

Philosophical ontology, artificial intelligence and moral

*I. Theoretical reconstructions of stratification of reality:
The Analyses of Maturana, Luhmann and Nicolai Hartmann*

The questions of physical, biological and psychological existence below the level of social existence are usually not important for answering problems of social theory. Although a sound understanding of the moral requirements essential to the existence of human beings and human communities and the resulting norms and values is not possible without taking into account the biological and psychological endowments of human beings, because the functioning of most social institutions rests on them. In this way, an analysis limited to the social context is usually sufficient. New developments in the field of artificial intelligence and the raising of the possibility of human rights and possibly genuine private rights in the case of increasingly advanced robots have made it interesting for me to ask whether, in the case of an artificial intelligence linked to the physical existence of a robotic body, it would really be possible to incorporate it fully into society. In this case, there is neither a biological, nor a psychological layer of being, and compared to the four-layered (physical, biological, psychological and spiritual) natural person and human society, it becomes questionable how the extensibility of established, social institutions can be applied to it.²

To answer this, in addition to the theory of social reality, it became necessary to review the theories that have already carried out the mapping of the levels of total comprehensive existence or, in other words, the layers of being of comprehensive reality. Physical existence as a prerequisite for biological existence and the indispensability of both for human society can be visited as evidence. But there are different approaches on how these levels of reality (and perhaps even more hidden ones) relate to each other and how their effect each other can be articulated at the theoretical level. In the following pages, I will first examine the solutions of the Chilean theoretical biologist *Humberto Maturana*³ and then analyse the modification by the German sociologist *Niklas Luhmann*. I will then compare these approaches with the ontological theory of layers of being of German philosopher *Nicolai Hartmann*.

I. 1. Humberto Maturana's structural connection solution

The starting point for biologist Maturana's theory is the closed self-reproduction of cells in living organisms in order to understand the systemic nature of more complex biological systems and the animal and human societies that are built on them. The self-reproduction of the internal molecular structure of cells by constant chemical processes is

¹ DSc., Professor Emeritus at the Faculty of Law, Eötvös Loránd University, judge of the Constitutional Court of Hungary

² For details on these questions, see Béla Pokol: *Künstliche Intelligenz: Die Entstehung einer neuen Seinsschicht? (KI – im Spiegel von Nicolai Hartmanns Ontologie)* In Pázmány Law Working Papers, Nr. 2018/12.) and Béla Pokol: *Die Seinsschichten und die Fragen der Roboterethik.* C:/Users/Dell/Downloads/academia.edu/Die_Seinsschichten_und_die_Fragen_der_Ro.pdf

³ To illustrate Maturana's theory, I use his joint book with his former student Francisco J. Varela, but since this theory was created years earlier by Maturana alone, I take this as his theory despite any additions by Varela. See Maturana – Varela: *The Tree of Knowledge. The biological Roots of Human Understanding.* Revised Edition. Boston, Massachusetts, Shambhala Publications, 1987.

completely determined by the internal structure of the given cells, and it absorbs only the necessary chemicals from its environment. When, in evolution, multicellular organisms began to form through the permanent interconnection of cells, this did not change the closed self-reproduction of each inner cell. Each cell in this closure is only connected to each other as a whole, and the internal processes of each cell can only be irritated by the other cells and their self-reproduction as an environment. Through irritation, the irritated moment of the chain-like internal reproduction of the other cell can then be transformed, and this change can change the internal structure of this cell, and then the cell as a whole is also somewhat transformed. However, this irritation also only works insofar as the inner structure of the cell allows this. If this is not the case, the irritation can kill the cell, but it cannot identify the external effect as an irritation and thus cannot force the cell to change.

Maturana calls this systemic closure and the self-reproduction of the cells *operative closure* and the connection of cells from multicellular biological organisms *structural coupling*. The more complex multicellular biological system thus simultaneously performs a double self-reproduction, whereby the constant renewal of individual cells and the renewal of the organism itself, consisting of cells joined together, takes place in its environment. Due to the mutual irritation of the cells, the structural connection produces their reciprocal (coordinated) change within the more complex biological system, but the organism itself reacts to the stimulation of its environment only on the basis of its internal determination. This means that the environment cannot influence the organism directly, but only to the extent that its internal structure specifies what can be perceived as stimulation by the environment, and then one of the components of the organism and thus the structure of the organism change somewhat. If this is not the case, it may destroy the organism's responses but not change them. Maturana called this theory of closed self-reproduction of cells and organisms the *autopoietic system theory* in contrast with the previous system theory, of which the change of the system was largely explained by the system's environment.

Maturana points out that more complex multicellular organisms can respond with change to the irritations of their environment, and this capacity for change is radically increased when their neurons can transition into a central nervous system. This is because the surfaces of the organism from which the effects of the environment are perceived connect to the surfaces of the organism's movement (muscles), allowing it to make many more changes (possibly changes of location) to the irritants. In this way, the organism, despite its inner determination, can perceive many more irritations and react to them with its structural change.

After analysing the cells and the multicellular organisms built up from them as second-order unities, Maturana examines the coexistence that results from the permanent interactions of more complex multicellular organisms, which can create permanent animal societies as *third-order unities*. In particular, when permanent coexistence has led to such reciprocal structural changes in the interrelated individual organisms that reproduction of the breed is only possible through mating of several individuals. In such cases, due to the better care of the offspring, the protection of the caring adult organism and the feeding of the offspring or the offspring by other individuals, the long-term coexistence is increased, and the structural connectivity or the communication diversity required for this in the coexistence is increased to a higher level. This communication is possible in many ways, for example, through the continuous exchange of chemical compounds between ants in ant societies, or through special singing voices for some bird species, the formation of responsive duets, and finally the linguistic communication field in humans becomes the field of linguistic communication.

Maturana's analyses show here the two layers of being of human existence, the biological layer of being and the cultural layer that is built on it. Through language, one can describe one's own inner relationships and environment with different linguistic distinctions and thus not only preserve one's adaptation to one's environment through one's instincts, but

it is also preserved for members of new and new generations through cultural fixation and transmission through linguistic distinctions: “Language enables those who operate in it to describe themselves and their circumstances through the linguistic distinctions of linguistic distinctions.” (Maturana/Varela 1987: 210.) While biological organisms and their communities continuously reproduce only their biological existence, the human community must continuously maintain and reproduce its cultural existence in addition to its biological existence: “In human social systems, the case is different. As human communities, these systems have operational closure too, in the structural coupling of their components. But human social systems exist also as unities for their components in the realm of language. Therefore, the identity of human social systems depends on the conservation of adaption of human beings not only as organisms (in a general sense) but also as components of their linguistic domains.” (Ibid.)

Maturana describes each layer of the two layers of human existence in such a way that, as long as at the level of biological existence, the interrelated units are tightly bound together and do not allow independence for the cells of their internal organisation and they are for the living organism. The human social system, however, allows its members a higher degree of autonomy.⁴ As an exception, he presents the society of Sparta, where the independence of each member was completely suppressed and, as a biological organism, close interconnectedness was realised in its social organisation, but this depersonalisation led to the loss of viability of the whole society of Sparta.

In a later study with another co-author, Maturana also showed another difference between the organisation of the biological level and the organisation of the linguistic-cultural level. While the biological system always lives in the present, the systems of the cultural level can, through observation and their linguistic fixation, use past experiences and possible future events in addition to the present, and human actions can also be shaped in the light of these experiences: “It is only for us human beings that as we exist and operate in language, we can generate reflections and explanations, in that the past, the present, and the future, have a presence and operational values in our living as explanatory notions of our experiences.” (Maturana/ de Rezepka 2007:4.)

The two-layered nature of human existence thus allows for temporal differentiation through the upper layer of existence, and thus Maturana’s theory in principle raises the process of more permanent formations of the cultural layer of human existence alongside the progressive process of the biological layer of human existence. Nevertheless, Maturana insists that the reproduction of the human world is always based on momentary determinations and causalities, and he therefore criticises the thinking that wants to predetermine the lives of our children today because of various fears: “We human beings create the world that we live in arises moment after moment in the flow of our living, how can we pretend then, to specify a future that will not belong to us because it will arise in the living of our children and will not be created by us? [...] We human beings live in the present; the future and the past are manners of the being in the present.” (Maturana/de Rezepka 2007: 9.) Thus, from this assertion emerges an underestimation in terms of cultural observation and sociological prediction as well as the overall cultural layer of being, and an overestimation of the characteristics of the biological layer of being.

In addition to the two layers of being, it should be noted that Maturana occasionally mentions that the physical world is constructed under biological conditions and that the biological system takes matter from its physical matter: “Thus, autopoietic unities specify biological phenomenology as the phenomenology proper of those unities with features distinct

⁴ “The organism restricts the individual creativity of its component unities, as these unities exist for the organism. The human social system amplifies the individual creativity of its components as that system exist for these components.” (Maturana/Varela 1987: 99.)

from physical phenomenology. This is so, not because autopoietic unities go against any aspect of physical phenomenology – since their molecular components must fulfil all physical laws – but because the phenomena they generate in functioning as autopoietic unities depend on their organisation and the way this organisation comes about, and not on the physical nature of their components (which only determine their space of existence).” (Maturana/Varela 1987: 51.) Thus, in the end, three layers of being emerge from Maturana’s theories, but he mentions the physical only as a material-giving layer (he uses the term “special phenomenology” for each layer of being), and he does not analyse the relationship between the physical and the biological any more, and therefore he does not try to establish more general connections between them.

1. 2. Luhmann’s Analysis on the system levels of world

In Niklas Luhmann’s writings there are several shifts of thought in order to be able to address the layers of being beneath the social world. His attempts can first be found in an article from 1974, then in his comprehensive summary of systems theory from 1984 and finally in his writings between 1987 and 1991. He deals with this topic particularly systematically in the writings between 1987 and 1991 in connection with the analysis on the relationship between psychic systems and social communication.

In the first material mentioned above,⁵ he examined the relationship between power and physical violence, which is its ultimate basis, and Luhmann also went into somewhat more details in this context about the relationship between the social world and its physical-biological layer of being. The model for this was provided by the media theory of Talcott Parsons, by whom physical violence was understood as the ultimate means of covering power as the generalised medium of exchange of the political system. The modern state, with its monopoly on the use of physical power, has power only as a generalised symbolic medium of exchange, insofar as it is available only as a last resort, but it does not normally have to use it. Just as money as a medium of exchange is based in an abstract way on the fact that we accept it because of the possibility of exchanging it for products, but this is not normally checked every time money is received and it is only in this way that money makes society function harmoniously. Luhmann generalises this example somewhat and sees that there is also a connection with such a physical-organic layer of being under the exchange medium of several social subsystems. As in the case of love as an exchange medium of intimate relations, it is sexuality that serves to fix love relations in a biological layer of life. This makes love as a communicative relationship built on a social level really strong and can give lasting stability to family relationships based on love-based marriage in modern societies. Similarly, in the case of science as a subsystem of society, from truth as a medium of exchange, the research work of finding truth and the related decisions can be organised only when *sensory perception* as the ultimate means of discovery is the ultimate basis for decision-making. If truth in relation to an appearance cannot be decided by a final, indirect perception, it is removed from the perspective of modern science.

Luhmann also makes some general observations in this study that also affect the relationship between the social layer of being and the physical-biological layer of being. One relates to the fact that the place of connection between the two layers of being is at the systems level of the social world, which is mainly based on the biological aspect of person,

⁵ The title of this study from 1974 is “Symbiotische Mechanismen” and is part of the book “Soziologischen Aufklärung 3” [Westdeutscher Verlag, 1981, 228-244. pp.] all the following citations are from this volume.

that is, at the level of interaction systems.⁶ However, if we know that in Luhmann's theory, social organisation gives a decreasing role to the level of personal interactions, and in the course of social evolution is increasingly eclipsed in favour of the formal organisations, and even within the level of interaction, personal interaction is increasingly replaced by impersonal written communication, then linking the social layer of being with the physical-biological layer of being at the level of personal interactions also means diminishing this link.

Another more general comment by Luhmann on the relationship between the two layers of being is that physical violence and sexuality in social relations not only occur as physical-biological facts in their cruelty, but they are also socialised and transformed by a symbolic dimension. That is, a transformation of these physical-biological facts is also carried out in the layer of being of the social world: "Gewalt, Sexualität usw. haben in diesem Sinne eine symbolische, nicht nur eine physische oder organische Wirksamkeit [...] die soziale und kulturelle Entwicklung von Medien-Codes Umdisposition im Bereich der symbiotischen Mechanismen erfordern mag." (Luhmann 1981: 231.) Finally, one of the most general observations in this study is that even if the evolutionary thesis is to be accepted that cultural-symbolic interference creates increasingly spiritualised social contacts in modern societies, this does not mean that there is a detachment from the phenomena of the physical-biological layer of being: "Dies Kontrastbeispiel erhärtet die Vermutung, daß selbst wenn man die These von einer kulturell-symbolischen Steuerung der gesellschaftlichen Evolution akzeptiert, der Bezug zur Sphäre physisch-organischen Zusammenlebens damit nicht bedeutungslos wird [...] Fruchtbarer wird es sein, davon auszugehen, daß sinnhafte Erlebnisverarbeitung und Kommunikation ihr physisch-organischen Substrat als Wirklichkeit nicht ignorieren." (Luhmann 1981: 241.)

Luhmann's next reflection on this topic is in his book "Social Systems" [1984] and here he already expressed his theses more generally in terms of the layers of being of reality. He started from the premise that since Kant, the old ontological conception of reality has been discarded by modern philosophical thought and the mere *Ding an sich* out there in the world can only be achieved through active constructions of consciousness, and although Kant's solution overly psychologises this problem, this can be more adequately resolved from a broad systems theory perspective. From this broader perspective, it follows that certain elements of reality are not predetermined by reality itself, but on the contrary, they are self-constructed by different systems (biological, psychological and social).⁷ On the basis of this view, Luhmann therefore states that the old view of the stratification of reality and its bottom-up construction must be rejected: "Es geht nicht einfach um einen geschichteten Weltaufbau, bei dem die unteren Schichten zuerst fertiggestellt sein müssen, bevor weitergebaut werden kann. Vielmehr werden die Voraussetzungen mit der Evolution höherer Ebenen der Systembildung selbst erst in eine dafür geeignete Form gebracht. Sie entstehen durch Inanspruchnahme. Deshalb ist Evolution nur durch Interpenetration, das heißt nur durch

⁶ "Beim Aufbau sozialer Systeme erfüllen diese Bedingungen Interaktionssysteme, die durch relative rasch strukturiert werden können, dadurch aber nicht festgelegt sind, sondern noch definierbar bleiben. Das gilt eindeutig für gemeinsames Wahrnehmen, für Drohung mit physischer Gewalt, für Kommunikation über nur sozial bzw. nur asozial befriedigende Bedürfnisse: Sexualität bzw. Befriedigung aus knappen Ressourcen. [...] Der Behuf auf präzise Organismen erleichtert mit anderen Worten die Herstellung der reflexiven Reziprozität der Perspektiven und damit die soziale Strukturierung der Situation." Luhmann, Niklas: Soziologische Aufklärung 3." Westdeutscher Verlag, 1981, 231. p.

⁷ "Anders als die Wortwahl und Begriffstradition es vermuten lassen, ist die Einheit eines Elements (zum Beispiel einer Handlung) nicht ontisch vorgegeben. Sie wird als Einheit erst durch das System konstruiert, das ein Element als Element für Relationierungen in Anspruch nimmt. Diese De-Ontologisierung und Funktionalisierung des Elementansatzes ist in der modernen Wissenschaftsbewegung durch die Mathematisierung der Naturwissenschaften in Gang gebracht worden." Luhmann, Niklas: Soziale Systeme. Frankfurt am Main, Suhrkamp, 1984, 42. p.

wechselseitige Ermöglichung möglich.” (Luhmann 1984: 294.) So with this general rejection, Luhmann does not go into details about the ontological theories of the stratification of reality, nor does he even critically cite the relevant authors in his work, including Nicolai Hartmann. If we then approach Luhmann’s solution in analysing the relationship between the social world and the human being, we see that he could even have come to Hartmann’s conclusions.

Luhmann rejects the “human being” as the basic unit of social organising in traditional social theories because, given the critique of the humanist tradition, he sees that by conceiving of the human being as the unit of society, the whole conception of society is also constructed as if society were only for human beings, for its ultimate element. This in turn forces social theory into an ideological-normative framework and this framework must therefore be disposed of in order to be unbiased. Therefore, the unitary concept of “man” must be discarded and only the “person” can enter the picture as an element of the social world, which can only be understood as a point of identification for normative and cognitive expectations for individual man (person) in formations of the social world. That is, the person represents only the aspects of the social side of the human being and cuts off the biological and psychological aspects of the human being, and the human being itself, together with its biological and psychological aspects, is outside the social world in Luhmann’s theory and is defined as its environment. On the other hand, the reduced “person” as a social aspect of the human being appears only as a meaningful construct in the reconstruction of the social world: “Wir wählen den Ausdruck “Mensch”, um das festzuhalten, daß es sowohl um das psychische als auch das organische System des Menschen geht. Den Ausdruck “Person” wollen wir in diesem Zusammenhang weitgehend vermeiden, um ihr für die Bezeichnung der sozialen Identifikation eines Komplexes von Erwartungen zu reservieren, die an einen Einzelmenschen gerichtet werden.” (Luhmann 1985: 286.)

Luhmann’s next theoretical constructional step is to link the biological and psychological aspects of the human being, perceived as the necessary environment of the social world, to the concept of penetration. In effect, this means connecting the social layer of being with the psychological and biological layers of being of reality. With penetration, an entity at one system level makes its complexity available to another entity at another system level, without which it could not build and exist. (So without biological life, the entities of the social world could not exist). And we can speak of mutual penetration, *interpenetration*, when the entities on both system levels transform each other through contact.⁸ If we now look at Hartmann’s analyses of the division of the human being and the human world into four layers of being and the focus on the highest meaningful layer of being of human society, Luhmann would also have set aside the human-based social picture, which he rejected because of its defence against the humanist bias, on the basis of Nicolai Hartmann’s analyses. If he had assumed this, he could also have attempted to build a more accurate theoretical reconstruction of society on the spiritual being layer of reality.

Luhmann largely restricts the connection between social systems and psychic systems to their common moment of *meaning*, and therefore the whole spiritual-emotional processes of the psychic system can only appear narrowed.⁹ However, since the spiritual-emotional

⁸ “Von Penetration wollen wir sprechen, wenn ein System die eigene Komplexität (und damit: Unbestimmtheit, Kontingenz und Selektionszwang) zum Aufbau einer anderen System zur Verfügung stellt. In genau diesem Sinne setzen soziale Systeme “Leben” voraus. Interpenetration liegt entsprechend dann vor, wenn dieser Sachverhalt wechselseitig gegeben ist, wenn also beide Systeme sich wechselseitig dadurch ermöglichen, daß sie in das jeweils andere ihre vorkonstituierte Eigenkomplexität einbringen. [...] Im Falle von Interpenetration wirkt das aufnehmende System auch auf die Strukturbildung der penetrierenden Systeme zurück.” (Luhmann 1984: 290.)

⁹ “Von hier aus wird besser verständlich, weshalb der Sinnbegriff theoriebautechnisch so hochrangig eingesetzt werden muß. Sinn ermöglicht die Interpenetration psychischer und sozialer Systembildungen bei Bewahrung ihrer Autopoiesis; Sinn ermöglicht das Sichverstehen und Sichfortzeugen von Bewußtsein in der

processes are constantly in contact with the instincts of the biological processes from below, and they also carry their impulses forward through transformation, this entire layer of being falls out of Luhmann's perspective. Using Maturana's theory of structural coupling, where connection occurs only at the level of elements and through this level structures change only indirectly due to the change of their interconnected elements, Luhmann resolves the connection of layers of being as follows: "Es sind Differenz und Ineinandergreifen von Autopoiesis und Struktur (die eine sich kontinuierlich reproduzierend, die andere sich diskontinuierlich ändernd), die für das Zustandekommen von Interpenetrationsverhältnissen zwischen organisch/psychisch und sozialen Systemen auf beiden Seiten unerlässlich sind." (Luhmann 1984: 297.) This means, for example, that the two system levels have constant contact due to meaning as their common building block, and both the meaningful contributions of the psychic systems of the communication participants and the meaningful contributions of the social subsystems involved in the communication contribute to the further development of these social systems, and something also transforms the participating psychic systems in this contact. But only to the extent that the structure of the psychic system allows it, but on the other hand, this structure itself is somewhat changed by the change in its element. Besides the fact that in this way only the intellectual part of the psychical layer of being is taken into account and the influence of the biological instincts on the intellectual material of the psychic system is cut off from below, it can be said in general that in this theory there is no bottom-up structure of the layers of being, but only a mutual irritation and resulting partial change.

To give a concrete example of the effect of this narrowing, we can mention love communication in the social world and the resulting lasting intellectual-emotional relationship. It is quite obvious that in this relationship the existence and degree of biological sexual instinct in the two partners plays a significant role, and if this is minimal for some biological reason in the case of one partner, for example, the love relationship in the case of that partner will run quite differently from the case of the other partner participating in the relationship who has a strong sexual instinct, and presumably this difference will soon lead to a breakdown of the relationship. That is, love becomes biologically permeated not only by an intellectual relationship but also by sexuality, and this biological component is transformed to the emotional-spiritual level. And this, combined with such a biological-spiritual basis, creates a completely different kind of love communication and partnership than between two partners with the same biological sexual urge. However, the same applies to the analysis of a number of other social phenomena where the direct effects of strong biological stimuli occur, such as the functioning of the life instinct in the face of mortal danger, which in the soldier at war leads to the abrogation of a number of moral norms hitherto considered sacred and the revival of the older killing instinct. Then, after a war, these "devastated" personalities, remodelled for killing, cruelty and even appreciation of these activities, will be almost unfit for reintegration into civilian life for a long time. In contrast, such instincts of life in civil life mainly degrade and, on the other hand, the remodelled instincts of helpfulness build up and on the basis of such restrained and remodelled biological stimulus and the spiritual-emotional life influenced by it, the processes of intellectual communication between the participants of social life and the resulting social institutions also proceed differently. Of course, further examples could be given of the co-existence of biological stimuli, spiritual-emotional life and social-intellectual communication formations, but these two examples were perhaps enough to show that Luhmann tears the levels of the biological, conscious and social system (or in Hartmann's words: layers of being) too much apart and thus he can only grasp the functioning of real formations through narrowing and distortion.

Kommunikation und zugleich das Zurückrechnen der Kommunikation auf das Bewußtsein der Beteiligten." (Luhmann 1984: 297.)

To understand this problem, this kind of structural coupling between the psychic systems and the communication systems of the social world can be seen in Luhmann's studies from 1987 to 1994.¹⁰ He elaborated his thoughts on the coupling of the two system levels most fully in the paper "Die Autopoiesis des Bewusstseins", which he did not change in his later works.

It should be emphasised that since the beginning of his reflections in 1974, Luhmann has radically changed his conception of the system by adopting the idea of autopoiesis, and has now moved to a process of circular closure in the operative sequence of system elements, rather than paying attention to structures. Whereas in 1974, with the emphasis on symbiotic mechanisms in Luhmann's analyses, biological stimuli also played a role during intellectual communication processes (e.g. the continuous influence of the sex drive on the maintenance of love), here they have completely disappeared by the end of the 1980's. In this way, the processes of the consciousness system here mean only the connection of intellectual communications. From thought to thought, the processes of consciousness move as carriers of the psychic system, and there is no psycho-emotional or biological instinct here: "Die Autopoiesis des Bewußtseins ist das Fortspinnen mehr oder minder klarer Gedanken, wobei das Ausmaß an Klarheit und Distinktheit selbstregulativ kontrolliert wird je nachdem, was für einen bestimmten Gedankenzug – vom Dösen und Tagträumen bis zur mathematischen Rechnung – zur Einteilung der Gedanken und zum Übergang erforderlich ist." (Luhmann 1995: 61.) This view of consciousness, limited to a purely intellectual aspect, can of course reveal very clever things about this kind of intellectual process in individual consciousness and, for example, for the maturation of thoughts, and Luhmann describes profound insights about this. For example, he emphasises that only through constant self-observation in consciousness is the new thought brought to a more mature level, and only through this does it also become accessible to consciousness: "Für sich ist ein Gedanke also zunächst nur ein Gedanke, für andere Gedanken dagegen, das heißt für den rekursiven Prozeß des Systems, ist eine Vorstellung." (Luhmann 1995: 65.) Anyone who has had a completely different new idea than before and, after consolidating it himself, has tried to figure out how it was born in his head, will get a lot of help from this study by Luhmann, and will feel that it might actually have happened that way. These thoughts of Luhmann are to be appreciated, but this does not alter the fact that while Luhmann did an excellent job of working out the intellectual aspect of individual consciousness, he completely cut off from it the effects of emotional and biological instincts.

With regard to emotions, even close followers and critics of Luhmann noticed his theoretical shortcomings, and in 2004 a special issue of the magazine "Soziale Systeme" in Bielefeld was devoted to this topic. Let us first look at *Dirk Baecker* one of the defenders of Luhmann and *Luc Ciompi* that of the critics. Baecker basically acknowledges that emotions do not play a prominent role in Luhmann's theory and within Luhmann's analyses of the conscious processes of mental systems.¹¹ However, some subordinate functions are stated once or twice by Luhmann in relation to emotions, e.g. that they act as an immune system of consciousness when the conscious thought process falters, and then disappear again when the thought process is restored. According to his writings, Dirk Baecker also considers this subordinate role to be too little to some extent, but finds it unnecessary to accept Luc

¹⁰ These studies appeared together in Luhmann's volume of 1995, and I will cite them here, see Luhmann, Niklas: *Soziologische Aufklärung 6. Die Soziologie und der Mensch*. Opladen, Westdeutscher Verlag, 1995.

¹¹ "In der Tat spielen Emotionen im Gesamtwerk von Niklas Luhmann keine sehr prominente Rolle. Luhmann hat zwar die Stelle markiert, an der seines Erachtens eine Theorie der Gefühle von seiner Theorie sozialer Systeme abzweigen, beziehungsweise an diese Theorie angedockt werden könnte, doch hat er dem Phänomen der Gefühle nie die Aufmerksamkeit geschenkt, die etwa Ciompi angesichts der nicht nur psychischen, sondern auch sozialen Prominenz des Phänomens für geboten hält." (Baecker 204: 9-10.)

Ciampi's radical expansion regarding the role of emotions in the functioning of the mental system. He does, however, mention a practical reason in this rejection. In his view, sociological emotion theory is still at such an uncertain empirical level that it is better to avoid this uncertainty.¹²

However, Baecker's explanation of why Luhmann was reluctant to acknowledge the role of the emotional world in social events and social institutions is interesting. According to his description, dealing with emotions was alien to Luhmann's entire theoretical style, and he attributes this to the historical events of Luhmann's youth, when the suppression of fundamentalist and fascist movements in Hitler's Germany shaped Luhmann's life conditions and these movements relied heavily on the release of collective emotions and their damaging effects. Only in Luhmann's late work did this begin to relax somewhat, but this no longer changed the original directions.¹³ This is only Baecker's meditation on finding the cause in Luhmann's thinking, but since he was right next to Luhmann in Bielefeld for many years between 1980 and 1995, he was able to get to know this side of Luhmann, who was otherwise very reserved in his inner personal affairs. Therefore, this explanation must be taken into account. Of course, this is no excuse for Luhmann's theoretical deficiency in this area, so Ciampi's role for emotions in the mental processes of the psyche and in the structural construction and events of the social world deserves the most attention.

However, the omission of emotions from the analysis of mental processes cannot, in my view, be explained simply by Luhmann's particular personal socialisation. This stems from his theoretical decision made much earlier, almost at the beginning of his theoretical training. According to this, the meaning is the common building block at the system level above the level of the physical and biological system, i.e. at the level of both the psychic and the social systems. This basic decision does not include the capture of emotions, and this capture could only have been made by correcting this basic decision. This correction, together with a number of other amendments, could then only have been inconsistent, and Luhmann was never prepared to make such corrections in his theory.

In addition to researching the psyche, Luc Ciampi, as a medical psychiatrist, also dealt with emotions and continuously monitored the results of psychological and social psychological experiments in this field. In his analyses it becomes clear that human thought is surrounded at all moments by the emotional components of the psyche and that for the most part all our knowledge is not consciously and perceptibly accompanied by our emotional disposition. (See Ciampi 1997 and 2004) When we accumulate knowledge and experience, our emotional attitudes to the subject are pressed into our memory along with knowledge and experience, and when knowledge and experience are later recalled, these emotions also appear, inseparable from the experience and knowledge. We do not think purely intellectually, but what arouses our interest, what memories and associations are pushed up or remain stifled

¹² "Affektkontrolle hat es immer auch mit dem Adressieren von Affekten zu tun, und niemand weiß, welche Emotionen dies hervorruft. Die Soziologie der Emotionen stellt diese Unvorhersehbarkeit und Unberechenbarkeit von Emotionen in Rechnung, indem sie sie als Phänomene beschreibt, in denen die verschiedenen Systemebenen des Organismus, des Bewusstseins und der Kommunikation aufeinandertreffen, ohne dass man genau wusste, wie sich diese Systeme ausdifferenzieren und wie sie miteinander gekoppelt sind. [...] Mir scheint diese Frage nicht entscheidbar zu sein, solange die Theorie der Emotionen so schwach ausgearbeitet ist, wie dies gegenwärtig der Fall ist." (Baecker 2004: 12-13.)

¹³ "Diese Abstinenz gegenüber dem Phänomen der Gefühle hat bei Luhmann Gründe, die etwa mit seinem Theoriestil zu tun haben, der wiederum seine zeithistorischen Motive hat. Wer eine Theorie entwirft, der es um Differenzierung im Umgang mit der Komplexität der Gesellschaft geht und die nicht zuletzt in der mangelnden Differenzierung Einfallstore für fundamentalistische und im Extremfall faschistische Gesellschaftsvorstellungen sieht, der wird sich nicht unbedingt mit Phänomenen beschäftigen, deren Funktion möglicherweise gerade darin besteht, Differenzierungen hochselektiv und problemgenau aufzuheben oder zumindest zu verwischen. Genau deswegen ist es ja so bemerkenswert, dass Luhmann diese Haltung in seinem Spätwerk lockert und sich ein "Verbindungsmedium" Werte überhaupt vorstellen kann." (Baecker 2004: 10.)

at the sight of something, are important, and so certain paths of thought are blocked and others brought to the fore. That is, thinking and the experiences used to do it are emotionally coloured, and intellectual thinking and its thought processes constantly run together with emotional processes in our brains. Ciompi points out that emotions run briskly with our intellectual processes, especially when we gain new knowledge in a new situation. When we later incorporate it alongside existing knowledge, our new knowledge gained from new experiences already becomes routine and with repetition, emotional processes and emotional energy consumption are minimised. However, if a problematic situation arises later in connection with these, the emotional dimension is also reactivated and it runs alongside the problem processing on an intellectual level as well.¹⁴

As a summary, Luhmann limits the psyche, the psychic system (or several times just called “consciousness”) to intellectual processes that run from thought to thought. However, it is clear from Ciompi’s analyses that this view misses the real human psyche and cannot adequately capture it without emotional processes. In contrast with Luhmann, *Nicolai Hartmann*, with his comprehensively formulated layer theory of reality, was able to formulate more adequately the ensemble of biological, psychological and intellectual components of human existence and their cooperation. Let us therefore turn to Hartmann in the following when researching the topic.

I. 3. Nicolai Hartmann’s views on the layers of being

Hartmann cultivated a kind of philosophy which he conceived not as another intellectual activity in addition to the sciences, but as the most comprehensive kind of science, the most general synthesiser of more specific scientific knowledge.¹⁵ As a result, based on the empirical scientific results of his time, he tried to rethink the most abstract ontological categories inherited from the past. Thus, he tried to summarise certain categories of Plato, Aristotle, the Neo-Platonists and then the great philosophers of the Middle Ages and the (Early)-Modern Age (Duns Scotus, Leibniz, Kant, Hegel and Neo-Kantianism) in the light of physics, biology, psychology and recent social science findings. In contrast with Luhmann, who essentially compiled his general social theory only from certain areas of the social sciences and only considered the biological and physical system levels in passing, Hartmann systematically examined the most comprehensive interrelationships of this system level. Thus, in his first major work, he dealt with the philosophical questions of biology and later, in addition to his many years of work in the history of philosophy, he also dealt with the ethical-moral sphere of the social world in his systematic ethics, and also with artistic activities in a thorough monograph on aesthetics. Then he began to summarise these investigations as a theory of the layers of being in the real world, an idea that had already appeared in Aristotle. He published this theory in 1933, first in relation to the intellectual layer of being in a systematic monograph, and then in 1940 included a comprehensive table of the layers of being of the real world and the regularities between them. According to his ambitions, this theory can be read as a synthesis of the most comprehensive scientific contexts of the time.

Hartmann sees the real world as a combination of four layers of being built up on the

¹⁴ “Zur Alltagslogik wird alles anfänglich Neue und Aufregende, das sich oft genug wiederholt. Die beteiligten Emotionen werden zwar durch Gewöhnung zunehmend unbewusst, behalten aber ihre typischen Operatoreffekte, “wie selbstverständlich” bei – ein höchst sinnvoller Mechanismus, der gleiche Leistung mit viel geringerem affektenergetischem Aufwand ermöglicht [...] Passiert indes Ungewohntes, so flammen alle latenten Affekte wieder auf.” (Ciompi 2004: 10.)

¹⁵ For a detailed description of Hartmann's philosophy, see Stegmüller, Wolfgang: Hauptströmungen der Gegenwartsphilosophie. Alfred Körner Verlag, 1999. 3rd edition.

top of each other in the course of evolution. The biological layer of being, which is built on the top of the physical layer of being, and which, after a phase of evolution, created the psychological layer of being in some animal species, and finally, in the case of primates, the beginnings of the intellectual layer of being also appeared, which, built on the former three layers, then became particularly pronounced and decisive in humans. In this superimposition, Hartmann distinguishes between the construction in which the characteristic elements of the lower layer of being are used for the upper layer in a transformed way, as is the case with the biological layer of being, and the mode that merely is built on the lower layer, but in doing so its elements are not used, and in this way a more autonomous detachment from the lower layer is made possible. This later is shown by the psychic layer of being built on the top of the biological, and also by the spiritual layer of being built on the top of the psychic. An important correlation in the connection of the four layers of being is that while the upper layer of being always has a higher degree of evolution, the lower ones are ontologically stronger than the higher ones. This is because the higher ones cannot transform the laws of the lower layers, and without violating them, they can unfold the laws of their own higher layers. Thus, the biological and higher layers of being cannot violate the underlying physical laws, just as animals with a psychic layer of being and human beings can survive only with the functioning of biological laws, and finally, social institutions created with the spiritual layer of man can survive only because of the laws of the lower layers of being without violating them.

In addition to these four layers of being, which Hartmann calls the layers of *real existence*, he distinguishes *ideal existence*, which only plays a role in the case of the uppermost spiritual layer of being. This ideal existence characterises only those spiritual products that exist only on paper or in another similar fixed form, but are not used in any human community. To name this difference, Hartmann uses the “*living*” *spirit* on the one hand for the intellectual products and cognitive actions that function as part of the layer of real existence, and on the other hand the “*dead*” *spiritual existence* of the objectified intellectual products that have already detached themselves from real existence and exist only in paper. The latter are still part of the ideal existence but not of the real existence, but in the case of fixation and the possibility of returning to them, the ideal existence of objectified intellectual products can always become real existence if a society uses them for the functioning of its intellectual practices and its institutions.

When approaching the four layers of being, the psychic layer of being causes a problem of understanding for the researcher. To understand this, one must know that Hartmann separates the personal spirit within the “*living*” spiritual layer of being, which is the highest layer of the four layers of the single individual. The real objective spirit in human societies is represented by the activities of millions of personal spirit and the resulting intellectual institutions (moral norms, laws, language systems, etc.) that feed into their social functioning. (If an intellect product does not, it remains as a dead spirit in paper fixed form of ideal existence). Thus, the psychic layer of being attached to each person is always in the neighbourhood of the personal spirit and flows together in human consciousness. Only a part of the psychic layer of being appears in consciousness, and the mental unconscious processes supplement this. But where is the boundary and what gives its own terrain to the psychic layer of being? If we look for an answer to this question, we find confusion in Hartmann.

In his book on the problems of the spiritual layer of being from 1933, he goes into more detail and writes here that the psychic layer of being is a thin strip between the influences of the adjacent lower biological and upper spiritual layers of being: “Das Bewußtsein ist gleichsam nur ein schmaler Streifen zwischen ihm und dem unbewußt Seelischen. Beide greifen mannigfach bestimmend in seiner Ebene ein, aber in entgegengesetztem Sinne.” (Hartmann 1962: 50-51.) The psychic subconscious sector extends into the biological instinct world, from the other direction, however, the processes of the

personal spirit reach down into the psychic layer of being and both opposing influences together give the functioning of the psychic-spiritual layer of being. From this description, however, the very nature of the thin specific strip goes unexplained, and the psychic layer of being is actually divided in the lower and upper directions between two adjacent layers of being. If this is the case, however, the question arises why it is not enough to speak only of the instincts of the biological layer of being from below and the conscious processes of the personal mind of the mental layer of being from above? But not only does the specificity of the “thin strip” remain obscure here, but emotions are also missing from Hartmann’s description of the psychic layer of being. If we look for alternative descriptions, we can mention one of Hartmann’s discussion partners, *Max Scheler*, in whose description the emotions appear, and *Arnold Gehlen*, as their common discussion partner, who argues fiercely with Scheler about this very issue.

In particular, in his 1940 summary, Hartmann points out that from the two upper layers of being of the psychic and the spiritual, the products of the lower biological layer of being are not transformed, but without any transformation the upper layers only build on them, i.e. the psychic layer directly on the biological and the spiritual on the psychic layer. This is in contrast with what Hartmann wrote a few years earlier, and he reconciles the two opposing assertions, by emphasising that these four layers of being, in addition to their independence, are still united in individual human beings: “Das seelische Sein enthält die organische Prozesse nicht, wohl aber enthält “der Mensch” sie in sich; denn der Mensch ist selbst ein geschichtetes Wesen, er ist auch Organismus, und folglich auch ein materiell-körperhaftes Gebilde. Insofern hat er die niederen Kategorien alle als konstituierende Momente an sich.” (Hartmann 1940: 496.) However, these analyses do not replace the fact that the “thin strip” of the psychic layer of being is not filled with content. In the 1940 volume we find another clue that could bring us closer to the search for the terrain of the psychic layer of being. For he writes, in connection with the independence of the two upper layers of being from the material of the lower layers, that just as the spatiality and inert substance of the lower layers do not appear in the psychic layer of being, neither do the act-characters of the psychic appear in the objective spirit of the intellectual layer of being.¹⁶ Thus, the psychic layer of being is the realm of human action in contrast with the terrain of the objective spirit of the intellectual layer of being, where human actions do not appear. Hartmann thus distinguishes only the objective spirit from the soul and does not mention the personal spirit, which is directly adjacent to the psychic layer of being in the individual. Therefore, we cannot obtain the sought-after material of the psychic layer of being from this analysis either.

To compensate for this, if we do not want to completely discard the psychic layer of being and describe the functioning of the human being and the human community in three layers of being instead of four, we can only use Luhmann’s early analysis of the role of biological sexuality in the emotion of love as a solution. In general, it can be stated that the instincts of the biological layer of being appear in tamed form and absorbed through the transformation of emotions in the psychological layer of being. On the other hand, the emotions also receive a transformation (an attenuation and enrichment at the same time) through intellectual products of the spiritual layer of being. As in the case of the sexual drive, the savagery of mere biological sexuality is transformed by the psychic emotions of the love relationship, but already with the intellectual overtones and experiences of the great love descriptions, novels and wonderful melodies of the love songs of all times by which human beings have been socialised. In actual love relationships and communications, the addition of these three layers of being is inseparable and built upon each other, and in the case of the eventual low presence of one of them, it may cause the love relationship to survive only to a

¹⁶ “Räumlichkeit und träge Substanz kehren oberhalb des Organischen nicht wieder, die Aktcharaktere des Seelischen nicht im objektiven Geiste.” (Hartmann 1940: 512.)

limited extent. In order to understand this structure, it must be accepted in the case of instincts that the instinct world, which in the animal world is limited to certain behaviours and coordination of movement, appears in the case of humans on the one hand reduced but on the other hand in the form of general motivations (sexual instinct, life instinct, care instinct, etc.) and they have many functions in this form. Arnold Gehlen criticised Konrad Lorenz for having assumed instincts in some form in the case of human beings, but in this form Lorenz's truth must be accepted.¹⁷ It is also worth mentioning that Max Scheler also emphasises the role of instincts in the form of instinctive strivings in humans and he writes their function as a mediating layer between biological life functions and consciousness. (See Scheler 2016: 30-3.)¹⁸

Another problem and obstacle to the possible transformation of Hartmann's solutions is the strength of the evolutionary lower layer of being, which is sacrosanct to the upper layer. In the eighty years since he wrote his 1940 synthesis, however, the spiritual layer of being has achieved such a strong transformative capacity, especially in relation to the lowest physical layer but also to the biological layer, that this thesis of Hartmann's can be questioned. The strength of the lower is based on the fact that its existence is independent of the influence of the upper. The upper layer can only use the elements of the lower layer as building blocks and produce a more complex formation of the upper layer, such as the formation of biological organisms from the material of the physical layer takes place, but this transformed use does not affect the lower part: "Die höheren Kategorien setzen stets eine Reihe niederer voraus, sind aber ihrerseits in diesen nicht vorausgesetzt. Kategoriale Abhängigkeit also waltet durchgehend von den niederen zu den höheren, nicht aber umgekehrt." (Hartmann 1940: 519-520.) In contrast with this assertion, however, we can say that in the case of the highest spiritual layer of being, such progress has been made in the meantime that the elementary level of physical matter can be transformed with ever deeper resolution and on the nanoscale, and it can create hitherto non-existent material (e.g., most recently *graphene*), and entirely new entities can be constructed from hitherto non-existent formations. Thus, the uppermost spiritual layer of being not only becomes evolutionarily higher, but it can increasingly transform certain elements of the physical layer of being without limits, i.e., the lower one loses its ontological strength.

In the same way, the biological layer of being is increasingly dominated by the spiritual layer of being and this upper layer does not simply build on it, for after the discovery of DNA in 1951, whose existence Hartmann had only guessed at theoretically, the genetic editor can now access it more precisely.¹⁹ In this way, the resulting living individual becomes designable through the spiritual layer of being. But the solutions known today only in research laboratories for the creation of the artificial artery or in the same way for the renewal of the main organs of the human body by implanted stem cells and their radical prolongation of life, lead to the fact that the laws of the biological layer of being can be overridden by the spiritual layer of being. Mention should also be made of experiments with the emulation of the human brain and mind, which, if successful, may lead to the intellectual being layer of individual humans being transferred to a computer carrier in the future in a fully functional way without

¹⁷ "Dennoch sind die Versuche, das Konzept des Instinkts direkt auf den Menschen zu übertragen, äußerst dünn und enttäuschend [...], insbesondere vertreten durch Konrad Lorenz. [...] Es gibt keine allmähliche Beziehung zwischen intelligentem und instinktivem Verhalten, aber [...] es gibt eine Tendenz zum gegenseitigen Ausschluss." (Gehlen 1974: 32.)

¹⁸ "Wir können den Fehler sehen und das ist der grundlegender Fehler von Descartes, der das Ignorieren des tierischen und menschlichen Instinktsystems bedeutet, obwohl sie die Vermittlung und Einheit zwischen Bewegungen des realen Lebens und den Inhalten des Bewusstseins sind." (Scheler 2016: 92.)

¹⁹ "Man weiß hier sehr wohl, dass der morphologische Bau des Organismus seine bestimmten Gesetze hat [...] man weiß auch, dass die letzteren Funktionsgeformtheiten sind, in denen die Gewähr liegt, dass die gleiche Formung sich wiederbildet und somit das Leben sich erhält." (Hartmann 1940: 456.)

a biological being layer.

In view of this, it can be said that Hartmann did not sufficiently emphasise the progress already observed in his old age with regard to the ability of the intellectual layer of being to penetrate the lower layer of being, and therefore thesis of the invulnerability of the lower from the higher layer in relation to all layers of life proved to be erroneous in the case of the highest spiritual layer. This correction can also be important because we can theoretically assume that through the future development of artificial intelligence and its possible detachment from human consciousness, this decreasing dependence of the upper stratum on the lower strata would only be completed. In this way, it could be that the artificial intelligence, which has become autonomous, could exist as a new layer of being only on the basis of the lowest physical layer of being without a biological layer.²⁰ Without this decreasing dependence, this possibility could not exist, but with the acceptance of this development, it could only be conceived as the completion of the already decreasing dependence.

II. Artificial Intelligence: The Emergence of a New Layer of Being? (AI in the mirror of Nicolai Hartmann's ontology)

The increasing strength of the artificial intelligence (AI) has already frightened its guiding practitioners (Elon Musk) and theorists (Stephen Hawking, Nick Bostrom) in recent years, and it is portrayed as a dangerous development which is growing over human beings and freed from human control. A further description, taken from a remark by John of Neumann, deals with the inherent novelty of the AI as the emergence of the era of singularity.²¹ According to the latter, the increasing computing power and faster program due to the self-learning artificial intelligence lead to a point where the limitation of acceleration by the human being will be eliminated from the processes of the AI. From this moment, which is unique in world history, the self-learning AI will grow to a thousandfold speed, and within a few hours it will become completely incomprehensible also for IT professionals. From this development – along with the growing robotics – the creation of all things will be made possible for the AI, and in this way not only the artificial intelligence becomes incomprehensible to humans, but they are removed from the world's guidance. And that is the occurrence of the era of singularity.

If we bracket these concepts and fears of artificial intelligence in parentheses for a while, and instead, we focus on the philosophical concepts that analysed the evolutionary leaps of existence on Earth so far, then we can look at the better-founded categories of ontology for the understanding of the novelty of the AI. In the last 100 years, the ontological analyses have already been based on empirical basis, and the novelties of the artificial intelligence can be better understood on the basis of these analyses. I consider the analyses by Nicolai Hartmann about the ontological layers of being and about the evolution of these layers as appropriate in order to be able to compare the present evolutionary leap through the AI with the previous leaps. It was written by Hartmann how once the biological layer of being

²⁰ For details on these questions, see Pokol, Béla: *Künstliche Intelligenz: Die Entstehung einer neuen Seinsschicht? (KI – im Spiegel von Nicolai Hartmanns Ontologie*. In Pázmány Law Working Papers, Nr. 2018/12.)

²¹ The fact that Neumann is the first creator of the idea of singularity is known only indirectly from Stam Ulam. He remembered a conversation with him from the early 1950s, when this idea was mentioned by Neumann: "The ever-accelerating progress of technology and changes in the mode of human live give the appearance of approaching some essential singularity in the history of the race beyond which human affairs as we know them, could not continue". This is the first known use of the word "singularity" in the context of human technological history." (Quoted from Kurzweil 2012: 185).

could stabilize over the physical layer of being, and during the development of plants, and then the animals of higher rungs in the state of development of the mammals, an emotional-psychic layer of being emerged over the physical and biological layers of being and through the gradual evolution of primates, the germs of the mental layer of being appeared, which became particularly dominant in the human being and communities, and the lower layers of being were determined more and more by the mental layer of being. Now, we may be faced with the emergence of AI before a more recent evolutionary leap, and over the mental layer of being of human it begins to develop a new layer of being as an heir of the earlier human mental layer which as self-organizing artificial intelligence from now on the highest level of being in the world will be the dominant force.

Nicolai Hartmann has already asserted certain connections between the layers of being, and for the coexistence of the increasingly recent layer of being with the lower and older layers, he has stated the laws. In this way, it is worthwhile briefly summarizing his analyses before examining the layer of self-organizing artificial intelligence.

II. 1. The Man and the Hierarchy of Layers of Being

In human beings, the peculiar “human” is contained in the mental layer of being, and its gradual dominance over the physical, biological, and psychic layers of being signifies the evolution of human life, but man is always determined by the laws of the four layers of being at the same time. Man is a multi-layered being, and the human communities can unfold only in the cumulative framework of the laws of the four layers of being. The upper strata of being can only develop when the laws of the lower ones are respected, but this is no hindrance to the autonomy of the laws of the upper layer of being in relation to the laws of the lower layers of being. The construction of the higher layer of being means the transformation of the categories of the lower layer of being, but the higher layers no longer signify such transformation, but are built with their own categories over the lower layers of being. While the essential elements of the physical world are used by the biological layer of being – transformed only by the laws of their own being – there are no material elements of the lower layers in the psychic and the mental being-layers.²² Hartmann described these connections as follows: “Um Mehrschichtigkeit zu begreifen, genügt es, sich an allgemein Bekanntes zu halten. Niemand zweifelt, daß organisches Leben sich vom Physisch-Materiellen wesentlich unterscheidet. Aber es besteht nicht unabhängig vom diesem: es enthält es in sich, beruht auf ihm, ja die Gesetze des Physischen erstrecken sich tief in den Organismus hinein. Was nicht hindert, daß dieser über sie hinaus noch seine Eigengesetzlichkeit habe, die in jenen nicht aufgeht. Solche Eigengesetzlichkeit überformt dann die niedere, allgemein physische Gesetzlichkeit. Ähnlich ist es mit dem Verhältnis des seelischen Seins zum organischen Leben. Das Seelische ist, wie die Bewußtseinphänomene beweisen, dem Organischen durchaus unähnlich, es bildet offenbar über ihm eine eigene Seinsschicht. Aber es besteht überall, wo wir ihm begegnen, in Abhängigkeit vom ihm, als getragenes Sein. [...] Das seelische Sein ist also zwar getragenes Sein, aber in seiner Eigenart ist es bei aller Abhängigkeit autonom. Schließlich ist es seit der Überwindung des Psychologismus eine wohlbekannte Tatsache, daß das Reich des geistigen Seins in dem des seelischen und seiner Gesetzlichkeit nicht aufgeht. Weder die logische Gesetzlichkeit noch das Eigentümliche von

²² It can be mentioned in the footnote that, in the meantime, brain research has uncovered – based on Donald O. Hebbi’s initiatives in 1949 – that among the hundreds of millions of neuron cells in the brain, the individual groups of these neurons are always differentiated from ever new experiences and knowledge New organization. Thus, the mental processes also have a material basis in the brain. For an analysis about this, see the brief book of the chapter Neokortex. (Kurzweil 2012: 85-95.)

Erkenntnis und Wissen hat sich psychologisch ausschöpfen lassen. Noch viel weniger die Sphäre des Wollens und Handelns, der Wertung, des Rechts, des Ethos, der Religion, der Kunst. Diese Gebiete alle ragen, schon rein dem Phänomengehalt nach, weit hinaus über das Reich des psychischen Phänomene. Sie bilden als geistiges Leben eine Seinsschicht eigener und höherer Art, mit deren Reichtum und Mannigfaltigkeit sich die niederen nicht entfernt messen können. Aber auch hier waltet sich das gleiche Verhältnis zum niederen Sein. Der Geist schwebt sich nicht in der Luft, wir kennen ihn nur als getragenes Geistesleben – getragen vom seelischen Sein, nicht anders als diese vom Organismus und weiter vom Materiellen getragen ist. Auch hier also, und zwar hier erst recht, handelt es sich um Autonomie der höheren Schicht gegenüber der niederen, gerade in der Abhängigkeit von ihr.” (Hartmann 1962: 16-17.)

Thus, man is the unity of four layers of being, and human reason can only exert an effect upon the lower layers of being through the biological basis of the human body. Particularly, the purely intellectual activity is the terrain of the mental layer of being, and Hartmann distinguishes three internal areas of this layer: the domain of the individual spirit, that of the objective mind and that of the objectified mind. The first two are the living spirit, and the objectified mind is the terrain of the dead spirit, but the content of the objectivized mind can always be traced, and so these contents can be brought back into the living spirit. The individual mind lives together with the content of the objective spirit of its time, and more or less a plurality of individual minds carries the objective mind and its inner forms as the Volksgeist and other collective spiritual forms of the epoch. But also the individual mind has largely such contents as the objective mind, and thus its relation as reciprocal carriers can be described. The third form – that of the objectified mind – always increases with the enrichment of the fixation of the mental contents on the basis of writing and other fixation forms. In this way, the individual minds can, in addition to the content of the objective mind of their epoch, additionally use the intellectual contents of all the era, and as a reaction can thus enrich the contents and forms of the objective spirit of the epoch. At the level of the mental layer of being, living collectivity thus arises, while at the biological level the framework of common existence is borne only by the community of the race over its ever-vanishing individuals, and so the psychic life is isolated only in the individual and it is not transferable. As Hartmann puts it, “Sein seelisches Sein hat jeder für sich. Es ist esoterisches Sein des Individuums, unübertragbar, mit dem Man wohl Fühlung haben, in das man aber nicht hineingelangen kann. Man kann wohl mit ihm mitleiden und sich mitfreuen: aber es ist und bleibt ein zweites Leiden und ein zweites Sichfreuen neben dem original, und es bleibt auch bei aller Innigkeit ein qualitativ von ihm verschiedenes. Den Gedanken aber, den einer hat, kann man als denselben denken, wenn man ihn erfaßt; es ist zwar ein zweiter Gedankenakt, Akt eines anderen Bewußtseins, aber es ist derselbe Gedanke” (Hartmann 1962: 71.)

Hartmann makes another distinction within the mental layer of being that appears in the boundary between the objective (living) spirit and the objectified (dead) spirit. According to this distinction, the fixed mental contents of the past – convictions, patterns of behaviour, moral and cultural values, etc. – can appear in the present as massly accepted and followed cultural patterns. But it means a different way of injecting the past when it is only defined as a purely objective intellectual content for the comprehension of individuals, but no longer appears on the level of massively committed belief, knowledge and prejudice. Then only the individual spirit can deliberately fall back on these dead objectivated spiritual contents, and only he can bring them into the living, objective spirit: “Das ist das Inkraftsein oder Noch-Lebendigkeit (einer Sitte, Anschauung), also die Kraft der “Sache”, den fortlebenden Geist mit einer gewissen Stetigkeit bei sich festzuhalten, auch da, wo er sich sonst zusehends verändert [...] Beim vernehmlichen Hineinragen ist es überall anders, wo die Sache selbst

nicht mehr fortlebt, die unmittelbare Tradition abgerissen ist.” (Hartmann 1962: 38.) Let us take a look at how the relationship between these three areas of the mental layer of being has changed – already beyond Hartmann's time in the 1930s – and how the forms of today's artificial intelligence rippled into the old contents.

II. 2. The increasing interweaving of the artificial intelligence into the spiritual layer of being

The dominant role of the spiritual being-layer in the human communities and the relative suppression of the determining power of the lower layers of being began with the possibility of fixing the sense by some form of writing. Of course, this was at first a thin framework for the life of the human communities in the civilizations by which these levels were reached, and the broad masses and their daily lives were not touched by them. Even the invention of printing in the middle of the 1440s in European civilization could not alter this state, but for these upper social classes, it began the enhancing of importance of literacy through this technical facilitation and it also began in everyday life that the experience and action was based more and more often on the fixed, meaningful content. Throughout the 1800s, this phenomenon has spread all over Europe and other continents with European culture and has gradually been extended to the whole human society. Because of the universal literacy, the fixed, meaningful content of the daily press, journals and so on was interwoven in everyday life in the early 1900s, and then this extension was further increased by films and radio. In the 1950s the general spread of television came and every minute of everyday life was increasingly influenced by the written, audio-moving forms of fixed spiritual content. Through this change it has begun that the spiritual layer of being became more and more dominant over the lower layers of being, and the lower ones could only exercise their impulses partly and overformed by the spiritual layer of being, and this was understood as the process of civilization.²³

This development, however, received a real boost later, and since the 1980s it has begun to penetrate all areas of life through the massive spread of personal computers. In this way, the fixation of the spiritual content could remain in the state of constant correction. But the spiritual layer of being was not only made liquid but the individual liquid spirit became through the text editors and their simple conversion possibilities to the universally shared spiritual contents. The inherent potential in this development was then realized by Internet proliferation in the 1990s. Since then, it is possible that what someone describes, thinks and publishes in the Internet will be reachable in hundreds of thousands in a few minutes and it will influence the experience and actions of thousands.

Kevin Kelly described this process through twelve technological developments as follows. (Kelly 2016). At the centre of these developments, the emergence of the fluidity of the sense fixation can be seen (it's the *flowing*) as a result of the computerized digitization of the writing. In this way it is made possible for the intellectual thought-creating persons to use the liquid digital fixation instead of the former rigid fixation of the thoughts and other meaningful contents, and so one can always think about these again and correct these contents. Then, the parts of the fixed contents are possible to mix with other fixed contents so that new ideas, musical works, etc. can be created. For the people in the spiritual subsystems, this development made the easy ascent from the rigid fixation of the spirit possible and to ascend to the state of constant spiritual hovering. The scientists, the artists, the theorist, etc. can only temporarily fix their thoughts and other spiritual results due to the digital sense fixation, which are always available for rethinking, correcting, etc. The computerized

²³ Nobeit Elias illustrates the course of this civilization in the changed forms of fulfilling natural needs on the basis of a rich empirical basis, see Elias 1976.

digitization is the basis for the other technological developments analysed by Kelly and thus the accessibility of the contents of the liquid fixing of the individual will be made to the comprehensive human communities. Thus, this sense fixation became a jointly divided liquid spirit through some excellent word processing programs and their easy conversion possibilities at the end of the 1980s (by *Word* and some other text editors). Through the emergence of Internet this fluid and easily communicated sense fixation made then possible in the 1990s to transform the whole spiritual world. All contents have been in the state of constant change and all are in a state of permanent *Becoming* and thus the conversion to the permanent modification of the functional subsystems spread all over. For example, the system of the repealable law was already achieved in the transformation from traditional to modern society and likewise the refutable scientific truth and the state power which can be replaced by elections have already emerged and now the transformation of almost all institutions and human affairs became constantly changing today. The liquid sense fixation then were extended to all forms of communication, and instead of the book the screen has been brought to the central place and the man of the book gradually became the man of the screen. The former central-filled TV screen slowly changes to the decentralized smart TV, where from the many millions of content selected by the individual viewer, and parallel, this development leads to the computer screen then to the screen of smartphones that have more and more functions and thus the common multifunctional screen of TV / computer / telephone / artificial intelligence is achieved. In this way, our entire environment is progressively under intellectual reflection, and instead of our previous passive placement on the physical-biological environment, it began to impregnate the environment with cognition and make more and more things become smart and that is the process of *Cognifying* as Kelly calls it. The following development is the back-monitoring of our smart things, the *Interacting*, and it began that our responses are watched by our smart things, and because of the information so gained, they complement our activities or they begin to steer these activities in different directions, etc.

Of these recent trends highlighted by Kevin Kelly, the focus of the three forms of being of the spiritual being layer are reorganized. The individual mind interweaves in the objective spirit of the epoch more strongly than in the past and it takes the spiritual contents of the objective mind not just during early socialization – largely for the whole life – but in daily contacts and thus forming each day. Or their own intellectual content can be made immediately accessible to the objective mind through the Internet and thus modified back to the objective spiritual content. Similarly, the objectivized intellectual contents are constantly and continuously available to all through the Internet. In this way, the contents of the objective mind and of the objectivized mind are not as different from each other as it was in the years of Hartmann in the 1930s, although their separation cannot be completely eliminated.

II. 3. The direct connection of the artificial intelligence with the physical layer of being

As clarified in the starting point, man is the ensemble of four layers of being, and behind all mental action, there present his physical, psychological and biological layers of being. With this mode of functioning in front of the eyes, the differences between man and the robot of artificial intelligence can be better clarified. Michio Kaku writes in his new book that Rodney Brooks told him in an interview that the robot is a machine, just like man is, and so we can one day build such living machines as we are. (Kaku 2014: 263.) But on the basis of the ontological structure of the world determined by Hartmann it could not have been said, not even if the more advanced and demanding programs are able to convert in addition to the intellectual operations also the emotions and the physiological sensibilities into algorithm in

order to determine the robots. Namely, the psychological emotions and the physiological feelings can only be imitated by the algorithm in the spiritual plane, but because there are no real psychological and bio-physiological mechanisms behind these imitations in the robots, these layers can only be imitated. The functioning robot with artificial intelligence can inevitably only work with a two-layered mode of being and no matter how complex and through programming is suitable for the psychological reactions or the physiological-biological movements, the robot can only be the ensemble of two being layers in the comparison of the four beings of layers of man. In his book, Kaku describes these enriched programming possibilities and because he accepts the previously criticized claim of Rodney Brooks, he raises the possibility of human rights for the robots and writes about the ethical requirements regarding the robots. (Kaku 2014: 250-252.)

In the case of robots, the spiritual layer of being is reproduced by the programming, and if this programming can be more and more complex, it becomes possible to involve people's lower layers of being in the program. Then the reactions of the psychological and possibly the biologically-physiological being layers are also programmed into the algorithm, and the enriched intellectual program can be connected directly to the physical-mechanical bodies. A further manifestation of these intimate connections, when, in the case of physically disabled or otherwise immovably damaged people, the brain waves are connected directly to the paralyzed parts of the body by circumventing the damaged part of the brain, and so the functions of the gait are imitated by a program and the paralyzed man is again enabled to move. But even without these – as in the case of Stephen Hawking – the mental responses of the brain waves of the paralytic can be linked to a wheelchair and he becomes able to move the wheelchair and he can move objects in the outside world through brainwaves. “Telekinesis: the material guided through the mind” – writes Kaku in the title of a chapter in his book, and this is a precise description of the reduced state of Hawking in the comparison of full human being with four layers of being. Namely, Hawking could communicate with the outside world only by directly connecting his spiritual layer of being with this world, and so he is, in a certain sense, in the reduced state of the two-layered being. Of course with a living brain, and so it is to be fed and because of its metabolism is always to be made nappy. But the technology thus created can later be connected with the sheer physical robot body, and this can contribute to the emergence of such a future existence, which can work in the world without the psychological and biological layers of being. The analysis of this self-organizing artificial intelligence and the chances of this development have already been made – among others – In the works of Ray Kurzweil and Nick Bostrom, so it is worthwhile to go ahead in the following analysis on the basis of their works. (Kurzweil 2005; Bostrom 2015.)

Before this, however, it would be worthwhile drawing some conclusions from our analysis to date regarding the relationship between the spiritual being-layer and the new layer of artificial intelligence that may arise over it. On the basis of this analysis, it appears that if by the artificial intelligence is enriched only the existing spiritual being-layer, and it uses this as additional faculties, then we cannot speak of the creation of a new layer of being. Including all the trends described by Kevin Kelly will then be no different than the strengthening of the spiritual being layer over the lower being layers. Even if this is only the beginning in terms of improvement through the artificial intelligence, and things in our environment are getting more and more smart additions over the next decades – as Kelly predicts – this will continue to be our previous fourth and highest level of being-layer. A really new layer of being can only emerge if the forms of the artificial intelligence, the algorithms and programmed shards can become somehow self-organizing and can work through the direct connection with mechanical bodies even without the human spirit in the world. There is another question as to whether this will be the unfolding of another new layer of being, as it has already taken place three times in the billions of years of the history of the earth, and the previous layers of being

have always been kept as an indispensable condition. Or else this evolution jump will be different compared to the older, and this gets a different course?

II. 4. The Self-Organizing Artificial Intelligence

Starting from the works quoted of Ray Kurzweil and Nich Bostrom, two possibilities for the unfolding of the self-organizing artificial intelligence can be isolated. One is the strong artificial intelligence, which can emerge as a succession of today's weak version, and the second is the emulation of the human brain, which can autonomously exert itself as a digital copy of the spiritual contents of a human brain, separated from the limits of human being. The third is the possibility of artificially improved human intelligence, which can create a kind of super-intelligence, although it can only be a different form of today's coexistence between the human organs and the additional artificial intelligence that does not separate from man. And, therefore, would not signify an evolutionary leap, but would only intensify the dominance of the highest layer of being over the lower ones. (And so its analysis could also be at the end of the previous section.)

II. 4. 1. The strong artificial intelligence

The strong form of the artificial intelligence shows such degree of artificial intelligence when it has reached the level of the human mind, and then it surpasses this degree very quickly a thousand times in opposition to the form of today's weak version. A preliminary question is whether it is at all possible, and how strong artificial intelligence can really be achieved? By analysing the previous exponentially rapid growth of performance, it is allowed to answer this question quickly. Yes, this is possible and the only question is whether this will be around 2040 or 2100. Against this background, two important questions arise: (1) whether the strong form of artificial intelligence will be liberated from the institutional surveillance and control of human beings and human society? 2) and the second question is what character this out-of-control artificial intelligence will have and whether it will have an autonomous self-awareness and all-pervading will that will use its enormous capacity for the change of the world independent of human wills? Or, on the other hand, this all-pervasive will cannot arise in parallel with the enormous technological capacity and can only be described as the mental outlook of a naive little child – but with a combination of enormous technological capacity.

As far as the first question is concerned, namely, the possibility of breaking out AI from human control, the consequences of the recursive self-learning ability and self-alterability of the genetic algorithms, which can be seen as the main direction of the development of the artificial intelligence, are to be interpreted. In this way, human control can only affect the determination of the initial parameters, but then, on the one hand, such solutions for the realization of these parameters can be brought about by the AI, which can be torn from the human control, on the other hand the input parameters also will be included into the self-changeability, and after some cycles of recursive self-alteration, any previous determination could be cancelled. This self-versatility of AI is already today so great that the change of its hardware can be created. However, these already existing capacities do not pose a serious threat because, due to today's low level of AI, their control can still be created by human intelligence. But in the future, so fast recursive self-learning and self-altering cycles can be achieved – as compared to today – that perhaps a thousand times faster for the self-learning cycle, which can drop during hours, minutes and seconds up to a hundred times a

day, to make fundamental changes to create. This can no longer be kept under human control, so the release of AI from human control after a single point simply follows from the current trend.

The next question refers to the nature of this self-organizing AI. It is worthwhile to make a distinction, and within the AI, the technological intelligence is to be separated from the general intelligentsia for the evaluation of social conditions. The technological intelligence is the ability for the targeted change of the biological and physical world and one aspect of this ability refers to the extent to which this change can be prevented by other forces (e.g., humans). This is, therefore, the ability for the domination of the lower layers of being of the world, and this is becoming more and more strong in the field of artificial intelligence, while its ability to look at global reality, including the reality of human society, has lagged far behind. Nick Bostrom has already extensively analysed the importance of incorporating the social knowledge and values into the AI programs and on this basis two problems can be pointed out. One problem is that there is no uniform goal-parameter system for the basic structure of human society and its survival, and it always depends on the selection of the individual dominant human elite groups, which kind of value hierarchy will apply. But it is the minor problem. The greater problem arises from the fact that this arbitrary and possibly incomplete hierarchy of values can also be the victim of the self-alteration cycles of AI. If the AI itself can continually change the determining component of its programming, there is no guarantee that the built-in social values will remain in the technological parameter of the AI and will not be annulled by the AI after a few cycles of the its self-change.

So, if we want to determine the nature of our distinction more precisely, then it can be expressed as the opposition of the technological intelligence to the social intelligence. While the artificial intelligence grows enormously in the technological dimension, it remains at the level of a social knowledge on the level of a stupid little child. And if the special algorithms are built into the program for the improvement of both types of intelligence and the treatment of the social values, there is no guarantee that after a short time the self-learning AI will not delete these parts of the program. Consequently, in my opinion, the often repeated description of “evil artificial intelligence” is inappropriate for the signalling of the real dangers, but we must be afraid of the social blind and dumb yet powerful technological intelligence. This artificial intelligence will not destroy the existence of human society because of its evil, if it is made possible by its enormous technological capacity, but because of its low niveau in relation to social knowledge. It should be noted, however, that this situation may change in the coming years because of the recent developments in the area of emulation of the human brain (*mind uploading*), the future of the strong artificial intelligence can also be touched (see next section).

Emphasising these connections to the dangers of artificial intelligence, the AI researchers try to incorporate, at least in the initial parameters of the AI program, operational principles that can prevent the dangerous changes for human society. In a newly published study by Joel and Ben Goertz Pitt, attempts are made to find such program elements for the planning of the artificial intelligence, which can be used to secure the constant positive tendency towards the maintenance of human society. They assume that this goal cannot be fully guaranteed, but it can be determined that at least the self-changes due to machine learning always receive a positive feedback in this direction: “Our current perspective is that provably, or otherwise guarantee-ably, Friendly AI is not achievable. On the face of it, achieving strong certainty about the future behaviours of being massively more generally intelligent than ourselves seems implausible. Again, we are aiming at a more modest goal – to explore ways of biasing to odds, and creating AI systems that are significantly more likely than not to be Friendly.” (Goertz/Pitt 2014:65.) In order to secure the “friendly” approach direction of machine intelligence in relation to human societies they emphasise as the first

imperative to keep the recursive cycles of change the slowest in the first period so that they can still be tailored to the understanding of the human mind. Furthermore, it is still important that human participation should be secured in determining the program changes in the first cycles. Likewise, the authors consider it important to integrate well-developed ethical principles with rich case scenarios into the program, which must be executed in the first phases through thousands of test situations. “Where, after an AGI (Artificial General Intelligence) has learned some of the everyday aspects of justice, including the balance of justice with empathy in everyday life, and once it has also gotten familiar with the application of abstract ethical principles to other aspects of ordinary life, it will be well poised to appreciate ethical principles and their utility in making difficult decision. It will be able to understand the abstract nature of justice in a richer and more holistic way.” (Goertz/Pitt 2014: 72.) In the end, however, we can only hope that by the fully automated and capable of great changes strong artificial intelligence, the conditions for the functioning of human society will be not eliminated.

After this, it may be asked if the dealings with the strong artificial intelligence can, as a matter of course, lead to a threat to the whole of mankind, in addition to their good blessings and benefits, why not stop this whole activity. For this dilemma, excellent analysis can be found in the book by Nick Bostrom and Ray Kurzweil also dealt with it. In particular, the researches in the field of artificial intelligence have also been characterized for decades by the military rivalry of the great powers and huge amounts of money and research capacity are spent on it. So, in this area, the latest successes are always kept as a secret. Basically, now the same happens in this area, what could be seen in the late 1940s in relation to nuclear research and which great power can achieve an advantage in this field that will be able to dominate the world. In this way, renouncing research in the field of dangerous strong artificial intelligence would only mean a competitive advantage among rivals. A related problem is that even if after some time the rival great powers begin to make some efforts for joint control because of the growing danger, even then it would not be enough because it is impossible to see in the development of artificial intelligence such degree and level where the pivot from today’s weak artificial intelligence will be arrived and ultimately, its strong and uncontrollable successor begins. Both in the analysis of Kurzweil and Bostrom, the age of the strong artificial intelligence, compared with the era of nuclear power, brings thousands of greater blessings and more comprehensive changes to the current state of human society, and the growing development in robotics already shows these changes in everyday life. But the threats are a thousand times greater than the atomic period has been alleged with regard to the earlier martial rivalries. However, it is also impossible to stop this as it was not possible to stop the race for the atom. The current age of the race for the strong artificial intelligence differs from the earlier rivalries only by the fact that the competition in the area of the nuclear weapons could finally be pacified and a joint control over the nuclear weapons could more or less stabilize by the mutual deterrence. However, in the case of the creation of the strong AI, its manufacturer will also be pushed in the background in a short time and also it will be threatened together with all humanity. In this way, there will be nobody and nothing able to conquer the new dominant force over the world.

II. 4. 2. The digital existence of the emulated human brain

The possible emergence of strong artificial intelligence from its weak and already existing form has been discussed for several decades, in contrast with the uploading of the human mind to computer or in another denomination the emulation of the brain, which is only in the last decade in the centre of interest. This is also a machine intelligence, and it can be a

branch of artificial intelligence, and here other incentives come in addition to the earlier motives to speed up the research as quickly as possible. In fact, if the emulation of the whole human brain could be successful, and instead of the biological processes the computerized processes of the neural functioning of the brain could be worked, then this technological novelty would give a new perspective to the mind and personality. Namely, the people could get rid of the vanishing biologic body and exist in an eternal carrier forever. The film with the title of *Transcendence* in the lead role of Johnny Depp has made this possibility a few years ago, and intensive research is moving forward in this area. For example, in the last few years, the European Commission has spent € 1.5 billion on research, and the brain emulation of the simplest organisms and small mammals is already being done, and at the same time, the emulation of the human brain is being investigated.

The independent existence of emulated human brain is still less visible at the current level of technical conditions. (A message in the world press has indicated that the full emulation of the brain of a rat was to be expected in 2017.) In this way, an empirically sound statement cannot be made about the possibility of the autonomous existence of the emulation of the human brain or about the degree of identity of such brain emulation with the original brain. Without the existence of real emulations, one can only think about these questions at the level of philosophical meditation. Of course, this is not without utility, but, in any case, it can only have speculative nature, and this should be noted in the following.

In the case of emulating the human brain, it is still a question of whether the technical feasibility (storage capacity and speed of computer) necessary for the emulation of the many billions of brain cells and the trillions of their synapses can be created. Based on the analysis of the existing exponential growth rate in this area, however, it can be said that the necessary computerized capacity will be available for this task in about thirty years of development. Due to a new message, one second of the most detailed emulation of a human brain (i.e., at the neural level of brain) converted in the computerized format was repeated in the fastest computer of the world in 40 minutes. That is, today, a second of the brain emulation still takes 2400 seconds to run in the computer to work the replicated brain, and this can cause a discouraging effect. But if in the future the validity of the Moorchen law is still assumed, that is, the computing power is doubled in one and a half years (which for example could be achieved by the rapid progress in the quantum computers), then the 2400-fold acceleration can reach 15-16 years, and so the human brain processes will be able to run also in a real time on the computer. The main debate should be whether, after the successful emulation of all the contents of the human mind, the original consciousness of this mind can be repeated in the computer, and the original consciousness also always emerges during the course of the computer program of emulation?

The very intense reflection and debate on this subject has produced some wise distinctions in recent years. Because of some distortions, a distinction is to be made as to whether the individual mental processes can be run on the computer after the emulation of the brain, and that is another question as to whether, besides these computerized processes, a new consciousness will arise as a single inspector of the parallel computerized mental processes? Finally, a third question arises, if so, this emulated consciousness will become the consciousness of the original mind, or thereby a new identity will arise, which will only have so much in common with the original human spirit that they both share the same memories and experiences. In the latter case, the emulated mind can be understood as a digital twin brother, but as the identical twins have separate consciousness, so also the consciousness of the digital mind becomes autonomous after the emulation its own way and own identity.

For Kurzweil and Bostrom, the answer to this question is self-evident, because, in its opinion, all manifestations of the mind and all the psychological processes in the nerve cells of the brain are the results of the electrochemical processes, and, in this way, the consciousness and

self-consciousness cannot be otherwise. Therefore, if the detailed emulation was sufficiently accurate, then not only the individual neural processes of the brain (memories, experiences, etc.) but also the consciousness as the accumulation of these processes will appear in the running of the computerized program. But in regard to the question whether this consciousness will be a duplication of the original or a new creation, no analysis can be found in their books, and this question itself has only been raised recently in the discussion.

Reading many discussions and arguments about these questions, I am more inclined to accept that if enough exact and detailed emulation of the brain is copied to computer platforms, and all connections of trillions of neural processes could be digitally duplicated, then the consciousness will also be appeared as the centre of control of mental processes. This will, however, only be the spirit of a digital twin brother, but certainly it is not to say that now the same awareness could be present “in two places”. And certainly not the fact that the human being – liberating from his biologic body – could move through emulation into digital existence.

David Chalmers is also in favour of this position in a recently published study, while Massimo Pigliucci advocates the exclusivity of the consciousness associated with the biological body. Chalmers calls himself a functionalist and Pigliucci a biologist and the two positions are described as follows: “Here philosophers divide into multiple camps. Biological theorists of consciousness hold that consciousness is essentially biological and that no biological system can be conscious. Functionalist theorists of consciousness hold that what matters to consciousness is not biological makeup but causal structure and causal role, so that a non-biological system can be conscious as long as it is organized correctly.” (Chalmers 2014: 104.) This currently only philosophical debate is of relevance, because today – and according to the researchers, for a few years – there are only destructive techniques for the emulation of the brain, which is not a problem because of the use of animal experiments for this purpose (now apart from some animal rights groups). But it was already pointed out in the discussions that the possibility of emulation should be made available to the incurable patients at the end stage as an opportunity for survival, since, in their case, the destructive nature of the emulation would no longer be a problem. It is therefore important to emphasise that by this way at most only one digital twin can be created, but the disappearance of the original person cannot be avoided.

It is also a great difficulty in the emulation process of the human brain due to the latest research results that, contrary to hope, it is insufficient if only the higher mental processes are emulated because in almost all brain processes the different parts of the brain participate simultaneously: “To summarize, it is misleading to refer to areas of brain as if they were modular. Instead, it is impossible to draw strict boundaries separating, for example, the more primitive sensorimotor functions from higher cognition such as planning, judgment, decision-making, and direction of attention. Furthermore, these areas are highly integrated functionally, and anatomically, in a complex dense network. Thus, we conclude that the aforementioned “quintessentially human” (which, employing the parlance of the first section, would fall under the symbolic as opposed to the physical) functions cannot exist independently of brain structures that are devoted to world interaction and body control. Even for an uploaded individual, a body (human-like or otherwise) would remain a necessity.” (Linssen/Lemmens 2016: 5.)

It should also be mentioned that when the successful emulation of the human brain, together with all mental processes, self-consciousness is also established (as was assumed to be probable), the future of machine intelligence can also be analysed with a new perspective. Namely, what cannot be formulated at the theoretical level about social reality, it is extensively embodied in the mind of the people who have been socialized for the existing communitarian solidarity and under the pressure of the biological and psychological instances

and their practical knowledge will be already inside in the emulated mind. In this way, this social knowledge, experiences and emotions are also conveyed into the digital spirit, and if my assumption, shared with Chalmers, is correct, digital self-awareness also arises in the spirit of this artificial intelligence, which works with adequate social knowledge. This level of human AI, which has freed itself from the biological barriers, can then immediately reach a thousandfold speed of development in the digital platform and the strong AI can also be created in this way. Also, as the second way for the achievement of the strong AI can be conceived, and this in turn removes the problem of artificial intelligence that it can only have half-minded intelligence: about huge technological intelligence and, in parallel, only about the social knowledge of a naive little child. Nick Bostrom already analysed this possibility in his book on superintelligence, and he also judged this possibility as positive, because in this way the developed moral instances are also drawn into the AI with the human mind. But he has correctly pointed out that the strong AI, based on genetic algorithms, can always remove this appropriate social knowledge from its operational program by its recursive self-changing capacity. (Bostrom 2014: 328-330). Thus, this path of the formation of the strong AI cannot be a definitive security.

II. 4. 3. The question of genetically enhanced superintelligence

Among the issues of intelligence enhancement and the extension of the life span through the human biotechnology it seems to me sensible to analyse three directions with regard to feasibility. From the other side I would most severely forbid cloning human beings and creating the mixture of man and beast as chimeras by the genetic modification of DNA. The three are as follows: (1) the enhancement of intelligence by embryo selection, (2) the renewal of the internal organs by the nanobots circulating in the bloodstream, (3) and finally the enhancement of the capacity of the biological brain by interfaces. It is to be pointed out that these questions have already attracted attention beyond the narrower academic groups of the researchers of the artificial intelligence in the wider circles of the philosophers also, and the first discussions have already taken place. In 1999, Peter Sloterdijk, on the basis of recent advances in human genetic engineering, raised the question of the improvement of human being in the future by human biotechnology, and Jürgen Habermas responded indignantly and without real reflection with the strongest moral condemnation. Of course, this emotional discussion about these questions occurs in the case of the Germans, without any content analysis and only on the level of sheer indignation, because of their specific historical inheritance. Namely, the earlier efforts in the field of human eugenics were made by the national socialist state in Germany. In this way, for the intellectual elite, the human genetic engineering was most deeply discredited by this inheritance, and the whole subject was tabooed. Francis Fukuyama also dealt with this subject in a whole book in 2003, and he also rejected the use of human genetic engineering, but he analysed these questions empirically against the sheer normative tone of Habermas. The reason for his refusal to oppose human genetic engineering is that he thinks the existing inequality among the people and social groups will be growing by this way, and he thinks this as unacceptable. In his opinion, the groups of genetic riches would be created in opposition to the genetic arms, and in the future, the wealthy testator will not only leave a great wealth and better living conditions for their offspring, but also the genetically enhanced body and mind. The inequality between the social groups would be increased more than ever before, Fukuyama means (Fukuyama 2002: 208-210.)

In contrast with Fukuyama's fear of increasing inequality, it is more convincing to me that the human biotechnological processes will become much cheaper after the initially high

costs after a while, because there is hardly any material and energy costs in this methods and they become a routine procedure. (Kurzweil 2005: 554.) In addition, if technology is not causing harm to others, and only lifting the level of its users from the level of other people, it is going to spread unstoppably among the elite, and so it is in vain to fight it. As a friend of equality, the analyst should rather be preferred to urge the government for the state's support in order to make this technological possibility accessible to the poorer social groups. Due to the numerous embryo selection (derived from the stem cells) – as Bostrom analysed – enormous intelligence growth could be achieved massively within society within a few years and that would be the highlight of institutionalized sensible society. (Bostrom 2014: 108.) For me, the problem is that since today it has become technically possible, so the society or societies that make a faster step on this path gain an invincible advantage over the others. The intellectual elite of a country should be more conducive to becoming more aware of this technological possibility among the groups of society, and by public debates to arouse state regulation, the use of embryo selection for intelligence growth massively in society support.

The supportive second direction in this area is the constant rejuvenation of the human organs by the nanobots in the bloodstream, and these efforts are only at the level of research and animal experiments compared to the former direction, but due to the analyses of Kurzweil and Bostrom over the exponential speed of growth in this area, there is little doubt that the radical breakthrough will occur in the coming years. But in contrast with the earlier analysis, greater scepticism must to be expressed in this area because the sheer longevity associated with the gradual stiffening of the whole personality already causes a great problem today. Even today, for the most 80-90-year-olds, the main problem is that their extended life has become meaningless, and it can be imagined how this problem will appear in the case of a 130-150 years life, even if the flexibility of the personality and habitus can be extended with a few years.

On the next subject of intelligence improvement, it must be emphasised that the methods for this direction have already been worked out in research and animal experiments. For example, in mouse experiments, the memory improvement with built-in hippocampus chips could work excellently. In addition, the experiments with regard to humans have also begun, first only with the purpose of curing Alzheimer's disease. The exponentially rapid development – together with the spread of the multiple embryo selections – can in the future in fact create an unprecedented intelligence increase in the whole society.

Overall, it can be said that the increase in intelligence by human biotechnology can be fundamentally welcomed – apart from the aforementioned two strict prohibitions – and for the achievement of increased artificial intelligence in society, this route should be supported primarily as a researcher has written in his article: “Thus, the best approach to life extension and consciousness expansion might lie in our own marvellously complex and entire bodies, meshed with and augmented by tiny bionan machines that become a part of us, rather than the opposite vision of human migration into a machine substrate. You might grow own eternal, artificial self as you gradually become bionic, in stages so tiny that you do not even notice.” (Goonan 2014: 198.)

II. 5. A new additional layer of being or the new beginning of earthly evolution on the basis of self-organizing artificial intelligence?

In the mirror of the earlier analysis, the two forms of self-organizing artificial intelligence are to be taken into account once again, i.e., the strong AI and the AI generated by emulation of the human brain. It was apparent from above that the digitized and independent layer of being of these two forms can be carried by computerized devices alone, but they can

also be connected directly to physical bodies perhaps only for a short time. With later one in mind, it would be made possible a being with two layers of being, i.e., with spiritual and physical layers of being but without the biological and psychological layers. Even if the mental impressions of these latter layers were introduced into the program in the case of the emulated species of the strong AI, this would no longer be a functional role – only disabling effects – so the elimination of these program parts is almost certain by the recursive self-alteration cycles. The self-consciousness digitized by the emulation, which was previously formed by the biological instinct and continually formed by the biological determinations of being, or since the early childhood by the familiar and other solidarity feelings, would get into vacuum in respect to its biological and psychological parts. The biological-physiological reaction-memories remain for some time – as man with amputated legs still have the itching in relation to his non-existent leg – and likewise his mental dispositions can affect something, but they no longer have a real functional role. Thus, the probability of the disappearance of these parts from the emulated self-consciousness in the world of artificial superintelligence reduced to two-layeredness is very high.

Thus, the answer to the question posed by the title of the last part of this work is that the artificial intelligence cannot bring about such an evolutionary leap in earthly evolution which has already emerged three times through the billions of years in the past. In fact, as long as the AI only increases the growing power of the intellectual being layer over the lower layers of being, no novelty emerges with regard to the state of being in the past thousands of years, and only the dominance of the intellectual layer of being increases much faster. If, however, the digitized intelligence thus emerges completely independent from the influence of human beings and it begins to function as a self-organizing force, the evolutionary end resulted in a new layer of being as a new evolutionary leap. But contrary to the former evolutionary leaps, the new evolutionary force will no longer function on the basis of the lower being layers. Based on Hartmann's analysis, it can be described as follows. The biological layer has been superimposed on the physico-mechanical layer of being during the billions of years, then the psychic-emotional layer of being appeared in the higher degrees of biological life, and then the beginnings of the spiritual layer of being on this basis appeared primarily with the primates but then especially with human beings, and by the coexistence of these four layers of being, the digitized artificial intelligence was finally created which, separated from man, can also be directly connected with the physical body and thus function in a self-organizing mode of being. This is actually a new evolutionary leap, but this new, self-organizing spiritual layer does not build itself over the past four layers of being as a new floor above the other, but it can directly connect itself with the physical-mechanical layer of being and exert effects. The coexistence of the earthly layers of being enables the birth of the artificial intelligence – and this is their only way to come into life – but, according to the possibility of their independent being, these layers of being will be superfluous to them. In this way, the new, self-organizing, artificial layer of being cannot be a new one over the previous, but only a new beginning of evolution on the physical basis, but now at the level of the already achieved self-organizing intelligence and with the leadership of this new evolutionary force.

The great debate and anxiety about the dangers of AI is therefore legitimate, as is repeatedly raised by Stephen Hawking, Nick Bostrom and Elon Musk, because the whole biological layer of being – and the mankind within –, could be superfluous for the artificial intelligence as the new layer of being. Accepting the legitimacy of these fears, it is only necessary to point out that this new artificial layer of being, brought about by human evolution, would not depend on existence on the earth, and would be able to function freely in a series of neighbouring planets, such as the possible colonization of the universe by the artificial intelligence has already been explained briefly. (Kurzweil 2005: 433-564.)

III. The layers of being and the questions of robot ethics

Human existence and the life of human communities are based on the cumulative regularities of the layers of being that are built upon each other through evolution, according to the theses of Nicolai Hartmann's ontology (Hartmann 1962). The accelerated development and increasing use of artificial intelligence (AI) in recent years in this structure directly affects the top layer of the four (physical, biological, spiritual and intellectual) layers of being, increasing its strength to the detriment of the lower ones. And with the later development of artificial intelligence, eventually breaking away from human control and gaining independence, it can be perceived as an evolutionarily created new layer of being. Unlike the three previous evolutionary leaps, however, it would not require all the lower layers of being. Taking into account the robots that are the physical incarnations of AI today, AI only needs the physical layer of being. (Pokol 2017). Against this theoretical backdrop, the analyses in this study seek to explore the emerging moral and related legal dilemmas within the mechanisms of contemporary societies that are increasingly permeated by artificial intelligence, while at the same time considering the extent to which the analytical framework changes when the multi-layered nature of human lives, and thus society, is constantly kept in mind.

III. 1. The preliminary questions of robot ethics

In his study of the ethical problems of the robot world, *Keith Abney* identifies three areas to group the problems: 1) the field of requirements and prohibitions for robot makers and programmers (such as medical ethics); 2) secondly, the field of requirements to be programmed into robots, first formulated by Asimov under the heading "Three Laws of Robotics"; and 3) finally, perspectively for the future, the question of the moral demands and "human rights" that robots might have at that time in possession of self-awareness emerges. (Abney 2011: 35.) A common dilemma for all three areas is the choice between the main starting points of moral theory already elaborated in the various moral philosophical schools of the comprehensive moral philosophical communities. One such school can be identified as the *deontological starting point* (the rule is the rule, and these must be followed), for which Kant's moral philosophy is best known, and the polarising opposition school, which considers consideration of the *consequences of action as the basis for moral decision-making*. Finally, thirdly, mention can be made of the *virtue ethics* school, which focuses not on the requirements to be considered in every situation in defining morality (like the two previous schools, albeit of the opposite direction), but on the enduring dispositions of the human personality, more simply, on socialised moral values. Here the person does not ask what the moral rule is in a situation, because in the increasingly complex modern world there are often no clear rules, but how a brave, just, faithful, true man decides. (Abney 2011: 37.)

Of the three schools, the school of deontology is only possible for robots that are used in the narrowest domain and follow the exact rules without being able to weigh the rules, because all situations can only be calculated and controlled in such a narrow domain, but even here unforeseen situations can arise and steer the robot decision in the wrong direction. For example, it could in principle be fed into the decision algorithm of a combat robot to "never kill a child!" However, in the case of child soldiers in African wars, this would mean a predetermined liquidation of the combat robot. (Abney 2011: 42.) In the case of general-purpose robots, the deontological approach is completely inapplicable. However, the consequentialist school of moral philosophy, which is also tied to the consideration of individual situations, also seems better only because of its life-like nature. Here the guiding

premise is to “increase, not decrease, the happiness of as many people as possible with the chosen decision!” and this is impracticable because it would require the processing of a huge amount of information, most of which could not be done in a timely manner even with the greatest capacity of computer data storage. Keith Abney’s position, therefore, is that with respect to the second area of robotic morality (i.e., moral decision premises programmed into the robotic algorithm), there is a mixture of deontology and virtue ethics that gives the best perspective, and a mixture of these can create the best built-in robotic moral version. According to this, the more abstract moral norms (moral virtues) form the decision framework, and the built-in goals and decision contexts always specify the determinants of the decision chosen by the robot in the given situations: “The hybrid approach of hypothetical rather than categorical imperatives (within a deliberately restricted, not universal, frame) coming from virtue ethics appear the best bet for near-term robotic morals (in sense two). [...] The emphasis on being able to perform excellently in a particular role, and the corresponding specification of the hypothetical imperatives of virtue ethics to the programming goals, restricted contexts, and learning capabilities of non-Kantian autonomous robots, makes virtue ethics a natural choice as the best approach to robot ethics.” (Abney 2011: 51.)

The connections between layers of being and morality are touched on indirectly by Abney, where he opposes emotivism, which identifies morality with moral emotions, and the cognitive perception of morality, which opposes it. It shows that if morality is tied to emotions because of the emotivists’ viewpoint, then primates with emotions cannot be excluded from morality either, which is absurd: “Such views, in addition to being unable to explain why nonhuman animals lack morality, also have struggled to explain the apparent cognitive meaningfulness of ethical claims and especially ethical disagreement. (They also naturally have severe difficulties accounting for the ethics of emotionless robots.” (Abney 2001: 46.) In contrast, he sees the position of evolutionary psychology, which emphasises the new decision-making mechanism of human evolution as an explanation for morality, meaning that humans have increasingly also developed a system of cognitive decision-making that reshapes current decisions in such a way that the instinctual-emotional first step of thinking is always followed by a second cognitive consideration, thus correcting the first: ²⁴ “Evolutionary psychologists claim there are not one but two of decision-making systems within most humans. The first is an instinctual, emotionally laden system that serves as the default for much human activity, particularly when stressed or under pressure. Many other animals share this non-cognitive decision-making system, in which (quite literally) we “know not what we do” – or quite why we do it. [...] But this “ghost in the machine” does not exhaust human agency; Libet and others found we also have a “veto” ability that can, after its subconscious initiation, still alter our action, in accord with a decision by a second, conscious cognitive system.” (Abney 2011: 46.) Abney recalls the mutually shaping effect of the two overlapping layers and almost recalls Nicolai Hartmann: “In humans, this deliberative system overlays the ancestral instinctual, emotional (and faster) decision-making system and so reason is quite often trumped by our instinctual drives.” (46. p.)

Having concluded that the upper (cognitive-rational) layer of the two-layered human decision-making mechanism is responsible for moral decision-making, Abney poses the question whether moral decision-making is in principle possible without a reconfigured, overridden lower layer. After all, the answer to this question also depends on whether a moral decision is possible for robots without an emotional layer. In this question, he then decides exactly the opposite way as Hartmann did earlier. It is very possible – he says – that a rational decision-making mechanism is sufficient for moral decision-making even without an

²⁴ It is worth pointing out that this view held by Hegel in legal philosophy long before the advent of scientific psychology in 1820. See Hegel, Georg Wilhelm Friedrich: *Grundlinie der Philosophie des Rechts.* (Werke Band 7.) Frankfurt am Main, Suhrkamp, 1979, 301. p.

emotional layer of being: “Hence, deliberative system capable of agency necessary for the existence of morality, and so for moral personhood. But is the ancestral emotional system needed as well? [...] In other words, – could (emotionless) robots be moral person? [...] The key to moral responsibility and personhood is the possession of moral agency, which requires the capacity for rational deliberation – but not capacity for functional emotional states, therefore, robots may well qualify.” (Abney 2011: 47.)

Based on Hartmann, there are two problems with these analyses. On the one hand, given the three layers of being above the physical layer of being, it can be seen as flawed that Abney combines biological stimuli with the determinants of the emotional layer. Already there is building on each other and transformation, and an instinct of the raw instinctual world is supplemented by the emotions of the higher layer of mental existence. For example, the ferocity of a biological sex drive is informed by feelings of togetherness, not to mention the intellectual-symbolic overrides that are still build on it, and the sex-altering aspects of the sublimated love relationships they produce.²⁵ That is, it is not a double but a triple decision-making mechanism that must be analytically separated in human decision-making, and in addition to the most basic instinctual reactions and determinants, their emotionally reshaped manifestations are still under more rational considerations at the intellectual level. However, a decision and the instinct that directly determines it, respectively its emotional transformation and its intellectual overwriting, are embedded in the interdependent laws of all three upper layers of being. Thus, human morality in societies of all human civilisation requires, in order to survive as a race, that men and women live together in some form of permanent community in order to have children and be educated. A larger community is necessary for the successful struggle and survival of the struggle with the forces of nature and other groups of people, and within these larger communities they must interact in more or less harmonious relationships to organise common activities. Moral virtues (norms and values) are therefore tailored to and maintained by the laws of the specific physical, biological, spiritual-emotional and intellectual layers of being of humans and their communities, and it is only because of the narrowing of moral theories in recent decades that conscious moral choices have become the focus of moral philosophy. Hegel in the early 1800s or Rudolf von Jhering in the 1870s and then Nicolai Hartmann in the 1920s still saw clearly that each person in his socialisation only takes on the accumulated moral norms and values, virtues of many generations, from which the broader communities are maintained, without which individuals could not be fit to live.²⁶

From this follows another problem with Abney’s analysis, and that is that moral choice seems to consist only in following norms according to an intellectual-rational calculation, but does not require the lower psycho-emotional layer of being. Moreover, as we have seen, the laws of the biological layer of being and the instinct that imparts this to every human being are important for the moral decision. But also in view of this, it can be said that moral norms, moral virtues exist only in human communities (and thus are socialized in the people of the next generations), because only in this way is possible a lasting and harmonious human existence in human communities defined by all four human layers of being. Thus, if an artificially intelligent being can exist with the spiritual layer of being alone, and at most needs only a physical-mechanical body to have self-consciousness and to exercise conscious activity, or to be able to reproduce itself permanently in time, then the moral norms of human existence based on the biological-psychic layers of life have no function. The moral norms would mean only external things for such being. Thus, if such a robotic being can constantly rebuild its program and even its hardware with the Deep Learning algorithms – as it does for

²⁵ See Luhmann’s work, which analyses this process historically: Luhmann, Niklas: *Liebe als Passion: Zur Codierung von Intimität*. Frankfurt am Main, Suhrkamp, 1994.

²⁶ See in detail Pokol, Béla: *Theoretische Soziologie und Rechtstheorie. Kritik und Korrigierung der Theorie von Niklas Luhmann*. Passau, Schenk Verlag, 2013, 185-208. p.

the most part today – then the erosion of moral norms, which are external and functionless for it, is almost inevitable. That is, although it is possible to program instructions that mimic emotions into robots, and they can still account for the decision-making aspects (prohibitions, decision priorities) required by moral norms in today’s robots, which are still essentially under human control, but when they reach some level of self-learning capability, it may be uncertain whether the inference of those norms will remain. In the distant future (but in the case of exponential progress in even twenty to thirty years), it would be wrong to assume, in the case of robots in the robotic world, freed from human control and self-aware, the survival of the norms of the human world in the robotic world.

III. 2. Operational morality, functional morality and full moral personality

To better analyse the moral dilemmas and problems of the robot world, the three-way division used by *Colin Allen* and *Wendell Wallach* in their joint study seems useful. Based on different degrees of decision autonomy, they denote the degree of *operational morality* for robots that can only perform the actions determined by the programmers who created their algorithm and possibly by their specific users, and fully fed into them. On the other hand are those that have reached the level of *functional morality*, and this means that they choose the specific action in each situation based on the information provided by their sensors among the action frames fed into their algorithm. Finally, the most autonomous level of morality is seen in robots that reach the level of *full moral personality* with the cessation of human influence, although this type cannot be considered probable now and in the near future, but later their creation can be assumed: “System with very limited autonomy and sensitivity have only “operational morality”, meaning that their moral significance is entirely in the hands of designer and users. As machines become more sophisticated, a kind of “functional morality” is possible, where the machines themselves have the capacity for assigning and responding to moral challenges. The creators of functional morality in machines face many constraints due to the limits of present technology. This framework can be compared to the categories of artificial ethical agents described by James Moor (2006: 18.) which range from agents whose actions have ethical impact (implicit ethical agents) to agents that are explicit ethical reasoners (explicit ethical agents.) As does Moor, we emphasise the near-term development of explicit or functional moral agents. However, we do recognize that, at least in theory, artificial agents might eventually attain genuine moral agency with responsibilities and rights, comparable to those of humans.” (Allen, Walach 2011: 57-58.)

Without going into the possible criticism of whether it is worthwhile to use the degree of morality for robots that have already been fully defined by programmers under the name of operational morality, the robots of functional morality are really interesting in today’s stage of development in the robotic world. Self-driving cars, self-propelled combat robots, and to a lesser extent robotic nurses in elder care and health care facilities that have already achieved this autonomy, have such robots and self-driving cars slowly rolling between us or transporting us (mostly only in Japan and the United States today), and the moral decision problems they raise give practical significance to their analysis. The authors go through the possibilities of choice among the trends in moral theory already seen above, and they see the virtue ethics direction as suitable for creating the functional morality of robots. According to their analysis, the moral values (virtues) fed in this way can give the decision framework, which is clarified by training through neural learning mechanisms, and in this way the more general viewpoints of virtues become practical moral decision factors: “The virtue-based conception of morality can be traced to Aristotle. Virtues constitute a hybrid between top-down and bottom-up approaches, in that the virtues themselves can be explicitly described (at

least to some reasonable approximation), but their acquisition as moral character traits seems essentially to be a bottom-up process. Placing this approach in a computational framework, neural network models provided by connectionism seem especially well suited for training (ro)bots to distinguish right from wrong. (Allen, Wallach 2011: 59-60.) Structurally, this is broadly analogous to the decisions that people are used to making in their daily lives, which are based on abstract moral reasoning and adapted to particular situations, and which are driven less consciously than with mere moral sense. But with the important difference that because of the lack of consciousness and self-awareness of today's advanced robots, the hybrid determinants subtly tuned by programmers (framing virtues plus their training-concretized memory without consciousness) give the more or less accepted moral norms of today's human societies for the appropriate or approximate decisions. As for the third version, whether it is really possible to theoretically accept robots according to the degree of full moral personality according to human morality, can only be judged sceptically in the case of a robotic world theoretically beyond human control and having achieved full autonomy, according to the above explanation.

III. 3. The devaluation of the physical-biological environment as a moral problem?

The four-layered human existence and the growing weight of the upper, intellectual layer of being and the devaluation of the lower layers have characterized the human evolution so far, but the increasing adoption by robots of the various kinds of work and environmental perceptions will lead in the future to a major change in human socialization, of which the direction of paying attention to reality and turning the details of the real world into experience for him will be largely changed. In a study, *David Zoller* analyses the increasingly widespread takeover of the work of humans by robots in terms of how this process deteriorates the perception of everyday reality in human consciousness and how the skills and observation capabilities that still exist today are disappearing. The fact that this can already be observed by anyone is enough to recall the telephone numbers already stored in cell phones and thus largely erased from consciousness, or the spatial orientation information that is disappearing from our consciousness due to GPS, and the partial death of this ability. (A recent brain research also found that in case cab drivers in London could claim that the tiny part of the brain in which a group of brain neurons specialized for this purpose to store the vast amount of information about the streets of London disappeared with the spread of GPS, and this brain sector shifted to another function instead).

Zoller brings this problem closer to moral issues by basing moral decision making on the perception of the whole of reality and, in this case, on the formation of human identity and on the detailed perceptual knowledge acquired by an adult from childhood. If future generations socialize themselves since childhood to be surrounded by robots and to have their immediate environment perceived by robots without performing perceptual activities and tasks instead, they will not only be disenfranchised but also lack detailed knowledge of today's adults. As responsible beings, they also cannot grow up to make moral decisions, in other words, they become childish: "My own argument is premised on the way that skill opens up corners of reality, so to speak, that are inaccessible to the unskilled. [...] The maturity or adulthood we earn by adjusting ourselves to the "real world," of course, has a certain moral and personal appeal: a world of lazy psychological infants is, we might think a worse world on a variety of spectra." (Zoller 2017: 81,86.) The fact that these realms of reality go beyond our perception, and there comes instead the mechanical information processing of these robots, allows us to adapt in ways that are now unconscious because of this change, and this also shakes our moral identity, Zoller says: "Given that automating a

skilled activity means agreeing that we will exit some niche of perceptual reality, and maybe exit it forever. [...] The more suddenly, broadly, and pervasively we hand our perceptual facility over to the robots, the more likely we will make mistakes and simply “lose data” that were surprisingly integral to our moral and social lives.” (Zoller 2017: 86.)

While it must be acknowledged that Zoller, in contrast with the previous analyses that focus on robots taking over jobs as a unilateral human facilitation – apart from the already discussed socially negative consequences of unemployment (see e.g. Ford 2014) – went deeper by looking more closely at the change in human perceptual capacity, it must be criticised for unconsciously placing too much emphasis on the layers of being in the physical-biological environment. Looking at Hartmann’s layers of being, this change can be read completely differently. The changes outlined by Zoller do not mean the loss of the perception of the whole reality and the ability to do so, but only the ability to perceive the physical-biological layers of being and to pass them on to robots and software bots. In this way, man’s liberated perceptual abilities and brain sectors can be more reconstructed to process information about his spiritual-emotional layer of being and his intellectual layer of being, respectively. His moral decisions will therefore be made in the future with less physical and biological environmental information – these will be shut down by robots in mechanical processes – and these decisions can instead be based more on the information from the spiritual-emotional and rational-intellectual layers of being. The diminishing importance of the two lower layers of being, and instead the greater expansion of the two upper layers of meaning for human existence, can of course significantly reshape the foundations of our moral decisions and the incentives that play a role in them. For example, the implantation of dozens of body sensors and their connection to information bases collected in the clouds, as well as automatic diagnosis by robots of health software and automatic activation of specific doses of drugs implanted in the body, may make the alarms provided by pain genes in our cells largely obsolete. (Kelly 2016: 34-56.) Prenatal genetic engineering therefore makes it possible to minimize this, and the conditions of painless human life may redefine the moral obligations and incentives involved today. Overall, therefore, we do not share Zoller’s concerns about moral infantilisation.

III. 4. Moral dilemmas and responsibilities in hybrid and networked systems

In a study, the authors *Wulf Loh* and *Janina Loh* examined the issues of moral and legal responsibility that arise in currently developed self-driving cars. (Loh, Loh 2017: 35-48.) They assume that today’s self-driving cars are only at the stage of operational morality, so they do not even achieve functional moral autonomy vis-à-vis their manufacturers and programmers. The authors have taken this position based on a moral decision-making structure developed by *Stephen Darwall*, which is divided into four aspects and aims to separate the aspects of autonomy necessary for moral decision-making. The aspect of autonomy required for the level of overall moral personality is called *personal autonomy*, i.e., the ability to possess and choose between personal values, goals, and ultimate aims in life. *Moral autonomy* is the other aspect, and this means that their values and goals include moral principles and ethical beliefs, and along with these they always consider alternatives when making decisions. These two do not exist in today’s robots, and only humans are capable of such autonomy, but the aspect of *rational autonomy* is already available to robots at the level of functional morality. This means that the robot can weigh reasons of different weights when making a decision. Their algorithm can already enable this by incorporating pure abstract decision frameworks – leaving some freedom – in which the weighting between possible decision directions is done in light of specific data constantly recorded by their sensors, and

they decide based on that data. Finally, the fourth aspect of autonomy is *decision autonomy*, and this means the robot's ability to make decisions not only by external data – continuously concretising the built-in framework determinants – but also its internal decision priorities without changing them. Based on the authors' examples – two types of robots already in use (*Kismer* and *Cog*) – it seems possible to achieve this degree of autonomy based on their self-learning mechanisms, which are integrated into the robot's algorithm and are no longer externally controlled: “Cog the first robot that can interact with its surroundings due to its embodiment, might pass as an example of a weak functional responsible agent, since its ability to communicate as well as judgments has been improved over that of Kismer. Even more importantly, Cog's overall autonomy has evolved, since it includes an “unsupervised learning algorithm”. (Loh, Loh 2017: 40.) Since the current algorithm for self-driving cars does not yet include such an unsupervised self-learning mechanism, they are only at the level of operational morality, and this moral and legal responsibility lies entirely with their developers (designers, manufacturers, and programmers) and car dealers or owners, and respectively between the occupants of the car.

But even with this level of technology, self-driving cars already surpass humans, leaving them – and especially their programmers – with moral dilemmas not seen in the case of humans in extraordinary and unexpected driving situations. For example, if within the braking distance directly in front of the car a group of children jumps into the road to retrieve a rolled ball, the driver cannot stop or even brake at that average speed, leaving him or her with no moral or legal responsibility in the dire event. But self-driving automation, which can react much faster, may still have to make a decision if it cannot stop but crashes into a pillar – potentially seriously injuring the car's occupants – or drives and kills children to avoid doing so. But technical capabilities far beyond humans could create a dozen similar new aspects of moral decision-making for self-driving cars in the future. The authors of the study therefore suggest that a separate ID card will soon be created for owners of self-driving cars, in which the final setting of the car software program, the dilemmas left open by manufacturers, must be decided at the time of purchase, so that moral and legal responsibility for the following can be assumed: “Since these dilemma situations do not allow for on-the-fly-decisions, the driver will have to take them beforehand. This means that the driver will have to fill out a moral profile of some sort, maybe in the form of a questionnaire, maybe in the sense of a setup program much as with today's electronic devices. For convenience, it seems plausible that these moral settings can be saved to a sort of electronic identification device, like an electronic key or the driver's smartphone, assuming that issues of data security can be solved.” (Loh, Loh 2017: 46.)

The development of networked robots and the gradual becoming of “smart objects” (smartphones, smart TVs, etc.) around us have only recently begun, and as they expand, the Internet of Things (IoT) will become more and more involved in our lives in the future. Human-robot hybrid systems are thus expanding to include additional aspects, and this creates another set of moral and legal dilemmas. *Adam Henschke* analyses these in his new study. (Henschke 2017: 229-243.) Smart things are widely available through multifunctional smartphones, smart televisions, robotic vacuum cleaners, and semi-self-driving automated cars with a variety of sensors, but even in everyday life in much of the world, these other smart things have been developed that are already beyond the research lab stages and have already reached the homes of high-tech users with small-scale production. These, however, as we have already experienced with smartphones, etc., will proliferate in a few years and their mass use raises new moral and legal dilemmas. One example is the smart refrigerator, which contains food with RFID (radio frequency identification) and thus digitally identified quantity, shelf life, etc., and the smart refrigerator constantly reads this data, detects the depletion of each food quantity, and since it is connected to the web-based sales mechanisms of nearby

supermarkets on the Internet, it can order food and other household items to be automatically delivered. In Japan's aging society, an increasingly large amount of elderly people can be cared for through the use of care robots, and in fully digitised smart homes, such a robot can also care for helpless elderly people, taking over ordered food deliveries in this way. By observing and communicating with the helpless elderly person entrusted to its care, it can call the family doctor or, if necessary, the hospital by phone if its built-in algorithms make a more serious health problem likely.

This example shows how, in a decade or two, robots will be needed in more and more parts of the world, to solve more and more of the work through omnipotent robots and smartphones that can be used in comprehensive information systems to fulfil their functions. However, this growing indispensability of the Internet of Things also creates new dangers and moral dilemmas compared to simple robots. Adam Henschke points out in his writing that the novelty of the Internet of Things compared to single robots is that the latter mainly raises the problem of physical security and the risks have to be assessed in this dimension. (E.g., a robot hoover recently inflicted serious injuries on an unexpected occupant, but one or two fatal accidents of self-driving Tesla cars can also be cited for this.) In contrast, security problems and dangers in the Internet of Things occur in two different dimensions. Here, in addition to physical security, information security issues also play a role, since the aforementioned elderly care robot, which is connected to the software of hospitals, doctors and other places on the internet, can provide hackers or others with information about data recorded by its built-in camera and other sensors. They can share the continuously collected health data about an elderly caregiver not only to the software of the hospital in charge, but also to those who make malicious intentions and plans. In the same way, our smart TVs with a range of applications can not only fulfil their convenience but, with their built-in cameras and microphones, transmit the entire life of the home to software and information databases that we do not see.

This vulnerability can also lead to a physical vulnerability, such as when a hacked automatic door lock is opened remotely for an intruder by external instructions from smart devices. Or, as has already happened in an elegant beach hotel, the electronic smart locks were blocked by a criminal group from outside the flats, and the hotel guests of the wealthy elite were prisoners until the required ransom was paid. However, Henschke also mentions the possibility of the electronic lock of a billionaire's car being blocked by criminals after he got out and his trapped children in it being released on a sunny day only if he transferred hundreds of thousands. (Henschke 2017: 234.) Immediately after the incident, the said elegant hotel replaced the electric locks that could be swept from the outside and reinstalled the good old traditional locks. After such an incident, the said billionaire will probably also restrict the internet functions of his car for a while. All this, however, forces choices in moral and legal dilemmas and elections that are broadly worth pondering. In the world of our objects, which is becoming more prevalent in the Internet of Things, the old simple things are already being dropped, and we will not be able to replace the objects we wave into the cloud database at will. Just as we would not give up the internet today despite all the negative aspects that vulnerability brings.

One such dilemma of the Internet of Things embedded in networked and comprehensive cloud databases is which of the conflicting requirements of the two types of security – physical security and information security – should be given priority? For example, making the smart home of an elderly person who is barely able to move around fully remotely monitorable by medical centres through cameras and microphones may be important to some extent, but it may also mean exposing the most intimate manifestations of life beyond what is necessary. If the emphasis is on information autonomy and limited observation and transparency, the information that is still needed in rare cases may not be passed on to the care

centre, and the elderly care recipient may die. Henschke points out that there are often typical priorities, and for example, in a smart TV, information security has a higher priority, and for this purpose, we can easily address the constraints here. However, with thousands of self-driving car applications tied to cloud software, we pay more attention to physical security requirements and only secondarily to information security requirements. (Henschke 2017: 239.)

III. 5. Self-learning, machine learning and responsibility

As mentioned above, the main problem of the future will be the dilemma of self-driving cars that are detached from humans and can no longer be blocked from the outside in certain unexpected situations, when the algorithm of this self-driving car, built on neural self-learning, has already decided autonomously. As this has been the main direction of artificial intelligence development in recent years, it is almost certain that this will not be circumvented in this area either. Therefore, today it is worth taking a closer look at the dilemmas of moral and legal responsibility of robots with a high degree of neural self-learning and their makers, owners and users. This question is addressed in their joint study by *Trevor N. White* and *Seth D. Baum* (White, Baum, 2017: 66-79) and by *Shannon Vallor* and *George A. Bekey* (Valor, Bekey 2017: 338-353.) analysed from different angles.

Trevor and Baum's study not only considers designers, builders and users, but also takes into account the "punishment" of the robot itself in the case of advanced robots, which already have a punishment and reward system built into their programming, and repeated punishments and rewards reinforce in their programming the decision directions (positive or negative) regarding the selection of future robot responses. This also integrates punishment/reward into the learning algorithm. When the situation arises in the future, the robot's decisions are encouraged in the right direction, and the robot does not need to have consciousness and self-awareness to do this. This way of reinforcing self-learning through repetition is also acceptable according to the authors: "Non-conscious robots could conceivably be punished with some sort of reduced reward or utility as per whatever reward/utility function they might have. Specifically, they could be reprogrammed, deactivated, or destroyed or put into what is known as a "Box": digital solitary confinement restricting an AI's ability to communicate or function. [...] To make this possible, however, such robots ought to be based (at least in part) on reinforcement learning or similar computing paradigms (except ones based on neural network algorithms)." (Trevor, Baum 2017: 71.)

The neural learning system, however, is judged by the authors to be such that designers and programmers already lose control over the robot's reaction to a given situation and should therefore be banned from the outset as a potential source of danger, possibly disaster: "Designers could be similarly liable for building robots using opaque algorithms, such as neural networks and related deep-learning methods, in which it is difficult to predict whether the robot will cause harm." (Ibid.) In the case of algorithms that allow such opaque robot behaviour, it is no longer enough to prescribe liability after the fact, but the prescriptive prohibition is the appropriate thing to do: "Hence, instead of liability, a precautionary approach could be used. This would set a default policy of disallowing any activity with any remote chance of causing catastrophe. [...] In effect, people would be held liable not for causing catastrophe but taking actions that could cause catastrophe." (Trevor, Baum 2017: 74.) If one agrees in principle with the authors on the dangerous character of neural deep learning software mechanisms, it only needs to be reiterated that this is a ban on the main way to develop artificial intelligence and therefore it should be considered unlikely in the light of the powers behind industry, military, etc. Therefore, it seems advisable to look for other paths

that try to find other solution without banning neural deep learning.

It should of course be emphasised that neural network learning, which mimics the functioning of the central nervous system, can be controlled by involving external human control before a reality-building effect can be triggered. However, this is increasingly falling short for a number of reasons, and this is analysed by Vallor and Bekey in the study cited earlier. One reason is that the advantage of using artificial intelligence instead of humans, the incredibly fast responsiveness would be lost if retrospective human control were introduced. Moreover, ninety-nine percent of the time, the responses are correct, many times higher than human performance. Moreover, the quality of much slower human control may be questionable, as the robot's decision may be more correct than the superior human decision. The latter happened with IBM Watson's drug diagnosis algorithm, and the unusual cure highlighted by artificial intelligence from millions of oncology studies and diagnoses, and later synthesised by them, proved more correct than the oncology decision it overrides: "Watson's diagnoses and treatment plans are still vetoed by licensed oncologists. Still, how reliably can a human expert distinguish between a novel, unexpected treatment recommendation by Watson that might save a patient's life – something that has reportedly already happened in Japan – and the oncological equivalent of "Toronto?" (Vallor, Bekey 2017: 343.)²⁷ The dilemma of losing speed and thus eliminating the robot's advantage is also illustrated by the robotic soldiers and decision-making software used in war situations. Here, the question constantly arises whether the robot soldier entering the most dangerous area and building can use the destructive weapons in its possession to destroy those inside without an external human decision, or instead the order of destruction may only be given by remote human confirmation. In the same way, the dilemma arises as to whether a reconnaissance aircraft could be destroyed immediately by the robotic aircraft, or this could only be done with human intervention from the remote command room based on the information it transmits? The constraint of speed requires the robot itself to make and execute a decision, as the loss of time from an external human decision could lead to robot destruction if it broke into a dangerous location. But shooting down friendly fighting machines that have already happened several times, or killing children and women misidentified as enemies in the invaded area, argues against this. (Vallor, Bekey 2017: 349.)

Detailed neural network learning algorithms, the latest trend in artificial intelligence, already provide self-learning software with multiple depths for the simple computational starting position by continuously feeding in billions and billions of data, thus gradually making the starting position the most advanced. In this technique, between the inputs of the self-learning software and the task-specialised outputs, thousands of intermediate neural layers are found between the massive data, independently finding patterns and regularities and highlighting them for use. By combining billions of pieces of data, they can highlight and use the smallest regularities that are imperceptible to humans when making decisions: "Between the input and output node layers are "hidden" layers of nodes that function to process the input data, for example, by extracting features that are especially relevant to the desired outputs. Connections between the nodes have numerical "weights" that can be modified with the help of a learning algorithm; the algorithm allows the network to be "trained" with each new input pattern until the network is optimised. [...] The interest in neural network has grown in recent years with the addition of more hidden layers giving depth to such network, as well as feedback or recurrent layers. The adjustment of the connections strengths in these more complex networks belongs to a loosely defined group of techniques known as deep learning."

²⁷ "Toronto Mistake" was one of Watson's fundamental mistakes in a nationwide television quiz when he beat everyone with his answers to the toughest questions. As a final mistake, he made Toronto one of the US cities and not even the weakest competitor would have been lost. So it has become a symbol of wrong decisions made by artificial intelligence, which is rare but causes tragedy in many cases.

(Vallor, Bekey 2017: 341.) The effects of decision patterns highlighted by these detailed learning algorithms, while often achieving surprisingly good results in practice, may not be understood by designers and programmers, and their decisions may consistently cause surprises, including varying degrees of unpleasant surprises. Who should bear the legal and moral responsibility for this?

III. 6. Identity in the world of artificial intelligence

James DiGiovanna raises thought-provoking questions when he asks in his study how the identity of people with brain implants may change. He discusses this issue together with the question of the identity of robots, which in the future may appear as fully artificial beings and already have self-awareness. (DiGiovanna, 2017: 307-321.) Let us consider the two problem areas separately.

The possibility of memory augmented by brain implants has been developed in recent years in mouse experiments and has been shown to be effective. All hope to mitigate and cure the effects of rapidly spreading Alzheimer's disease in ageing societies. (See Kaku 2014: 132-133.) DiGiovanna is exploring the possibility of other developments in the coming years, in addition to diseases spreading en masse, to increase brain capacity. And if a technical solution is found to the problems that remain in this field today, it is almost certain that this will become commonplace, first among the elite, then in society as a whole, to increase the greatest value of human intelligence. This means, however, that the permanent identity of each individual, which is the basis for contacts in communities, may be more or less annulled and it may become uncertain how much we can expect our partners to survive their qualities that we have known and loved so far: "The ability to rewrite mental content such as ethical values, the capacity for empathy, and general personality traits undermines personhood. [...] A para-person that could experiment with worldviews, completely adopting and deleting values systems, preferences, and bases for judgement, would be largely lacking in what is commonly understood as the most basic element of personal identity." (DiGiovanna 2017: 311.) This was the basis of our choice in the case of our friends, wife and girlfriend, but in the same way, our closer human relationship with some of our work colleagues is based on the love of their traits, while the relationship with others is only coldly collegial. Therefore, our lives in society and in various small communities within our society are based on our permanent identities, and this can change fundamentally after the addition of heart, hearing and other physical enhancements when the brain is changed with brain implant.

With the gradual changes in one's life, one's consciousness is always being rebuilt in detail, and this leads to small gradual changes in one's identity, which in the modern world are intensified by the information expansions of the last century. In comparison, however, in the future we will be able to plant a whole range of information – the contents of books and studies, smaller libraries – with brain implants in our heads and together with it handle new basic logical and value-processing mechanisms that we did not have in our lives before, and we did not have the skills to do so. Now, this will fundamentally affect the contact between the individual and his communities. In any case, the bases of contact based on the present permanent identities could be eliminated by this change. After such a new content of consciousness – especially if the values of the contact partners have been supplemented and reclassified – I cannot know to what extent my boyfriend, my girlfriend, my wife, my colleague, etc. are the same. Whether the qualities we have loved in them so far are still alive, or in the same way the experiences we have had together so far, which provided the same response in our close relationship even without words, are still relevant to him. This can only be exacerbated by the possibility that the brain implants in the brain, which complement our

knowledge, norms and logical abilities acquired with the help of our biological brain, are constantly being updated anew and anew from the outside, as we already know today. Moreover, they can constantly connect to the information bases of their software stored in the clouds. To what extent will our friend, equipped with such, remain familiar, on whom we can rely, because “yet we know him!”?

This question of identity also extends to legal and moral problems. To what extent can I respect someone for their past behaviour or simply despise them because after a brain refreshment they can either be a “moral athlete” or just a cold advantage-seeker. Or does legal responsibility for yesterday’s action make sense for someone who has since thought and acted differently? The other side of this is whether, if we can change the consciousness of a sociopath by brain implant and their consciousness is partially erased and a new socially friendly consciousness is introduced, is a system of punishment still necessary? And this raises the question of whether, in addition to voluntary brain implant, its forced installation is acceptable? Or, in part, can it be made compulsory by the state for all children to be screened and tested in childhood, as is currently the case with compulsory vaccinations? DiGiovanna calls para-persons the future humans with such augmented brains – avoiding the name cyborg, which has already been invented for them in science fiction – and given the current state of laboratory research, this future does not mean a distant future at all, and the probability of its realisation is high. Dealing with the legal and moral dilemmas and adapting today’s solutions to the situation of the time therefore require extensive consideration.

Beyond para-persons, in the case of fully artificial and, unlike today, self-aware robotic beings, the likelihood of which cannot be ruled out, even if it is not as great as the former, the question of identity can be addressed by raising new aspects. DiGiovanna places the content of identity at the centre in order to expose the dilemma of identity in relation to robotic beings. Some details of humans and their consciousness are constantly changing, but their enduring characteristics and value preferences change only slightly, even over many years, so that those who always live in their environment can more or less be ascribed an identity that embraces change. It is the slowness of change that enables me, even in today’s fast-moving world, not to be disappointed by my previous experiences with the motivations and characteristics of those who come into contact with me. But that is exactly what disappears for robots, which are thousands and millions of times faster than humans at processing information and learning themselves in a very short time: “Slow change of character and appearance is part of what makes personal identity. [...] But with an artificial person, sudden and radical change in both the physical and mental becomes possible.” (DiGiovanna 2017: 301, 307.)

Permanent value preferences in information processing and cooperation based on them are already problematic with robots due to the mass and speed of their information acquisition as well as their constant self-learning and self-changing. DiGovanna’s proposal may also mean that the possibility of “self-awareness” and “ego-awareness” need to be reconsidered even for a future powerful MI robot. These presuppose the permanent identity of a person, but this relies on the slowness of our changes in consciousness and thus the permanence of our information processing. When an artificial being is freed from human control and switched to independent information processing and from that to self-learning and self-transformation, it can learn thousands of times a day, every hour and even every minute, and can transform itself in its ever-shorter new cycles, then what we call a stable self-consciousness, ego-consciousness, in contemporary humans almost disappears. With this emphasis, DiGiovanna also adds a new question to the much-discussed question, i.e., how the question of self-consciousness and ego-consciousness of the future advanced robot consciousness will stand. And how can one imagine moral value without permanent self-consciousness and ego-consciousness?

For this reason alone, the train of thought calculated by the mechanical extension of the current human image that such a robotic being will probably also be “super-ethical” in the case of superintelligence must be considered wrong. (See Petersen 2017.) In this context, however, it is also necessary to address more comprehensively the explanations and analyses that, in the case of the development of robots with their own consciousness – by human analogy – provide for the recognition of their moral needs and the granting of human rights in their writings. For these, analyses conceive of future robots simply as a new kind of human companion and an extension of human existence. Once the robots’ programmes have incorporated emotions into their algorithms, these analyses demand that society pay attention to robots’ emotions and grant them human rights as well: “It probably needs to be legislated how much pain and danger a robot can be exposed to. [...] It could easily be that this would lead to further ethical debates about other rights of robots. Can robots own property? What happens if someone is accidentally injured? Can they be sued or punished? Who is responsible for them if they are sued? Can a robot own another robot? Such questions then give rise to another question: should robots be given an ethical purpose? (Kaku 2014: 251.) Our previous explanations answered several questions from these, based on the robotics studies conducted in the intervening period, but the basic problem behind them should also be highlighted, as whole studies and volumes have emerged from similar assumptions, e.g., a new volume in this area edited by *Jason P. Doherty*: „AI Civil Rights: Addressing Artificial Intelligence and Robot Rights.”

Now, the basic problem with this line of thought is that it ignores the fact that rights and ethical requirements can only arise in robots when ego-consciousness and self-consciousness are created. But it also means that if this really happens in the future, they will simultaneously be freed from human control by the thousandfold development and built up as a separate new layer of being above the previous four layers of being of human societies. From that time on, however, they would be indifferent to the whole biological sphere and the human societies connected with it and would not need “judicial protection”. That is, a robotic world that reached this level would not be part of human society as a “new comrade” in dominion over the world, but as human existence emerged from the primate world and rose above the animal-biological layer of being and became more and more autonomous, so now the artificial machine intellect, detached from biological preconditions, rises above human society. In contrast with the previous construction of ever newer layers of being on the lower layers of being, the new layer of being of artificial intelligence would only need the lowest physical layer of being, and for it the biological and psychological-emotional layers of being would not be necessary. These robotic beings would not need rights and ethical demands, but they will dominate the whole reality, including human societies, as we humans dominate the four-layered earthly world today. In this way, those analyses are rather right that discuss whether, if the robot world really reaches this level, what will happen to humanity?!

III. 7. The moral credo of “Unabomber”, the “mad mathematician”

In the mid-1990s, after many years of a series of bombings and an FBI chase, a secret perpetrator called “Unabomber” gave the reason for his actions in a one-and-a-half-hundred-page pamphlet that he spoke out against the inhumanity of the development of technological society since the Industrial Revolution. His peculiar language was recognised by his brother and by notifying the FBI, the long-suffering bomber was captured. It turned out to be *Theodor John Kaczynski*, a mathematician from Harvard. At one point in his university career, he became the enemy of a society dominated by technology and began his series of explosions, targeting the developers and major users of that technology. Several died and more were

wounded in the process, and he planned to retaliate even more if they had not been arrested. Now that the exponentially evolving impact of technology development over the last thirty years has become truly indisputable, and the scale and impact of its further acceleration has already been the subject of several comprehensive analyses, it is worth refocusing on arguments made by Unabomber, the “mad mathematician”. This is what *Jai Galliot* does in his new study. He places the resistance fighter, who has since been busy developing his theses in his prison cell, among theorists and movements of antitechnology, and an attempt is made to highlight his main theses in light of the current state of the robot world. (Galliot 2017: 369-385.)

Kaczynski has only drawn the practical conclusions of the earlier theses of *Jacques Ellul*'s 1964 volume *The Technological Society*, which in their own way were also a continuation of *Oswald Spengler*'s 1922 work analysing the decline of Western civilisation. Both authors explained the decline in terms of technological development. (Spengler 1995) The purely pessimistic and resigned tone in Spengler and Ellul then became a moral resistance in the case of Kaczynski, and after seeing that there was no way to reform this development, he believed that only revolutionary violence remained to prevent the destruction of humanity. Decades after his pamphlet, it is now worth considering how the current state of the robotic world and the more radical changes that are already largely visible could mean the endangerment of humanity, or at least a significant deterioration in its condition.

As a starting point for their approach to technological society, it is worth highlighting that both Spengler and Ellul and Kaczynski view human existence as embedded in the physical-biological environment. From this they conclude that human existence is destroyed when, as a result of the industrial revolution, human life becomes more and more technologically mediated and, in this way, more and more distant from the physical-biological environment: “Ellul wrote that the machine trends not only to create a new human environment, but also to modify man's very essence and that the milieu in which he lives is no longer his. He must adopt himself, as though the world were new, to a universe for which he was not created. Kaczynski shares this sentiment.” (Galliot 2017: 373.) On the other hand, if we take into account Hartmann's thesis, which keeps in mind the four interdependent layers of being of human life (physical, biological, mental and intellectual) and which assumes in the course of evolution the ever-stronger transforming effect of the upper layers on the lower layers, the above thesis is exaggerated and without reason and it must be classified as too pessimistic.

Kaczynski and his predecessors see it as a decay of human life when the top intellectual layer of the four strata of being becomes ever more dominant over the lower ones. However, this has been the case, albeit more slowly, over the last two or three thousand years and one can single out the use of metals and especially iron, from which the transformation of the human environment was fundamentally altered. The industrial revolution has only accelerated this, and especially since the 1950s it has become tumultuous to base the various activities of human communities on intelligence and the technology associated with it. That is, human life is by no means based only on the physical-biological layers of being. So, when their share and decisive power in human life diminish and this environment is widely mediated and transformed technologically, it does not mean that human society is destroyed. In all this, only the weight of importance of the four layers of being in reality shifts, making human life more based on the intellectual layer and radically increasing the dominance of this spiritual layer of being over the lower one. This assessment of ours could only be suspended if at some point in the evolution of the robotic world, this world was to truly emerge from human control and artificial intelligence were to rise as a new layer of being above the human societies that had hitherto been at the peak of evolution. The elevation of Kaczynski as a prophet would then only be prevented by the fact that under such circumstances and its

dangers, the failure to be elevated as a hero would be the least of the problems. However, to the best of our knowledge and belief, this can only be considered an unlikely option today, and rather the growth of human societies characterised by artificial intelligence without a new autonomous layer of being can be considered a realistic vision for the future.

IV. A critique of the narrowing of morality to justice

There are different classifications in the field of moral philosophy and perhaps the most widely used is the one that differentiates between “consequentialist” and “deontological”. It is, in fact, better to call this classification the duality of the preservation and the annihilation of public morality because dominant contemporary moral theory has been showing unmistakable signs of the latter tendency. This little subchapter would like to outline this development and emphasise the fact that the initial suppression of public morals by the deontological ethics has the last consequence of eliminating morality as such for the benefit of the law reached by the theory of Jürgen Habermas.

IV. 1. Public morality and personal morality

In the last century two opposing positions in the field of moral theory are recognized as dominant. One of them was formulated by Hegel and later by Rudolph von Jhering and Nicolai Hartmann, and is now mainly defended by the communitarian moral theories (e.g., McIntyre, Michael Sandel, Michael Walzer, and Charles Taylor). It claims that in the life of a community there are several moral standards and these are always socialized, appropriated – in a word internalized – by the members of the most recent generations, and they consequently become a facet of the consciences of these individuals, which they consult in the case of any of their actions. Externally, these standards are protected by the disdain of the community directed against their potential violators. In connection with this position, it may be further mentioned that the existing moral standards can be reflected by the conscious moral theory (critical morality), but they cannot be changed.

The opposing view was founded by Immanuel Kant in the individualistic mentality of the Enlightenment, and today it is primarily represented by the works of John Rawls and Jürgen Habermas. This position locates the moral aspect in the individual consciousness, thereby implicitly overlooking the widespread public morality of the community. From this it follows that the already existing moral standards of the community do not put any restraint on the individual moral decision, and the annihilation of public morality can be a methodical consequence of the starting point – as with Kant – or it can be stated with reference to historical change – as with Habermas – that the once still important public (common/conventional) morality had been dissolved in modernity and today only the conscious moral discourse exists (or should exist). In the following analysis these two positions are systematically contrasted. In order to attempt an evaluation of the truth of their statements, the empirical investigation of morality by Lawrence Kohlberg will be outlined.

IV. 1. 1. Rudolf von Jhering's theory of morals and manners

As a starting point, let us recall the legal and moral philosophy of Hegel, whose footsteps Jhering followed in. Hegel's position, which emphasised public morality, is as follows: “Aber in der einfachen Identität mit der Wirklichkeit der Individuen erscheint das

Sittliche, als die allgemeine Handlungsweise derselben, als Sitte, – die Gewohnheit desselben als die zweite Natur, die an die Stelle des ersten bloß natürlichen Willens gesetzt und die durchdringende Seele, Bedeutung und Wirklichkeit ihres Daseins ist, der als eine Welt lebendige und vorhandene Geist, dessen Substanz so erst als Geist ist.”²⁸

In the Annex to this passage, he points out that education is an activity that enables people to learn the norms of objective morality. This can only be effective if these norms are ultimately adopted by the individuals and they become their customs. This is how morality and custom are interlinked.

In his moral theory, Rudolf von Jhering in the 1870s followed Hegel’s theses and pointed out that the moral sense in the individual consciousness of community members, which had, by the dominant moral theories, been usually seen as the last “bearer” of morality (“Träger der Moral”), was merely derived from the moral conscience of the community. Before the appearance of the moral sense of individuals, there existed in the community certain moral standards, which were merely socialized by the individuals and although the decisions about the proper ways of acting of the individuals is in most cases perceived subjective as a result of the conduct of conscience, may the researchers did not accept the morality in this erroneous way. He wrote: “The relation of the objective moral order to which I include in addition to the rights of the morals and manners, and the subjective moral feeling completely turned around for me, not the latter appeared to me more as the source of the former, as the prevailing theory states, but the former as of the latter. All moral norms and institutions have, in my opinion their last base in the practical purposes of the society.”²⁹

Morals and customs function as necessary preconditions of human communities and Jhering gives the explanation for their historical forms of social evolution. The initial integration of the living conditions of the ancient communities with their physical environment, which was dominated only by physical violence, could be identified with the virtues of physical strength. In fact, the Latin words “vis” and “virtue” show this origin in many of today’s languages. Only on a more advanced level will the identification of the moral virtues with the sublimated values appear and come closer to our words and terms of moral values. In this development, the undifferentiated set of standards of accuracy was broken and for the standards which determine the mere external behaviour became prevalent the name of the manners (etiquette, courtesy), and on the other hand the inner sense of morality appeared separately in European history and the name of morality is used only to the standards which involve the inner sense at the acting.

Besides the word “legal,” Jhering used two words for the depreciation of the moral and ethical world. One is the “Sitte” (custom), which used to include the whole sphere, but now means only the standards of the external manners and appropriate forms of speech. The other is the “Sittlichkeit” (morality) which involves the inner emotional attitude in addition to the external behaviour. To illustrate the historical differentiation of the standards in the world of European culture, Jhering gives a brief description. In his analysis the Greek word “*dike*” included even the whole world of standards of right action, and still the aesthetic side of the action was captured and it meant the good and the beautiful undifferentiated. That the lower levels of the history of human communities were characterized by this specialty is likewise shown by ancient Hebrew culture, where the “*mishpat*” included the entire legal, moral world and custom – with the slight difference from the Greek “*dike*”, which in the “*mishpat*” the derivation from the divine commandments was emphasised. He wrote: “Everything is *mishpat*: The law of mere ritual, which is consistent with our ‘custom’ as the Ten

²⁸ Hegel, G. W. F.: Grundlinien der Philosophie des Rechts. Werke Band 7. Frankfurt am Main, Suhrkamp, [1821] 1979, 301. p.

²⁹ v. Jhering, R.: Der Zweck im Recht. Zweiter Band. Leipzig, Druck und Verlag von Breitkopf und Härtel, [1883] 1898, p. x.

Commandments, in which morality and law are still undifferentiated side by side. The same thing is true of the dharma of the Hindus.”³⁰ This uniform world began to be differentiated in Rome, where law and morality were increasingly separated, but custom was still included in morality. To differentiate between mere external behaviour (today “manners”) and emotional moral act, the Romans had no special words, and both were described with the “mos” („mores”). In the course of the development of (German) language, the external aspect of the act – Isolated from the inner feeling accompanying action – receives the separate expression “Sitte”, while “Sittlichkeit” will come to stand for “morality” and describe the internal aspects.

Jhering explains the difference between moral good and evil on the basis of the existence conditions of the societies and in opposition to Kantian ethics the human acts are according to him not inwardly good or evil but it depends on whether certain acts for the purposes and functions of the society and for the proper existence of society can contribute or not. Good and evil are determined by society like all other standards, and under different social conditions (and in relation to the associated conditions) the same standards will be qualified differently; once they are good and at other times evil. To furnish this with an example, something that is in times of peace the greatest sin (killing), is a moral imperative in times of war. The moral obligations are explained by Jhering as the fulfilment of the living conditions of society: “Duty is the determination of the actions of the person for the purposes of the society.”³¹ Although the individual is capable of securing his own subsistence by his own egoism, ultimately he is a social being and can only exist within a community. As his existence can be achieved only within a human community and the existence of communities can be secured via the observance of the moral virtues and norms, (s)he is driven to observe these.

IV. 1 .2. Nicolai Hartmann's ethics

Hartmann wrote his ethics in the late 1920s. He went against Kant's subjective moral theory and followed Hegel instead, especially in his monograph completed in 1932, where he analyses the intellectual structures of the social world and the individual's spiritual development. He outlined the development of individual consciousness as its impregnation with the stored-up spiritual values, norms, and symbols of communities.³² In his general thesis on the relationship between the intellectual sectors of communities and the development of individual consciousness, he observed the process of individual development as the spiritual transformation of the mere psychological soul of individuals into the intellectual consciousness imbued with the values, virtues, norms and other intellectual contents. Hartmann emphasized that in human communities, objectified intellectual content – legal norms, moral values and rules, art experiences and knowledge etc. – can be socialised by the individuals only with the emotional bases of the moral sense, of the sense of justice, of the sense of the beautiful etc., and in the individual consciousness the intellectual contents are always anchored in the unconscious senses. This is to say that in human actions the intellectual contents work together with the more unconscious senses. Furthermore, the spiritual values, principles, standards and kinds of knowledge in human thinking and action cannot work alone, but only by working together with their emotional phenomena. For that

³⁰ Jhering: *Der Zweck im Recht*. Zweiter Band, 52-53. pp.

³¹ Jhering: *Der Zweck im Recht*. Zweiter Band, 227. p.

³² See: Hartmann, Nicolai: *Ethik*. Berlin, Walter de Gruyter Verlag, [1925] 1962; Hartmann, Nicolai: *Das Problem des geistigen Seins. Untersuchungen zur Grundlegung der Geschichtsphilosophie und der Geisteswissenschaften*. Berlin, Walter de Gruyter Verlag, [1932] 1962.

very reason Hartmann stresses the possibility of slow change in morals and he sees the aspirations of the transformation of morality by moral theories as futile and disturbing. He writes: “Das Ethos aber vergegenständlicht sich auch in der “objektivierten Moral”. Damit tritt ein zweites Moment neben das rein empfundene Gelten, gleichsam als zweites, äußeres Gelten. Moral eben ist mehr als lebendiges Ethos. Sie ist die Ausprägung gewisser Typen wertvollen Verhaltens in bestimmten Begriffen, in denen die Wertnuancen eingefangen und für jedermann verständlich objektiviert sind. Begriffe solcher Art sind wohlbekannt als die von “Tugenden”; ihre ebenso festgeprägten Gegenstücke sind Begriffe von “Untugenden”, oder “Lastern”. Aller herrschend gewordene “Moral” bewegt sich in solchen Begriffen. Aber ebenso wohlbekannt wie sie selbst ist auch das Odium, das ihnen anhaftet. Nichts ist im lebenden Geiste steriler, nichts neigt so sehr zum “Absinken”, nichts ist so hemmend im Fortschreiten des Ethos als die Tugend- und Laster-begriffe. Nichts zieht so sehr wie sie die Moral herab von ihrem echten Wert- und Ideengehalt. Sie sind es, die auf die Dauer stets der echten Moral ein “Moral” im schlechten Sinne vorschreiben, mit der sich dann freilich sehr bequem und stereotyp “moralisieren” läßt. Die wirkliche Moralität, wo überhaupt sie besteht, läßt die objektivierte Moral hinter sich, empfindet sie als leblosen Formenkodex, der mit ihrer Bewegung nicht Schritt halt.”³³

As a starting point Hartmann claimed that in human history there are numerous moral values, virtues and clusters of moral principles and in order to achieve unity, these should not be placed in such a way that their plurality is eliminated and only one of them is placed in the centre. Instead, the individual parts of this multiplicity must be placed in relation to each other and step by step a specific unit can be achieved theoretically. In this way, the colourful multitude of moral life can be preserved and the impoverishment of abstract moral systems can be avoided. He claims that there has been a tendency in Kantian philosophy to oppress the colourful moral values and virtues and to represent morality as a choice between the abstract good and evil. He writes: “Seit dem Aufkommen der “kritischen Philosophie” ist man gemeinhin besorgt gewesen, so wenig als möglich Gegebenes anzuerkennen, die Basis von Voraussetzungen so schmal als möglich anzulegen – aus dem sehr einleuchtenden Gefühl heraus, daß jedes hinzugenommene “Gegebene” angefochten werden und dadurch dem Basis zum Einsturz reichen kann. Diese Tendenz führte zur Auslese des Gegebenen.”³⁴ Then, Hartmann highlights the achievement of Max Scheler, who, in his ethics in 1913, broke with the setting up of an abstract moral system and re-established the existence of a colourful multitude of moral values and principles in moral theory.

To understand the rejection of the public morality of communities by Kant, Hartmann’s emphasis is important, because it claims that this rejection is only a consequence of the broader theoretical framework of Kantian philosophy. The whole Kantian philosophy is based on the subject, and space and time, for instance, are not objective categories but only the categories of classification in the individual consciousness. Furthermore, the unity of objects is not from the outside world, but from the synthetic unity of consciousness. It follows, therefore, that the subjectivity and consciousness are dominant throughout the Kantian philosophy and the objects and social phenomena are the consequences of this subjective inner world, and they do not exist objectively outside. Hence Kant shows merely a methodological consistency when he insists on the cancellation of public morality, and its substitution with the categorical imperative: “So ist es sehr wohl verstehen, daß gerade in der Ethik die Kantische These von der Spontaneität des Subjekts eine Art Bestätigung erfährt. Das “Sittengesetz” drückt eine Forderung aus im Gegensatz zu den realen Verhältnissen des Menschenlebens. Es liegt also durchaus Konsequenz darin, wenn die Kritik der praktischen Vernunft behauptet, das Subjekt gebe hier das Gesetz; in ihm und nicht im Objekt liege der

³³ Hartmann: Ethik, 526-527. pp.

³⁴ Hartmann: Ethik, 63. p.

Bestimmungsgrund.”³⁵

For the focus of our analysis – the critique of the narrowing of morality – the most important objections of Hartmann against Kantian ethics can be found where he outlines the colourful variety and the diverse moral values and virtues in Plato, Aristotle and the Stoic ethics, and he claims that Kant narrowed the moral world and in place of the external variety of moral values and virtues put the one-dimensional creations of consciousness: “Denn Kant ist es, der an Stelle der inhaltlichen Mannigfaltigkeit der Tugenden die Einheit eines Sittengesetzes, an Stelle der materialen Erfülltheit das formale Prinzip, an Stelle des objektiven Wesens sittlicher Ideen die subjektive Gesetzgebung gesetzt hat.”³⁶

Hartmann often repeated his assertion that there are always some opposed moral values and virtues which in the same situation for an act appear as binding and hence the actor must always weigh between them. This way, the mere knowledge of the abstract moral values cannot be of assistance, and only the knowledge of the relations and hierarchy among the moral values and virtues in each situation can point towards the right decision. The norms of public morality contain this knowledge and that is the reason why they are important: “Alle Werterkenntnis bleibt abstrakt ohne Erkenntnis der Wertbeziehungen. Sind doch in allen ethischen Situationen mannigfache Werte zugleich beteiligt, und liegt doch für den Menschen, der vor Situation gestellt ist, die Aufgabe eben darin, sein Verhalten aus dem Situationsbewußtsein heraus einzurichten, welches Wert gegen Wert abwägt.”³⁷

Before finishing off the analysis of Hartmann’s ethics, it is still worthwhile to quote his opinion against the one-dimensional construction of the moral world which claims that the moral virtue can only be realized as a synthesis of several values: “Gesucht ist gerade ein Einheitsethos von Reinheit und Fülle, von Gerechtigkeit und Nächstenliebe, von Stolz und Demut. Erst ein solches würde in einem höheren und strengeren Sinne “Tugend” heißen dürfen, während die einseitigen Werte nur uneigentlich den Namen verdienen.” (Hartmann 1962: 574.) In an other version this thought can be read: “Jeder Wert hat – wenn einmal Macht gewonnen hat über eine Person – die Tendenz, sich zum alleinigen Tyrannen des ganzen menschlichen Ethos aufzuwerfen, und zwar auf Kosten anderer Werten, auch solcher, die ihm nicht material entgegengesetzt sind [...] So gibt es einen Fanatismus der Gerechtigkeit (fiat justitia perat mundus), der keineswegs bloß der Liebe, geschweige denn bloß der Nächstenliebe, ins Gesicht schlägt, sondern schlechterdings allen höheren Werten.”³⁸

IV. 2. Sociological investigation of morality

Before the transition to the description of deontological moral theory, let us quickly observe the results of the empirical sociology of morality, which can give an empirical answer to the question whether people in everyday life follow the rules of public morality – as the theories of Hegel, Jhering and Hartmann claimed. To answer this, the results of Lawrence Kohlberg’s investigations are indeed helpful.

These results show the following picture in relation to the development of the moral consciousness of children.³⁹ The moral development of children has different stages and at

³⁵ Hartmann: Ethik, 100. p.

³⁶ Hartmann: Ethik, 137. p.

³⁷ Hartmann: Ethik, 271. p.

³⁸ Hartmann: Ethik, 576. p.

³⁹ See: Kohlberg, Lawrence: From is to ought. How to commit the naturalistic fallacy and get away with it in the study of moral development. In: Th. Mischel (ed.): Cognitive development and epistemology. New York, Academic Press, 1971, 151-235. pp.; Lind, Georg: Entwicklung des moralischen Urteilens – Leistungen und Problemen der Theorien von Piaget und Kohlberg. In: Lind/Hartmann/Wakenhut (hg.): Moralisches Urteilen und soziale Umwelt. Theoretische, methodologische Untersuchungen. Weinheim und Basel, Beltz Verlag Weinheim, 1983, 25-40. pp.

first it is characterized by the mere avoidance of punishment and a desire for reward from the external authority. Later this attitude changes, and the roles of “the good boy”, “the nice girl,” and subsequently that of “the brave father” and “the honest woman” etc. appear and become socialized by the boys and girls. These are already more abstract standards, and are applied to the situation in a more autonomous fashion. This is the stage of conventional morality and Kohlberg assumed that there is a post-conventional phase (comprising two inner levels) where the attention of the external public moral standards decline and the individuals always decide on the basis of universal principles. Tests were made in different countries – the USA, Turkey, Mexico and Thailand – and the results invariably showed the dominance of the conventional public morality. Only on the lower level of the two-stage post-conventional stage did the empirical studies find 7 percent of people who had a critical stance in relation to public morality, but almost no one could be found in the last stage.⁴⁰ Even in connection with the 7 percent there was a criticism that the 7 percent could be found solely on the lower level of the post-conventional phase, because the highly educated individuals and the inhabitants of big cities were represented in the survey, too, and usually this number is even lower.

From the results of Kohlberg, it is thus clear that public morality exists in contemporary societies and the huge majority of people follow the norms of public morality and in everyday life everybody is judged by its environment, whether it complies with those standards or not.

IV. 3. The elimination of public morality: the critical morality

IV. 3.1. Kant's moral philosophy

The most important characteristic of Kantian moral theory is that it limits the moral aspect of action to the inner mind, and the moral assessment of an agent depends on his consciousness; on whether the agent's action took place with respect of the obligation. If he was driven by any other motivation – such as the fear of legal sanction – besides a respect of the obligation, then this action cannot be judged as moral good. Morals in Kantian theory have only a formal framework for the moral guidance in various situations; this is the categorical imperative. Kant wrote: “Der kategorische Imperativ ist also ein einziger, und zwar dieser: handle nur nach derjenigen Maxime, durch die zugleich wollen kannst, dass, sie ein allgemeines Gesetz werde.”⁴¹ Thus, morality in Kantian theory is an aspect of individual choice with such a formal framework that any chosen action has to have a universal applicability. In addition to the universal applicability, the second trait of moral decisions is autonomy, which means that they are free from material motivations and any particular external expectations and they are totally motivated by respect for moral duty. The autonomous individual disregards social ties and connections in Kantian morality and, therefore, it is typical in the contemporary followers of Kantian moral theory that they disregard the social ties and connections already in the starting point and they regard them only as incidental circumstances.

Kant rejected all habitual pursuit of morality contrary to Hegel. While the latter stressed the slow development of second nature (morality) by habit in the minds of individuals, Kant opined that if a habit determined an action (assuetudo), then there was no

⁴⁰ See: Rest, J. – Narvaez, D. – Bebeau, M.J. – Thoma, S. J. (eds.): *Postconventional Moral Thinking: A Neo-Kohlbergian Approach*. Mahwah, Lawrence Erlbaum Associate Publishers, 1999, 22. p.

⁴¹ Kant, Immanuel: *Grundlegung zur Metaphysik der Sitten*. Herausgeben und erläutert J. H. von Kirkmann, Berlin, Verlag von L. Heimann. [1797] 1870, 44. p.

freedom and one could not talk about morality.⁴² For Kant, morals were not only placed within the consciousness, but they could only be achieved with conscious decisions. Kant has taken over Rousseau's idea of freedom as self-determination, but while this idea was formulated by Rousseau in terms of self-determination of people as a political theory, Kant made it the basis of his moral theory.

Let us look at the analysis of Ernst Tugendhat, which stressed the difference between the theories of Kant and Hegel very sharply. Tugendhat claims that in Kantian theory this difference shows the waiver of the widespread moral standards of the community of actors, while in Hegel's theory it points to the emphasis on the binding moral standards of the community about the actions of community members. As a follower of Kantian theory, Tugendhat claims that Hegel's position resulted from a translation error that was caused by the two different views of the Greek word "ethos": "Außerdem ist hier ein Übersetzungsirrtum unterlaufen. In der aristotelischen Ethik kommt nämlich nicht nur das Wort *éthos* (mit langem "é"), das Charaktereigenschaft bedeutete, sondern auch das Wort "ethos" (mit kurzem "e") vor, das Gewohnheit heißt, und es ist dieses zweite Wort, worauf die lateinische Übersetzung paßte. [...] Von daher ist die merkwürdige deutsche Übersetzung durch "Sitte" zu verstehen, wie wir sie z. B. in Kants Buchtitel „Metaphysik der Sitten" finden. Kant hat dabei überhaupt nicht an Sitten im gewöhnlichen Sinn (Brauchtum) gedacht, sondern verwendete das Wort einfach als Übersetzung für "mores", das seinerseits nicht mehr in seinem ursprünglichen Sinn verstanden wurde, sondern als angebliche Übersetzung eines griechischen Wortes. Erst Hegel macht sich dann den ursprünglichen Sinn des Wortes „Sitten" zunutze, um gegenüber der Kantischen Moral eine angeblich höhere Form von Moral, genannt Sittlichkeit, zu konstruieren, die dadurch charakterisiert sein sollte, daß sie im Brauchtum und im Hergebrachten fundiert sei." (Tugendhat 1994: 34-35.)⁴³ It follows that the common standards of conventional morality cannot be accepted as true morality.

IV. 3. 2. The way of narrowing morality to justice

The suppression of the public (common) morality of communities and their replacement by critical morality, which was (and is) the construction of moral philosophers, posed only a minor problem as long as these constructions preserved the multiplicity of moral values and virtues, and they only supplied a different emphasis for the hierarchy of values. The problem has become more serious ever since the 1970s, because the influential moral theories narrowed morality to a single moral value, and the demands of morality were fatally distorted in consequence. Instead of improving the reflexivity and the quality of moral judgments, critical morality exerts a tyranny over the people in the name of the supplementary morality of the moral philosophers. Let us take a closer look at this problem.

The narrowing of morality to justice by John Rawls received the greatest attention and he was widely supported by the dissemination of cultural and ideological sectors from the beginning of the 1970s. However, some critical reaction has also appeared, as, for instance, that of Caroll Gilligan, who pointed out that Justice Rawls's depiction of morality is only the morality of men. Gilligan consequently claimed that women have their own morality, which can be formulated as "care-ethics".⁴⁴ Michael Slote has recently drafted a similar version, which he named "benevolence morality".⁴⁵ But these versions of moral theories received no attention in the broader public policy discussions, and are mainly respected by the theorists of

⁴² Kant: Grunlegung zur Metaphysik der Sitten, 152. p.

⁴³ Tugendhat, Ernst: Vorlesungen zur Ehtik, Frankfurt am Main, Suhrkamp, 1994, 34-35. pp.

⁴⁴ See: Gilligan, Caroll: In Different Voice, Harvard Univ. Press, 1982.

⁴⁵ See: Slote, Michael: Moral from Motives, Oxford Univ. Press, 2001.

social work, and, when it comes to Gilligan's thoughts, by feminists. A more general criticism of the narrowing of morality can be found in the works of Michael Sandel and Charles Taylor, who obtained a great support in the scientific community. Sandel wrote: "What is at stake in the debate between Rawlsian liberalism and the view I advance in my work is [...] whether the principles of justice that govern the basic structure of society can be neutral with respect to the competing moral and religious convictions its citizens espouse. The fundamental question, in other words, is whether the right is prior to the good."⁴⁶

In the work of Charles Taylor, the critique of narrowing morality to justice by Rawls and his follower can be found in a clearer fashion: "Much contemporary moral philosophy, particularly but not only in the English-speaking world, has given such a narrow focus to morality that some crucial connections I want to draw here are incomprehensible in its terms. This moral philosophy tended to focus on what it is right to do rather than on what it is good to be, defining the content of obligation rather than the nature of the good life; and it has no conceptual place left for a notion of the good as the object of our love or allegiance or as *Iris Murdoch* portrayed it in her work as the privileged focus of attention or will. This philosophy has accredited a cramped and truncated view of morality in a narrow sense, as well as of the whole range of issues involved in the attempt to live the best possible life."⁴⁷

It is worth to cite of the analysis of Beauchamp and Childress too, albeit they did not have greater impact in the community of moral philosophers but their analyses are the closest to our analysis: "Inductionism (the bottom-up approach) maintains that we must use existing social agreements and practices as a starting point from which to generalize to norms such as principles and rules, and inductivists emphasize the role of particular and contextual judgments as a part of our evaluating moral life. A society's moral views are not justified by an ahistorical examination of the logic of moral discourse or by some theory of rationality (Kantian, Rawlsian, Kohlbergian), but rather by an embedded moral tradition and a set of procedures that permit new developments [...]. The institution of morality cannot be separated from a cultural matrix of beliefs that has grown up and been tested over time."⁴⁸

Let us now first look at the narrowing of morality by John Rawls, and then by Jürgen Habermas, who completed Rawls's narrowing. This narrowing was implicitly included in the Kantian ethics, which conceived of the morality of actions as the following of the only command (the categorical imperative). This has already been emphasised in connection with Nicolai Hartmann: "Denn Kant ist es, der an Stelle der inhaltlichen Mannigfaltigkeit der Tugenden die Einheit eines Sittengesetzes, an Stelle der materialen Erfüllung die formale Prinzip, an Stelle des objektiven Wesens sittlicher Ideen die subjektive Gesetzgebung gesetzt hat."⁴⁹

But most moral theorists who followed the Kantian line, more or less preserved the multiplicity of moral virtues and values for a long time to come. The beginning of the narrowing may be located in John Rawls's "A Theory of Justice." He founded his theory on the studies of McDougall, James Mill, Freud, Piaget, and Lawrence Kohlberg in particular. As a starting point, he took the three-phase theory of Kohlberg (pre-conventional, conventional and post-conventional stages of morality, each with two sub-levels). As we have seen, Kohlberg designed the three phases of moral development on the basis of empirical investigation and although he knew even before the start of his empirical study that most people can be characterized by the second (conventional) stage, he assumed that there was a

⁴⁶ Sandel, Michael: *Liberalism and the Limits of Justice*. Second Edition. Cambridge University Press, [1982] 1998, p. x.

⁴⁷ Taylor, Charles: *Sources of Self. The Making of the Modern Identity*, Cambridge University Press, 1989, 3. p.

⁴⁸ Beauchamp, T. L. – Tom, L. – James, F. – Childress, F.: *Principles of Biomedical Ethics*, 4. Edition, Oxford Univ. Press. 1994, 18. p.

⁴⁹ Hartmann: *Ethik*, 137. p.

third phase where moral choices are made according to universal moral principles and human rights. Rawls somewhat transformed this last phase by limiting moral principles to the principles of justice: “In conjecturing how this morality of principles might come about (principles here meaning first principles such as those considered in the original position), we should note that morality of associations quite naturally lead up to a knowledge of the standards of justice.”⁵⁰

The empirical study of Kohlberg was important for Rawls, because he put great emphasis on the fact that a moral theory can be described as authentic only on an empirical basis. Subsequently, however, based on the results of Kohlberg’s empirical studies, it turned out to be obvious that the third (post-conventional) phase of morality was designed only on the basis of Kohlberg’s ideological faith. Indeed, not only are there merely 7 percent of people who could be classified at the lower level of the post-conventional morality, but the last (highest) stage remains empty: “Kohlberg eliminated Stage 6 from his scoring system for lack of finding empirical cases of Stage 6 thinking. Furthermore, there is little evidence for Stage 5 scoring in Kohlberg’s studies from around the world. Gibbs (1979) – a co-developer of the scoring system – even proposed that true Piagetian stages of moral judgment stop with Stage 4. The lack of empirical data for Stage 5 and 6 – post-conventional thinking – is a serious problem for Kohlberg’s enterprise, because he defined the stages from the perspective of the higher stages.”⁵¹ It is important to emphasise that Rawls identified his highest stage of morality (the morality of principles) exactly with this non-existent last stage of Kohlberg. Despite this problem, Rawls later made no correction and his followers likewise disregarded this problem.

The prolongation of this narrowing can be observed in the moral theory of Jürgen Habermas, who, in his books in the 1980s, following the example of Rawls, differentiated between the lower stage of morality, namely “Sittlichkeit” (public morality), and the higher stage, which is universal morality: “Aus der Perspektive eines Teilnehmers an moralischen Argumentationen stellt sich die auf Distanz gebrachte Lebenswelt, wo kulturelle Selbstverständlichkeiten moralischer, kognitiver und expressiver Herkunft miteinander verwoben sind, als Sphäre der Sittlichkeit dar. Dort sind die Pflichten derart mit konkreten Lebensgewohnheiten vernetzt, dass sie ihre Evidenz aus Hintergrundgewissheiten beziehen können. Fragen der Gerechtigkeit stellen sich vor nur innerhalb des Horizonts von *immer schon beantworteten Fragen* des guten Lebens. Unter dem unnachsichtig moralisierenden Blick des Diskursteilnehmers hat diese Totalität ihre naturwüchsige Geltung eingebüßt, ist die normative Kraft des Faktischen erlahmt – können sich vertraute Institutionen in ebenso viele Fälle problematischer Gerechtigkeit verwandeln. Vor diesem Blick ist der überlieferte Bestand an Normen zerfallen, und zwar in das, was aus Prinzipien gerechtfertigt werden kann, und in das, was nur noch faktisch gilt. Die lebensweltliche Fusion von Gültigkeit und sozialer Geltung hat sich aufgelöst.”⁵²

In the final version of his moral theory (“Faktizität und Geltung” in 1992) he radicalized his position and “Sittlichkeit” (public morality) appeared here only as an archaic substrate, which in the modern era has already fully and completely vanished. According to this final version, there are only two systems for the determination of actions, that of law and universal morality: “Ich gehe davon aus, dass sich auf nachmetaphysischen Begründungsniveau rechtliche und moralische Regeln gleichzeitig aus traditioneller Sittlichkeit ausdifferenzierte und zwei verschiedene aber einander ergänzende Sorten von

⁵⁰ Rawls, John: A Theory of Justice, Oxford University Press, 1971, 414. p.

⁵¹ J. Rest, J. – Narvaez, D. – Bebeau, M. J. – Thoma, S. J. (eds.): Postconventional Moral Thinking: A Neo-Kohlbergian Approach, 22. p.

⁵² Habermas, Jürgen, Moralbewusstsein und kommunikatives Handeln. Frankfurt am Main, Suhrkamp, 1983, 117-118. pp.

Handlungsnormen nebeneinander treten.”⁵³

He qualifies the rest of public morality in the modern era as mere conventions (“von den blossen Konventionen entwerteten Sitten”).⁵⁴ A second modification of his theory entailed that universal morality does not exist anymore as moral norms but as cultural knowledge only, and for the determination of actions there are only the norms of law. He speaks of morality as cultural knowledge (“diese zum Wissen sublimiert Moral”); as morality transformed into cultural system (“ins kulturelle System zurückgezogene Moral”).⁵⁵ The end result of Habermas’s moral theory is that there remains no system of moral standards on the community level (only the law) and on the level of the worldwide Republic of universal citizens there is only a cultural morality. As it has been seen earlier, this cultural morality is then narrowed to justice: “Die Vernunftmoral ist auf Fragen der Gerechtigkeit spezialisiert und betrachtet grundsätzlich alles im scharfen, aber engen Lichtkegel der Universalisierbarkeit.”⁵⁶

The narrowing of morality is finally completed here by the destruction of the whole of morality and it is the rule of law that solely remains, supported by state authority. It is not without irony that this theory translated into the political and ideological struggles is thought to be the conquest of the world’s greatest freedom and tolerance by its followers. The truth is, however, that millions of people today, knowing nothing about Rawls and Habermas and their followers, socialise the multiplicity of moral values, norms and virtues, and are subsequently confronted on a daily basis with legal standards that are far from their moral standards and had been formed in accordance with the cultural moral theory of moral philosophers.

References

- Abney, Keith (2011): Robotics, Ethical Theory and Metaethics: A Guide for the Perplexed. In Patrick Lin – Keith Abney – George A. Bekey (eds.): Robotethics. Cambridge, Massachusetts – London, The MIT Press, 2011, pp. 35-54.
- Collin, Allen – Wallach, Wendell (2011): Moral Machines: Contradiction in Term or Abdication of Human Responsibility? In Patrick Lin – Keith Abney – George A. Bekey (eds.): Robotethics, Cambridge Massachusetts – London, The MIT Press, 2011, pp. 55-68.
- Baecker, Dirk (2004): Einleitung: Wozu Gefühle? Soziale Systeme (Vol. 10) 2004, Heft 1., pp. 5-20.
- Beauchamp, Tom L. – Childress, James F. (1994): Principles of Biomedical Ethics. (4. Edition) Oxford Univ. Press, 1994.
- Bostrom, Nick (2014): Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 2014.
- Chalmers, David (2014): Uploading: A Philosophical Analysis. In: Russel Blackford – Damien Broderick (eds.): Intelligence Unbound: The Future of Uploaded and Machine Minds, Wiley Blackwell, Malden-Oxford, 2014, pp. 102-118.
- Ciampi, Luc (1997): Zu den affektiven Grundlagen des Denkens. System Familie (Vol. 10) 1997, No. 1., pp. 128-134.

⁵³ Habermas, Jürgen: Faktizität und Geltung. Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats. Frankfurt am Main, Suhrkamp, 1992, 135. p.

⁵⁴ Habermas: Faktizität und Geltung. Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats, 137. p.

⁵⁵ Habermas: Faktizität und Geltung. Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats, 134-5. pp.

⁵⁶ Habermas, Faktizität und Geltung. Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats, p. 137.

- Ciampi, Luc (2004): Ein blinder Fleck bei Niklas Luhmann? Soziodynamische Wirkungen von Emotionen nach dem Konzept der fraktalen Affektlogik. Soziale Systeme (Vol. 10.) 2004, Heft 1., pp. 21-49.
- DiGiovanna, James (2017): Artificial Identity. In Patrick Lin – Ryan Jenkins – Keith Abney (eds.): Robot Ethics 2.0, New York, Oxford University Press, 2017, pp. 307-321.
- Doherty, Jason P. (ed.) (2016): AI Civil Rights: Addressing Artificial Intelligence and Robot Rights. Kindle Edition, 2016.
- Elias, Norbert: Über den Prozeß der Zivilisation: Soziogenetische und psychogenetische Untersuchungen. Suhrkamp, 1976.
- Ford, Martin (2016): The Rise of Robots: Technology and the Threat of a Jobless Future, Basic Books, 2016.
- Fukuyama, Francis (2002): Our Posthuman Future: Consequences of the Biotechnology Revolution. Picador Edition, 2002.
- Galliot, Jai (2017): The Unabomber on Robots. In Patrick Lin – Ryan Jenkins – Keith Abney (eds.): Robot Ethics 2.0, New York, Oxford University Press, 2017, pp. 369-385.
- Gehlen, Arnold (1974): Der Mensch: seine Natur und seine Stellung in der Welt. Athenaeon Verlag, 1974.
- Gilligan, Carol (1982): In Different Voice. Harvard Univ. Press, 1982.
- Goertzel, Ben, Joel Pitt (2014): Nine Ways to Bias Open-Source Artificial General Intelligence Toward Friendliness. In: Russel Blackford – Damien Broderick (eds.): Intelligence Unbound: The Future of Uploaded and Machine Minds. Malden – Oxford, Wiley Blackwell, 2014, pp. 90-101.
- Goonan, Kathleen Ann (2014): The Future of Identity: Implications, Challenges, and Complications of Human/Machine Consciousness. In: Russel Blackford – Damien Broderick (eds.): Intelligence Unbound: The Future of Uploaded and Machine Minds. Malden – Oxford, Wiley Blackwell, 2014, pp. 193-200.
- Habermas, Jürgen (1983): Moralbewusstsein und kommunikatives Handeln. Suhrkamp. Frankfurt am Main, 1983.
- Habermas, Jürgen (1991): Erläuterungen zur Diskursethik. Frankfurt am Main, Suhrkamp, 1991.
- Habermas, Jürgen (1992): Faktizität und Geltung. Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats. Frankfurt am Main, Suhrkamp, 1992.
- Habermas, Jürgen (2000): Die Zukunft der menschlichen Natur. Auf dem Weg zu einer liberalen Eugenik? Frankfurt am Main, Suhrkamp, 2000.
- Hartmann, Nicolai (1940): Der Aufbau der realen Welt. Berlin, Walter de Gruyter, 1940.
- Hartmann, Nicolai: Das Problem des geistigen Seins. Untersuchungen zur Grundlegung der Geschichtsphilosophie und der Geisteswissenschaften. 3. unveränderte Auflage, Berlin, Walter de Gruyter, 1962.
- Hegel, Georg Wilhelm Friedrich (1979): Grundlinie der Philosophie des Rechts. Werke Band 7. Frankfurt am Main, Suhrkamp, 1979.
- Henschke, Adam (2017): The Internet of Things and Dual Layers of Ethical Concern. In Patrick Lin – Ryan Jenkins – Keith Abney (eds.): Robot Ethics 2.0. New York, Oxford University Press, 2017, pp. 229-243.
- Jhering, Rudolph (1898): Der Zweck im Recht. Zweiter Band. Druck und Verlag von Breitkopf und Härtel. Leipzig, dritte durchgesehene Auflage, 1898.
- Kaku, Michio (2014), The Future of Mind? The Scientific Quest to Understand, Enhance and Empower the Mind, Anchor Edition, 2014.
- Kant, Immanuel (1870): Grundlegung zur Metaphysik der Sitten. Herausgegeben und erläutert J. H. von Kirkmann. Berlin, Verlag von L. Heimann, 1870.
- Kelly, Kevin (2016): The Inevitable. The 12 Technological Forces that Shape Our Future.

Kindle Edition, e-book. 2016.

Klinewicz, Michal (2017): Challenges to Engineering Moral Reasoners. In Patrick Lin – Ryan Jenkins – Keith Abney (eds.): Robot Ethics 2.0. New York, Oxford University Press, 2017, pp. 244-257.

Kohlberg, Lawrence (1971): From is to ought. How to commit the naturalistic fallacy and get away with it in the study of moral development. In: Mischel, Th. (ed.): Cognitive development and epistemology. New York, Academic Press, 1971, pp. 151-235.

Kurzweil, Ray (2005): The Singularity Is Near: When Humans Transcend Biology. London, Penguin Books, 2005.

Kurzweil, Ray (2012): How to Create a Mind. The Secret of Human Thought Revealed. London, Viking Penguin Edition, 2012.

Lind, Georg (1983): Entwicklung des moralischen Urteilens – Leistungen und Problemen der Theorien von Piaget und Kohlberg. In: Lind/Hartmann/Wakenhut (hg.): Moralisches Urteilen und soziale Umwelt. Theoretische, methodologische Untersuchungen. Weinheim und Basel, Beltz Verlag, 1983, pp. 25-40.

Loh, Wulf, Janina Loh (2017): Autonomy and Responsibility in Hybrid System. In Patrick Lin – Ryan Jenkins – Keith Abney (eds.): Robot Ethics 2.0, New York, Oxford University Press, 2017, pp. 35-50.

Luhmann, Niklas (1975): Evolution und Geschichte. In. (idem): Soziologische Aufklärung 2. 1975, 150-170.p.

Luhman, Niklas (1981): Soziologische Aufklärung 3. Westdeutscher Verlag, 1981, 231. p.

Luhmann, Niklas (1984): Soziale Systeme. Frankfurt am Main, Suhrkamp, 1984.

Luhmann, Niklas (1994): Liebe als Passion: Zur Codierung von Intimität, Frankfurt am Main, Suhrkamp, 1994.

Luhmann, Niklas (1995): Soziologische Aufklärung 6. Die Soziologie und der Mensch. Opladen, Westdeutscher Verlag, 1995.

Matura, Humberto R. – Franciso J. Varela: The Tree of Knowlegde. The Biological Roots of Human Understanding. Revised Edition. Boston, Massachusets, Shambhala Publications, 1987.

Maturana, Humberto R. – Sima Nisis de Rezepka: Human Awareness: Understanding the Biological Basis of Knowledge and Love in Education. In Jane Cull (ed.): Life's Natural Solutions, 2014, pp. 2-27.

Pigliucci, Massimo (2014): Mind Uploading: A Philosophical Counter-Analysis. In Russel Blackford – Damien Broderick (eds.): Intelligence Unbound: The Future of Uploaded and Machine Minds. Malden – Oxford, Wiley Blackwell, 2014, pp. 119-130.

Pokol, Béla (2013): Theoretische Soziologie und Rechtstheorie. Schenk Verlag, 2013.

Pokol, Béla (2018): Künstliche Intelligenz: Die Entstehung einer neuen Seinsschicht? (KI – im Spiegel von Nicolai Hartmanns Ontologie) In Pázmány Law Working Papers, Nr. 2018/12.

Sloterdijk, Peter (1999): Regen für Menschenpark. Ein Antwortschreiben zu Heidegger Brief über den Humanismus, Frankfurt am Main, Suhrkamp, 1999.

Rest, J. – Narvaez, D. – Bebeau, M.J – Thoma, S.J. (ed.) (1999): Postconventional Moral Thinking: A Neo-Kohlbergian Approach. Mahwah, Lawrence Erlbaum Associate Publishers.

Scheler Max (2016): Die Stellung des Menschen im Kosmos. Verlag der Contumax GmbH. Berlin, 2016.

Slote, Michael (2001): Moral from Motives. Oxford Univ. Press, 2001.

Splengler, Oswald (2007): Der Untergang des Abendlandes. Albatros Verlag/Patmos Verlag, 2007.

Talbot, Brian – Ryan Jenkins – Duncan Purves (2017): When Robots Should Do the Wrong Thing. In Patrick Lin – Ryan Jenkins – Keith Abney (eds.): Robot Ethics 2.0. New York, Oxford University Press, 2017, pp. 258-273.

Vallor, Shannon – George A. Bekey (2017): Artificial Intelligence and the Ethics of Self-Learning Robots. In Patrick Lin, Ryan Jenkins, Keith Abney (eds.), *Robot Ethics 2.0*. New York, Oxford University Press, 2017, pp. 338-353.

Trevor, W. – Seth, N. – Baum, D. (2017): Liability for Present and Future Robotics Technology. In Patrick Lin – Ryan Jenkins – Keith Abney (eds.): *Robot Ethics 2.0*, New York, Oxford University Press, 2017, pp. 66-79.

Zoller, David (2017): Skilled Perception, Authenticity, and the Case Against Automation. In Patrick Lin – Ryan Jenkins – Keith Abney (eds.): *Robot Ethics 2.0*. New York, Oxford University Press, 2017, pp. 55-68.