Mind, Agency, and Biosemiotics

Alexei A. Sharov

National Institute on Aging, Baltimore, MD USA
sharoval@mail.nih.gov

Abstract

Development of artificial cognition, one of the major challenges of contemporary science, requires better understanding of the nature and function of mind. This paper follows the idea of Searle that mind is more than computation, and explores the notion that mind has to be embodied in agency that actively interacts with the outside world. To avoid anthropocentrism and dualism, I develop the concept of agency using principles of biosemiotics, a new discipline that integrates semiotics (science on signification and meaning) with biology. In evolutionary terms, human cognition is an advanced form of agency that emerged from simpler ancestral forms in animals, plants, and single-cell organisms. Agency requires autonomy, informed choice, and goal-directedness. These features imply a capacity of agents to select and execute actions based on internal goals and perceived or stored signs. Agents are always constructed by parental agents, except for the most simple primordial molecular-scale self-reproducing agents, which emerged from non-living components. The origin of life coincides with the emergence of agency and primitive communication, where signs are not yet associated with objects, and instead used to activate or regulate actions directly. The capacity of agents to perceive and categorize objects appeared later in evolution and marks the emergence of minimal mind and advanced communication via object-associated signs. Combining computation with agential features such as goal-directedness, adaptability, and construction may yield artificial systems comparable in some respects to human mind.

Keywords: biosemiotics, constructivism, multi-levelness, autonomy, protosemiosis, eusemiosis, autonomous learning.
1. Introduction

The quest for artificial cognition (intelligence) is one of the major challenges in contemporary science and technology (Crowder, Carbone, & Friess, 2013). It has been proposed that cognition (or thinking) is a kind of computation (Fodor, 1975; Putnam, 1979). However, computers fail to replace humans in such cognitive tasks as driving a car, translating a book into another language, or coding of new software. John Searle suggested that computation cannot replace mind, because it is a syntactic process that lacks semantics or meaning (Searle, 1980).

If we take the argument of Searle seriously, then the next question is what are the functions essential for cognition that are missing in a computer? If the lack of semantic relationships is the only problem, then it can be addressed by connecting a computer with a semantic module that establishes meaningful links between computer states and the human language. If semantic module is supplied with a human thesaurus, ontology, and linguistic rules (syntax and paradigmatic models), then it can serve as an interface for direct communication of humans with computers. Computer equipped with a well-designed semantic module is a useful tool, but it certainly falls short of real cognition. The main weakness of such system is the absence of goal-directedness, active interaction with the outside world, personal experience, autonomous learning, and capacity of second-order understanding (i.e., understanding of own cognitive states, assumptions, limitations, and errors, as well as these features in the cognition of others). Here I emphasize that cognition requires an autonomous agency capable of following certain goals or values, selecting actions accordingly, and learning how to make meaningful representations of the outside world. Such kind of agency may acquire a capacity to repair or replace malfunctioning components and modify and even restructure its thesaurus and ontology. The idea that cognition requires agency matches well with the methodology of constructivism and enactivism, which assume that learning requires active interaction with the world and with other people, experimenting, and

---

1 Here I do not mean radical enactivism, which attempts to explore “minds without content” (Hutto and Myin, 2013).
construction of new theories and models of the world (Gallagher, 2017).

The idea that cognition requires agency is widely accepted in the domain of Artificial Intelligence (AI), but mostly in anthropocentric terms. For example, Russell and Norvig (Russell & Norvig, 2010) consider AI as a capacity to think and/or act either humanly or rationally, where rationality is viewed as the logical component of human cognition. However, attempts to expand the notion of agency beyond humans were met with strong resistance in science. Physics and Darwinism both spent their major efforts in fighting the notion of agency, and showing that agency is just an illusion and not needed for understanding nature (Dawkins, 1986; Dennett, 1995). Deviations from this dogma were considered as compromising science and opening a back door to religion (Dawkins, 1986), mysticism, or “folk psychology” (Dennett, 2013).

In this paper I approach the notion of agency from the evolutionary perspective, and present arguments that it can be naturalized within the framework of a new interdisciplinary field of research, called biosemiotics, which integrates biology with semiotics, a science on meaning and signification (Hoffmeyer, 2008; ISBS, 2016). According to biosemiotics, all living organisms and even their parts (e.g., cells) are autonomous or semi-autonomous agents that have certain freedom in selecting their actions based on the incoming sensorial input and internal factors such as developmental and behavioral models and subjective preferences (Sharov, 2010; Sharov, 2017). Activities of agents are goal-directed, sign-dependent, and cannot be explained adequately by physical laws, although they do not violate these laws. The pattern of action may seem highly reproducible over time but it is not deterministic in a physical sense. In certain situations, agents change their goals and priorities and this may cause unpredictable dramatic shifts in the pattern of activity (e.g., during stress). From the evolutionary view, human cognition is an advanced form of agency – the tip of an iceberg of biological organization that started from self-reproducing molecular complexes at the origin of life and continued to prokaryote cells, eukaryote cells, early multicellular animals, fish, reptiles, and mammals (Fig. 1).
A natural question is then, how can we approach the development of artificial cognition? Computation is important but not enough, as discussed above. It may be tempting to try developing mechanisms that emulate brain structure and function in every detail. This strategy, however, will likely fail because we know too little about the functionality of brain, and human technology is not sufficient to emulate nanoscale molecular processes that occur in neurons and their organelles (e.g., axons, dendrites, synapses, and microtubules). A more realistic path towards artificial cognition seems to be in augmenting computers with elements of life and agency as I discuss in section 6.

2. Biosemiotics

The rise of Biosemiotics can be explained by three major factors: (1) dissatisfaction of biologists with mechanistic explanations of evolution, development, and behavior, (2) insights in humanities that human cognition is grounded in the biological nature of organisms, and (3) logical necessity
to accept the idea that humans are not the only organisms with such agential powers as freedom to select actions, capacity to create, follow, and update the goals, and develop meaningful representations of the world. Biosemiotics is a synthesis of biology with semiotics (Barbieri, 2008; Sharov, Maran, & Tønnessen, 2015), and its major challenge is to reformulate basic terms in each “parental” discipline in order to find mutual understanding. Participants of this movement often express different opinions on the importance of various semiotic and biological processes and components in the behavior and evolution of agents. Thus, I present a short overview of major theoretical platforms in biosemiotics.

2.1. Theoretical Biology

Biology has been monopolized by mechanistic thinking for more than half a century. Mechanistic and structural research became the gold standard representing true progress in biology that deserved publication in leading journals. However, the revolt against mechanistic tradition is growing in many biological disciplines. The Extended Evolutionary Synthesis has rejected the dogma of neo-Darwinism that growth, development, and reproduction of organisms are passive mechanical processes (Pigliucci & Müller, 2010). Progress in epigenetics contradicts with a belief that the genome is a blueprint of an organism that determines the phenotype (Sharov, 2014). Behavioral biology does not accept any longer the idea that human and animal behavior can be fully explained in terms of reflex and conditioning. Additional factors include goal-directedness (Stout, 1996), meaning (DeGrandpre, 2000), and thinking (Overskeid, 2000). Neither information nor self-organizing is enough to explain life (Brier, 2008). Self-organizing behaviors have to be combined into meaningful and robustly-functioning whole organisms capable of self-repair, self-reproduction, adaptation, and evolution.

Biosemiotics offers a non-mechanistic paradigm in biology because it is focused on the agential properties of life and emphasizes the autonomy and goal-directedness in the activity and communication of organisms. Thus, biologists hope to use the ideas of biosemiotics as a platform for building theoretical biology (Emmeche & Kull, 2011). In particular, “The predictive power of biology is embedded in the functional aspect and cannot be based
on chemistry alone” (Kalevi Kull, Deacon, Emmeche, Hoffmeyer, & Stjernfelt, 2009), p.169. Ethologist Jacob von Uexküll, one of the predecessors of biosemiotics, suggested that animals and other organisms develop meaningful models of their environment (Umwelten), where objects are associated with living functions, and thus become functional tokens such as food, shelter, or enemy (Uexküll, 1982). Different species interpret the same environment differently, and thus they have distinct Umwelten tailored to their living needs and affordances. Uexküll viewed organisms as subjects competent to interpret their sensorial inputs and chose appropriate actions. Meaningful activities (habits) are thus organized via choice-dependent (i.e., non-mechanistic) regulatory feedbacks that connect sensors and effectors into closed “function circles” (Uexküll, 1926, 1957).

2.2. Code Biology

Progress in molecular biology prompted applications of biosemiotic ideas to molecular functions within cells. One of the pioneers of this approach was Giorgio Prodi (Prodi, 1988) who proposed the term “protosemiosis” for molecular signaling processes that include DNA copying, transcription, translation, and signal transduction. In contrast to biochemistry that views molecular processes as purely mechanistic, Prodi emphasized that these processes are meaningful and thus semiotic. Marcello Barbieri continued exploring the semiotic nature of the genetic code as well as other molecular codes such as splicing codes and signal transduction codes (Barbieri, 2003). He explored the role of molecular agents in copying of coding sequences and translating them into immediate molecular-level meanings (e.g., the sequence of nucleotide triplets is translated into the order of aminoacids in the protein polymeric molecule). The former task is performed by copymakers (e.g., DNA polymerase), whereas the latter task is performed by codemakers (e.g., ribosome interacting with tRNA adapters loaded with aminoacids). Because many codemakers include RNA, Barbiery developed the notion of ribotype as a linker between genotype and phenotype. He also emphasized a qualitative difference between code-based semiosis at the molecular level and interpretational semiosis that required neural system and brain (Barbieri, 2009).

Barbieri called his project “organic codes” (Barbieri, 2003), and lately,
“code biology” (Barbieri, 2015). This project was initially integrated with other research platforms in biosemiotics including theoretical biology and humanities-related studies. In 2007, Barbieri organized journal “Biosemiotics” (Springer), which became a major periodical for interdisciplinary papers that integrate biology and semiotics. Later, code biology became a fully independent project and its interactions with biosemiotics declined.

2.3. Connection with Humanities

Biosemiotics is heavily influenced by humanities, in particular by linguistics and human semiotics (Favareau et al., 2017). Charles Morris (Morris, 1964) proposed to expand the semiotic theory of Charles Peirce to the behavior and communication of animals. In particular, he assumed that semantic and pragmatic aspects of sign relations can be found in the animal world. This idea was further developed by Thomas Sebeok who attempted to integrate Peirce’s semiotics and the Umwelt-theory of Uexküll (Thomas A. Sebeok, 1972). According to Peirce (Peirce, 1998), sign is a triadic relation between a sign vehicle (or representamen), object, and interpretant (i.e., result of interpretation of a sign by mind) (Fig. 2).

Fig. 2. Peirce’s concept of sign as a triadic relation between representamen (or sign vehicle), object, and interpretant.

Before Peirce, the dominating theory of signification was structural linguistics (Saussure, 1959), where each word in a human language was viewed as a signifier that corresponds to a signified (meaning). Peirce’s theory of signification is a more general approach than structural linguistics,
and it can be applied to perception and non-verbal communication in humans and in animals. The main advantage of Peirce’s theory is that semantic links between representamen and their interpretants are based on the association with *objects*; and thus, appear grounded in the real world. For example, smoke is caused by fire; and thus, it is a sign of fire, which is an object. When humans see smoke they think of fire and interpret it as heat or danger.

Peirce’s theory is generally applied to mental interpretation of signs; and thus, it can be naturally extended to the behavior and communication of animals with brains. However, it is challenging to apply the theory to organisms without brains. To make a connection, it has been assumed that mind-like interpretations may exist in animals without nervous system (Hoffmeyer & Stjernfelt, 2016) and even in plants (Krampen, 1981). Most biosemioticians agree that sign processes (or semiosis) are coextensive with life (T. A. Sebeok, 2001). In other words, interpretation (in a broad sense) of signs exists in all living organisms but not in the physical world without life.

Bruni (Bruni, 2008), and Hoffmeyer and Emmeche (Hoffmeyer & Emmeche, 1991) argued that living cells (including bacteria) have enough complexity to interpret signs in a Peircean way. But these authors did not apply the notion of signification to molecular processes such as DNA replication, transcription, and translation on the grounds that these processes are mechanistic, and therefore, not semiotic. Thus, they considered cells as minimal living and semiotic systems, which are composed of non-living and non-semiotic parts. This view is not compatible with the ideas of protosemiotics and code biology, and it contributed to the separation of code biology from biosemiotics. The denial of signification at the level of functional molecular complexes and organelles is a regrettable mistake that not only breaks the unity of the biosemiotics movement, but also stalls any attempts to explain the origin of life. Whole cells are too complex to emerge by pure physics and chance. Thus, the only possibility is that Darwinian-type biological evolution based on hereditary signs started long before the emergence of living cells. The origin of signs and meanings should be moved back in time to the origin of life in the form of simple functional and heritable molecular networks (Sharov, 2016a). This approach brings a new meaning to the thesis that life and semiosis are
coextensive.

### 2.4. Constructive Biosemiotics

A new trend in biosemiotics is establishing links with “sister disciplines”, such as second-order cybernetics, autopoiesis, enactivism, cognitive science, and constructivism, and shifting the focus from structural relationships to the dynamical aspects of agency and signification at the evolutionary, developmental, and behavioral time scales (Cariani, 1998; Sharov, 2016b; Sharov, 2017). This approach can be formulated as constructive biosemiotics because of its close links with the methodology of constructivism taken in a broad sense (Vehkavaara & Sharov, 2017). Constructivism generally denotes a theory of human knowledge that emphasizes the importance of active involvement in knowledge-building and rejects the idea that knowledge comes via passive imprinting or copying (Riegler, 2006; Tobias & Duffy, 2009). In biology, it is not only knowledge or know-how (e.g., Umwelt) that is constructed. As I wrote before, “Everything [in agents] has to be constructed: sense organs – to detect signals; networks – to integrate and analyze signals; effector organs – to respond; memory – to store information; subagents – to perform downstream tasks including lower-level construction; body – to integrate all functional units; niche – to live in; tools and resources – to increase functional efficiency; and signs – to support communication between parts of an organism and with other organisms” (Sharov, 2016b).

The most characteristically constructivist approaches in biology are “second order cybernetics” (Von Foerster, 2003), “autopoiesis” (Maturana & Varela, 1980), and “practopoiesis” (Nikolić, 2015). These theories are consistent with the principles of “radical constructivism” (Von Glasersfeld, 1984) by assuming that agential knowledge fits to reality in the Darwinian sense, but does not necessarily include a representational relation to it. Second order cybernetics study organisms together with their environment as self-constructed unities characterized by “operational closure” (Von Foerster, 2003). The idea of closure comes from mathematics, where certain states in a multidimensional vector space, called eigen-vectors, appear self-sustained after repeated applications of the same linear transformation that is interpreted as a model of world change through time. Although the
rules of world change are not fully known, it is assumed that certain agential
states (eigenforms) and/or dynamic patterns (eigenbehaviors) become
stabilized in the long term due to their reinforcement by the self-production
network. Eigenbehaviors depend on the state and dynamics of the outer
world, and therefore represent the world to a certain extent, but they also
depend on the internal self-production network of the agency. Thus changes
of internal relations may lead to novel emergent representations of the
world, and if these changes are heritable, they may become favored by
natural selection yielding selected self-organization (Rocha, 1998). At a
higher level of complexity, agents may learn to switch between distinct
dynamic attractors (eigenforms or eigenbehaviors) on demand, resulting in
the emergence of new sign relations that link memory units with dynamic
attractors in a context-dependent way (Rocha, 1998). Autopoiesis and
practopoiesis explore the rules by which biological systems construct
themselves, and apply the principle of self-construction to the nervous
system and cognition. Practopoiesis considers multiple levels of
organization (at least three levels), and each higher level is constructed by
subsystems that belong to the preceding lower level (Nikolić, 2015).

I see the merit of radical constructivism in focusing on the pragmatic
relation between knowledge and reality, which is indeed a fundamental
feature of agency. This is especially important for understanding the living
world of primitive organisms such as bacteria, which apparently have no
capacity to access (e.g., categorize, track, and handle) objects in the environment
(Vehkavaara & Sharov, 2017). Nevertheless, bacteria developed methods
for coping with their environment without representing it at the object level.
Radical constructivism, however, appears unnecessarily restrictive in its
agnostic approach to reality when applied to higher level organisms, such as
mammals and birds, which have competence to categorize, track, and
manipulate the objects of the world (Sharov, 2013).

3. Agents and their features

3.1. Agency is a core notion of biosemiotics

Constructive approach in biosemiotics (section 2.4) shifts the emphasis
from signs and meanings to agents, because signs and meanings are
grounded in the recurrent goal-directed activity of agents. Thus, the core notion appears to be *agency* rather than sign or meaning. Agents include all living organisms from bacteria to humans. However, the notion of “agent” is more general than the notion of organism and includes parts of organisms with autonomous functions, such as organs, cells, intracellular organelles, and functional molecular complexes, as well as super-organism agents such as colonies, populations, or symbiotic consortia (Sharov, 2017). Agents also include automated human devices (robots, computers), parasitic viruses, and computer viruses (Sharov, 2017). What is common to all agents is that they make meaningful and sign-dependent choices in their goal-directed activities.

The notion of agency offers a new interpretation of mechanisms. The common view is that mechanisms are systems that change according to physical (mechanical) laws; thus, mechanisms have dynamics but no functions. This interpretation works well for explaining the movement of parts within a mechanism, but it fails to explain why mechanisms do exist at all. From the agential point of view, mechanisms exist because they are constructed on purpose by parental agents (e.g., humans) to perform specific functions. The function of a thermostat is to keep stable temperature in the room, and the function of a mouse trap is to catch mice. Parental agents design and configure mechanisms to perform specific goal-directed activities, and thus, goals become embedded in mechanisms in the form of action networks, circuits, and/or electronic software. Thus, mechanisms can be viewed as subagents (or slave agents) that are carriers of parental goals and corresponding functions. Human-made mechanisms can be compared to intracellular mechanisms, such as enzymes or organelles that are manufactured by cells to perform specific functions. For example, ribosomes are programmed constructors of proteins, which are designed and tuned for this function by evolving lineages of cells.

Although autonomous mechanisms have agential properties, their functions are way too simple in comparison with living organisms, and

---

2 Note that not all constructed mechanisms are agents. Agents require autonomous activity and the use of signs (e.g., sensor outputs). Non-agential mechanisms are tools that lack autonomy and goal-directedness.
especially with humans. Most mechanisms have no capacity to develop internal representations of outside objects\(^3\), and thus, their way of handling signs is limited to protosemiosis. Thermostat does not learn or adaptively adjust its response to the temperature sensor. Nevertheless, it would be a mistake to consider thermostat a deterministic device. Most thermostats can be adjusted to different temperatures, which indicates the absence of direct efficient causation between the heat and thermostat state. If thermostat is not adjustable, then its response is programmed during production. The relation between heat and state of a thermostat is not physical but chosen by parental agents. Moreover, a thermostat can stop functioning, whereas physical laws of nature responsible for efficient causation cannot be stopped. In this respect, mechanisms are not deterministic but follow certain imprinted or acquired rules in selecting actions. It seems more reasonable to view mechanisms as components of life rather than something opposed to life. It is also conceivable that life originated from simple mechanism-like self-reproducing agents, which were recursively constructed and programmed by parental agents (Sharov, 2016a; Sharov, 2017).

Agency requires three interrelated features: autonomy, informed choice, and goal-directedness. **Autonomy** is a capacity of agents to select and execute actions using available resources and tools, or recruit and program other agents for certain activities. Due to autonomy, agents can be initial causes of chains of subsequent events. **Informed choice** means that agents use information (signs) to initiate and regulate their actions. Finally, **goal-directedness** means that agents tend to select actions that are expected to be beneficial and avoid actions that may damage or disrupt functional capacities of agents in the foreseeable future. Functional capacities represent agential identity (i.e., the living state), and thus, goal-directedness is targeted at preserving and, possibly, propagating and improving the agential identity. Let us discuss these agential features and check their compatibility with the existing scientific methodology.

### 3.2. Autonomy

Autonomy of agents contradicts to the principle of *universal causality*/

---

\(^3\) Here I do not consider advanced robots with trainable image-recognition system.
determinism which claims that each event has a cause or a set of causes that happened before\textsuperscript{4}. This principle is a useful heuristics in such tasks as car diagnostics or finding a bug in a computer program. However, the universality of the principle is questionable because it does not provide an operational rule for finding a cause and proving that the relation between a hypothetical cause and the observed event follows from universal laws. The search for causes may appear unfeasible or impractical, especially when applied to living organisms. For example, most human actions cannot be predicted from the physical state of the brain.

The notion of autonomy is supported by the subjective feeling of free will experienced by most humans (except some cases of mental disorder). Every conscious moment of life we perceive as a chance to start or change an action, and these choices are not illusive because they have objective physical consequences. Of course, subjective feelings can be deceiving as happened with the flat earth belief.

3.3. Informed choice

Autonomy implies freedom in selecting and processing informational inputs (i.e., signs) in the process of making a choice between actions. Freedom is real even in cases when alternative actions are not used. For example, car drivers stop on red light despite of their freedom to run through. Therefore, the choice of action is generally not blind but informed. This example can be used to expand the notion of autonomy to subconscious actions that are executed following reflexes or habits. Subconscious habits are automatic and thus we do not feel them as active choices (Bargh & Ferguson, 2000). For example, driving to work is often automatic and we do not remember making turns at intersections, because these actions are repeated every day and eventually go unnoticed in the “autopilot” mode. In this case, conscious choices were programmed earlier when the person was learning to drive to work, and then the program was transferred to the subconscious level. However, unexpected problem (e.g., a roadblock or car malfunction) terminates the subconscious routine, and the choice of action is transferred to the conscious level.

\textsuperscript{4} Universal causality follows from Newton’s physics and was supported by Kant.
Humans originated from animal ancestors; thus it is natural to assume that animals have a comparable capacity to use sensorial, neural, and heritable information to select relevant actions. This notion is supported by complex and meaningful behavior patterns in many mammals and birds (Thomas A. Sebeok, 1972). For example, ravens demonstrate problem-solving abilities in the use of tools even without previous practice (Heinrich, 1995). Informed choices can be found in simple organisms that change their metabolism and activity depending on environmental factors. For example, bacteria use chemotaxis to move towards areas with increased concentration of glucose (Porter, Wadhams, & Armitage, 2011). Activities of bacteria are automated and do not require object-level representation of the outside world (Sharov & Vehkavaara, 2015).

Informed choices exist even in the activity of cellular subagents: molecular complexes and organelles. For example, a ribosome selects which aminoacid to append to the synthesized polypeptide based on the triplet of nucleotides in a messenger RNA (mRNA) molecule. Each triplet of nucleotides encodes a certain aminoacid, and this correspondence is known as genetic code. The choice of aminoacid is regulated by another molecule called transport RNA (tRNA), which carries 3 nucleotides (anticodon) that can be paired with a triplet in the mRNA at the current position of the ribosome. After adding aminoacid at the end of the synthesized polypeptide, the ribosome shifts along the mRNA by 3 nucleotides to process the next triplet. To prove that the choice of aminoacids is real, scientists generated artificial tRNA with altered recognition triplet of nucleotides (Xie & Schultz, 2005). After this procedure, bacterial cells started synthesizing proteins that included an additional novel aminoacid.

3.4. Goal-directedness

Goal-directed behaviors are well studied in respect to conscious and rational thinking of humans (Emmons, 1996). Humans and higher animals perceive goals as mental representations of desired events and projected pathways towards their achievement. In contrast, primitive organisms with no mind (e.g., bacteria) have no such capacity. Here I use the term “goal-directedness” in a broad sense to make it applicable to agents of
various complexity levels ranging from functional molecular complexes to humans. In these cases, goal is a sign of success in specific activity (e.g., capturing resources, producing offspring, or escaping predators). Terrence Deacon’s term teleodynamics (Deacon, 2011) can be used for primitive forms of goal-directedness that do not require mind. The existence of mind does not necessarily mean the existence of mental goals, which represent anticipated objects. The primary function of mind is categorization of objects (see section 5), whereas mental support of goal-directedness appeared later in evolution, most likely it emerged independently in animals with brains in Protostome and Deutorostome clades.

Goal-directedness should not be confused with teleology, which is a method of logic that uses goals for predicting the actions of agents or results of these actions. In this method, goal is treated as a final cause, following the terminology of Aristotle. A closely related term teleonomy, is used to explain biological evolution based on organism needs (Pittendrigh, 1958) or non-mental programmed activities (Mayr, 1965). I agree that goals can be used as predictors, but in this paper I focus on more fundamental questions: what is the nature of goals and why goal-directedness is the essential capacity of agents.

Goals can be roughly divided into three categories: (1) recorded goal is a sign that the goal is attained; (2) goal efficacy (or self-efficacy) represents perceived likelihood (and/or degree) of success in the future as conditioned by the current state; and (3) goal anticipation represents an association of possible actions of an agent with anticipated change of goal efficacy after each action. These types are well established for human mental goals. For example, according to Pyotr Anokhin (Anokhin, 1974), an essential component of neural functional system is the acceptor of action results which evaluates recorded goal from sensorial inputs. Assessment of goal efficacy can be found in various aspects of human cognition (Albert Bandura, 1982; A. Bandura, 2001). According to Edmund Husserl, goal anticipation is an important component of human logic (Schutz, 1967).

These three aspects of goal-directedness also exist in protosemiotic agents that lack mind, where they are represented by inherited (or imprinted) signaling networks. For example, cell division should not start before the completion of DNA replication, and therefore, signals are needed that
indicate the success of DNA replication (recorded goal). In eukaryotic cells, such signals include molecules ATR, ATM, and Chk1. Examples of goal efficacy in bacterial cells are various SOS signals, indicating that life is threatened by current conditions. Many of them are signs of absence or deficiency (Deacon, 2011). For example, the lack of glucose in the environment is detected by bacterial membrane-bound glucose transporters. If a transporter is idle, then the concentration of cyclic AMP (cAMP) is increased in the cytoplasm sending a signal of starvation (Park, Lee, Seok, & Peterkofsky, 2006; Steinsiek & Bettenbrock, 2012). Goal anticipation in bacteria is represented by initiation of emergency actions in response to SOS signals. For example, starvation signals (cAMP) directly activate consumption and processing of alternative organic resources (e.g., lactose) (Carpenter & Sells, 1975), and promote cell motility to find glucose (Dobrogosz & Hamilton, 1971). Although goals in bacteria are automated, the goal-setting mechanisms are contingent in the evolutionary time scale because they emerged and changed adaptively in the long-term.

Goals differ in their time scale. A short-term goal of an agent can be to execute a function. Because executing a function often involves a feedback regulatory loop, it was proposed that function is just a feedback loop (Keller, 2010). I do not agree with this idea because self-regulatory systems such as tornado or flame are not agents, and thus they have no goals and no functions. Functions are “artificial” processes that do not exist in the physical world without agents. They emerged via learning and/or evolution and empowered agents to solve problems and reach new goals. Most agents, including all living organisms, require recurrent execution of functions on demand within their life span and even in the following generations of progeny agents. Thus, agents have a long-term goal to keep the capacity of repeating functions on demand, which includes securing resources and tools for performing future functions, preserving a fully-functional body, avoiding injuries, and improving functions via learning and evolution.

4. Agency without Dualism

Cartesian dualism separates human mind from the body and physical processes, and thus supports a taboo for exploring mind-like processes in
nature and the nature of human mind. The mechanistic worldview of contemporary science breaks dualism but at the expense of denying agency and the creative capacity of mind and life. Biosemiotics and related approaches (autopoiesis, practopoiesis, enactivism, and constructivism) provide a theoretical platform that overcomes dualism in a gentle way by introducing the notion of agency as follows. First, agents vary in their complexity from simple functional molecules to complex organisms and even super-organism entities. Thus, instead of a dichotomy (yes/no) there is a range of agential organization. Second, it is assumed that simple agents (primordial life) originated from physical non-living components. These first agents were likely more simple than it is expected in most scenarios of the origin of life (Sharov, 2016a), which makes the assumption of abiogenesis more plausible. And third, complex agents are always multi-level, and they construct, recruit, or culture subagents or dependent agents of lower complexity. This assumption avoids the well-known homunculus paradox that leads to an infinite regress. In particular, functions of higher-level agents are supported by their semi-autonomous subagents without being determined by them; and higher-level agents can regulate or constrain their subagents to carry out higher-level functions (Sharov, 2017). The latter effect is known as downward causation (Campbell, 1990) and heterarchy (Bruni & Giorgi, 2016; Crumley, 1995). Mechanism-like agents at lower levels are then not opposed to life but instead function as subagents of living organisms.

Agents cannot emerge spontaneously from non-living matter, except extremely simple primordial agents. Thus, agents are (almost) always constructed by some parental agents, and the class of agents is closed relative to the operation of construction (Sharov, 2010; Sharov, 2017). It is also logical to expect that agents do not increase their complexity instantly because it takes a long time to develop each new function via trial and error. Because the simultaneous emergence of a large number of entirely new functions is very unlikely, we expect that construction of agents satisfies the principle of gradualism: the functional complexity of descendant agents cannot be much higher than the complexity of parental agents (Sharov, 2010). This principle was supported by Darwin, but was later challenged by the theory of “punctuated equilibrium” which assumed rapid emergence of
new evolutionary lineages (Gould & Eldredge, 1977). Paleontological records indeed show sudden appearance of new lineages, especially following mass extinction events. However, these episodes of rapid evolutionary change did not yield organisms with a much higher functional complexity (Sharov & Gordon, 2017); and thus, the observations of Gould and Eldridge do not seem to contradict to the principle of gradualism as it is formulated above.

Not all agents are self-reproducing, but all agents are products of self-reproducing agents such as living organisms. Thus, it makes sense to distinguish between primary agents that are self-reproducing, and secondary agents that are not capable of self-reproduction (e.g., ribosomes, DNA polymerase, and human-made robots) (Fig. 3).

![Fig. 3. Primary agents are self-reproducing, whereas secondary agents are produced by primary agents or other secondary agents.](Fig3.png)

Self-reproduction is often misinterpreted as “passive copying” because this view was promoted in neo-Darwinism (Dawkins, 1986; Dennett, 1995). There is no doubt that DNA molecule is physically passive when it is replicated by DNA-polymerase. However, genes (functional segments of chromosomes) are active agents that recruit transcription factors, participate in regulation of transcription, and facilitate production of descendental agents such as proteins. The activity of genes is more informational than physical. It is also wrong to consider self-reproduction of whole organisms as passive. Self-reproduction activities include development of sex organs, maturation of sperm and oocytes, courtship, fertilization, development of an embryo, birth, and postnatal growth and development of the progeny. Replication
and transfer of genetic information (DNA sequence) is only one component of reproduction, which by itself does not make another organism. Genetic information has to be interpreted by cellular subagents and whole cells in order to guide differentiation and movement of cells in developing embryos. Although production of subagents is encoded in the genome, the cell requires a network of physically existing subagents to interpret the DNA (Sharov & Gordon, 2013). In particular, heritability of phenotypes across generations cannot be explained by the transfer of genetic information. Another and equally important channel of heredity is represented by agents capable of reading and interpreting genetic information together with necessary tools and scaffolds within cells (Sharov, 2014) (Fig. 4).

**Fig. 4.** Two channels of heredity: Informational and agential. From: (Sharov, 2014).

Agents have a *different ontological status* than physical objects because they do not just exist, but exist for their internal purpose. In particular, they are produced by parental agents for the purpose of preserving functional identity within the lineage and offering chances for future expansion and optimization of agential competence (Austin, 2017; Sharov, 2017). Agential ontology is grounded in production relations between parental and progeny agents.
5. Two Major Types of Semiosis

The complexity of semiosis (i.e., processing and interpretation of signs) increased in parallel with the evolution of living organisms and their functions. Besides a quantitative increase in the number of supported functions, these changes were associated with a qualitative reorganization of the meaning of information. Thus, we can delineate phases in the evolution of semiosis. For example, Prodi (Prodi, 1988) suggested a term “protosemiosis” for primitive forms of signaling at the molecular level. Krampen (Krampen, 1981) used the term “phytosemiosis” for semiotic processes in plants. Barbieri (Barbieri, 2009) distinguished between code-based semiosis at the cellular level and interpretational semiosis that is based on the use of includes representations and requires brain activity. Kull (K. Kull, 2009) described three levels of semiosis: vegetative, animal, and cultural, which he linked with three types of signs defined by Peirce: icons, indexes, and symbols, respectively.

Sharov and Vehkavaara (Sharov & Vehkavaara, 2015) combined approaches of Prodi, Kull, and Barbieri, and suggested to distinguish protosemiosis, where signs are linked to actions of agents either directly or via simple logical gates, and eusemiosis, where signs are linked with classifiable objects and only then - with possible actions. Molecular proto-signs are not associated with objects because they are processed by cellular subagents (e.g., ribosomes) that have no capacity to classify and track objects. It seems natural to associate a triplet of nucleotides in the mRNA with aminoacid as an object. However, a ribosome has no internal representation of aminoacid as object and it does not “know” that it makes proteins. Instead, a ribosome determines if a triplet of nucleotides in the mRNA matches to the anticodon sequence of the incoming tRNA loaded with aminoacid, and then makes a peptide bond. Humans (i.e., biologists) know the chemical structure of these components and understand the details of their interaction, but a ribosome simply gets a signal that indicates readiness for the reaction and then uses the catalyst tool to finish the action.

As the number of proto-signs increased in evolution, they became integrated into networks via logic gates with contextual control (Sharov & Vehkavaara, 2015). For example, transcription of a gene is often activated
by a simultaneous binding of two or more transcription factors to the promoter of a target gene. However, this innate logic is fixed genetically and cannot be modified within the life span of an organism even if it fails to produce beneficial effects. To overcome this limitation, organisms developed epigenetic mechanisms that modify logic gates on demand. For example, a gene with multiple regulatory modules in its promoter may initially carry open chromatin at all modules. However, after some “memory triggering” event, the chromatin may become condensed at all regulatory modules except the one that was functional at the time of the event that had a successful outcome (e.g., captured food). This can be viewed as a primitive mechanism of adaptive learning (Sharov, 2010).

As organisms advanced in perceiving internal and external signals, they eventually acquired a capacity to integrate them into meaningful categories representing real objects and situations (e.g., food items, partner agents, and enemies), and predict events using models (Sharov, 2013). This change is essential for eusemiosis. The categorization toolbox can be visualized as a set of fixed point attractors in the phase space of mind, where attractors represent object categories (Fig. 5). Each trajectory in this space corresponds to a process of recognition that leads from the initial sensory input to some attractor: tree, berry, bird, or predator (Fig. 5). Ontologically, each category corresponds to a type of objects whose perception by humans converges to the same attractor. This kind of recognition algorithm, explains the categorization capacity of the brain and was implemented in computational neural networks (Amit, 1989). Classifications and models comprise the knowledge of an agent about itself and its environment, which roughly corresponds to the Innenwelt and Umwelt of an organism, following the terminology of Uexküll (Uexküll, 1982). Information processed at the eusemiotic level does not necessarily induce physical actions of an organism, but it causes mental actions (e.g., accumulation of knowledge) which may affect future physical actions. This preparedness for future actions was called “disposition to respond” (Morris, 1964).

Eusemiosis relies on the functional support of protosemiotic subagents that perform elementary sensing (e.g., membrane-bound cell receptors), signal transduction, and internal communication. Thus, eusemiosis emerged after integration of protosemiotic agents into an autonomous
higher-level agency. A somewhat similar idea of multi-levelness was independently developed in the theory of practopoiesis, which assumes that first-level systems (T1) support control and computation, second-level systems (T2) perform supervision, deduction, and learning, and third-level systems (T3) use abduction and improve their learning methods (Nikolić, 2015).

Fig. 5. Two major types of semiosis: protosemiosis is the use of signs/signals for automated control of actions, whereas eusemiosis requires signs that point to objects interpretable by mind. A portion of the phase space of mind with four attractors is shown on the right.

Individual memory is not heritable, but, the accumulation of knowledge in newborn organisms is facilitated and directed by heritable features of the body, such as effector organs (e.g., legs, tail, and mouth), sense organs, and neural system that connects all organs with the brain. Thus, the mind of newborns is not a tabula rasa (blank state). The structure, sensitivity, and resolution of senses determine which patterns an animal can learn in its individual life. The size of the body, life span, and movement speed also contribute to the perception and interpretation of the world. Thus, animals of different sizes (e.g., a cow and ant) perceive and use the same environment
(meadow) in entirely different ways (Uexküll, 1982). Because of these heritable constraints, Umwelten of conspecific organisms are similar to each other, which is a pre-condition for the emergence of language to transfer the knowledge of objects within a society of individuals. Language requires references to abstract objects such as qualities, quantities, and relations that generalize the features of real objects. Thus, languages are based mostly on symbols, whereas icons and indexes play only supplementary roles. Language is fully developed only in humans, but higher animals also have a limited capacity for language-like communication (Čadková, 2015).

In pre-language communication, messages carry information only about the sender (e.g., emitting pheromone means “I am ready to mate”) or its immediate environment (e.g., bird warning call means “I see a predator”). In contrast, language messages carry information about other things that are not necessarily perceived and can be just named. Teaching appears as new kind of communication activity, which supports the development of language skills. In contrast to other species, humans actively change their environment which becomes progressively artificial, a noosphere (Vernadsky, 1945).

Protosemiosis appeared together with the origin of life, and it is still present in biochemical signaling networks within cells (Fig. 6). It is automated and does not include adaptive learning, although habituation and calibration of receptors are possible. Eusemiosis appeared much later in evolution, together with the origin of minimal mind, which likely happened during the transition from prokaryotes to eukaryotes. However, eusemiosis reached much higher complexity and reliability in animals with neural systems, and especially in those with brains. Although eusemiosis relies on protosemiosis at the lower level, it cannot be reduced to protosemiosis.

In a highly-redundant system, where signals can travel by thousands or millions of alternative pathways in order to perform the same function, it is meaningless to study individual signals. Thus, the topology of system-level attractors appears more important in eusemiosis than specific molecular-level signaling pathways.
Fig. 6. Protosemiotics started with the origin of life, and eusemiotics stated with the origin of minimal mind.

Mind exists as a functional subsystem in an organism, but it is not separated from the body as it was assumed by cartesian dualism. Organs that specialize in mental functions are parts of the body and they would not work without support from protosemiotic (non-mental) activity of their subagents (e.g., cells and molecular complexes). It seems equally unproductive to view mind as an abstract system without content, as proposed in radical enactivism (Hutto & Myin, 2013). Mind does not operate in a thin air, it requires properly organized, interconnected, and functional body; and the body cannot be formed without mind-like regulation of embryo development (Levin, Pezzulo, & Finkelstein, 2017). Thus, minimal mind should be present in the egg and early embryo, long before the body and brain develop and become functional. Learning of animals is facilitated by a large number of reflexes, which are components of the innate mental content. Newborn animals would never survive without these reflexes.

6. Agency and computation

The notion of computation is often unreasonably generalized by assuming that life is just another form of computation, or that the whole Universe is a giant computer. Any form of computation is represented by a universal Turing machine, which is a deterministic automaton carrying a
potentially infinite tape with discrete positions. A position on the tape can be blank or may carry a symbol from a finite alphabet. The machine reads a symbol at the current position, then, depending on the symbol and the internal state of machine, it may do several actions such as replacing the symbol, changing internal state, and/or moving to an adjacent position. What a Turing machine cannot do, however, is construction: it cannot make a knot on the tape, repair itself, make a copy of itself, build a wall, or make a connection to an alternative source of energy.

Computers are slave agents of humans, they operate on a syntax level without semantics, and thus their functionality fits to the definition of protosemiosis. Computers have no internal representation of objects around them or inside them. Humans often use object-oriented programming, but a computer does not distinguish a pointer to an object from any other memory address. Robots certainly have more autonomy than regular computers. In addition to a Turing machine, they are usually equipped with sensors, cameras, and various effectors (e.g., hands, feet, wheels, engines, or weapons), which help to establish a bidirectional interaction with the world. Their software may include cognitive architecture – an internal behavioral model that includes the internal ontology interconnected with sensing, acting, and goal-seeking. Robotic cognitive architecture can be formalized and programmed (Kurup & Lebiere, 2012). The main problem in developing artificial cognition is that machine cognitive architecture is currently designed by humans; thus, it remains constrained by the scope of human knowledge. I am not aware of robots that have constructed a substantial portion of their own cognitive architecture; thus, even advanced robots are still protosemiotic. Progress in artificial cognition seems to require breaking the “umbilical cord” with human designers and obtaining competence for advanced learning and supporting extensive interaction with unpredictable objects and events. An important step towards computer autonomy was the development of programs capable of reinforcement learning without human guidance (Silver et al., 2017). It is conceivable that this step can be achieved by augmenting computers with some features of agency such as goal-directedness, adaptability, and construction, which are briefly explained below.

Goal-directedness is a top-down organizational principle of agents (see
section 3.4), which is especially important in human cognition (A. Bandura, 2001). Although higher levels functions depend on the execution of lower-level functions, the higher level needs to be empowered to control, organize, and even construct lower levels. Computers were not designed to be goal-directed agents, however evaluation of goal efficacy and anticipation appeared necessary for developing the new generation of computing devices. For example, AlphaGo Zero computer program for playing a complex Chinese board game utilized an automated strategy evaluation algorithm that allowed self-learning without human input (Silver et al., 2017).

Adaptability of agents depends on the existence and accessibility of “adjacent possible” states (Kauffman, 2000), where agents may have a chance to survive in critical situations or environments. The number and diversity of such states is often limited; thus only those lineages of agents/organisms survive and diversify during long-term evolution that managed to expand sets of adjacent possible states. Goal-directed agents can memorize paths (or recipes) to perspective adjacent possible states and thus gain adaptability or evolvability (Conrad, 1983). This means that eusemiotic agents interpret objects as “affordances” (Gibson, 1975). In other words, they create sign relations that empower them to access adjacent possible states on demand. Donald Favreau (Favreau, 2015) called this phenomenon “creation of relevant next”. As a result, agents gain “semiotic freedom” (also known as free will if applied to humans), which is a higher form of adaptability. Hoffmeyer (Hoffmeyer, 2010) defined semiotic freedom as a “... capacity for responding to a variety of signs through the formation of (locally) ‘meaningful’ interpretants”. Meanings include recipes for construction as well as possible activities that may include each object type. There is no depth limit for generating new meanings; one recipe for construction or utilization of an object is never enough, there is always a question: what else can be done? In contrast to computation, generation of meanings requires external activity of agents, such as making and modifying objects in the outside world and, possibly, self-construction. This activity is then evaluated as success or failure in respect to agential goals and used to update the cognitive architecture.

Advanced generation of meaning also includes simulations, and it may be
argued that simulation can replace real activity as a path towards expanded knowledge. This argument however, ignores the fact that simulation requires a well-validated model of reality, and development of such models includes extensive experimenting with the object of study. Thus, simulation cannot replace activity.

Construction is the essential component of living organisms, which are primary agents and need recurrent self-construction and self-reproduction. Construction of functional components is the main business of a living cell, and it is conceivable that the same will be true for “constructive computing” in the future. But construction is almost absent in human computation technology, which is a serious obstacle for developing self-repairing, self-organizing, and analog computers. So far, 3D printing is the only computer-guided construction technology. It needs to be improved in various ways, such as miniaturization, combining with assembly robots, and self-assembly. Traditional silicon chips are too slow for semantic learning, and they were not designed for this task. Thus it may be necessary to develop self-organizing computers that combine digital and analog computing with self-construction. Computing modules (e.g., neural networks) can be assembled or disassembled depending on demand and relative efficiency. Analog computing devices change during their function; thus, they have to be replaced or recycled continuously by construction. Constructive computing may appear most advantageous in multilevel systems. Multilevelness is a fundamental feature of all living organisms, which is necessary to support robustness and evolvability (Sharov, 2016b). Damage of body parts can be meliorated by regeneration or compensation by remaining subagents. Moreover, mistakes at one level (e.g., mutations) are corrected and creatively reinterpreted at another level of hierarchy. I see the future of multilevel computing in converting mini-processors into mini-agents with productive goal-directed activity and semantic interaction with their neighbors. Mini-agents may become specialized in producing specific computational resources and scaffolds or in delivery of these products to other agents.

In summary, future artificial agents will likely become goal-directed, adaptive, semantically empowered, analog, and constructive. Hopefully, these features will support autonomous learning and self-enhancement. In
this respect, artificial agents will start resembling living organisms, and their cognitive capacities may reach or even exceed human mind at least in some areas of knowledge.

Acknowledgements

I thank Dr. Kenneth Augustyn for organizing the Special Issue and anonymous reviewers for valuable suggestions that helped me to improve the manuscript. I declare no conflict of interest in publishing this manuscript.

References


