamental epistemic barriers discussed in this work. Modern VLMs, for instance, despite their high performance in object categorization, primarily operate on statistical correlations between visual and linguistic patterns. From a Kantian perspective, as analysed in this paper, they remain within the realm of establishing relationships between phenomena without achieving the capacity for autonomous conceptualization—the ability to create the idea behind the category. Furthermore, although affordance-learning agents represent a significant step toward functional scene analysis, they often lack a formal, transparent mechanism for verifying the adequacy of their internal models. The cybernetic approach presented here suggests that genuine understanding requires the agent to not only detect a functional property, but also to verify it through a goal-oriented process where predicted internal states are compared with achieved states. Without this self-reflective loop and the capacity for conceptualization, even the most advanced contemporary systems remain at the level of complex signal processing and pattern recognition rather than achieving the higher-order scene understanding characteristic of cognitive subjects.

The paper is organised in the following way. In the next section, the problem of robotic cognition is discussed using robotic cognitive vision systems as an example. In particular, the levels of cognition are specified in the frame of cybernetics. Furthermore, the open problems are presented. In Section 3, the current status of the scene understanding problem is discussed on the basis of selected examples of vision cognitive systems of the robots as well as the problem of learning. Philosophical aspects of the problem are analysed in Section 4.

2. The formulation of the problem

An autonomous cybernetic agent operating in a given environment has to be equipped with the system of sensors. The signals detected by sensors are then processed and analyzed at various levels. From a cybernetic point of view, the following levels in this analysis should be specified:

1. **Signal processing.** Signals received from the agent's sensor network, especially when they are complex in nature, such as camera images, must be pre-processed in order to

enhance or extract their essential features. Denoising, contrasting, binarization, skeletonization can be given as examples of such operations.
2. **Signal analysis** is the lower level of sensor data analysis. It allows the agent, for example, to quickly detect changes in a data stream. The detection of motion on the image from a camera monitoring the surroundings of the house can be put as an example.
3. **Simple classification.** At this level, the object is classified into one of the predefined classes on the basis of a feature that can be verified directly. Classification of a fruit as ripe or unripe based on its color is an example of such classification (Kurpaska et al. 2021).
4. **Pattern recognition.** At this level, a structure of single objects is analyzed. First of all, a representation of the pattern should be defined, usually formally. Then, the features that characterize the object are required. Usually, the preselected features are transformed into the more optimal form, for instance by reducing their dimension. The selection of the features, that have the most discriminative power and are sufficient to describe the recognized objects unequivocally, is the last stage of the recognition system design—see (Flasiński 2016, Section 10.1). In complex objects, individual features often have a structural character and in order to specify them for a given object, a structural analysis of the object should be performed, for instance an analysis of the contour of its image. Such complex analysis is often done by using syntactic methods—see (Flasiński 2019, Section 1.3). In order to analyze a given object, the preprocessing of the image of the object must be done. It consists in, for instance, normalization and scaling. Usually, classification of an object is the result of pattern recognition.
5. **Scene analysis** includes the identification of objects and the relationships between them. In the case of complex scenes, it covers complex issues that require multi-level analysis. An example is the problem of deciding whether a given line is the edge of an object or the edge of its shadow. For this purpose, it is necessary to analyze the positions of light sources and obstacles that reflect, absorb or scatter light. Scene analysis is a necessary step to be able to move on to the level of understanding the scene.
6. **Scene understanding** consists in associating its structure with certain meanings. They can be both structural and functional. For example, detecting the presence of other agents may mean the possibility of establishing contact with them, forming a coalition in order to perform the task collectively, but it may also mean a threat if the analyzed scene is a war zone.

One subtlety should be noticed. Namely, the division into objects (pattern recognition level) and the scene is common in computer science in the context of artificial intelligence. From a philosophical point of view, it is ontological in nature and is, in a way, arbitrary. The related philosophical problems are analysed in (Krzanowski and Polak 2022).

At the current stage of development of computer science, in particular artificial intelligence systems, the algorithms implementing points 1–3 are well developed, and the algorithms implementing points 4–5 are able to solve a wide class of tasks quite effectively. However, the algorithms implementing point 6 are in the very early stages of development. Accordingly, the following questions arise:

1. What are the perspectives for the development of scene understanding systems?
2. To what extent can systems of this type be based on the modelling of structures and processes in biological agents? In particular, how helpful in the development of scene

understanding algorithms can be the achievements of psychology? These two questions are legitimate and interesting from the scientific point of view in the light of the fact that the problem of understanding the scene is closely related to the autonomy of the agent, and this, in turn, is discussed in the context of biology. It should be noted that an in-depth analysis of the autonomy of biological agents is available (Rosslenbroich 2014) as well as the problem of biological basis of the autonomy, first of all development of neural systems and, as a consequence, sensorics and information processing (Scott 1995; Tadeusiewicz 2010).

3. What are the possible limitations in artificial systems of this type?

Discussing these issues in the context of artificial autonomous agent vision systems, is presented in the next section. Analysis is based on exemplary vision systems of robots and is conducted in the context of epistemology.

3. Cognitive scene understanding by artificial agents—the current status and immediate prospects

In this section, the problem of cognition that reflects the structure and nature of the environment in which the artificial autonomous agent operates is studied. In general, the philosophical systems in which truth as such is analyzed can be divided into *closed*, such as mathematics, existing *for itself* and open, i.e., those that operate in a certain environment, which is their *world*. In open systems, truth is related to learning about the world. Creating the models of the world is an effect of this learning. Cybernetics provides tools that shed significant and new light on the procedures of cognition by autonomous (in the cybernetic sense) agents, creation of models of the world, and verification of these models by autonomous agents that are the cognitive subjects.

In this section, we discuss the specifics of scene models and algorithms for their construction and verification of their adequacy. The presentation of the exemplary existing partial solutions is the starting point of the analysis. The possibilities of their development, generalization and integration are then discussed.

3.1 Cognition

The creation of an adequate model of the analyzed scene or, more generally, a model of the environment in which the agent operates is one of the crucial steps in the problem of the scene understanding. If the agent is fully autonomous, it must create the said models on its own. The models can be created on the basis of the signals obtained from the environment as well as on the cognitive frames defined in the agent's cognitive module. The model creation includes the following steps.

1. **Creating models of individual objects** by using the information created on the basis of signals received from the agent's sensory network. An important and often overlooked problem is the generation of information based on the acquired signals. A significant problem is the lack of an adequate theory of information. Existing theories, known as information theories (for instance, Shanon information theory, Kolmogorov information theory), are generally concerned with a method of measuring or calculating the amount of information rather than its generation or its nature, which is essential for future

development of algorithms for generating information from signals. However, it should be noted that contemporary frameworks, such as active inference and information-theoretic active perception, provide a bridge between signal processing and semantic understanding. In these approaches, information is not merely a passive input but a result of the agent's actions aimed at minimizing uncertainty (entropy) regarding its environment (Bajcsy, Aloimonos, and Tsotsos 2018; Friston 2009). Consequently, the algorithm for active investigation of the object by an agent can be viewed as a mechanism for maximizing information gain. This process allows the agent to update its internal cognitive frames and verify the adequacy of its scene models through purposeful interaction, effectively linking action with semantic exploration. The concept of structural information (Bielecka et al. 2025; Bielecki and Schmittel 2022; Bielecki and Stocki 2023), as well as the currently discussed theories of physical information (Barreiro et al. 2020; Krzanowski 2020, 2022; Roterman and Konieczny 2023), are a good starting point for the analysis of the problem of information generation. Such an algorithm should be based on active investigation of the object by an agent by using its sensors. After integrating the signals from various sensors, the model would be built. Such an approach was tested both for the two-dimensional and three-dimensional models in the context of unmanned aerial vehicles and primary results are promising (Bielecki et al. 2016; Bielecki, Buratowski, and Śmigielski 2013, 2014). For instance, let us briefly recall the algorithm for creating a 3-D representation of a single building by an unmanned aerial agent (Bielecki, Buratowski, and Śmigielski 2014). It is assumed that the agent operates in an urban-type area. The general idea is to construct the set of walls that constitutes 3-dimensional model of the analysed object. To enrich the model and make it more complete, the algorithm also creates representation of holes in the investigated building. To create 3-dimensional representation, the proposed algorithm is based on the idea of composing the solid from projections that consists of photographed and vectorized sides of a building. One photo is taken from above in order to show the top side of the investigated building. Next, the building is photographed horizontally towards half of the sides represented by the convex hull of the top side shape. This allows the agent to detect holes that are regarded as the object features. Such an algorithm allows the agent to create the solid representation. The mentioned integration may include, in addition to the image from the camera, also e.g., lidar signals.

2. **Modeling the spatial relationships between individual objects** is a key element of scene representation, and thus its analysis and, as a consequence, understanding. The said spatial relations are usually modelled by using graph-based models (Flasiński 2019), section 12.1, sometimes aided with algebraic methods (Chaib-draa 2002). Let us consider the case where the agent flies at high altitude—see Bielecki and Śmigielski (2017) for details. Since the artificial flying agent operates at high altitudes, from its point of view, the studied scene is two-dimensional. It is assumed that the robot operates in the urban-type environment. The scene is represented as a neighborhood graph that encodes data about the buildings shapes, locations, and spatial relations. The robot understands parts of the examined scene by analyzing the context of the objects' neighborhoods. Such a scene representation can be used effectively for recognition of the object, while a few objects of the same shape exist in the scene because during the recognition process of an individual building, not only the information about the shape of the object is exploited, but also the spatial relations with other objects in its close neighborhood are analyzed.

The assumption is that the flying artificial autonomous agent has to find the concrete building in real urban-type environment. A bitmap image that represents the scene photographed from high altitude is used as the reference representation of the scene. The bitmap is not taken by the agent itself, but it is previously uploaded to the agent's memory and is used by the agent during the scene examination. The agent takes photos of the terrain beneath. Then, representations of single objects as well as the neighbour graph are created. Each single object is compared with the one that is marked as the wanted one on the previously uploaded map. Then, a comparison of the graph fragments is performed that represent the neighbour closest to the analyzed objects pre-selected as identical to the searched one on the created map and the neighbourhood on the object on the uploaded. If the compared neighborhoods are different, then the object is classified as different from the searched one, despite the identical shape. If only one object in the analyzed scene has the same shape and the closest neighborhood as the searched object, it is classified as the searched object. If more than one object in the analyzed scene has the same shape and the closest neighbourhood as the searched object, then in the graph, not only the nearest neighbourhood of a given object is taken into account, but also the nearest neighbours of its nearest neighbours and the graphs obtained in this way are compared. In a such way, the extended neighbour graph is obtained. If necessary, the neighbour graph is extended by the neighbourhoods of subsequent neighbors.
It should be mentioned, that the possibilities of the scene representation go beyond the tools currently used in artificial intelligence. For instance, by rubbing their antennae, ants are able to convey information about the way to the food source, at least in the case when the route has the structure of a branched tree (Ryabko and Reznikova 2009, 2012). It should be stressed that the experiment was designed in such a way that a predefined food route was set up on the water so that the ants could not deviate from it, and after a single scout ant found food, the route was replaced so that the ants could not use the pheromone trail. Other ants were let into the litter box with the scout that returned. After exchanging information, the scout was removed from the litter box, and the ants went straight to the food without fail. This means that the ants can accurately encode the crucial features of the scene although it is not yet known what the nature of this scene representation is. Bees also have specific capabilities related to learning about the environment and transferring information about its structure (Frish 1954; Frish and Lindauer 1956). Namely, bees see in ultraviolet light, and for orientation in space, they use observations of the angle of incidence of sunlight and the polarization of light. Bees use a complex information system an important element of which is the ability to encode information about the location of the food source by means of certain characteristics of the flight of a single bee in the vicinity of the mother hive— the orientation of the circular movements indicates the direction in which the food is located. Although the way of communication of bees is much better studied than the way of communication of ants described above, so far it has not become an inspiration for the development of scene representation algorithms. Similarly, the way of perceiving the environment by dogs in which olfactory stimuli are the basic source of information about the world has not yet been studied in the context of stimuli integration in order to generate an integrated model of the scene.

3. **The analysis of the structure of single objects** is crucial for giving meanings to them (see the next point). In the context of the harvesting robot (Kurpaska et al. 2021), discussed in more detail in the next point, this type of analysis may concern a single fruit in the context of fruits sorting during intelligent harvesting. The strawberry can be mechanically damaged, for example, by biting by a snail. In such a case it changes shape without changing colour. Applying the syntactic contour analysis by using string languages of Jakubowski type (Bielecka 2018; Jakubowski 1990) enabled structural description of damaged and undamaged fruit and, as a consequence, successful qualification (Kurpaska et al. 2021).
4. **Giving meaning to individual objects** involves, among other, assigning specific functionalities to the objects, e.g., whether the analyzed object is a residential building or a defensive bunker, whether a small mound is a natural slight elevation or a large termite mound, etc. Let us analyze the issue more precisely in the context of the vision system of the autonomous robot for harvesting strawberries - see (Kurpaska et al. 2021) for details. In the context of the said robot, the assignment of meanings may, for example, consist in determining whether the detected human figure is a human standing on the robot's collision course, or just a shadow of a human that the robot can overrun. In the context of navigation, it is also important whether the detected obstacle is fixed to the ground or is a light object, e.g., a bucket, that can be moved. The vision system of the robot should perform two various tasks that are mutually independent. First of all, it should ensure navigation on the scene. The robot was supposed to operate in large greenhouses, so it could be assumed that it was a human-made environment. Therefore, the previously developed methodology of scene representation using a neighbourhood graph (Bielecki and Śmigielski 2017) was adapted to represent the scene. The only difference is that the robot operates on limited area. It should be mentioned that the field harvesting robot faces slightly different challenges in the context of the scene representation as well as the trajectory planning and optimizing (Jia et al. 2020). The scene representation is used by the robot to plan its trajectory. An important element is the anti-collision module, which allows for quick recognition of people and animals that the robot may hit while moving. Precise determination of the location of the fruit on the plant and inside is the second task. Furthermore, the analysis of its ripeness phase will have to be done. Furthermore, since the harvesting is assumed to be intelligent, it is necessary to determine whether the fruit is rotten, mouldy, or mechanically damaged. As a consequence, the system should make a decision if the fruit is suitable for collecting or should be discarded due to damage or disease. The scheme of the cognitive vision module of the robot is presented in Fig.1.
5. **The analysis of the relationship between the structural and functional properties of objects and the functional conditions of the agent,** especially in the context of the task being performed and the achievement of the assumed goals, is the next level of the scene understanding. Let us discuss this problem using the example of an autonomous flying agent (Bielecki and Śmigielski 2021). Identifying whether the structure of any examined objects allowed the robot to have collision-free flight under or through the object was one of the tasks performed by the robot. The strategy for the performed test scenario was that the rectangular convex hull is calculated for each side of the representation of the object. Then, the agent verified whether it is possible to navigate through the space

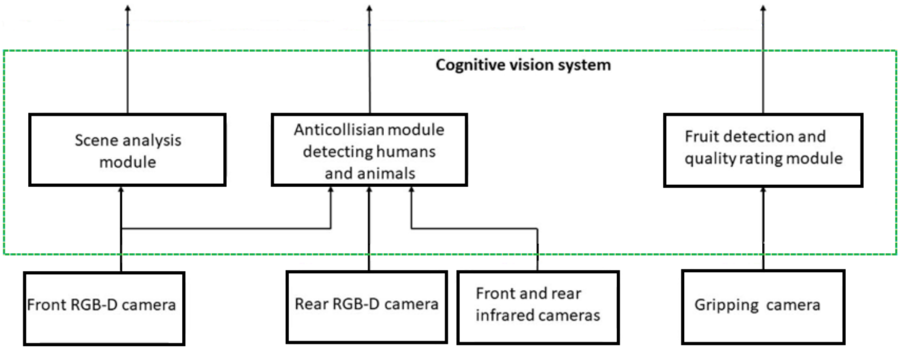

Figure 1. The scheme of the cognitive vision module of the robot for harvesting strawberries (Kurpaska et al. 2021).

bordered by this rectangle without collision with the object. Field trials positively verified the proposed approach.

6. **Analyzing the meaning of the entire scene in the context of the actions performed by the robot is the most general level of cognitive understanding of the scene.** The problem whether the analyzed scene is a fragment of a city after a catastrophe such as an earthquake or whether it is a military training ground for practicing combat tactics in the city is an example of such issue. Without doubt this level of scene understanding, characteristic of humans, is currently unapproachable for autonomous robots.

3.2 *Learning process*

The functionalities described in subsection 3.1, related to the understanding of the scene by autonomous robots, model the functionalities displayed by humans. Human cognitive abilities, in turn, are the result of the organization of the psyche at the level of cognitive structures. The organization of cognitive structures is known only superficially, and even this aspect that is known, is able to be modelled only in a very simplified way. Nevertheless, models of this type are created and their usefulness in certain applications has been positively verified. One of such models, which was created for the purpose of implementing a mechanism analogous to functional automation, is shown in Fig.2—see Bielecki (2014) for details. The phenomenon of functional automation is analysed as a process as a result of which cognitive structures are replaced by reflexive structures that realize the same task reflexively. Functional automation is a beneficial optimization mechanism when a given activity is performed with high frequency. Thus, when the supervisory module detects that a certain part of the cognitive module is used frequently and the same response pattern is generated as a result, it starts the process of automation. It is implemented in such a way that the reaction generated by the cognitive module, in which knowledge is explicitly encoded, is compared with the reaction generated by the reflex module connected in parallel, which implements the functional stimulus-response relationship. The process is terminated when the output reactions of the reflexive module are equal with sufficient accuracy, to the reactions generated by the cognitive module. Until the automation process is not stopped, the reaction of the

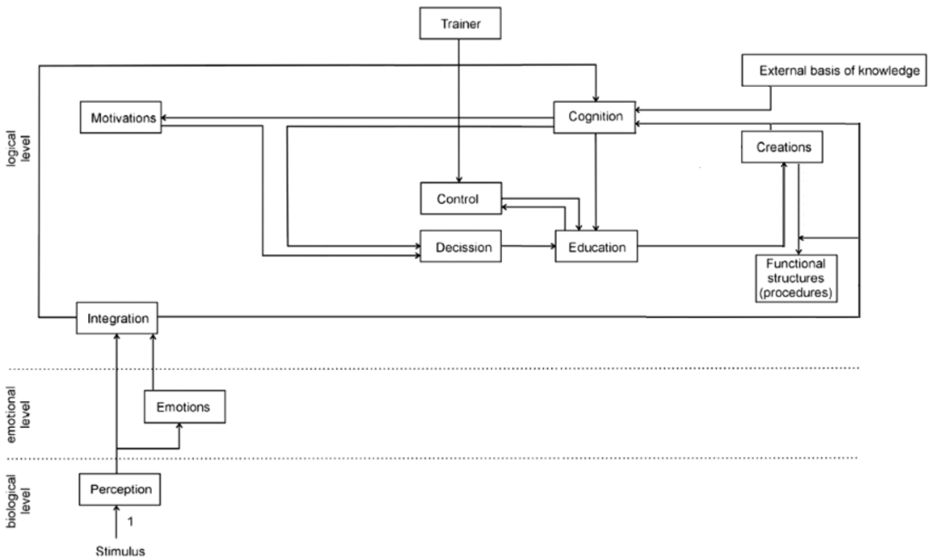

Figure 2. Scheme of the cognitive module of artificial agent. The module is based on the structure of the organization of human cognitive mechanisms related to learning (Bielecki 2014).

cognitive module is the reaction of the whole system— see Fig.3(a). When the learning process is terminated, the cognitive module is disconnected from the reactive loop which means that it does not take part in a direct processing of the input signals—see Fig.3(b). This

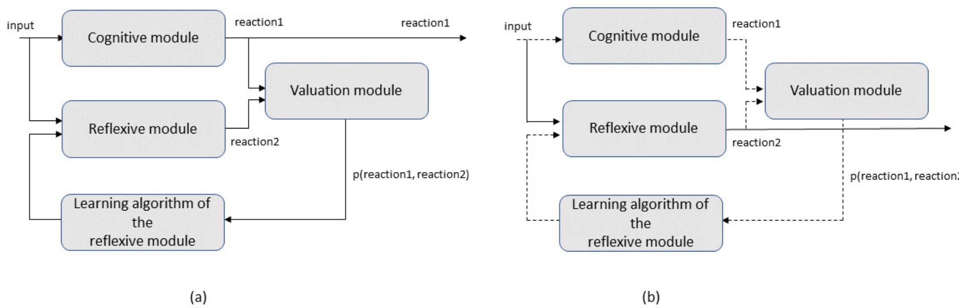

Figure 3. The model of human activity automatization (Bielecki 2014). (a) The agent before automation. (b) The agent after automation.

module becomes the monitoring unit that is activated when reflex responses are no longer satisfactory. Then, the cognitive module is reconnected to the reactive loop and the training of the reflexive module is repeated.

It should be stressed that scene understanding can be considered both in broad and narrow meaning. In the first case, the inner representation of the environment encodes it explicitly, modelling directly, for instance, spatial relations between objects – see (Bielecki and Śmigielski 2017). Structural and syntactic methods are adequate tools for creating such representations. In the broader sense, however, scene understanding can be considered as all tools that enable the agent to act effectively in the environment. Contemporary AI systems based on computational AI represent the pinnacle of functional automation as defined here. Despite their high operational efficiency, from an epistemological perspective, they remain at the level of signal correlation rather than structural understanding. The absence of explicit syntax in these models precludes the transition from reaction to notion-based cognition, which, in our hierarchy, distinguishes advanced data processing from genuine scene understanding.

4. Epistemic analysis of the robotic cognition abilities

Artificial intelligence, both in general meaning and applied to autonomous robotics, is defined as a feature of artificial systems that enables them to perform tasks that require intelligence (Flasiński 2024). It should be stressed, that intelligence is not understood narrowly, only as a property of the human psyche, but also as natural processes occurring in biological structures and systems that exhibit certain characteristics, such as, for example, the ability to learn, the ability to generalize, and adaptability. In this sense, immune systems, evolutionary processes, insect swarm behaviour, or signal processing in nervous systems, not necessarily related to conscious processes but, for example, to the formation of conditioned reflexes, are also manifestations of intelligence (Flasiński 2016). Cybernetics provides the tools to implement the following scheme for applying this type of natural solutions in artificial intelligence, particularly in autonomous robotics. Namely, the first step is to analyse the biological phenomenon in the context of its functionality. Next, we isolate those aspects that are key to ensuring this functionality. The subsequent step is to create a formal model of the given phenomenon, i.e., the process and biological structures that ensure its implementation. The final step is to implement the formal model in a computer or robotic system. Historically, the biologically inspired AI systems mentioned above were created according to this framework (Bielecki 2024a, 2024b). Let us notice that the mentioned biologically inspired AI systems do not address such concepts as logic, psychology, consciousness, or self-awareness. Nevertheless, they shed light on certain fundamental epistemological issues, for example, those related to aletheology. One of the classic aletheological problems is the problem of truth verification by cognitive entities that have reference to the physical world in which they operate. This is the problem of the correspondence theory of truth, which states that true propositions are those that are consistent with the actual state of affairs in the physical world. However, the correspondence theory of truth, by its very nature, cannot provide a direct algorithm for truth verification. First, we can compare two facts or two propositions, but not a fact with a proposition. Furthermore, significant problems arise in defining correspondence, related, among other things, to the nature of the representation of reality and the manner of expressing this representation. Scene models, created and residing in cognitive modules of autonomous agents, should be verified in terms of their adequacy. This means that for such agents the reference to physical world, considered in the frame of a certain correspondence, plays a crucial role. As a classic philosophical problem, this issue has so far been considered in the context of the human process of learning about the world, including the epistemic procedures of the natural sciences. It seems that the publication (Bielecki 2021) is the first one in which this problem is considered from the perspective of cybernetics, in particular systems theory, taking into account artificial agents as cognitive subjects. The proposed algorithm of verification of the adequacy of a model is presented in Fig.4. Verification of whether the

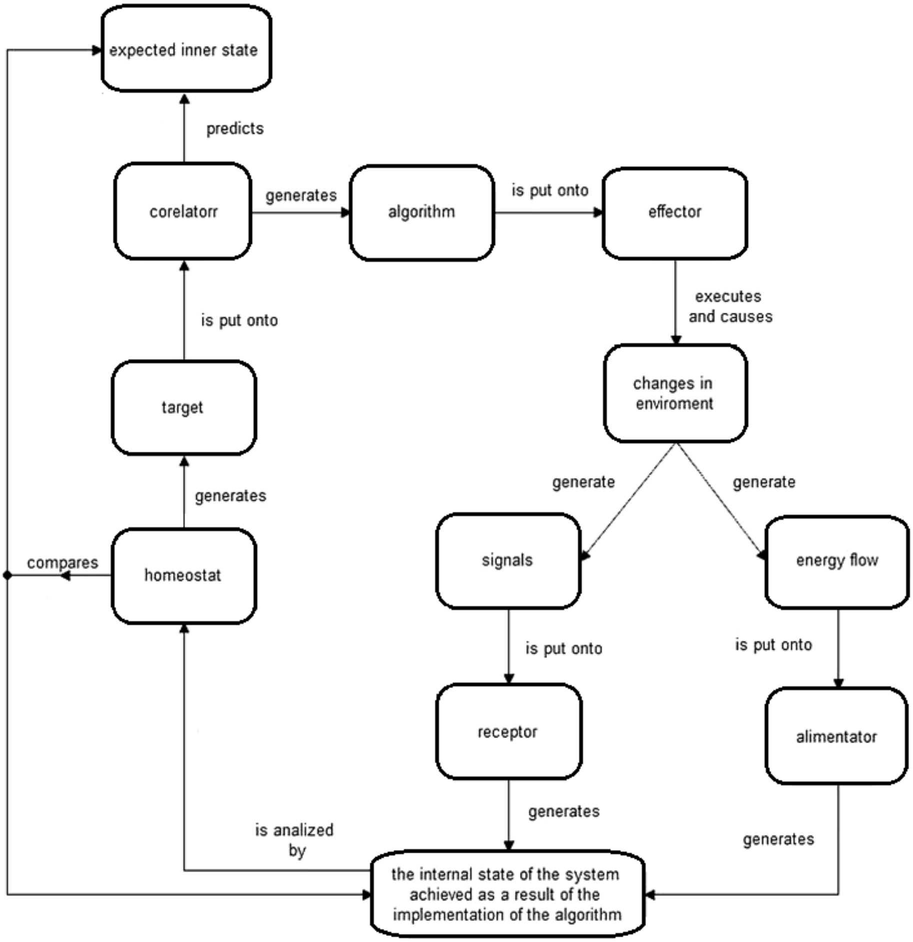

Figure 4. The algorithm of verification of adequacy of a model (Bielecki 2021).

assumed goal has been achieved is the basis of the presented concept. The precise definition of the achievement of the goal is a prerequisite for enabling the aforementioned verification. For an autonomous agent, both biological and artificial, that operates in a physical environment, achieving a goal implies both some change in the environment or the agent's relationship to the environment, as well as the achievement of a certain internal state by the agent. This internal state is predicted by the agent as one of the effects of the successful execution of the algorithm created by the agent. The advantage of the proposed approach is that the internal state achieved by the agent can be directly compared with the state that was predicted as one of the effects of the algorithm execution. On the other hand, the possibility of verifying changes in the environment is limited by the indirect perception of the environment with the help of sensors in the case of artificial agents and the senses in the case of biological agents. In other words, we do not know to what extent our senses deceive us and we do not know to what extent true information is provided by sensors. The ability to verify the internal state, however, is given to the agent directly. The algorithm presented in Fig.4 can be regarded as a cybernetic rule of inference about the adequacy of the model. Formally, this can be stated as follows—see Bielecki (2021, Section 3). The cognitive subject *S has reason to believe that* ϕ if:

- ϕ is the fragment of the model such that there exists at least one goal, there exist conditions, for instance a range of parameters, and there exists the algorithm created on the basis of the ϕ such that the goal is achievable by using the algorithm provided that the conditions are satisfied;
- ϕ is encoded in the cognitive module of the autonomous agent *S*;
- *S* performed the algorithm and, as a result, *S* achieved its goal, which manifests, among others, by achieving the inner state the agent assumed before performing of the algorithm.

It should be emphasized that while the *cybernetic truth verification loop* is presented in a formalized manner, its individual components are grounded in established paradigms of visual information processing in autonomous systems. The implementation of the scene understanding level utilizes the structure of neighbourhood graphs, which has been previously validated as an effective tool for the representation and recognition of objects within complex urban environments (Bielecki and Śmigielski 2017). Furthermore, the operational efficiency of the system in unstructured settings is supported by outdoor experimental data obtained from unmanned flying agents using monocular vision, where algorithms successfully performed three-dimensional analysis of structures under varying lighting and geometric conditions (Bielecki and Śmigielski 2021). In the present work these methods are theoretically analysed by introducing a rigorous truth verification mechanism that integrates raw sensory data with higher levels of cognitive reasoning. In a such way, the above cybernetic approach sheds new light on the problem of truth verification.

Turning to the problem of the limitations of cognitive capabilities of the autonomous robots, it should be emphasized that it is difficult to find a valuable philosophical analysis of this issue. There appear to be no methodological limitations to adapting biological mechanisms for AI and autonomous robotics. There are no fundamental problems in researching them, creating their models and implementing them. It should be also noted that both AI systems themselves and their learning algorithms can be biologically inspired—see, for example, the functional automation mechanism described in the previous chapter. It should be mentioned that these topics are included in biomimetics, called also biomimicry or bionics, that is an intensively studied branch of science (Bhushan 2009).

When it comes to psychologically inspired AI systems, however, the situation is different. A thorough discussion of the problem in the light of Aristotle, St. Thomas Aquinas and, first of all, Kant's approach was conducted in a series of works by Flasiński (1997, 2016, 2024). Let us sum up briefly his analysis. Namely, according to the philosophy of mind formulated by Kant (1838), the intellect is carefully distinguished from the reason. According to Kant, the intellect realizes the operations of the logical type, establishing relationships between phenomena and creating categories. Reason, on the other hand, is capable of creating transcendental ideas and primary principles on the basis of which greater premises are created in syllogistic reasoning.

Some, but not all, functions of the intellect can currently be implemented by artificial intelligence systems. In addition to the already mentioned logical inference, artificial intelligence systems are able to group objects into categories based on their characteristics, but they are not able to create concepts corresponding to these categories. For example, they can group round objects into one collection, but they will not create the idea of the circle. In general, reasoning is satisfactorily realized by AI systems due to the development of the contemporary mathematical logic that is fully formalized and, as a consequence, algorithmized.

Pronouncing judgements, however, can be partially modelled by using knowledge stored in knowledge bases or by realization of inductive learning based on examples. Such schema is realized, for instance, in artificial neural networks (Bielecki 2024a). Nevertheless, since the contemporary psychology does not describe the mechanism of conceptualisation, this cognitive functionality is unattainable for contemporary AI systems. As a consequence, formulation of notions, propositions and judgements remains beyond capabilities of AI systems. Arguably, consciousness and self-awareness are also unattainable for AI systems.

It should be noted that the claim regarding the inability of artificial systems to conceptualize, as raised in this paper, may be perceived differently from a functionalist perspective. From the standpoint of functionalism, cognitive processes can be realized through mechanisms entirely distinct from human psychology, provided they lead to equivalent results. However, within the context of cognitive epistemology based on Kantian theory, a crucial distinction must be made between functional categorization and essential conceptualization (Flasiński 1997, 2024). While AI effectively groups objects into sets based on features, which externally may resemble the formation of concepts, the absence of a formalized psychological theory concerning the emergence of ideas prevents the implementation of a mechanism that would grant these categories the status of abstract concepts and judgements. Thus, this barrier is not merely technical but stems from fundamental differences in how formal systems and living cognitive subjects constitute meaning.

5. Concluding remarks and comments

In this paper, we discussed the problem of representation and analysis of the scene and building its models in the context of cognitive possibilities of future, fully autonomous, artificial agents, primarily robots. With all the current, considerable possibilities in the field of implementing cognitive modules in artificial intelligence systems and with the perspectives of development of such systems, the natural question arises: "Do such systems have significant barriers compared to human capabilities?" The analysis of this problem goes beyond the framework of computer science and cybernetics, at least at the current stage of development of these sciences, entering the area of philosophy. At the current stage of artificial intelligence development in general and the problem of scene understanding in particular, it seems that there are not fundamental barriers to development of biologically inspired AI systems. Furthermore, they can shed light on some fundamental epistemological problems, such as truth verification. However, the problem of the possibility of constructing thinking or conscious machines remains open (Bremer and Flasiński 2022). Perhaps the consciousness and self-awareness are these barriers, which by their very nature artificial systems cannot cross. The distinction between advanced biological mimicry and true consciousness is often perceived as "porous". A rigorous analysis of formal foundations of AI systems suggests, however, a fundamental categorical barrier. This limitation is rooted in three primary areas. First, as noted by Flasiński (Flasiński 1997, 2024), current AI paradigms lack formalized models for the basic cognitive operation of the autonomous creation of concepts (*indivisibilium intelligentia*). While AI excels at reasoning (*ratiocinare*) through logic and pattern recognition through induction, it cannot perform the qualitative leap of abstraction required for true conceptualization, as this process has not been operationalized in a way that allows for algorithmic implementation.

Second, the perceived "porosity" of the boundary between machine performance and human reason is frequently a result of a misuse of language (Bremer and Flasiński 2022). Following the analytic tradition of Wittgenstein and Chomsky, it is argued that ascribing mental qualities like "self-awareness" to a system based on its behavioral output (e.g., passing a Turing-like test) constitutes a category mistake. AI demonstrates performance without competence; it simulates the results of consciousness without possessing the underlying mental states.

Finally, the transition from weak AI to strong AI remains blocked by the gap between syntax and semantics. As it is emphasized (Flasiński 2024), the computational nature of AI, whether symbolic or connectionist, is limited to the manipulation of representations according to formal rules. Without a biological or ontological foundation for understanding, the system remains an imitation of intelligence. Thus, the barrier to AI consciousness is not a temporary technological hurdle but a structural limitation of the current computational paradigm.

To sum up, the analysis conducted in this paper demonstrates that the problem of scene understanding is not merely a technological challenge, but a profound theoretical inquiry into the nature of cognition and epistemology in artificial systems. By situating robotic vision within the framework of cybernetics and Kantian philosophy, we have shown that while biological inspirations provide a robust path for developing functional autonomy, they do not automatically resolve the fundamental limitations related to conceptualization. The technological solutions described herein serve as a practical verification of theoretical assertions regarding the correspondence theory of truth and the adequacy of world models. Ultimately, the distinction between intellect, i.e., capable of logical operations and categorization, and reason, i.e. capable of creating transcendental ideas, remains the primary theoretical boundary that separates current artificial intelligence from human-like cognitive subjects.

It should also be mentioned that a very important and often overlooked problem is the impact of emotions on the learning process and the problem of emotions as a learning mechanism. These issues are also considered in the context of artificial intelligence and cognitive abilities of artificial systems (Abbate 2023; Assunção et al. 2022; Bielecki, Bielecka, and Bielecki 2017; Rutar, Wiese, and Kwisthout 2022). This topic, however, goes beyond the scope of this article.

References

Abbate, Francesco. 2023. Natural and Artificial Intelligence: A Comparative Analysis of Cognitive Aspects. *Minds and Machines* 33 (4): 791–815. https://doi.org/10.1007/s11023-023-09646-w.

Assunção, Gustavo, Bruno Patrão, Miguel Castelo-Branco, and Paulo Menezes. 2022. An Overview of Emotion in Artificial Intelligence. *IEEE Transactions on Artificial Intelligence* 3 (6): 867–886. https://doi.org/10.1109/TAI.2022.3159614.

Bajcsy, Ruzena, Yiannis Aloimonos, and John K. Tsotsos. 2018. Revisiting active perception. *Autonomous Robots* 42 (2): 177–196. https://doi.org/10.1007/s10514-017-9615-3.

Barreiro, C., Jose M. Barreiro, J. A. Lara, D. Lizcano, M. A. Martínez, and J. Pazos. 2020. The Third Construct of the Universe: Information. *Foundations of Science* 25 (2): 425–440. https://doi.org/10.1007/s10699-019-09630-7.

Bhushan, Bharat. 2009. Biomimetics: lessons from nature–an overview. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367 (1893): 1445–1486. https://doi.org/10.1098/rsta.2009.0011.

Bielecka, Marzena. 2018. Syntactic-geometric-fuzzy hierarchical classifier of contours with application to analysis of bone contours in X-ray images. *Applied Soft Computing* 69:368–380. https://doi.org/10.1016/j.asoc.2018.04.038.

Bielecka, Marzena, Andrzej Bielecki, Aleksander Suchorab, and Igor Wojnicki. 2025. Hierarchical Structural Information – Theory and Applications. In *Computational Science – ICCS 2025. 25th international conference: Singapore, July 7–9, 2025: proceedings, Pt. 3,* edited by Michael H. Lees, Wentong Cai, Siew Ann Cheong, Yi Su, David Abramson, Jack J. Dongarra, and Peter M. A. Sloot, 15905:48–59. Series Title: Lecture Notes in Computer Science. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-97632-2_4.

Bielecki, Andrzej. 2014. A Model of Human Activity Automatization as a Basis of Artificial Intelligence Systems. *IEEE Transactions on Autonomous Mental Development* 6 (3): 169–182. https://doi.org/10.1109/TAMD.2014.2319740.

Bielecki, Andrzej. 2021. The Systemic Concept of Contextual Truth. *Foundations of Science* 26 (4): 807–824. https://doi.org/10.1007/s10699-020-09713-w.

Bielecki, Andrzej. 2024a. Artificial Neural Networks. In *Cognitive Science. Vol. 13,* edited by Józef Bremer, 299–314. Social Dictionaries. Kraków: Ignatianum University Press.

Bielecki, Andrzej. 2024b. Cybernetics. In *Cognitive Science. Vol. 13,* edited by Józef Bremer, 281–298. Social Dictionaries. Kraków: Ignatianum University Press.

Bielecki, Andrzej, Marzena Bielecka, and Przemysław Bielecki. 2017. Conditioned Anxiety Mechanism as a Basis for a Procedure of Control Module of an Autonomous Robot. In *Artificial Intelligence and Soft Computing. 16th International Conference: ICAISC 2017 Zakopane, Poland, June 11–15, 2017. Proceedings, Pt. 2,* edited by Leszek Rutkowski, Marcin Korytkowski, Rafał Scherer, Ryszard Tadeusiewicz, Lotfi A. Zadeh, and Jacek M. Zurada, 10246:390–398. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-59060-8_35.

Bielecki, Andrzej, Tomasz Buratowski, M. Ciszewski, and Piotr Śmigielski. 2016. Vision based techniques of 3D obstacle reconfiguration for the outdoor drilling mobile robot. In *Artificial intelligence and soft computing. lecture notes in computer science,* edited by Leszek Rutkowski, Marcin Korytkowski, Rafał Scherer, Ryszard Tadeusiewicz, Lotfi A. Zadeh, and Jacek M. Zurada, 9693:602–612. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-39384-1_53.

Bielecki, Andrzej, Tomasz Buratowski, and Piotr Śmigielski. 2013. Recognition of two-dimensional representation of urban environment for autonomous flying agents. *Expert Systems with Applications* 40 (9): 3623–3633. https://doi.org/10.1016/j.eswa.2012.12.068.

Bielecki, Andrzej, Tomasz Buratowski, and Piotr Śmigielski. 2014. Three-Dimensional Urban-Type Scene Representation in Vision System of Unmanned Flying Vehicles. In *Artificial Intelligence and Soft Computing. 13th International Conference, ICAISC 2014, Zakopane, Poland, June 1-5, 2014, Proceedings, Part I,* edited by David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Alfred Kobsa, Friedemann Mattern, John C. Mitchell, et al., 8467:662–671. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-07173-2_56.

Bielecki, Andrzej, and Michael Schmittel. 2022. The Information Encoded in Structures: Theory and Application to Molecular Cybernetics. *Foundations of Science* 27 (4): 1327–1345. https://doi.org/10.1007/s10699-022-09830-8.

Bielecki, Andrzej, and Piotr Śmigielski. 2017. Graph representation for two-dimensional scene understanding by the cognitive vision module. *International Journal of Advanced Robotic Systems* 14 (1): 1729881416682694. https://doi.org/10.1177/1729881416682694.

Bielecki, Andrzej, and Piotr Śmigielski. 2021. Three-Dimensional Outdoor Analysis of Single Synthetic Building Structures by an Unmanned Flying Agent Using Monocular Vision. *Sensors* 21:s21217270. https://doi.org/10.3390/s21217270.

Bielecki, Andrzej, and Ryszard Stocki. 2023. The concept of structural information and possible applications. *Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce),* no. 75, 157–183. https://doi.org/10.59203/zfn.75.654.

Bremer, Józef, and Mariusz Flasiński. 2022. The Turing Test, or a Misuse of Language when Ascribing Mental Qualities to Machines. *Forum Philosophicum* 27 (1): 6–25. https://doi.org/10.35765/forphil.2022.2701.01.

Chaib-draa, B. 2002. Causal maps: theory, implementation, and practical applications in multiagent environments. *IEEE Transactions on Knowledge and Data Engineering* 14 (6): 1201–1217. https://doi.org/10.1109/TKDE.2002.1047761.

Flasiński, Mariusz. 1997. "Every Man in His Notions" or Alchemists' Discussion on Artificial Intelligence. *Foundations of Science* 2 (1): 107–121. https://doi.org/10.1023/A:1009687513096.

Flasiński, Mariusz. 2016. *Introduction to Artificial Intelligence.* Cham: Springer Nature.

Flasiński, Mariusz. 2019. *Syntactic Pattern Recognition.* New Jersey-London-Singapore: World Scientific.

Flasiński, Mariusz. 2024. Artificial Inteligence. In *Cognitive Science. Vol. 13,* edited by Józef Bremer, 315–330. Social Dictionaries. Kraków: Ignatianum University Press.

Floreano, Dario, and Robert J. Wood. 2015. Science, technology and the future of small autonomous drones. *Nature* 521 (7553): 460–466. https://doi.org/10.1038/nature14542.

Frish, Karl. 1954. *The Dancing Bees.* Berlin-Gotingen-Heidelberg: Springer.

Frish, Karl, and Martin Lindauer. 1956. The "Language" and Orientation of the Honey Bee. *Annual Review of Entomology* 1 (1): 45–58. https://doi.org/10.1146/annurev.en.01.010156.000401.

Friston, Karl. 2009. The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences* 13 (7): 293–301. https://doi.org/10.1016/j.tics.2009.04.005.

Hu, Cheng, Farshad Arvin, Caihua Xiong, and Shigang Yue. 2017. Bio-inspired embedded vision system for autonomous micro-robots: The LGMD case. *IEEE Transactions on Cognitive and Developmental Systems* 9:241–254.

Jakubowski, Ryszard. 1990. Decomposition of complex shapes for their structural recognition. *Information Sciences* 50 (1): 35–71. https://doi.org/10.1016/0020-0255(90)90004-T.

Jia, Weikuan, Yan Zhang, Jian Lian, Yuanjie Zheng, Dean Zhao, and Chengjiang Li. 2020. Apple harvesting robot under information technology: A review. *International Journal of Advanced Robotic Systems* 17 (3): 1729881420925310. https://doi.org/10.1177/1729881420925310.

Kang, Soo-Han, and Ji-Hyeong Han. 2023. Video Captioning Based on Both Egocentric and Exocentric Views of Robot Vision for Human-Robot Interaction. *International Journal of Social Robotics* 15 (4): 631–641. https://doi.org/10.1007/s12369-021-00842-1.

Kant, Immanuel. 1838. *Critick of Pure Reason.* Translated by Francis Haywood. London: William Pickering.

Katsev, Max, Anna Yershova, Benjamín Tovar, Robert Ghrist, and Steven M LaValle. 2011. Mapping and Pursuit-Evasion Strategies For a Simple Wall-Following Robot. *IEEE Transactions on Robotics* 27 (1): 113–128. https://doi.org/10.1109/TRO.2010.2095570.

Krzanowski, Roman. 2020. What Is Physical Information? *Philosophies* 5 (2): 10. https://doi.org/10.3390/philosophies5020010.

Krzanowski, Roman. 2022. *Ontological information: information in the physical world.* World Scientific series in information studies, vol. 13. Singapore: World Scientific.

Krzanowski, Roman, and Pawel Polak. 2022. The meta-ontology of AI systems with human-level intelligence. *Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce),* no. 73, 197–230. https://doi.org/10.59203/zfn.73.610.

Kurpaska, Sławomir, Andrzej Bielecki, Zygmunt Sobol, Marzena Bielecka, Magdalena Habrat, and Piotr Śmigielski. 2021. The Concept of the Constructional Solution of the Working Section of a Robot for Harvesting Strawberries. *Sensors* 21 (11): 3933. https://doi.org/10.3390/s21113933.

Li, Jun, Xue Mei, Danil Prokhorov, and Dacheng Tao. 2017. Deep Neural Network for Structural Prediction and Lane Detection in Traffic Scene. *IEEE Transactions on Neural Networks and Learning Systems* 28 (3): 690–703. https://doi.org/10.1109/TNNLS.2016.2522428.

Polak, Paweł. 2023. Philosophy in technology: A research program. *Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce),* no. 75, 59–81. https://doi.org/10.59203/zfn.75.682.

Quintas, João. 2023. Improving Visual Perception of Artificial Social Companions Using a Standardized Knowledge Representation in a Human-Machine Interaction Framework. *International Journal of Social Robotics* 15 (3): 425–444. https://doi.org/10.1007/s12369-021-00859-6.

Rosslenbroich, Bernd. 2014. *On the Origin of Autonomy: A New Look at the Major Transitions in Evolution.* Cham: Springer International Publishing AG.

Roterman, Irena, and Leszek Konieczny. 2023. Protein Is an Intelligent Micelle. *Entropy* 25 (6): 850. https://doi.org/10.3390/e25060850.

Rutar, Danaja, Wanja Wiese, and Johan Kwisthout. 2022. From representations in predictive processing to degrees of representational features. *Minds and Machines* 32 (3): 461–484. https://doi.org/10.1007/s11023-022-09599-6.

Ryabko, Boris, and Zhanna Reznikova. 2009. The Use of Ideas of Information Theory for Studying "Language" and Intelligence in Ants. *Entropy* 11 (4): 836–853. https://doi.org/10.3390/e11040836.

Ryabko, Boris, and Zhanna Reznikova. 2012. Ants and bits. *IEEE Information Theory Society Newsletter* 62 (5): 17–20.

Scott, Alwyn. 1995. *Stairway to the Mind: The Controversial New Science of Consciousness.* New York: Springer New York.

Siagian, Christian, and Laurent Itti. 2009. Biologically Inspired Mobile Robot Vision Localization. *IEEE Transactions on Robotics* 25 (4): 861–873. https://doi.org/10.1109/TRO.2009.2022424.

Tadeusiewicz, Ryszard. 1992. *Vision Systems of Industrial Robots.* Warszawa: WNT Press.

Tadeusiewicz, Ryszard. 2010. New trends in neurocybernetics. *Computer Methods in Materials Science* 10 (1): 1–7. https://doi.org/10.7494/cmms.2010.1.0271.

Tejera, Gonzalo, Martin Llofriu, Alejandra Barrera, and Alfredo Weitzenfeld. 2018. Bio-Inspired Robotics: A Spatial Cognition Model integrating Place Cells, Grid Cells and Head Direction Cells. *Journal of Intelligent & Robotic Systems* 91 (1): 85–99. https://doi.org/10.1007/s10846-018-0852-2.

Umbrico, Alessandro, Riccardo De Benedictis, Francesca Fracasso, Amedeo Cesta, Andrea Orlandini, and Gabriella Cortellessa. 2023. A Mind-inspired Architecture for Adaptive HRI. *International Journal of Social Robotics* 15 (3): 371–391. https://doi.org/10.1007/s12369-022-00897-8.

Wang, Heng, Bin Wang, Bingbing Liu, Xiaoli Meng, and Guanghong Yang. 2017. Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle. *Robotics and Autonomous Systems* 88:71–78. https://doi.org/10.1016/j.robot.2016.11.014.