METAZİHİN YAPAY ZEKA VE ZİHİN FELSEFESİ DERGİSİ
METAMIND: JOURNAL OF ARTIFICIAL INTELLIGENCE AND PHILOSOPHY OF MIND

ISSN: 2651-2963 Cilt: 1, Sayı: 2, Aralık 2018, 167-181
www.dergipark.gov.tr/metazihin Volume: 1, Issue: 2, December 2018, 167-181

Free Will and Artificial Intelligence

[Özgür İrade ve Yapay Zeka]

Aziz Fevzi ZAMBAK *
Middle East Technical University

Received: 10.12.2018 / **Accepted:** 26.12.2018
DOI: ……………………….
Research Article

Abstract: Free will is one of the main challenges against artificial intelligence. Free will is considered to be one of the unique characteristics of human being which cannot be represented in any artificial intelligent system. Free will is a special issue in the philosophy of mind. There are two main approaches to free will; namely, compatibilism and incompatibilism. We construct two main compatibilist arguments in order to overcome this challenge and these arguments show that the deterministic and computational structure of machine intelligence does not empirically prevent artificial intelligence from possessing free will. In addition to that, we claim that the agentive action is the only condition for the occurrence and analysis of free will. And these occurrence and analysis conditions can be modelled and simulated in machine intelligence. Therefore, AI can possess the tools through which it can realize its autonomous free choices. Simply stating, AI can have a free will.

Keywords: artificial intelligence, compatibilism, free will, determinism, computational decision.

M e t a Z i h i n Y a p a y Z e k a v e Z i h i n F e l s e f e s i D e r g i s i

1. Introduction

Attributing a machine the faculty of thought, intelligence and consciousness requires showing that machine intelligence can possess free will. But in order to do that artificial intelligence (hereafter, AI) should deal with the commonsensical idea that a machine [computer] is a purely mechanistic [computational] artefact which is essentially slavish

* **Author Info:** Assoc. Prof. – Middle East Technical University, Faculty of Arts and Sciences, Department of Philosophy, Üniversiteler Mah. Dumlupınar Blv. No: 1, Çankara-Ankara, TURKEY.
E-mail: zambak@metu.edu.tr / Orcid Id: https://orcid.org/0000-0003-2905-1218

To Cite This Paper: Zambak, A. (2018). "Free Will and Artificial Intelligence" *MetaZihin, 1*(2): 167-181.

and servile[1]; and that free will and a power of self-determination, regarded as a special faculty, only belongs to the human mind. In modern AI, there exist certain views which claim the possibility of constructing a machine that can have free actions. For instance, Sloman claims that being free and responsible can be understood in a computational model of decision-making. He states that "people are increasingly designing programs which, instead of blindly doing what they are told, build up representations of alternative possibilities and study them in some detail before choosing. This is just the first step towards real deliberation and freedom of choice." (1978: 266). To construct machines that have their own goals, criteria and principles seems to Sloman an achievable project. According to Sloman (1978: 267), a self-modifying program "could acquire not only new facts and new skills, but also new motivations; that is, desires, dislikes, principles, and so on." He mentions that if the self-developmental faculty of a machine is not accepted as the "free will of a machine," then what we mean by freedom would not be clear. Boden (1987: 432) is another philosopher who claims that the modern AI view can provide an analysis for the understanding of free will problem:

> A philosophical explication of freedom that does not deny it must clarify what sort of phenomenon it is and how it is possible. An acceptable analysis of freedom should illuminate whatever genuine psychological distinctions (and moral implications) are marked by the vocabulary of freedom, without falling into either a self-defeating indeterminism or an antimechanistic mysticism that ignores our psychological constitution and evolutionary origins. Such an explication is made more accessible by bearing computational concepts in mind, so that artificial intelligence has a role to play in helping us to understand what it is to be a human being.

Both Sloman and Boden defend the idea that machine intelligence representing all the functions and essentials of human cognition and performance must *in principle* be in the *possession* of free will. On the other hand, Fisher (1993: 79) repudiates such a principle and he claims that AI cannot have any philosophical contribution to the problem of free will.

We claim that AI and free will can co-exist. In other words, the collision between the concepts of free will and machine intelligence can be avoided. In AI, compatibilism must be the standpoint from where we can show the very possibility of the *possession* of free will in machine intelligence. The way we defend compatibilism determines the sense in which machine intelligence is said to have free will**.** Therefore, we propose two arguments in defence of compatibilism. In the first argument, called *methodological*

[1] Here, the terms "slavish" and "servile" refer to the idea that a machine only follows the orders. As MacKay (1951: 105) states, "The commonly-heard expostulation that 'a machine only does what you tell it to do' thus becomes virtually a tautology."

argument, we aim at showing that the compatibility of free will and determinism is an epistemological issue; and therefore, free will can, in principle, be the subject-matter of AI. The epistemological principles providing the compatibility of free will and determinism give the possible *conditions* of free will in AI. The *methodological argument* helps us to understand these epistemological principles. In the second argument, called *theoretical argument*, we aim at showing that the co-existence of free will and determinism can be interpreted in an *agentive manner* which indicates the action, behavioural, operational, and eventual based nature of free will.

2. The Methodological Argument

Premise 1: We live in a materialistic/physical world.

We live in a materialistic world and all physical events in the universe depend on natural laws. In other words, we do not accept the existence of any supra-physical (e.g. God) power that regulates and interferes to the order of physical world. In addition, our mind has a materialistic structure. Therefore, the free will problem must repudiate the dualistic approaches advocating for the distinction between the mind (i.e. soul) and the body. In that sense, all the epistemological and methodological assumptions of the free will problem should depend on materialistic theories.

Premise 2: At t1, the number of our free choices has a limited range.

An action occurs or happens in time and space. An event related to an action cannot be considered out of a spatio-temporal context (extent). An explanation of an event requires an indication of its spatial and temporal dimensions. Our lives are bound to physical and social conditions. Our choices are made as to the physical and social conditions of a certain place and time. That's why our choices have a limited range. Suppose that you are in a prison. Due to the physical conditions of prison, you are deprived of many activities such as visiting your friends, having a beer, driving a car, using a computer and etc. Therefore, a person, who lives in a prison, has limited choices due to the physical and social conditions of the prison. Although we have much more opportunities than a prisoner, the same principle also applies in our lives. The range of our choices is determined by physical and social circumstances.

Premise 3: All physical events are caused by prior events. [Neural events are caused by other neural events]

All physical events (e.g. neurophysiologic ones) are based on an *Order* in which natural phenomena occur under certain conditions. What we call this *Order* is "natural law." Newtonian mechanics postulates that there are necessities and regularities in the physical world and natural law is the description of these necessities and regulations. The idea of natural law originates from causality which has a deterministic structure. Bunge (1979: 22) gives the account of this postulation: "Every single fact is the locus of a set of laws." However, the relation between laws and regularities (or necessities) in a deterministic structure is not static and singular. A particular event is part of an *Order* which has a deterministic structure. Yet the deterministic structure is a complex, dynamic and asymmetrical system. Eigenhardt (1998: 119) describes the dynamic, complex and asymmetrical features of *Order* as "Deterministic chaos"[2] which means, for him, "bifurcation of order." Eigenhardt states that "The states of the system are deterministic, but in the subsequent course these states separate, split, diverge, and run in all directions, and so the relevant information concerning the initial state of order gets lost regarding to a calculation, that means regarding to an epistemic state not to the ontic state." Bohm and Peat (1987: 141) mention the dynamic character of deterministic structure in another way:

> The concept of order is, by itself, of very general interest. But one of its most fundamental and deepest meanings is that it lies at the root of structure, which is a key issue, not only in science, but life as a whole. Structure is often treated as being static and more or less complete in itself. But a much deeper question is that of how this structure originates and grows, how it is sustained, and how it finally dissolves. Structure is basically dynamic, and should perhaps better be called *structuring*, while relatively stable products of this process are *structures*. But even these latter structures should not be considered as basically static, for they are the results of processes which sustain them and keep them, for a time, more or less within certain limits.

Many discussions were held on the structure and *Order* of the physical world with the development of quantum physics and chaos theory. However, these discussions did not ontologically falsify the acceptance that all physical events are causally determined by prior events. Quantum physics or chaos theories show that it is very difficult to know all causal determinations of any physical event. For instance, the rise of cigarette smoke in the air is based on certain physical conditions. If we know all the physical conditions affecting the rise of smoke, then we can also know the way it will raise. However, since we cannot know all of these physical conditions, we cannot causally determine the way the smoke raises. However, it does not mean that the rise of the cigarette smoke has an ontologically indeterministic structure.

After 1980's, many studies have been done in order to show the relation between quantum mechanics and the mind. These studies aim at showing that mind is a

[2] The term "Deterministic chaos" is also used by Mark Stone (1989) in his article "Chaos, Prediction and Laplacean Determinism" in which he aims at overcoming the problematic relation between predictability and determinism.

quantum-mechanical phenomenon (See Beck and Eccles 1992, Eccles 1994, Herbert 1993, Hodgson 1991, Jibu and Yasue 1995, Lockwood 1989, Margenau 1984, Penrose 1989, 1994, Squires 1990, Stapp 1993, Wolf 1986, Zohar 1990). Many philosophers infer from these ideas that AI can never simulate brain states, for the quantum-mechanical structure of the mind cannot be represented in a computational manner.[3] The idea of the mind as a quantum-mechanical phenomenon gives way to an objection to AI which claims that free will is incompatible with determinism since it is the result of quantum-mechanical effects. In other words, quantum mechanics is seen as a physical phenomenon which has an indeterministic structure and cannot be explained by the classical (mechanistic) notion of *Order*. These indeterministic and inexplicable features are considered a ground for the incompatibilist ideas. In addition, chaos theory is also seen as a scientific ground for supporting the ideas of indeterminism and incompatibilism. In the incompatibilist approach, human action (e.g. free choice) is considered out of the predictability of physical laws due to the indeterministic property of quantum mechanics. Boden (1972: 333) describes this indeterministic position as follows:

> Recent neurophysiological research suggests that the response of individual nerve cells to stimulation is stochastic, or indeterminate, in that it requires statistical analysis in terms of probability of firing. Any meaningful statement of the relation between an individual neural response and a particular stimulus is thus impossible. This indeterminacy rests on spontaneous random activity at the synapses.

We have two main objections to the incompatibilist claim which depends on two theories, namely quantum mechanics and chaos theory. Firstly, chaos theory is not contradictory to determinism.[4] Chaos theory does not show the inadequacy of *Order*, but rather that of the methods we use in science. The deterministic structure of *Order* does imply any set of specific limitations. Chaotic systems are the result of a *range of huge variations* in a dynamic *Order*. Chaotic systems are sometimes inapplicable to the classical physics due to their very complex structures. However, these complex structures of chaotic systems can be considered as a different aspect of dynamic and asymmetrical features of the deterministic structure. In other words, a chaotic physical event is part of the *Order*. "Chaos theory is *the qualitative study of unstable aperiodic behavior in deterministic nonlinear dynamical systems*." (Kellert, 1993: 2). Therefore, there is not any chaotic physical event which can be considered out of natural laws. Chaotic

[3] The idea of "quantum computer" can be seen as a result of this inference. (For instance, see Lockwood, 1989: Chap. 14).

[4] Taylor and Dennett (2002: 271) argue that "In general, there is no paradox in the observation that certain phenomena are *determined* to be changeable, chaotic, and unpredictable, an obvious and important fact that philosophers have curiously ignored."

systems represent the dynamical and asymmetrical aspects of the deterministic structure. Here, the asymmetrical aspect of *Order* refers to the reinforcing influences of various subsystems which cause the chaotic behaviour of a system. In addition to that, the dynamic aspect of *Order* points out the difficulty of analyzing a chaotic event in a unique theoretical model. In that sense, chaos theory cannot be seen as a scientific ground for the indeterministic feature of free will.

Secondly, even if certain quantum mechanical events occur on a neurophysiologic level, it does not require an experiential and scientific understanding of the human mind as a quantum-mechanical phenomenon. Moreover, in cognitive science, quantum mechanics does not allow one to speak of mental processes as being random and indeterminate. The main reason is that quantum-mechanical events are micro-events which do not have any effect on macro-events occurring in mechanistic and deterministic structure. As Honderich (2002: 462) states, there are "a certain subclass of micro or atomic and subatomic events. They are quantum events of quantum theory. They, like all micro events, are far below the level of spoon movements and, more importantly, far below the neural events associated with consciousness and conscious choice or decisions in neuroscience." In addition to that Honderich (2002: 465) conceives brain states as a matter of only macro events which are parts of the dynamic and nonlinear deterministic structures. Therefore, quantum mechanics cannot be a reference point to discuss the free will problem. Hodgson (2002: 86) emphasizes the determinant role of macro events (i.e. deterministic structures) regarding the brain states from another point of view:

> […] the indeterminism suggested by QM [quantum mechanics] is mere randomness, which is hardly conductive to rational choice; and that in any event in systems hot, wet, and massive as neurons of the brain, quantum mechanical indeterminacies quickly cancel out, so that for all practical purposes determinism rules in the brain.

In principle, neural events are macro-events and parts of deterministic structures and they can always be understood in terms of causal interpretation.

Premise 4: If the 1st *Premise* is true, then it means that behaviours (and mental states) emerge[5] from the neuro-physiological events in the brain; and if the 3rd *Premise* is true, then these neuro-physiological events are caused by prior events and this causation is a part of the *Order* (i.e. deterministic structure).

---

[5] Here, we use the word "emerge" in the sense that "*Emergence* is the fundamental concept of a general theory which *explains* the way of the *transition to different degrees of order* in a hierarchy of complexity. It is the theory of the deep structure of complex dynamical systems and it looks for universal properties of these transitions." Eigenhardt (1998: 120).

Free choice includes thinking and behavioural activities. And neuro-physiological events are the source of behavioural activities and thoughts.[6] They cannot be considered out of the physical laws since they are physical entities. In other words, the mental and the neural are closely related. Free choice as a behavioural and mental activity is both essentially the result of the causation of mental states and the effect of the further causation of mental states.

Conclusion: If the 2nd and 4th *Premises* are true, then epistemologically and in principle, free choice can be explained.

Methodologically and in principle, it is possible to explain our thoughts and behaviours by physical laws. Of course, our capacity of knowledge is not sufficient for this explanation yet. But this does not mean that it is not possible to explain the neurobiological motives behind our thoughts and behaviours. Here, a question might be posed: "How can the explicability of free choice support the compatibilist argument which argues that free will and determinism can co-exist?" After these premises and conclusion, we have to redefine what we understand from free will. Firstly, we understand that free will is not unlimited (See *Premise* 1). Secondly, free will is a kind of *power* and *ability*[7] to choose between certain pre-determined possibilities which are the results of natural laws (See *Premise* 2 and 3). Thirdly, this *power* or *ability* depends on neurobiological activities of the brain (See *Premise* 4). Therefore, here, the most important point for the redefinition of the free will is to acknowledge that "*power* and *ability* to choose something" as a characteristic of free will refers to a different kind of *possibility* in the neurobiological activity of brain. In that sense, free will is a kind of *possibility* of pre-determined possibilities of the natural world. Methodologically, this means that free will is a special kind of *possibility* which can occur in the neurobiological and physical conditions. Once we consider the range of this *possibility*, a great number of choices for the variety of our thoughts and behaviours come up. But this *possibility* should be considered as a part of an *Order* (i.e. deterministic structure). It is not an incapacitation of free will.

---

[6] Of course, here, we do not imply brain states as a unique cause for the behavioural activities of human mind. However, we notice the interactional, agentive and circumstantial conditions in human behaviour and thinking.

[7] Berofsky (1987: 70) considers these *power* and *ability* as an indispensable component of deterministic structure: "Compatibilists have advanced analysis of power…designed to permit judgments that an agent had the power to perform an action he failed to perform in a sense which permits one regard him as morally responsible for the act he did perform, even if determinism is true. There are, to be sure, senses of 'power', 'can' and 'ability' which apply in a deterministic world."

If anything is explicable, then it has a determined and computational structure. The explicability of free choices shows us that free will is determined.[8] On the other hand, the characteristic of this determination includes a specific type of *possibility* which can be a ground for the redefinition of the free will. Here, the most important point is that this *possibility* does not imply any predictability, and this is the main point that gives us an opportunity to conceive free will and determinism as compatible. Any attempt to predict the neural events active in decision-making (i.e. free choice) with long term exactness would fail completely since it is not practically possible to take into account all neurological data and environmental conditions. In other words, the unpredictability of decision-making processes results from the complex and dynamic features of deterministic structure of the human mind. For instance, Hunt (1987: 132) claims that "if the state of a system at a particular time is known then its state at a later time can be predicted. This prediction can be made using the deterministic laws of classical mechanics." On the other hand, in the *methodological argument*, we claim that the deterministic principle put forward by Hunt is not applicable to human mind and free will problem.

The co-existence of free will and determinism must no longer include the idea of predictability. The human mind as a complex and dynamic (chaotic) deterministic structure does simply imply the impossibility of the prediction for certain neural events, but this does not require introducing new indeterministic quantum-mechanical laws to human action (i.e. decision-making process). Hodgson (2002: 87) conceives the relation between determinism and unpredictability in chaos and complexity theory in which "differences in initial conditions can produce great differences in outcomes." In addition to that Stone (1989: 128) claims that explanation is possible in science without predictability:

> The Scientific Determinism is motivated in large part by the faith that nature is thoroughly predictable. The search for predictability thus constrains what will count as a complete explanation. Yet where predictability as conceived of by the Scientific Determinism fails, we need not abandon the attempt to provide a scientific explanation. Instead, we must redirect our expectations about what counts as a complete scientific explanation. This is precisely what physicists studying chaos have done. To say that systems are unpredictable is not to say science cannot explain them.

As a result of the *methodological argument*, we claim that free will is a mental *power* and *ability* to make decisions in certain situations. This mental *power* and *ability* are the result

[8] The deterministic structure of free will does not require a computational access to the mind. Moreover, "determinism does not entail that all our actions are done out of ignorance, accident, and so forth." (Haji 2002: 205). However, this ignorance and accidental condition is not out of the *possibility* condition that we have defined above.

of the dynamic and complex property of the deterministic structure of the human mind. If an AI model constructs the *conditions* of/for mind, then it will be possible for a machine to have the same[9] kind of mental *power* and *ability* of the human mind which is the source of his/her free will. Therefore, the deterministic and computational structure of machine intelligence does not empirically prevent AI from *possessing* free will.

3. The Theoretical Argument

We will defend four ideas in the *theoretical argument*. Firstly, we argue that the origin of free choice is the agent himself. Free will does not have any *conceptual* value without agentive action. In other words, free will is not the outcome of processes in the brain alone; and agentive action gives us an opportunity to analyze external and causal reasons of free will. Free will is the "state of action" which is guided through *interactions* in the environment by agentive *performance*. The "state of action" itself cannot be the sufficient condition for the conception of free will. Taylor (1968: 228) claims that free will "is not merely a congeries or series of states or events," but rather an act which is performed by an agent.

Secondly, an agentive action as the origin of the free choice is *caused* by prior events. *Rationalized* action is the source of this *causation*. Therefore, the *rationalization* process of an agent can provide a causal analysis of free will. Davidson is one of the philosophers who believe in the possibility of such an analysis.[10] Davidson (2001: 72) states that:

> The only hope for the causal analysis is to find states or events which are causal conditions of intentional actions, but which are not themselves actions or events about which the question whether the agent can perform them intelligibly be raised. The most eligible such states or events are the beliefs and desires of an agent that *rationalize* an action, in the sense that their propositional expressions put the action, in a favourable light, provide an account of the reasons the agent had in acting, and allow us to reconstruct the intention with he acted.

The agentive action in its *rationalized* form is the explanatory source allowing us to understand the event in the agent's decision-making process. However, it is an

[9] Here, the word "same" does not mean identical. It refers to a specific similarity.

[10] Like Davidson, Berofsky (1987: 73) believes in the possibility of giving the causal analysis of free will. He states that "we can formulate a statement that we might call a causal analysis of freedom along the lines of this proposal: P is free to do *a* if there is a set of conditions *c* and law *l* according to which people under conditions *c* do *a* intentionally if they have attitudes that rationalize the doing of *a* and P is under *c*. The analysis enables us to say P, who is under *c* and who does *a* intentionally because he has attitudes that rationalize the doing of *a*, that he does *a* freely, and also enables us to say of Q, who does not do *a* because he does not have an attitude that would rationalize the doing of *a*, but is under condition *c*, that he is free to do *a*."

epistemic difficulty (but not impossibility) for AI to formulate the whole *rationalization* process of an agentive action since we cannot expect to find a strict *law* of the *rationalization* process. That is to say that we cannot specify the *law* of *causation* in free will. This is the fundamental difference between analyzing the mechanical order of an action and the agentive *rationalization* of an action. An action happened in a natural law is necessitated in a mechanical way; but an action happened in human behavioural condition is necessitated in a *rationalized* way. As Honderich (1988: 483) states, "*no action but the action he* [*agent*] *performed could have been performed*" and this performance is open to a *causal* analysis in the *rationalized*-based agency.

Thirdly, our free choices are determined by prior events. An act is to be performed in time and space (extension). The temporal and spatial extension of an event gives us the actual position and the time of an event gives us the actual process. The actual event can be modelled by looking at its position and process. The *causation* of free choice can be analyzed in a model. In this model, we define two positions; namely "current action" and "actual event."

Current Action	Actual Event
A: To get up from the bed	A1: To get up from the left side of the bed A2: To get up from the right side of the bed A3: To get up by jumping from the bed…
B: To go to bathroom	B1: To go to the bathroom by wearing slippers B2: To go to the bathroom on bare foot
C: To wash your face	C1: To wash your face with hot water C2: To wash your face with cold water C3: To wash your face with soap C4: To wash your face with shampoo…
D: To prepare breakfast	D1: To make coffee D2: To make tea D3: To take the cheese from the refrigerator…
E: To get dressed	E1: To wear a red T-shirt E2: To wear a blue T-shirt E3: To put on jeans E4: To put on trousers
F: To leave the apartment	F1: To use the elevator F2: To use the stairs
G: To go to the office	G1: To travel by car G2: To travel by train G3: To travel by bus

In this example, A, B, C, D, E, F, and G are current actions. For the current action A, there are different possibilities to carry out the A (i.e. to get up from the bed) in various ways. In other words, there are many possible actual events (e.g. to get up from the left side of the bed) in the current action A. However, the possibility for each actual event is limited and pre-determined in each current action. Therefore, our free choices are the *rationalized* result of limits and pre-determinations of actual events. In other words, our free will is the *rational* choice of actual events in a limited and pre-determined condition. For instance, for the action F (i.e. to leave the apartment), we have limited conditions such as using elevator or using the stairs. Of course, jumping from the window can be another possibility in order to leave the apartment, but if you are living in the 14th floor, then your free choice of leaving the apartment must be situated in the actual events of

F1 and F2 which are the *rational* choices. Therefore, in principle, it is possible to give the *causal* analysis of your actual events of F1 and F2 in current action F.

Suppose that Tom carries out the following acts from T1 (8:00 A.M.) to T8 (9:00 A.M.): A2, B1, C3, D1, E2, E4, F1, G3. Here, Tom's actual events are modelled in a linear direction. While Tom's actual past between T1 and T8 includes his free choices, his current actions are determined. Therefore, his free choices are constrained by the possibilities of actual events. In that sense, a freely chosen actual event is *causally* determined by current actions. However, this determined form of free choice does not avoid free will. As a result, free will is the agentive *rationalized* action in the limited and determined conditions of actual events. In AI, it is possible to construct a *rationalization* model and give the analysis of *causation* of agentive actions. Moreover, AI can simulate these *causations* in machine intelligence.

Fourthly, the *causational* simulation and *rationalization* model of free will does not mean the ignorance of the autonomy of the human mind. A person lives in a *given* conditions and autonomy is the agentive and *rational* skill to survive his existence in this givenness. As Berofsky (1995: 57) states, "A person with intelligence, rationality, skills, and talents may or may not be lucky when he is thrust in to the world. But he at least possesses the tools he can utilize on behalf of the ends."

To sum up, in the *theoretical argument*, we have defended four ideas which indicate that agentive action is the only condition for the occurrence and analysis (in *causational* and *rationalized* form) of free will. And these occurrence and analysis conditions can be modelled and simulated in machine intelligence. Therefore, AI can possess the tools through which it can realize its autonomous free choices. Simply stating, AI can have a free will.

4. References

Beck, F. and Eccles, J. (1992). "Quantum Aspects of Brain Activity and the Role of Consciousness." *Proceedings of the National Academy of Science US*, *89*: 11357-61.

Berofsky, B. (1987). *Freedom from Necessity: The Metaphysical Basis of Responsibility*. New York: Routledge & Kegan Paul.

Berofsky, B. (1995). *Liberation from Self: A Theory of Personal Autonomy*. Cambridge: Cambridge University Press.

Boden, M. A. (1972). *Purposive Explanation in Psychology*. Cambridge, Massachusetts: Harvard University Press.

Boden, M. A. (1987). *Artificial Intelligence and Natural Man*. London: The MIT Press.

Bohm, D. and Peat, F. D. (1987). *Science, Order, and Creativity*. London: Routledge.

Bunge, M. (1979). *Causality and Modern Science*. New York: Dover.

Davidson, D. (2001). *Essays on Actions and Events*. Oxford: Clarendon Press.

Eccles, J. (1994). *How the Self Controls Its Brain*. Berlin: Springer.

Eigenhardt, P. L. (1998). "Chaos, Emergence and Complexity: On the Outset of Variety." In K. Mainzer, A. Müller and W. G. Saltzer (Eds.), *From Simplicity to Complexity: Information-Interaction-Emergence*, Part II. (pp. 117-126). Weisbaden: Vieweg.

Fisher, M. (1993). "A Note on Free Will and Artificial Intelligence." *Philosophia*, 13: 75-80.

Haji, I. (2002). "Compatibilist View of Freedom and Responsibility." In Robert Kane (Ed.), *The Oxford Handbook of Free Will*, (pp. 202-228). Oxford: Oxford University Press.

Herbert, N. (1993). *Elemental Mind: Human Consciousness and New Physics*. New York: Dutton, Penguin.

Hodgson, D. (1991). *The Mind Matters*. Oxford: Clarendon Press.

Hodgson, D. (2002). "Quantum Physics, Consciousness, and Free Will." In Robert Kane (Ed.), *The Oxford Handbook of Free Will*, (pp. 85-110). Oxford: Oxford University Press.

Honderich, T. (1988). *A Theory of Determinism: The Mind, Neuroscience, and Life-Hopes*. Oxford: Clarendon Press.

Honderich, T. (2002). "Determinism as True, Both Compatibilism and Incompatibilism as False, And the Real Problem." In Robert Kane (Ed.), *The Oxford Handbook of Free Will*, (pp. 461-476). Oxford: Oxford University Press.

Hunt, G. M. K. (1987). "Determinism, Predictability and Chaos" *Analysis*, *47*: 129-133.

Jibu, M. and Yasue, K. (1995). *Quantum Brain Dynamics and Consciousness: An Introduction*. Amsterdam: Benjamin.

Kellert, S. H. (1993). *In the Wake of Chaos: Unpredictable Order in Dynamic Systems*. Chicago: The University of Chicago Press.

Lockwood, M. (1989). *Mind, Brain and the Quantum: The Compound of "I."* Oxford: Basic Blackwell.

MacKay, D. M. (1951). "Mindlike Behaviour in Artefacts." *British Journal for the Philosophy of Science*, *2*: 105-121.

Margenau, H. (1984). *The Miracle of Existence*. Woodbridge, CT: OxBow.

Penrose, R. (1989). *The Emperor's New Mind*. Oxford: Oxford University Press.

Penrose, R. (1994). *Shadows of the Mind*. Oxford: Oxford University Press.

Sloman, A. (1978). *The Computer Revolution in Philosophy: Philosophy, Science and Models of Mind*. Sussex: The Harvester Press.

Squires, E. (1990). *Conscious Mind in the Physical World*. Bristol: Adam Hilger.

Stapp, H. (1993). *Mind, Matter and Quantum Mechanics*. New York: Springer.

Stone, M. A. (1989). "Chaos, Prediction and Laplacean Determinism." *American Philosophical Quarterly*, *26*: 123-131.

Taylor, C. and Dennett, D. (2002). "Who's Afraid of Determinism? Rethinking Causes and Possibilities." In Robert Kane (Ed.), *The Oxford Handbook of Free Will*, (pp. 257-277). Oxford: Oxford University Press.

Taylor, R. (1968). "Determinism and the Theory of Agency." In Sidney Hook (Ed.), *Determinism and Freedom in the Age of Modern Science*, (pp. 224-230). New York: The Macmillan Company.

Wolf, F. A. (1986). *Mind and the New Physics*. London: Heinemann.

Zohar, D. (1990). *The Quantum Self*. London: Bloomsbuty.

Öz: Özgür istenç yapay zekaya dair en temel karşı çıkış noktalardan birisidir. Özgür istenç insan olmanın ve insanı diğer canlılardan ayırt etmenin biricik unsuru olarak görülmektedir. Zihin felsefesi içerisinde özel bir konu olan özgür istenç problemine dair iki faklı yaklaşım bulunmaktadır. Bu yaklaşımlar bağdaşırcılık ve bağdaşmazcılıktır. Bu yazıda iki farklı bağdaşıklık uslamlaması üzerinden makinelerin belirlenimli ve berimsel yapılarının onların özgür istence sahip olabilmesi için bir engel çıkartmadığı gösterilmektedir. Buna ek olarak, eyleyici edimlerin özgür istencin çözümlenmesi ve ortaya çıkabilmesi için en temel koşul olduğu iddia edilmektedir. Bu çözümlemeler ve ortaya çıkışlar makine zekasında modellenebilinir. Bu modelleme aynı zamanda zihin felsefesinin bir konusu olarak özgür istenç problemine başka bir açıdan bakabilmemizi sağlayacaktır. Sonuç olarak, yapay zeka özgür kararlar alabilecek bir yapıya kavuşabilir.

Anahtar Kelimeler: yapay zeka, bağdaşırcılık, özgür istenç, belirlenimcilik, bilgisayımsal karar.

This Page Intentionally Left Blank