

JANUS-HEADED ROBOTICS: DILEMMAS AND PARADOXES IN ROBOT ETHICS

ENDRE E. KADAR

Department of Psychology, University of Portsmouth, Portsmouth, PO1 2DY, UK

In designing robots with high level of autonomy for safe and ethically acceptable interaction with humans, engineers need to have a profound understanding of human behaviour control including social interaction. The present paper was inspired by Janus, a god of transition in Roman mythology whose double face could be used as a metaphor to provide insights into current difficulties and ethical concerns of designing artificial autonomous agents. A critical review of biology and the theoretical background of ethical concerns in biology are used in order to gain insights in recent developments in robotic research and ethical concerns in the context of the history of Western Science. Aristotelian organic science was contrasted with the Newtonian-Cartesian science to discuss the dilemmas and paradoxes of robot ethics arising from insufficient understanding of the autonomy of an agency. These are presented from the two perspectives Janus head is facing: looking backwards to see the dangers and benefits of past and current limitations of our knowledge in robot designs; and looking forward benefitting from or abusing of our future deeper understanding of the principles of normal functioning of autonomous agents.

1. Introduction

In designing robots whose interaction with humans is safe and ethically acceptable, engineers need to understand human perception and action in behaviour control including those skills that are needed in successful interaction processes [1]. In other words, fluent human-robot interaction requires profound understanding of human social and non-social interaction skills. Biology provides examples of systems with complex action repertoire and high level of autonomy. In contrast, designing artificial systems with high level of autonomy is still a challenging task even though the field of robotics is developing fast. To enhance safety, one of the common strategies is to develop mixed control system that allows human interference if the robotic control system is malfunctioning. An additional advantage of the mixed control design is that human performance could also be monitored and human errors can be corrected. But this mixed control systems have their challenges too. They also imply having finely calibrated basic interaction skills because the performance of an automatic controller should be similar to human-like behaviour otherwise the human controller would have difficulties in detecting the need to take over the control when a possible error/malfunctioning of the automatic control mechanism occurs. Similarly, the

automatic controller (artificial agent) should be able to monitor human performance in order to warn the human agent and/or take over the control when obvious human errors are detected.

The present paper was inspired by Janus, a god from ancient Roman mythology, whose metaphoric role could provide insights into current difficulties and ethical concerns of designing artificial autonomous agents. Janus was the god of transition, beginnings, doorways, passages and endings. He is depicted with two faces, but the original meaning of these two faces is different than contemporary interpretation of the term “Janus-headed”. The two faces are traditionally representing the two directions of looking, that is, looking into the past and future rather than having a double identity, the common negative connotation of the contemporary interpretation. But Janus was also associated with a different negative meaning in ancient Rome. He was the god who presided over conflicts from the beginning to ending. These traditional meanings are useful to illustrate the state of the art in contemporary robotics. On the one hand, looking back at the problems and limitations of past robotics to learn valuable lessons from the past. On the other hand, we can look forward and try to anticipate future challenges to be prepared to overcome foreseeable difficulties. To put differently, we live in a transitional period of the history of robotics. The past few decades were dominated by intellectual efforts to overcome various limitations of our knowledge and technology. These limitations allowed the use of limited autonomy with limited danger only but the new era of robotics will be dominated by increasing level of autonomy together with increasing levels of both benefits and dangers.

The discipline of robot ethics has emerged as a new area of research to mark this transitional period with lots of foreseeable and unknown dangers, sources of fears and anxieties in contemporary societies. The research of robot ethics and safety can be regarded as a battleground for opposing forces of development such as the desire of using autonomous designs and the fear of losing the ultimate human control over all robot systems. This situation, however, is not a new phenomenon in the history of modern sciences. For instance, a century ago, Physics had a similar period at the birth of quantum physics when the new insights into nuclear forces created the fear of the danger of abusing this knowledge that could lead to a nuclear catastrophe. Similarly, a few decades ago, Biology entered a similar transitional period when genetics research opened up new possibilities and at the same time became a source of various concerns. Fabricating new life forms with genetic engineering such as genetic manipulation can be used to treat genetic disorders but can also be abused in many different ways. New life forms are new agents with not known or not fully understood behaviour patterns. Even a simple system such as a bacterium or a virus could be modified in such a way that they could be used as biological weapons to kill huge number of humans. These biological concerns are very similar to the problems of new robotics, because contemporary researchers are able to design robots with much higher-level autonomy than in the past. The obvious similarity of fabricating new life forms and fabricating autonomous robots has inspired the present paper to learn valuable lessons. Thus, we first look at how biology was influenced by the tradition of Western Science, and how

the machine metaphor constrained its developments while a new biology (including bionics, genomics, etc.) has emerged. Then, our paper highlights that the prospect of a radically new approach to biology outlined by Rosen [2] can be considered as an analogue to the age of new robotics dominated by principles of autonomous agency building on Rosen's [2,3,4,5,6] insights. Janus is called again to preside over "another war", an "intellectual war" between the old tradition and a new emerging field of research, robotics of autonomous agents. The paper concludes with the presentation of a few dilemmas and paradoxes to consider that help overcome the limitations of old robotics and facilitate the birth of new robotics.

2. The Problem of Agency in Science

The birth of Western Science is usually associated with Aristotle whose physics (Natural Philosophy) and metaphysics laid its foundation. His view on science, however, was profoundly different than modern science that is based on the works of Newton, Galileo, Bacon and various other prominent thinkers of modern era [7,8,9]. Aristotle's Science (Natural Philosophy) is often labeled as organic based on the fact that motion was its central concept and each motion is associated with a mover (Bk. 8 of the *Physics* argues for the additional thesis that for each motion, whether natural or contrary to nature, there needs to exist a mover.) To put differently, within this system agency is central because the notion of mover could be used to define agency, which includes the first mover (God), living creatures and could also be artificial mechanisms (e.g., tools, machines, etc.). Another important part of Aristotle's science is the way things, changes, non-changes in Nature are described to provide explanations of changes and non-changes. To be more exact, the famous doctrine of four causes is meant to provide explanation of why things are as they are or why changes happen the way they do. These four causes can be grouped into two pairs of causes. The more fundamental pair of causes is the material and formal causes. Material causes provide potentials, which are or can be actualized by the formal cause. Thus, these two are related rather than independent aspects of explanations. The other two causes are more closely related to changes in Nature. The efficient and final causes are beyond but not necessarily independent of the material and formal causes. The efficient causes initiate processes that lead to changes, while the final cause is the end of the process of changes, what efficient causes intended to achieve. Aristotelian approach could provide a relatively simple framework for the theoretical issues surrounding the problem of agency in robotics, in particular autonomous robot designs. Specifically, the chain of causes can be infinite but Aristotle emphasized that often this chain has a beginning, and that is associated with an agent (god, animal, human, etc.), which is a primary cause of the unfolding changes over a period of time [7].

Some of Aristotle's ideas are still appreciated but modern science developed on a radically different basis. Philosophers (Spinoza, Descartes, Bacon) and scientists

(Newton, Galileo) during the Renaissance and ensuing centuries challenged some of the assumptions of Aristotelian science [8,9]. Specifically, Newton was keen on eliminating the agency from Nature by elaborating the foundation of Natural Philosophy on the basis of mathematics and laws that do not require or imply mover (agency) in explaining how things are and why changes occur. Newton developed classical mechanics, which does not have an agency/mover but Newton's ambition of creating a systematic approach that could explain everything in Nature failed (e.g., gravitational attraction remained unexplained). Despite its limitation, the Newtonian classical mechanics became so popular that the world-view of the universe as a huge mechanism became widely accepted. Accordingly, quite often the machine metaphor was used more specifically and the universe is likened to gigantic clockwork mechanism. This view is deterministic in a limited sense with regard to Aristotle's four causes. In the Newtonian universe, the efficient cause became dominant and the material and formal causes of mechanisms remained still useful but played far more limited role than in the Aristotelian system. The "spooky" final cause that is mostly associated with an agency became meaningless and unnecessary because it was regarded as either an illusion or something that can be reduced to other causes or something that can be eliminated altogether. The role of agency as a primary cause was eliminated because it was mostly reduced to a specific version of efficient cause.

Descartes [9] was a strong supporter of this mechanistic view, but interestingly, he realized that humans do not fit into this universal clock-work mechanism. He postulated that humans have free will that is essential ingredient of his dualist view of humans who consist of matter and another substance that constitutes the mind. To put differently, human mind is different to matter, because the bodily machinery is void of agency but the mind can play the role of an agency. Thus, Descartes, perhaps not intentionally, brought back Aristotle's agency into modern science. This mind-body dualism survived centuries and it is still popular within Psychology (in particular, cognitive psychology promotes, for instance, the theory of mind, etc.). Two important aspects of the Cartesian view, however, have changed dramatically. One of these was the completion of the mechanistic view of the world that includes mental aspects. This was due to development in language and logic about a century ago, which led to the computational theory of mind that brought human (animal) mind back into the Newtonian universe by making the mind computational (a specific kind of mechanism) [10]. The other important trend is that scientists no longer believe that mental life is unique to humans. Most scientists accept the view that has already been anticipated by Darwin, who claimed that mental ability is not unique to humans and argued for the continuity of mental life in evolution. Darwin [11] published a famous research on emotion that clearly challenged the Cartesian view, which postulated that animals are mindless and non-sentient beings. He suggested that many other non-human species have mind, which could be different for different species but in many ways similar to human mind.

These two changes have important impact on the notion of agency in biology, psychology and robotics. Across these disciplines, two opposing and competing trends

emerged during the past 150 years. The dominant trend (mainstream view) is to accept the scientific stance that living creatures and their mental life can be viewed as part of the mechanist universe. In other words, agencies can be described and their functioning can be explained as fundamentally mechanistic even though in many ways they are simply a specific kind of mechanistic systems. The other, opposing view of agency in these three disciplines is to separate living creatures from the reductionist mechanistic view of the universe. The separation can be made on an ontological basis similar to the Cartesian dualism, except the dualism would be between non-living part of the universe and living creatures. But there are other possible separations such as differentiating organisms with or without nervous systems, organisms with or without consciousness, etc. It will be argued in this paper that there is a fundamental similarity the way biology was developed and raised ethical concerns several decades earlier before similar developments raised similar concerns in robotics. This is the motivation for reviewing the notion of agency and associated ethical concerns in biology.

3. The Problem of Function and Agency in Biology

3.1. *Vitalism, Organismic Biology and Relational Biology*

The mystery of life, what makes living creatures different to non-living things is a very old problem. It was already discussed in ancient texts, including the works of Aristotle. This problem was framed in different ways as human knowledge developed through the ages. During the 18th and 19th centuries there were two alternative approaches: 1) the Cartesian mechanistic reductionist explanations of living organisms as an agency based on non-living matter; and 2) the vitalist view that postulated additional extra material/substance or principle required for life. Bechter and Richardson [12] summarizes vitalism:

According to vitalists, living organisms are fundamentally different from non-living entities because they contain some non-physical element or are governed by different principles than are inanimate things. Various forms of vitalism have been developed. Some argued that living entities contain some fluid, or a distinctive ‘spirit’. Other, more sophisticated versions promoted the idea of the vital spirit becomes a substance infusing bodies and giving life to them, etc. Modern vitalist views were developed to oppose Cartesian mechanistic view of organisms and there were prominent advocates of this view early as recently as in the twentieth century (e.g., Hans Driesch (1867–1941), an eminent embryologist, argued for the presence of an *entelechy*, a substance that controls organic processes, Henri Bergson (1874–1948) argued for *élan vital* to overcome the resistance of inert matter).

Organismic biology was another approach that found both the Cartesian mechanical explanation and vitalism unsatisfactory in searching for an alternative explanation of the unity of living organisms. Ritter [13] is often considered as the father of this view but some would argue that Aristotle was not a vitalist and his approach was the first

version of organismic biology.

Ritter [13] argued that the totality of the organism is as essential to an explanation of the behavior of its elements as elements are to an explanation of the behavior of the organism. Woodger's [14] *Biological Principles* attempted to explain organismic unity via principles of hierarchical organization. At the lowest level, elementary physical particles can be described by physical laws. According to Woodger, a system is perfectly organized hierarchically if parts of one level are the sole constituents of the next higher level. Woodger [14] considered organisms to be fundamentally hierarchical, even though there are exceptions such as the cardiovascular system, which is a subsystem that is not hierarchically organized. Although its tributaries are hierarchically organized over the size dimension, ranging from major arterial and venal pathways to tiny capillaries but is a more-or-less closed loop system. Whilst it is true that not all organisms possess cardiovascular subsystem (e.g., obvious exceptions are protozoa, bacteria, etc.), this subsystem is closely linked to the fundamental metabolic processes that are characteristic of all organisms.

Rashevsky [15] has taken a different approach. He checked various physical principles used in the literature and reviewed all known attempts to capture the "essence of life" and he realized that all of these efforts were futile. He found the cardiovascular subsystem useful to gain insights into strange circular processes in living system. Rashevsky [15] developed an argument to capture the essential features of living organisms based on the fundamental processes. Mapping the interrelationships of all basic processes in graph theoretic terms he no longer had to rely on cardiovascular system and he could propose a radically new approach that could accommodate plasticity (i.e., differences in species as well as individual differences in organisms) as well as circularity in fundamental biological processes.

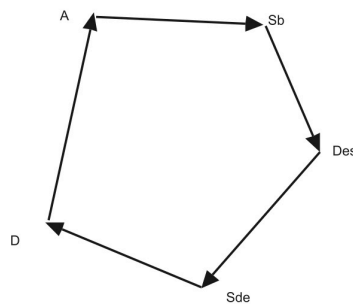


Figure 1. The common cyclic process of the relational structure of living organisms in Rashevsky's [15] study.

Rashevsky [15] decided to look at functional processes in a variety of organisms. Representing these basic processes by transformations using directed graph (digraph) techniques, he was able to develop a topological representation of

the organism as a functioning whole for a specific species. Figure 1 shows how the feeding process can be represented by a directed graph. Feeding begins with ingestion followed by digestion (D) and, in Rashevsky's [15] words:

This is followed by the absorption A of the digested food into the protoplasm of the cell, as well as by the rejection of indigestible waste, or defecation Df . The absorbed food is transported to various parts of the cell where a synthesis S_b of the body of cell follows. These general synthetic processes result, among other things, in the synthesis Des of digestive enzymes. This is followed by the secretion Sde of the digestive enzymes into the digestive vacuole, where they come into contact with the ingested food, and digestion D results. The directed path $IDAS_bDesSdeD$, which contains the cycle $DAS_bDesSdeD$, expresses a very fundamental property of every organism: in order that food may be assimilated and the body of the organism synthesized, the organism must be already present (pp. 326–327).

In sum, living organisms are shown to be different to non-living entities because they include a fundamental metabolic processing system that is common across all species. This fact could be framed by using Aristotle's notion of causality. What makes biological agents different? How biological agents maintain their lives? Why do living creatures differ to non-living entities? These and many other similar questions could be raised and the answers would include causations of various kinds that are linked to or based on the fundamental cycle revealed by Rashevsky [15]. This insight could provide the foundation of looking at the problem of autonomous agency in a new way.

3.2. Rosen's Complex System Approach

Rashevsky's student, Robert Rosen realized that the relational structure is actually a simplified form of more abstract mathematical structures, complex chain of mappings that are derived in category theory. This conclusion was based on a number of important steps in modeling and understanding living organisms, but the gist of his approach can be explained in simple terms. Rosen [2] noticed that biology is dominated by the mechanistic (Newtonian-Cartesian) view of the world that could only provide simple reductionist models that are inadequate to capture fundamental aspects of living organisms. He phrased the problem simply: What makes an organism different than a machine? Rosen's answer is seemingly simple: Machines are simple and organisms are complex. Machines are simple because they can be taken apart and they could also be put together from its pieces. Complex systems can be defined in contrast with simple ones. All systems are complex if they are not simple (i.e., the system is more than simply a some of its parts). Many of the physical systems that are known as simple mechanisms, in

fact, are complex. Organisms are all complex systems. This intuitive definition of complexity can be controversial because there are many other definitions of complexity, especially if the definition is based on the behavior pattern of the system. For sake of simplicity, the present paper uses Rosen's intuitive definition by contrasting complexity with simplicity and equating simplicity with mechanisms (machine metaphor).

Rosen has arrived at this position based on a critical review of modern science. He realized that Newtonian – Cartesian science lost the appreciation of Aristotle's four causes partly because the focus of modern science is on answering "how" rather than "why" questions. Even if the scientific analysis involves causes, the final cause is mostly missing or reduced to the other three. But final cause is clearly important in organisms, which are goal directed anticipatory systems [5]. Rosen argued that complex systems always involve all four causes that are mixed (interlinked) and cannot be separated as they can in machines.

The second important aspect of Rosen's approach is to focus on the functional process rather than structural components of the system, which is typical of machine-based views of the world. He realized that proper handling of time is a problem in dynamics, but he side-stepped this issue by using mappings. Mapping was also instrumental in revealing complex relationships based on the loops generated by mappings. With mappings, closed loops can emerge that would be similar to Rashevsky's closed loop in Figure 1.

Thirdly, following Rashevsky [15], Rosen shifted the focus from structural components typical of machines to functional components of complex systems such as organisms. He specifically investigated the problem of how to repair malfunctioning and noted that functional components could and should be replicated and this replication should be done by itself [3]. In other words, the replication will be self-replication. This issue of self-referential element in the system is an enigma for standard science (please note that Russell's effort to eliminate paradoxes from science resulted in elimination of self-referential statements from mathematical/formal logic of science). Rosen has developed a technique that addresses the problem at the proper level (high level of abstractness).

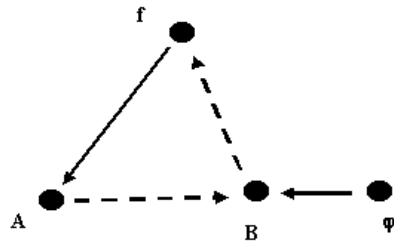


Figure 2. The broken lined arrows represent material causation and the solid arrows represent efficient cause. If we note that f has the meaning of some *relation* on the Cartesian product of sets A and B [$H(A,B)$]. This can also be written as a mapping: $f \varphi A \rightarrow B \rightarrow H(A,B)$.

In Rosen's approach, basic functional issues are modeled as relational processes that can be represented as a mapping diagram. For instance, f , represented as $f: A \rightarrow B$. A can be considered a set of initial conditions, B a set of final conditions, and f denotes the particular ways that some change was brought about to transform A to B . Rosen then assigns to this symbolism a set of entailments using the Aristotelian "why?" question. For example, to answer the question "why B ?" there are at least two answers. A is the material cause of B . The mapping, f , is the efficient cause. We might also ask "why f ?" and the answer would be "to bring A into B " assigning to f a final cause role.

Rosen tried and add repair function to the system represented in the mapping structure. This mapping, however, is just a simple one by adding another φ mapping component (see Figure 2) that could provide the cause of f , but this mapping could represent a simple machine that requires external interference (action). This is because the cause of φ is not going to be part of the system itself. By introducing another mapping to take care of the cause of φ , one could add another cause β and one could continue adding another cause to end up in an infinite regress [3]. Figure 3 demonstrates how this infinite regress could be avoided by creating a complex system that does not include external causes. To put differently, to eliminate infinite regress of causes, all causes should be made intrinsic rather than external. This would mean that the diagram would no longer have arrows that imply causes that are extrinsic (in other words, they are not entailed within the system if we use Rosen's terminology). This could also be expressed as a way to show how autonomy could be represented/defined with regard to a specific function. All functions could be used to achieve enhanced autonomy, which could be modeled by Rosen's mapping technique.

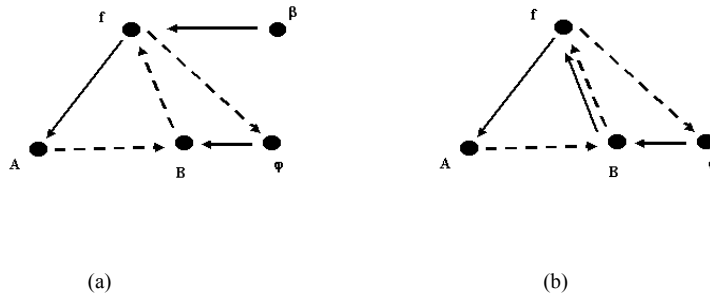


Figure 3. In a machine-based approach, diagram (a) shows that we can add the replication step β to entail (add a cause for) ϕ but it would have left β “unentailed” (not caused internally), for given that β is really $\beta: f \rightarrow \phi$, that is that it is the efficient cause of ϕ and that f is the material cause of ϕ , β is left “unentailed”. If it is established that it is possible that $\beta = b^{*-1}$, the inverse of $b^* : f \rightarrow B$, so that the diagram can be redrawn (b) to ensure that no external cause is necessary for repair function (i.e., repair became self-repair).

3.3. From Biology to Bionics and Ethical Concerns

Rosen was worried about the problems, including the danger and ethical concerns biological researchers may cause. Those concerns included bionics research (the study of *mechanical systems* that function like living organisms or parts of living organisms) raised contemporary societies were mostly worried. To put it differently, he was aware of the danger of fabrication and manipulation of life, but he was not deeply concerned because he argued that as long as machine-based thinking dominates bionics (*nomen est omen*), there is no real danger of causing major problems. Bionics is also defined as biologically inspired engineering, that is to say, this discipline is also associated with applying biological methods and systems in engineering. If bionics becomes a truly biologically inspired engineering, that is using insights derived by methods Rosen was working with, that will be alarming.

In our discussion of Rosen’s insights and theoretical works on complex systems, including organisms, the word autonomy was not central notion because that was not part of Rosen’s jargon in discussing his methodological and theoretical considerations. But it is obvious, that his notion of complexity implies that autonomous systems are complex. To put differently, without having a complex system embodied in a robot, the robot remains a simple machine only. Thus, recent concerns regarding the dramatic increase in the level of autonomy of contemporary robot systems can be directly linked to Rosen’s theory and can be discussed rigorously within his framework.

4. Lessons from Biology for Robotics and Robot Ethics

Clearly, robotics was and still is a discipline that is part of the mainstream of modern science. It is well-known that robotics research increasingly relied on biologically inspired performance during the past few decades. So what this fuss is about? Robotics researchers could be and are indeed often misguided by the fact that biology and psychology are also victims of Newtonian-Cartesian scientific tradition. When researchers are looking for insights in biology or psychology, they may find that mainstream views can be accommodated in robotics. But they do not realize that those methods and insights are not truly biological and natural, rather they are simplistic due to the limitations of mainstream science (mechanistic models). Specifically, when researchers are talking about autonomy in designing robotic system, they mean a fairly limited autonomy that would still leave the system a machine, that is a simple system rather than a complex one using Rosen's definition. This is due to the fact that robots are mostly dependent on external control. Also, the "internal" control mechanisms are not truly internal because they are part of the software that is either rigidly hard-wired (burnt in) or external to the hardware. Minimally, designing autonomous robots based on Rosen's conceptualization would require analogue systems, rather than the commonly used hardware-software digital systems that are based on the Cartesian dualist mechanistic tradition. To put it differently, most contemporary robot systems are the results of a historic continuity in robotics that can be summarize in a brief overview.

Early robots were designed based on clockwork-driven smart mechanisms but remained "mindless" systems until language processing became available with the help of mathematics and computers. In other words, with the advent of computers both mind and body became part of a mechanism and both are physical and computational. In the 1960s and 70s, the cognitive revolution in psychology was the byproduct of this development. Accordingly, the body of the agent and its surroundings are represented symbolically in the modular structure of the agent's Cartesian "mind". Information about the body of the agent and its environment is processed based on sensory data, movement plans are designed and executed based on the output of motor control (executive) modules. The idea of modular structure of the mind is also derived from the machine metaphor of mind, which is integral part of the Cartesian-Newtonian tradition of science.

During the past 50 years, artificial intelligence and robotics research adopted this cognitive architecture as evidenced by basic textbooks [16]. Based on our critical review of biology and Rosen's work, it should be clear that this cognitive architecture-based model of robotics still has major limitations. It is based on mainstream psychology that remained the prisoner of the Newtonian-Cartesian view of science, the science of mechanisms and machines but this approach has received several critical comments by prominent researchers. Brooks [17], for instance, noted that there is no need to create representations because the environment can itself be used for computations. Representational approaches are operating within the Cartesian tradition whereby the physical body (hardware) and

environment have a dual presence in the mind-software, which is extrinsic to the hardware and the physical environment. Others argued that Gibson's [18, 19] radical theory of perception should be used to eliminate representations. This approach had a few additional advantages including its intrinsic dynamic nature having the close link between perception and action by breaking the rigid separation of sensory and action control systems. These ideas could be adopted in robotics research [20]. Gibson's [19] theory could be helpful in paving the way towards a new dynamic and functional robotics that is fundamental in developing autonomous robots.

5. Conclusions

The present study is based on a critical review of biology and the theoretical background of ethical concerns in biology in order to gain insights in recent developments in robotic research and ethical concerns. Our review of recent developments in biology and robotics was situated in the context of the history of western science. Aristotelian organic science was contrasted with the Newtonian-Cartesian science that dominated the past few centuries. This comparison was based on two important aspects: The Aristotelian four causes and the problem of agency. Using Rosen's [2,3] theoretical work in biology, the importance of Aristotelian four causes were highlighted and category theory was introduced to demonstrate how abstract problems can be discussed in a formally rigorous way.

There are many studies on the problem of comparing natural and artificial agents with their intelligence. For instance, Dreyfus [10,21] published his philosophical studies in support of his arguments on the limitations of artificial intelligence. The author of the present paper has also argued in support of fundamental limitations of robotics based on the fact that engineers are not taking into consideration the difference between the life world and the abstract world of science [1,22]. However, these studies and their arguments are not sufficiently effective because of the lack of proper formalism that is suitable to discuss these abstract problems. The present paper tried to remedy this shortcoming and argued that Rosen's theoretical framework with the formalism of category theory is suitable to discuss agency at an abstract level. Mappings are representing various types of Aristotelian causes in Rosen's discussion of complex systems of biological organisms. The diagrams represent agencies that absorb external causes and driven by internalised causes with minimal external forcing (high degree of freedom, not dependence on external effect – high level of autonomy by definition).

The sources of fundamental ethical problem of bioethics/bionics and robot ethics/robotics are very similar (i.e., avoiding the creation/fabrication of agencies that could pose significant threat for humans). To protect humans from being victimized by the adverse affects of recent developments of biology or robotics similar ethical concerns can be associated with artificially created agents. It is

argued that the complexity of autonomy and associated dangers can be discussed with the help of Rosen's mathematical modeling strategy.

Overall, the state of the art in robotics can be described by a Janus head whose faces are directed towards the past and future. Looking backward we can still feel the presence of the burden of the inherited limitations of the machine universe of the tradition of modern Western Science. But if we look into the future, we can see the lurking dangers associated with creating robots with increasing level of autonomy. The apparent paradox of robot ethics is a new version of the paradox of knowledge, which is one of the well-known old paradoxes of human condition. Ignorance is bliss and knowledge is power, but humans can suffer from the consequences of both ignorance and knowledge. Our goal is not to benefit from ignorance like swindlers tend to do by relying false promises rather we should learn the lessons from the failings of the past, so we can be looking forward to benefitting from robot autonomy while facing an coping with the challenges they bring with them.

References

1. I.E. E. Kadar, A. Koszeghy, and G. S. Virk, Safety and ethical concerns in Mixed Human-Robot Control of Vehicles. In M. I. S. Ferreira, J. S. Sequeira, M. O. Tokhi, E. E. Kadar, G.S. Virk, *A World with Robots: International Conference on Robot Ethics: ICRE 2015*. (pp. 135-144). Springer. (2017).
2. R. Rosen, *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*. New York: Columbia University Press. (1991).
3. R. Rosen, Some relational cell models: The metabolism-repair systems, in *Foundations of Mathematical Biology, Volume II*, R. Rosen, ed. Academic Press, NY. (1972).
4. R. Rosen, *Fundamentals of Measurement and representation of Natural Systems*. New York: North-Holland. (1978).
5. R. Rosen, *Anticipatory Systems: Philosophical, Mathematical & Methodological Foundations*, New York: Pergamon Press. (1985).
6. R. Rosen, On the limits of Scientific knowledge in *Boundaries and barriers: on the limits to scientific knowledge*. (J. L. Casti and A. Karlqvist, eds.) pp. 199-214. Reading: Addison-Wesley. (1996).
7. Aristotle, *The Complete Works of Aristotle*. J. Barnes (Ed.). Blackwell. (1984).
8. A. Stinner, The Story of force: from Aristotle to Einstein. *Phys. Educ.*, 29 77, (1994).
9. D. Garber, Descartes, Mechanics, and the Mechanical Philosophy. *Midwest Studies in Philosophy*, 25, (2002).
10. H. L. Dreyfus, What computers can't do: The limits of artificial intelligence. (1972).
11. C. Darwin, The expression of the emotions in man and animals. (1872).
12. W. Bechtel, and R. C. Richardson, Vitalism. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge. (1998).

13. 13. W. E. Ritter, *The unity of organism*. Boston: R.G. Badger. (1919).
14. 14. J. H. Woodger, *Biological principles, a critical study*. London: Kegan Paul. (1929).
15. 15. N. Rashevsky, "Topology and life: In search of general mathematical principles in biology and sociology". *Bull.Math. Biophys.* Vol.16: 317-348. (1954).
16. 16. B. Siciliano and O. Khatib, *Springer handbook of robotics*. Springer. (2016).
17. 17. R. A. Brooks, Intelligence without representation. *Artificial Intelligence*, 47(1), 139-159. (1991).
18. 18. J. J. Gibson, *The senses considered as perceptual systems*. Boston: Houghton Mifflin. (1966).
19. 19. J. J. Gibson, *The ecological approach to visual perception*. Original work published 1979. New Jersey: Lawrence Erlbaum Associates. (1979/1986).
20. 20. A. P. Duchon, W.H. Warren and L.P. Kaelbling, Ecological robotics. *Adaptive Behavior*, 6, 473-507. (1998).
21. 21. H. L. Dreyfus, *What computers still can't do: A critique of artificial reason*. MIT Press. (1992).
22. 22. E.E. Kadar. Mind the gap: A theory is needed to bridge the gap between the human skills and self-driving cars. In press.