

THE KANTIAN NOTION OF FREEDOM AND AUTONOMY OF ARTIFICIAL AGENCY

LA NOCIÓN KANTIANA DE LIBERTAD Y AUTONOMÍA DE LA AGENCIA ARTIFICIAL

A NOÇÃO KANTIANA DE LIBERDADE E AUTONOMIA DA AGÊNCIA ARTIFICIAL

Manas Sahu

(Indian Institute of Technology Bombay)

manassahu23@gmail.com

Recibido: 22/06/2020

Aprobado: 04/04/2021

ABSTRACT

The objective of this paper is to provide a critical analysis of the Kantian notion of freedom (especially the problem of the third antinomy and its resolution in the critique of pure reason); its significance in the contemporary debate on free-will and determinism, and the possibility of autonomy of artificial agency in the Kantian paradigm of autonomy. Kant's resolution of the third antinomy by positing the ground in the noumenal self resolves the problem of antinomies; however, invites an explanatory gap between phenomenality and the noumenal self; even if he has successfully established the compatibility of natural causality and non-natural causality through his transcendental argument. This paper is also devoted to establishing the plausibility of the knowledge claim that Kantian reduction of phenomenality has served half of the purpose of the AI scientists on the possibility of Artificial Autonomous Agency.

Keywords: transcendentalism. phenomenal consciousness. artificial agency. naturalized rationality and autonomy.

RESUMEN

El objetivo de este trabajo es proporcionar un análisis crítico de la noción kantiana de libertad (especialmente el problema de la tercera antinomia y su resolución en la crítica de la razón pura); su importancia en el debate contemporáneo sobre el libre albedrío, el determinismo, y la posibilidad de autonomía de la agencia artificial en el paradigma kantiano de autonomía. La resolución de Kant de la tercera antinomia al colocar el fundamento en el yo nouménico resuelve el problema de las antinomias; sin embargo, invita a una brecha explicativa entre la fenomenalidad y el yo nouménico; incluso si ha establecido con éxito la compatibilidad de la causalidad natural y la causalidad no natural a través de su argumento trascendental. Este artículo también está dedicado a establecer la plausibilidad de la afirmación del conocimiento de que la reducción kantiana de la fenomenidad ha servido a la mitad del propósito de los científicos de IA sobre la posibilidad de una agencia autónoma artificial.

Palabras clave: trascendentalismo. conciencia fenomenológica. agencia artificial. racionalidad naturalizada y autonomía.

RESUMO

O objetivo deste artigo é fornecer uma análise crítica da noção kantiana de liberdade (especialmente o problema da terceira antinomia e sua resolução na crítica da razão pura); seu significado no debate contemporâneo sobre livre-arbítrio e determinismo, e a possibilidade de autonomia da agência artificial no paradigma kantiano de autonomia. A resolução de Kant da terceira antinomia postulando o fundamento no eu numênico resolve o problema das antinomias; no entanto, convida a uma lacuna explicativa entre a fenomenalidade e o eu numênico; mesmo que tenha estabelecido com sucesso a compatibilidade de causalidade natural e causalidade não natural por meio de seu argumento transcendental. Este artigo também se dedica a estabelecer a plausibilidade da alegação de conhecimento de que a redução kantiana da fenomenalidade atendeu à metade do propósito dos cientistas de IA sobre a possibilidade de Agência Autônoma Artificial.

Palavras-chave: transcendentalismo. consciência fenomenal. agência artificial. racionalidade naturalizada e autonomia.

Introduction

The overwhelming debate among A.I scientists, psychologists, neurologists, and philosophers on free-will and determinism have opened up a new perspective for the age-old problem of free-will and autonomy. The A.I Scientist claims that when the level of AI will be reached the level of A.S. I¹ (Artificial Super Intelligence), it can defeat human intelligence and have the autonomy to build the A.I machine itself without the intervention of human intelligence. The revolutionary advancement in the field of A.I has raised the possible threats of the dominance of Artificial Intelligence over human autonomy. The present paper is a critical inquiry of the Kantian paradigm of autonomy and evaluation of his paradigm, whether it is compatible with the knowledge claim of the autonomy of artificial agency.

We are going to deal with the following question in this paper: How does Kant has established the compatibility of natural causality with unconditional causality through abstraction? Whether the solution to the third antinomy of Kant can also be applicable to the contemporary debate on the possibility of the autonomous artificial agency? Can an artificial intelligent machine have autonomy? How the autonomy of A.I is different from human autonomy?

The first section of the paper delineates the conflict between natural causality and unconditioned causality (transcendental freedom), the Kantian solution to this antinomy, and its relevance in the current debate on the autonomy of artificial agency. The second section will focus on the possibility of artificial agency and autonomy, the conflict of phenomenality with transcendental idealism, the objection to autonomy of artificial intelligence- agency (hence after AAIA) from phenomenal realism and transcendentalism and in the concluding section, their limitation has shown.

1. Third Antinomy: Conflict of Natural Causality with Transcendental Freedom

The debate on “freedom versus determinism” is a traditional problem of philosophy. Every scientific theory is grounded on the laws of causality principles, which claim that whatever happens in the universe is based on the principle of causality. Every physical occurants in the universe (including human action) is grounded on the causality principle (cause-effect relationship); hence wholly determinable. The present state of something is caused/determined by the previous chain of events, which is based on

¹ The genealogical foundation for the argument of the possibility of artificial super intelligence can be traced back from the Alan Turing's (1950) highly cited paper “Computing Machinery and Intelligence” in which, he argued that an artificially constructed machine not only can *act like a human being rather think like a human being*. Bostrom (2014), Kurzweil (2006), and Copeland (1998) argued for the possibility of artificial super intelligence. They predict that there will be the *artificial singularity* in the future, the view that A.I agents will transcend the artificial general intelligence and have the ability to build to the similar level of A.I machines.

natural laws. Therefore, there cannot be any instances wherein the chain of causality principle can be ruled out.

On the contrary, if the deterministic view is correct, there cannot be anything in the universe that is not subject to the laws of physics. Hence, no human freedom is possible in the deterministic paradigm, which provide the foundation of moral philosophy². If the subject is not free to choose their action because their decision is predetermined by causal laws or causal chains of events, then the subject is not responsible for their actions. Because freedom is the precondition for moral action and moral responsibility. All moral principles are only applicable to those actions performed by the agent with the exercise of free-will. Therefore, determinism and freedom are incompatible with each other because both cannot be true together at the same time.

Immanuel Kant, one of the most influential philosophers of the 17th century, argues that although the conflict seems real to us, hence it needs attention to solve the issue, and either one must be true because both cannot be true or false at the same time, is, in fact, a pseudo-conflict. It is the consequence of the *Rhetoric use of reason*, which generates antinomies in Kant's language. An antinomy is a problem in which the debate between two superficial conflicting ideas seems real to the participant because both of their claims are seems to be supported by reason; however, in reality, is a pseudo conflict. It is the basic principle of an antinomy that the debate over two conflicting claims will be called antinomy in so far as either both to be true or false. If one became true or false, the conflict became real; hence nevertheless can be called antinomy. Kant, in his "Critique of Pure Reason," demonstrates that there are four pseudo conflicts that seem to us as a real problem due to *euthanasia of reason*. Our objective in this paper is to focus on the third antinomy of Kant which is popularly known as the conflict of freedom and determinism.

Thesis

“Causality in accordance with laws of nature is not the only one form which all the appearances of the world can be derived. It is also necessary to assume another causality through freedom in order to explain them” (A444/B472)³. The sensible world cannot be explained through causality only; transcendental freedom as a principle of causality is also necessary in order to explain the *world of appearances*. In other words, we have to accept that there are two kinds of causality for accommodating natural philosophy and non-natural philosophy. The natural causality and transcendental freedom as a causality. The natural causality governs the physical states of affairs of the phenomenal world, whereas the unconditional causality is responsible for moral action and the moral responsibility of the conscious individuals.

The argument for the thesis

Let us suppose that only the causality (the laws of nature) can explain the world of appearances. The principle of causality says that every occurrence must be sufficiently determined for being part of the causality. If the chain of causes is infinite, then a complete explanation of the chain of events or their causal account became impossible, which leads to the conclusion that the chains of the event as a totality cannot be established from a deterministic standpoint.⁴ As we know, in the laws of causation, every effect must have a predetermined (a-priorily determined) cause. Every event necessitated by a prior event, (i.e., x caused by w, w caused by v ... Infinite regress) without first beginning or origin. However,

² It is a noncontroversial claim for both the proponents of naturalized ethicists (Except Illusionist, reductive materialist) and non-naturalized ethicists to have a collective agreement that freedom is the necessary condition of morality. Illusionist argument for Illusion of free-will (Sam Harris, Dennett, Churchland, etc.) does not uphold the position that free-will is necessary for moral responsibility. The subject has the illusion of free-will. Everything is predetermined by predeterminate conditions and scientific laws.

³ See Guyer & Wood, (1998) trans. “The Critique of Pure Reason”. Hence after only the Acedemie number (for instances, A445/B443, A415/B44, A449/B447 etc., here A refers to the first edition of the original text which was published in 1781 and B refers to the second edition published in 1887) will be mentioned.

⁴ See, M. A. Gillespie (2014, p.3) "Philosophy and Rhetoric in Kant's Third Antinomy,"

as we know that no events can occur without sufficiently predetermined/ a-priorily determined principles; otherwise, it violates the laws of causality principle. Therefore, the view that there is only one kind of causality is inconsistent with its own principles, and the claim that everything is determined by causality only is self-contradictory. On the one hand, the unconditional causality is responsible for freedom of choice and secure the ground for morality. On the other hand, natural causality is responsible for the physical events of the world.

Antithesis

There is one kind of law that determines everything in the universe, and transcendental freedom is an *empty thought entity* (A445-7/B-473-5). "There is no freedom, but everything in the world happens solely in accordance with laws of nature" (A445/B473). Let inherits that there are two kinds of causality, natural causality in one hand and transcendental freedom as an unconditional causality on the other for the sake of the argument. The claim that an uncaused-cause (absolute spontaneity), is the antecedent of all other events, however, itself not caused by any other antecedent cause is inconsistent with the laws of natural causality (A415/B443). Every beginning of action/event presupposes a state, which must be depended on fixed laws of nature. There cannot be any causal connection between unconditional causality⁵ (Transcendental freedom) and natural causality for the reason that transcendental freedom is contrary to the laws of nature based on which the unity of experience is possible. If transcendental freedom is independent of causality, then it is impossible to establish the relationship between transcendental freedom and the phenomenal world, in contrast, if it is a part of nature or affected by the laws of nature, then it is bound to follow the deterministic principles of the empirical realm. In that case, there is no need to claim that transcendental freedom has an independent existence from nature. Because it is nature itself that guides the principle of the phenomenal realm, therefore, freedom and causality are contradictory to each other, and both cannot exist at the same time in the phenomenal realm. Freedom is independent of the laws of nature is illusory because it entails *lawlessness*.⁶ Even if we accept transcendental freedom in the epistemic sense, we cannot establish the *unity of experience* (A449/B477) because the laws of nature will always guide the experience. Though the causality always conditioned by its own law, however, be able to provide *the lawful unity of experience*, which cannot be established through the transcendental freedom.⁷

The current advancement in the field of neuroscience also argues for the validity of the antithesis. The neurological activities (follow the deterministic neurological pattern) in the brain generates the illusion of free-will⁸ and responsible for decision making and performing a particular action.⁹

Synthesis

If we look at the thesis and antithesis, then it seems that both of their claims are inevitable and grounded on rationality at the same time it is crystal clear that if the thesis (a) is true, then the antithesis (-a) must be false and vice-a-versa. Kant contends that the conflict may seem to be real because of the

⁵ Kemp smith (1918) termed it as "free *cause*" (which is not an object of possible experience), is an empty thought entity for determinist because of its nature (non-spatial and non-temporal) as advocated by the supporters of the thesis as the unconditional causality.

⁶ See Allison (1990, 'The Third Antinomy.' p.20), if freedom as a causality, is independent of laws of nature (hence lawlessness), then it cannot be a cause at all and have any effect on Will of the agent. On the other hand, if it is determined by the laws of nature, then there is no need to presume that freedom as a different kind of causality separate from natural causality; hence, it became part of natural causality.

⁷ This problem is a threat to all forms of idealism that owe to establish the relation between the transcendental idea with the experience of the phenomenal realm. It is crystal clear that by undermining the external world as an appearance, the relationship cannot be established.

⁸ The science wing on the project of *Big Questions on free-will* conducted the research at Harvard in 2011 for the support of the antithesis, which argues for determinism and illusionism.

⁹ Libel, B. (1999) argues for the deterministic theory of free-will. He has conducted a thought experiment popularly known as Libet's experiment for a scientific understanding of free-will and arrived at the conclusion that whatever happens in the brain's decision-making processes can be traced to the neurological brain processes through EEG techniques. One can predict whatever decision the subject is going to take before half a second, the subject being consciously aware of about the decision from the analysis of neurological activities of the brain of the subject gained through EEG techniques. Hence free-will is a unconscious process and has no role while taking decision and performing actions.

delusionalisation of rationality in their claims; however, it is a pseudo conflict¹⁰. Both the thesis and antithesis are correct from their perspective as far as physical and metaphysical necessity is concerned; however, they do not contradict each other because it is the same reason, which is the basis of their claim. The appearance of contradiction is due to the misconception of mathematical totality as the sole totality. When the nature views in terms of mathematical totality it provides the sensible condition, which is subject to laws of nature and space and time whereas when it views as dynamic causal series, then there is a demand for an unconditioned condition to condition the heterogeneous causal series which can only grasp in terms of intelligibility. The thesis needs a foundation on which unconditioned unity of totality can be established, whereas the antithesis requires a foundation for completeness of the causal series and universal unity of experience. Freedom is the prior condition for both experiences of nature and moral action; however, it must be outside the domain of the empirical realm.¹¹

Kant's resolution of the third antinomy through transcendental idealism is a controversial issue.¹² His attempt to preserve the sanctity of universal moral order has to be praised, however, not with the cost of suspension of the phenomenality.¹³ The reason for Kant to keep separate the transcendental freedom from the empirical realms is to protect the sanctity of the phenomenal realm, however, the intelligible realm, on which, the conceivability of the transcendental freedom is relied upon, is implausible (Kant's justification for transcendental world is unsatisfactory, controversial because it yields technical problem in philosophy) and promotes skepticism about the phenomenal reality and solipsism. There is no satisfactory response to the question that how transcendental freedom related with phenomenal realm by preserving the sanctity of the both realm? Generally, there are two categories of response from Kantian reader to the question, namely *two-world thesis*- the view that the object of things in themselves and appearances are numerically distinct object, belongs to transcendental world and phenomenal world respectively. The *two-aspect thesis* advocate that there is ontologically one world but two kinds of perspectival interpretation- The things really are (things in themselves) and how they are appear to us through the constituent of private sense data.

Allison¹⁴ one of the proponents of two-aspect view, holds that the same reality can be viewed from transcendental and empirical prospective. he argues for two-fold sense of viewing the same reality by rejecting the two-object thesis such as- phenomenal and noumenal object. He argues

“Man (along with rest of creation), is a part of nature, subject to its laws, and on the other hand knower, actor, rational being, transcendent nature in the sense that not completely determined by natural laws and along with that author of these laws”.¹⁵ It's the transcendental paradox of Kant that on the one hand, man is part of the sensible realm in the sense that subject has to follow the laws of nature, on the other hand, have the ability of intelligibility, hence knows the limits of sensibility and capable to transcend the limits of the sensible realm. Allison argues that this apparent perspectival dualism- transcendental and empirical distinction does not lead us to the noumenal-phenomenal dichotomy¹⁶ rather provides two perspectives of viewing *the ordinary empirical point of view; its limit and transcendental necessary conditions of possibility*.

Despite of the disagreement on things-in-itself, and appearance distinction, it is quite uncontroversial that Kantian treatment of appearance as the product of sensible realm cannot have separate ontological

¹⁰ For Kant, it's a conceptual confusion about the foundation of causality according to which natural causality is based on the phenomenal realm whereas transcendental causality demands its foundation for beyond the limits of sensibility (phenomenal realm) that is noumenal realm. Kant argues that both natural and transcendental causality are grounded on *rationality*. The laws of physics in its abstract form can't be grounded on sensibility. The discovery or creation of any laws of nature or moral laws is impossible without rationality. Rationality provides the foundation for both the condition of sensibility and intelligibility.

¹¹ Allison (1990, p.26), "The Third Antinomy," argues that reason has the tendency to transit its limit; therefore, it demands an unconditional totality of conditioned independent of the empirical realm for the sake of causal principles, which can be only intelligible and beyond the sensibility.

¹² He negates the content of experience as real, which is prone to promotes solipsism and skepticism on the existence of the external world.

¹³ This problem is popularly known as the boundary condition of knowledge.

¹⁴ See, Allison (1971 pp. 182-186) *Kant's Transcendental Humanism* for critical analysis of his two-view thesis.

¹⁵ Ibid, p.183

¹⁶ Ibid., p.188

status, and the locus of transcendental freedom cannot be sensible realm. Appearances has to be treated as the content of experience for the subject (things in themselves) which has only ontological status of being in noumenal reality; otherwise, transcendental idealism became indefensible and will lead to transcendental realism. The object of the external world can be expressed in two way- as it appears to the subject and really it is in its essence (indistinguishable from subject-noumenal reality). There can be two ways of interpretation of Kantian view of things-in-itself, the plurality of self or things-in-itself, in which there can be more than one thing-in-itself, and monistic self- the view that there is one inseparable and indistinguishable thing-in-itself. From the observation of the trajectory of *transcendentalism*, we have seen that in the interpretation of the Kantian self¹⁷ as the plurality of self always ends with an open-ended question about the relationship among things-in-itself and, has been the victim of skeptical attack. On the other hand, the monistic interpretation of things in itself has either leads to *dogmatic transcendentalism* or concludes with *epistemic phenomenalism*.¹⁸

In the *Critique of Practical Reason*, Kant contends that unlike pathological beings, the Will of the rational agent is conditioned by practical principles that are independent of subjective maxims, because subjective maxims are causally affected by deterministic laws of nature¹⁹. In other words, the Will has to be only conditioned by the objective maxims²⁰ or categorical imperatives alone for preserving the autonomy or freedom of the agents. The great difficulty for the Kantian philosopher is to establish the connection between higher-order reality with lower-level phenomena by keeping the purity of both realms. If there is freedom and any moral principle at all, that must be reflected in human action. If the higher order principles (freedom) can't have reflection on the lower-level phenomena then such kinds of higher order principles become redundant. On the contrary, if the lower-level reality (phenomenal realm) has an effect on the higher order reality (noumenal realm- the locus of transcendental freedom) then transcendental freedom bound to follow natural causality. Theoretical knowledge without any relation to the phenomenal world is utterly piffle, because a theory-building activity²¹ without any significance to the practical world is not going to solve any problem.

There can be two general approaches to connect the transcendental ideas of freedom with the phenomenal reality: the bottom-up approach and the top-down approach. Kant rejects the former and accepts the latter. Kant argues that rationality discovers the categorical principles because of which the status of human autonomy can be retained. The categorical imperatives are devoid of the content of experience and unaffected by the phenomenal realm in order to retain the purity of reason. In the top-down approach, certain moral principles (i.e., deontological principles, and cardinal values) are the ground of moral action.²² The higher level reality (noumenal realm-locus of transcendental freedom and morality) is independent of the causal nexus of phenomenal realm, hence, grounded on pure rationality which can never be naturalized, however, has an indirect effect on phenomenal realm due to which the effect of the transcendental realm became materialized and moral activities became possible.

¹⁷ There is no direct reference for the transcendental self as a substance in Kantian literature. The simple reason behind it might be Kant's commitment to establishing a non-dogmatic transcendentalism. There is *no metaphysical assumption of self* in Kantian literature advocated by Guyer, P. (1987, p.86), Walker (2006). They argue that any attempt to attribute the transcendental self will lead to dogmatic metaphysics. Hence, the inference of the existence of the transcendental self from the boundary of the phenomenal self can only be intelligible.

¹⁸ See, Willaschek (2001, p.683) for textual evidence.

¹⁹ The dependency of the Will on the representation of appearances means to give up the freedom of rational agent, which makes the action of the agent similar to the action of a puppet for Kant. The practical principles determine the condition of the Will. The conditioned Will generates an effect in the sensible realm, which can be traced through the action of the agent. Once the Will of the rational agent affected by inclination and sensualistic desires, it follows a certain kind of deterministic behavioral pattern in which non-deterministic action is not possible. The determination of the Will must be based on those practical principles which are independent of the empirical world or based on the self-determination of the Will of the agency in order to ensure freedom or autonomy of the agent.

²⁰ Kant makes the distinction between objective laws from objective maxims. If the Will is determined by objective law, then the autonomy or freedom cannot be protected. Therefore, the categorical principles are given in terms of ought rather than is. The determination of the Will is depended upon the subject, that is, whether the Will of the subject will be determined by the laws of causality of the phenomenal realm or by the categorical principles. Every human being by nature rational or has the potentiality to be rational which entails the fact that they are free or autonomous beings that's why Kant advocates the dignity of human beings is routed around this idea that every rational agent must be treated as an end and never as a means for the simple reason of recognition of the freedom of individuals.

²¹ One of the most influential analytic philosophers, namely Wittgenstein, attacked such kind of theory building activity in his work- *Tractatus Logico Philosophicus* (4.112).

²² Though all these ethical principles are arguing for top-down determination of action, however, each one has a unique way of establishing their point of view, and each theory has its limitations.

The bottom-up approach has quite a different standpoint. Unlike the former, it argues that transcendental freedom has direct effect on phenomenal realm, hence transcendental freedom is traceable in human action through the application of rationality. In this approach, one can ground moral principles on the phenomenal experience itself. The awareness about phenomenal experience of the subjects is the foundation of morality. It upholds the position that every living being has shared the ability to be aware about the phenomenal experience; therefore, it is universal and is the ground of morality. The value paradigms of this approach are depended on the awareness of phenomenal experiences of every individual. The lower level phenomenal- having phenomenal experience, is directly connected with higher order reality-awareness²³ about phenomenal experience. The bottom-up approach even though provide a verificationist account of transcendental freedom, however, is incompatible with transcendental idealism of Kant.

Critical Remark

Form the observation of Kantian solution to antinomy; his transcendental argument can be summarized in the following way

P1) Whatever belongs to the noumenal world is transcendental and unknowable²⁴.

P2) The locus of autonomy cannot be the sensible world²⁵ Hence must belongs to the noumenal world for its existence.

Therefore, the autonomy of agents is transcendental and unknowable.

The reading of Kant from an idealistic perspective may have no problem with this argument. However, there are implicit problems with the argument for transcendental realist. The mere logical deduction of unconditionality for autonomy is not the sufficient condition for establishing the certainty of the knowledge claim of the unknowability of autonomy. It also invites the skeptical attack (i.e., reductionist and illusionist position) for the agency's autonomy. If the locus of autonomy or free-will is the noumenal world, one can never ever know or be certain about the autonomy of human beings. The only option left is to accept the autonomy as a-priorily given, and the fundamental building block of the universe in order to explain the responsibility of the action of the agents, because as Kant argues that it can only be intelligible but not sensible, which leads to us for an explanatory gap between the relationship of intelligible realm and sensible realm.

One of the most influential works of Kantian ethics is that it construed moral order as universal. The attempt for reconciliation of natural causality and freedom through the internalist approach is a significant effort of Kant; however, Kant's approach to the external world generates lots of technical problems in philosophy. Kant, like Descartes, invited the danger of solipsism and skepticism by upholding the view that the intelligible world is the sole reality. Kant holds that the mind constructs the world through the representation of appearances. Hence, the world has no independent existence from the mind, and the external world is just a mere product of mental construction. The skeptical approach towards the world and the denial of phenomenological reality for the sake of transcendentalism is not an ideal approach. Kant failed to establish the transcendental freedom as a living reality by brushing off the reality of the phenomenal realm and subjective certainty about the experience.

The mind always takes part in structuring the sensory experience as well as in the active construction of thought or ideas, which might be fictional, non-existential, and objects of the possible and the real world.

²³ Here I have used both of the term consciousness and awareness interchangeably.

²⁴ Here the word unknowability is used in a special sense. The autonomy of agents cannot be traceable because of its unknowability condition, as far as the limits of the phenomenal perspective is concerned; however, the effect of the self-determination of the Will can be found in phenomenal realism.

²⁵ Either the locus of autonomy is the phenomenal world or the transcendental world. The unconditional causality (i.e., autonomy) cannot be part of the sensible world. Whatever exists in the phenomenal realm, is contingent in nature. Therefore, it's locus must be in the transcendental world.

The product of mental construction sometimes may correspond to the world sometimes not. The mind is not entirely detached from the experience and content of experience. Experiences provide the information based on which the future prediction of conventional reality became possible through the help of reason. It is the logical faculty of reason which makes the distinction between the possibility of the existential entity from the non-existential or fictional entity. It also plays a vital role in the environmental prediction, construction of ideas, and judgments. Therefore, both experience and reason have a significant role in moral judgment.

For Kant, theoretical reason secured the ground for human autonomy. He argues that we have sufficient reason to accept the transcendental freedom as the foundation of morality. Transcendental freedom could not rely on the sensible realm for its existence because of the demand of an unconditioned condition of unbounded reason. Kant also accepts that the effect of the Will, being affected by only intelligible realm- (moral laws²⁶), can be traceable through action only, which is itself belongs to the phenomenal realm. The underlying assumption behind Kant's transcendental idealism is that a universal moral order cannot be established, if the phenomenal realm is accepted as real due its nature of relativity and uncertainty. Hence, for Kant, it is not possible to find a universal moral order through the bottom-up approach. Rationality is the key factor in the self-determination of the Will of the subject. It is the criterion of autonomy, and it prevents the natural causality from affecting the determination of the agent's Will.

In light of the above evidence, it is crystal clear that the transcendental argument established the ground for transcendental freedom, however, is not plausible enough to receive universal recognition irrespective of the divergent paradigm. In the non-reductive naturalistic paradigm, the so-called human autonomy is an evolutionary product.²⁷ The redundancy of phenomenality for autonomy in the Kantian project and justifying it through some abstract principle has opened up the possibility for autonomy of artificial agents.

2. Artificial Agency and Autonomy

The tremendous advancement in the field of A.I. has brought the technological revolution and has changed life perspective. If intelligence is the sole criterion and almost everything is a product of intelligence (including moral principles and normativity of autonomy), then AAIA is possible. The dependency of autonomy on intelligence also affirms the fact that greater intelligence generates greater autonomy. If A.I. machines can be more intelligent than human beings; then, they can have more autonomy and be able to make better moral decisions in every situation.²⁸ If we accept the criteria of utility (maximization of utility and minimize the loss) and subjectivism (might is right principles) and not withhold speciesism,²⁹ then artificial super-intelligent machines become a reality that can not only be autonomous but also can build autonomous machines.³⁰ In fact, as Kahneman³¹ contends that human beings are less efficient in maximizing the utility because of the limitation of their cognitive power, though they are good at minimizing the loss. Hence, the A.I. machines can be a better performer

²⁶ Which are in itself abstract principles, are the criterion of moral action. The a-priori moral principles are being independently true and do not rely on the sensible realm, are necessarily true; therefore, universalizable, not because they are universalizable, therefore true. In other words, these moral laws are given in terms of abstract form as a criterion of moral action. However, the moment we try to understand the principles with the help of content, the self-contradiction within the moral laws arises. Therefore, Kant has tried to keep the distinction between the content of the moral law from its structure while talking about the moral principles. However, there is an ontological as well as explanatory gap between objective moral laws and its transformation in action; This has not bridged yet despite the several attempts to solve the problem. Another problem of the Kantian notion of moral agency is that, as he defended one kind of anthropocentric morality on the influence of rationalist slogan- *Man is the major of all things*, according to which, our moral obligation must be limited within rational agents (human beings alone). Rationality is the criteria for moral obligation and free-will. Non-human beings are devoid of rationality; therefore, they do not have moral rights and human species do not have any moral obligations towards them, which is a problematic issue and implausible with the argument of environmentalists, (i.e., Biocentrism, Cosmo-centrism) and feminism.

²⁷ See Mirazo and Moreno (2012 p.25-27) *Autonomy in evolution: From minimal to complex life*

²⁸ In certain circumstances, the biological rational agent became frozen (cannot take any moral decisions for a particular time being, i.e., in the trolley case thought experiment) due to limited intelligence, which might not be a problem for A.I. machines.

²⁹ Speciesism is the view that the dominance of one species over another is justifiable, grounded on the might is the right principle. It promotes human species centric ethics in which all the moral principles are rooted in favor of human interest only.

³⁰ Of course, it is a logical possibility that might happen in the future, as claimed by A.I. scientist. It is an inference or future prediction from the observation of current advancement in the field of A.I.

³¹ See, Kahneman, (2011) *Thinking*, for more details

proportionately than human beings. The ultimate goal of autonomy is to promote moral action. An A.I agent can have more capability of information processing and predicting the situation than human beings. It can also have the ability to follow the normative conditions of morality in order to generate desirable actions. A perfect moral agent always maximizes the desirable action and minimize the undesirable actions as per the utilitarian principles of morality. An A.I agent can serve the purpose in a better way in comparison to human agents.

Another approach towards autonomy of A.I. agency can be possible that as Kant argued, rationality³² is sole criterion that gives human beings the status of autonomy. If A.I. machines can be built in such a way that they can be rational or can have the potential to be rational, then the objective (Autonomy of A.I. agency) can be achieved without the violation of the principles of categorical imperatives of Kant.³³

Can A.I machine be rational?

The word rationality itself is an elusive and ambiguous term because we cannot have a particular definition of rationality.³⁴ Rationality can be interpreted in many different ways as per the needs of the context. There are two kinds approaches of rationality a) Non-naturalistic approach of rationality b) Naturalistic approach of rationality³⁵,

The non-naturalistic rationalist treats rationality as an ability; independent of normativity³⁶, and *apriorily given* to human species. The ability of rationalizing is objective in nature and plays the significant role in the civilization of human species. There can't be any moral principle possible without rationality. The rational ability plays the pivotal role in the determination of the agent's Will. Rationality provides the formal structure because of which synthetic a-priori knowledge and knowledge about the phenomenal experience became possible. It can't be the object of experience due to its non-reducibility to the world of experience, however, is the necessary condition for the knowledge of phenomenal experience.

There is another category of rationality; namely naturalized rationality proposed by Quine³⁷, Nozick³⁸, Dretske³⁹, in the 20th century- the view that the normative condition of rationality is dependent on the empirical world. The non-naturalist perspective of defining rationality (i.e., Kantian rationality) always keeps the equal distance from psychological reductionism and natural reductionism.

Let us calibratedly look into the non-naturalistic definition of rationality to examine their limitation of universalizability.

32 The formula of autonomy as stated by Kant in the Groundwork of the metaphysics of morals that 'the idea of the Will of every rational being as a Will that legislates universal law' (G4: 431; II, p.55.) The principle of autonomy is, therefore: "to choose only in such a way that the maxims of your choice are also included as universal law in the same volition" (G4: 440; II, p 80). See also Gerald Dworkin (1988, pp.4-15), *The Theory and Practice of Autonomy* for a comprehensive analysis.

33 The normative principles of categorical imperative can be reduced and set as the rules of programming language once it is derived to natural language. Further, the reduction of natural language to programming language is not an impossible task, which will determine the activity of the A.I. machine. The self-determination of programmed robots, like human beings, can make decisions without any human intervention.

34 Christine M. Korsgaard (2018, pp.5-12) illustrates that rationality sometimes defined as "the capability to perform prudently "(doing what is in your own best interests), acting with instrumental utility or acting with instrumental rationality (doing what will get you whatever ends you wish to achieve (doing what will get you whatever ends you wish to achieve)" which are easily compatible with AAIA. this definition of rationality cannot be construed as a correct definition for the Kantian thinker. The disagreement about the definition of rationality between Kantian and Anti-Kantian philosophers lies in the problem of normative conditions of rationality. Once we try to set the normative condition of rationality, it is quite possible to be fall in the trap of instrumentality of reason. Hence, reason as the precondition for establishing objective reality limited within human cognition.

35 The view that the normativity of rationality can be traced in the natural world. See, Sturm T. & Gigerenzer, G. (2012) *How (Far) Can Rationality Be Naturalised*, for defence of ecological rationality (one version of naturalised rationality) which attempt to establish the relationship between cognitive heuristics with the environment. See also Konzelmann-ziv A. (2008) *Naturalised Rationality: A Glance at Bolano's Philosophy of Mind*, for argument in the support of naturalised rationality.

36 See, Kolodny, N. (2015) *Why Be Rational* for the defence of non-normative rationalist.

37 See, Quine (1969) *The Naturalised Epistemology* for comprehensive analysis. Quine in this text provides a critical outlook towards the traditional formulation of rationality. He has defended one kind of replacement naturalism.

38 See, Nozick (1993, p.133) *The Nature of Rationality* for a version of normative rationality in which he argues that the norms of rationality can be traced from empirical resources

39 See, Dretske (1995) *Naturalising the Mind* for naturalistic project of rationality.

Df. - X is rational *iff*, has a universalizable logical explanation of Y for the action Z, which justify it as right.⁴⁰

The justification of rationality must have an isomorphic relation with logical structures and must follow the predetermined principles of logic in order to have a valid justification. Hence, the universal valid justification for the action is the criteria of rationality.

The anticipated objection to this definition of rationality from reductionist, can be the limits of human cognition for having an infallible universal logical structure. Therefore, we cannot have a concept of rationality, which is in itself universal and true forever. We have evidence from the observation of the past that the valid logical structure is not valid forever. So, the change in logical structure can also change our perspective of rationality. However, this objection cannot be applied for bounded rationality.⁴¹ If the ascription of rationality for agents is depended on certain abstract laws, rules, or logical principles and it necessitated by the abstract neural pattern of the human brain, then an artificial neural network can be created by replicating the neural network of the human brain which can generate rationality by following the same logical principles/rules as applied to the human brain⁴².

From the analysis of the debate on rationality, it is clear that the reduction of rationality to causal or functional effect of the neurological structure or abstract neural pattern of the brain entails the fact that AAIA is possible. If rationality is merely an inbuilt program to follow certain a-priorily given rules unconditionally, then also it is possible to have such kind of A.I. machines which are by default rational and being capable of making the decision independently. As we know, Kant argues that rationality entails the autonomy (or potential to be autonomous); the same logic can be applied to A.I. machines also if we accept the same normativity of rationality.

Now the question may arise that if the AAIA is possible then how the autonomy of A.I will be different from human Autonomy? The biological machine analogy construed the view that human beings are different from the machines as far the living biological organism is concerned; otherwise, functionally and behaviorally can be identical to machines. Machine acquires certain technique to perform an act, and do so as programmed by the programmers; similarly, an untrained brain (the newborn baby) learn from the environment (i.e., language acquisition, skills, environmental mapping, so and so forth), programmed itself as per the information collected from the environment, like the machine programmed. However, the difference is the capability of having multiple functions because of the more skillfulness of human beings than the machine, which is due to the limitation of programming. Human beings have multiple sources and techniques to be programmed than A.I. machines. The efficiency of programming determines the efficiency of a machine to perform the act in a skillful way.

The reductive-naturalist externalism also argues that the moral values are evolutionary products acquired by human beings through the process of natural selection. The newborn baby and animals cannot have these values because these values can only be acquired through moral training. However, Dretske⁴³, one of the proponents of naturalistic externalism, argues that unlike other representational systems, human beings have the privilege of a *metarepresentational* power- ability to conceptually represent the sensory representational information. Other biological representational systems can have only the natural-systemic indicative representational function through which the particular system is able to provide only sensory representational information. The non-biological representational system can have only the conventionally assigned representational function due to the particular representational function is

⁴⁰ Here 'X' is the agent, Y is the justification for the action, and Z is the action itself. Some of the Kantian thinkers may disagree with this definition of rationality because of their acceptance of the view that reason as the A-priori necessity for human beings to establish something as valid. Hence reason is not limited within its own boundary. Therefore, reason in the first place as the foundation for any kind of discourse has to be accepted as the primary condition of knowledge, because no reason can establish it for the simple fact that they cannot go beyond their own limitations. Asking the question that what is rationality is itself the question of meta-rationality (it is the rationality about the rationality). In order to define what rationality is, we have to apply the reason, which is prone to the circularity of reason.

⁴¹ Bounded rationality has accepted the flexibility of the normative condition of rationality. It claims for the complete rationality of A.I., unlike human rationality. See, Simon, H. (1982). "Models of Bounded Rationality" vol.1 for detail analysis

⁴² See, Steinhauser, M. et. al (2012) for neurological basis of rational activities and Decision making.

⁴³ See, Dretske (1995 pp.5-8) *Naturalizing the Mind* for detailed analysis of natural, conventional representational indicative function.

assigned by those biological representational system who has the privilege of a special kind representational function, called-conceptually acquired representational indicative function. In Dretske's naturalistic paradigm, even if these higher order functions can be naturalized, the non-biological representational system (i.e., A.I) can't have the metarepresentational power- conceptually represent sensory representational information. However, recent development in data analysis by A.I machines indicates that the limit of A.I can't be constrained within the information providing system. Because it can not only perform a specific assigned function (like provide a particular piece of information), but also can analyse and predict the future from the assigned indicative representational information. Hence, Dretske's objection to AAIA can be defeasible and it is logically possible that A.I machines can defeat human intelligence on conceptual representation. The naturalistic reduction of higher order properties to natural language is compatible with the possibility of A.S.I. Anything which can be converted to programming language can be artificially constructed and programmed. The ethical principles (i.e., ten commandments and other core value teachings of religion, categorical imperative so and so forth) can be programmed, consequently, the artificial moral agency is possible. We can have an artificial moral agent who may be more ethical than the human agent in their action in the Kantian paradigm of autonomy.⁴⁴

The argument for the possibility of AAIA can only succeed if it is only a matter of intelligence and rationality. Intelligence and rationality are not the sufficient condition for the criterion for autonomy even though, are the necessary condition for autonomy and moral actions. What makes human agents special is the phenomenal experience from the first-person perspective.⁴⁵ Only a living organism (i.e., human agents) can have privilege of having phenomenal qualitative experience⁴⁶, which an A.I. agent will always be deprived even though it can replicate all the human action, brain function and behavior. Human agents have not only phenomenal subjective experiences but also can have privilege of cognitive skill to use the information of phenomena or experiences for creative acts and can understand the content of the experience of others through analyzing their own content of the experience in the similar given environmental input, which cannot be possible in case of A.I. agents. An A.I. machine can act like an autonomous human agent by creating the delusion of the replicated-behavior of an autonomous agent; however, in reality, it is just a machine, subject to deterministic patterns of behavior and bound to perform assigned function. Even if artificial super intelligence can have rationality in a loose sense,⁴⁷ It will still be like a philosophical zombie. There can be nothing to experience something from inside in the first-person perspective for a zombie. It is quite uncontroversial that rationality and intelligence are not the only criterion of autonomy even though are the necessary conditions of autonomy. Phenomenality, as the sufficient condition, must be accepted for the sake of realization of free-will. Even if an A.I. agent can have rationality in a minimalistic sense without phenomenality; however, can't have autonomy without phenomenality; therefore, phenomenality becomes the criterion for autonomy, accessible to the living organism alone.

Another approach to transcendentalism can be the reading of Kant from the Transcendentalism-Naturalism Compatibilist (hence after TNC) perspectives, which advocate for one kind of

⁴⁴ The objection against this model can be raised that the AAIA is itself an imaginary friction. Even if we inherited it as a logical possibility, the determination of the action by human agents, as argued by Kant, will be different from A.A.I.A, for the simple fact that the determination of the Will of the agent through the categorical imperative is given in terms of self-determination of the Will and conducive to the principles of alternative possibility. To be moral or immoral is deepened upon the Will of the human agent, which means the possibility of doing the action otherwise. The AAIA agent (will either fall into the category of by-default to follow the moral laws or otherwise in which there cannot be autonomy), being the product of human intelligence, cannot have the choice to do it otherwise. The determination of autonomy will rely on human intelligence until it surpasses human intelligence, and as we know, human intelligence is determinant of A.I and remains forever. Therefore, the AAIA is imaginary friction.

⁴⁵ See, Sahu M.K. (2019, pp.77-83) "Qualia and Intentionality" for detailed analysis of phenomenal first-person accessibility. This paper defends the ontological subjectivity of phenomenal qualities of experience. Deborah, J.H. (2006) also argues against the possibility of artificial moral agency.

⁴⁶ See, Sahu M.K. (2020, pp.160-66) "Phenomenal Consciousness" for the argument of phenomenal consciousness.

⁴⁷ Kant construed rationality is a non-representational form, which is a-priorily given in the consciousness of human Agents, and cannot have a mechanistic explanation at all by virtue of the fact that it is the nature of abstraction and untenable in experience. Kantian thinkers tried to block the possibility of rational A.I. agency through this argument. However, the term rationality itself has elusive and ambiguous usage. They could not give a satisfactory account for rationality; therefore, A.I. rational agency can be defended by giving a mechanistic theory of rationality. Russell (1997), in his article "*Rationality and Intelligence*," defended such kind of position. He has classified four categories of rationality in this article and argues for Artificial Rational Intelligence.

transcendental realism⁴⁸. Only rational agents can have phenomenal pre-conditionality; therefore, it excludes the non-rational animal as well as artificially constructed machines from having the phenomenal pre-conditionality. The TNC's interpretation of the Kantian phenomenal self can succeed in blocking the knowledge claim of the possibility of AAIA; however, the problem in this approach can be easily detectable. For the sake of transcendentalism of the self, its interconnection with the world has to be diminished in order to preserve the sanctity; In that case, it will give rise to an unbreachable ontological as well as explanatory gap between phenomenal-transcendental self and its further interconnection with the world becomes impossible. Therefore, Kantian transcendentalism for AAIA is not a bigger challenge with comparison to phenomenal realism.

Furthermore, let's consider the compatibility of transcendentalism with phenomenal realism for a while—the view that the transcendental account of phenomenal pre-conditionality (the demand of reason for phenomenal consciousness as the precondition of the phenomenal qualitative aspect of experience) as a transcendental form (like other categories) discoverable through the application of transcendental method of reasoning, hence, provide an objective method of inquiry of the subjective mental states without any appeal to phenomenal perspectivism (as advocated by Berkeley and his successor theories, phenomenalism, and perspectivism). Phenomenality as an a-priori transcendental form can play the role of a synthesizer of the subjective-objective dichotomy of experience.⁴⁹ It can function as a-priori⁵⁰ condition for phenomenal experience without phenomenal content (enables us for defending non-reducibility of phenomenal qualitative experience in Kantian sense); because of which an organism has the privileged access of the phenomenal experience, is objective to the particular living organism; must be outside the domain of sensibility, hence, it is the demand of reason to accept the phenomenal pre-conditionality (phenomenal self) as the foundation for every phenomenal experience. then also it can raise the objection against AAIA, however, became inconclusive and violate the principles of transcendentalism.

3. Conclusion

The resolution of Kant's third antinomy through abstraction, though established the ground for both moral and natural principles, cannot give sufficient reason for the connection between the autonomy of agents with the phenomenal world. Kant has the ontological commitment to establish the existence of the self in the first place; therefore, he adopted the pure abstraction method, and denied the existence of phenomenal reality. Transcendental realism somehow manages to avoid this obscurity by arguing that the knowledge about the phenomenal reality is different from phenomenal reality. Kant has this misconception that reality has to be transcendental and unaffected by the sensible realm because of the fallibility of the phenomenal realm. Thomas Nagel, one of the most influential philosophers of the 20th century, concurs with Kant about the non-physiological concept of reasoning by arguing for the objectification of knowledge; however, he differs by holding “*objectification-without-reification*” strategy. Kant has an ontological commitment, which Nagel tried to avoid in his work. If we critically analyze the Kantian solution to the problem of autonomy, then we can find out that except the Kantian ontological commitment, his epistemological standpoint is compatible with AAIA. However, Kantian autonomy is incompatible with any naturalistic and non-naturalistic theory of autonomy because of its unique nature of self-legislation and self-determination of moral laws of the transcendental self. The floccinaucinihilipilification of phenomenal reality and the reduction of the external world into the appearance of representation by Kant makes the job easy for A.I. scientists.

48 See, Louden (2019, pp. 8-12) Kant the Naturalist for more details.

49 One has to be careful while interpreting the transcendental a-priori of reason for the defence of phenomenality as a new category either of the sensibility and understanding or an independent category from both sensibility and understanding. Because it is just one step away from falling in the trap of Berkelian empirical subjectivism. 0

50 Here the a-priori condition is used as the transcendental a-priori condition distinct from synthetic a-priori condition of reasoning. Allison (1971, p.187) has the similar kind of view that the misinterpretation of the a-priori conditions always leads to entanglement of Kantian transcendental subjectivism with Berkelian empirical subjectivism. Kant, unlike Berkeley, has rejected the view that the objects of experience are private sense data (as followed by the Russellian sense datum theory).

Rationality without phenomenality leads to abstract objectivism, whereas phenomenality without rationality leads to subjectivism and relativism. Phenomenality is the necessary condition for the realization of the effect of the self-determination of the Will even if we accept the Kantian thesis for a while as a plausible thesis. The only difference between naturalists and Kant is the logical implication of the transcendentalism of autonomy, which deviates Kant from the naturalistic camp. Therefore, except for the idea of transcendental self, the categorical imperative as a command can be implemented in the A.I system, hence, is compatible with the AAIA thesis. From the evaluation of the paradigm of rationality (as we have found that rationality sets its condition on normativity and certain principles can also be implemented in a rational, autonomous agent), it is crystal clear that the compatibility of transcendental naturalism with AAIA is undeniable. Therefore, the objection from phenomenal realism to AAIA is more plausible than the Kantian objection to AAIA.

Acknowledgements: I would like to thank my supervisor Prof. Ratikanta Panda for his constant help, support and encouragement, being aware about his responsibility of supervising without which the substantive revision of the manuscript might not be possible. I am also grateful to the anonymous referees for their quick review and constructive feedback. I would also like to thank Prof. Pravesh Jung for his valuable comment on the paper.

References

- Allison, H.E. (1971), Kant's Transcendental Humanism, *The Monist*, Vol. 55:2, pp. 182-207
- Allison, H.E. (1990), *The Third Antinomy, Kant's Theory of Freedom*, New York: Cambridge University Press
- Bostrom, N. (2014), *Superintelligence: Paths, Dangers, Strategies*, Oxford, UK: Oxford University Press.
- Copeland, B. J. (1998) *Super Turing Machines, Complexity*, Vol. 4, pp. 30-32.
- Christine, M. K. (2018) Rationality, In Lori Gruen (ed.), *Critical Terms for Animal Studies*, Illinois: University of Chicago Press.
- Deborah, J.H. (2006) Computer Systems: Moral Entities but Not Moral Agents, *Ethics and Information Technology*, Vol.8, pp.195-204
- Dretske, F. (1995) *Naturalizing the Mind*, London: MIT Press
- Dworkin, G. (1988) *The Theory and Practice of Autonomy*, New York: Cambridge University Press.
- Gillespie, M. (2014), Philosophy and Rhetoric in Kant's Third Antinomy, *The Political Review*, Vol. 30, pp.7-33
- Guyer, P. and Wood, W. A. (1998) *Critique of Pure Reason*, New York: Cambridge University Press
- Guyer, P. (1987) *Kant and The Claims of Knowledge*, Cambridge: Cambridge University
- Kahneman, D. (2011) *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux Press
- Kolodny, N. (2015) Why Be Rational, *Mind*, Vol. 114, pp.509-63
- Konstantin, P. (2013) Naturalism and Kant's Resolution of The Third Antinomy, In Margit
- Ruffing, Claudio La Rocca, Alfredo Ferrarin & Stefano Basin (eds.), *Kant Und Die Philosophie in Weltbürgerlicher Absicht: Akten des Xi. Kant-Kongresses*, 2010, pp. 731-742.

- Konzelmann-Ziv, A. (2008) Naturalized Rationality: A Glance At Bolzano's Philosophy of Mind, *Baltic International Yearbook of Cognition, Logic and Communication*, Vol.4:1, New Prairie Press
- Kurzweil, R. (2006) *The Singularity Is Near: When Humans Transcend Biology*, New York: Penguin USA.
- Louden, R.B. (2020) Kant the Naturalist, *Journal of Transcendental Philosophy*, Vol. 1:1, pp. 3-17
- Mirazo-Ruiz K. & Moreno A. (2012) Autonomy in Evolution: From minimal to complex Life, *Synthesis*, Vol.185, pp.21-52
- Nozick, R. (1993) *The Nature of Rationality*, New Jersey: Princeton University Press
- Reimer T. & Hoffrage U. (2004) Models of Bounded Rationality: The Approach of Fast and Frugal Heuristics, *Management Revue. The International Review of Management Studies*. pp. 437-459
- Sturm T. & Gigerenzer, G. (2012) How (Far) Can Rationality Be Naturalised, *Syntheses*, Vol. 187, pp. 243-268
- Quine, W.V.O. (1969) *Ontological Relativity & Other Essays*, New York: Columbia University Press
- Russell, E. (1999) Rationality and Intelligence, In Wooldridge M., Rao A. (eds) *Foundations of Rational Agency*, Applied Logic Series, Springer, Dordrecht, Vol.14, pp.11-33
- Sahu, M.K. (2019) Qualia and intentionality, *Journal of All Odisha Philosophy Association*, Vol. 5:4, pp.76-87
- Sahu, M.K. (2020) Phenomenal Consciousness: A Critical Analysis of Knowledge Argument, Inverted Spectrum Argument and Conceivability Argument, *Journal of Advances in Education and Philosophy*, Vol. 4:4, pp. 160-166
- Simon, H.A. (1982) *Models of Bounded Rationality*, Vol. 1. London: MIT Press
- Smith, N. K. (1918) *A Commentary to Kant's Critique of Pure Reason*, London: Palgrave Macmillan Press
- Steinhauser, M., Achtziger, A., Alós-Ferrer, C., Hügelschäfer, S. (2014) The Neural Basis of Belief Updating and Rational Decision Making, *Scan*, Vol. 9, pp.55-62
- Turing, A. (1950) Computing Machinery and Intelligence. *Mind*, Vol. 59:236, pp. 433-460.
- Willaschek, M. (2001) The ambiguity of the Kantian Distinction between things in themselves and appearance: On the debate about two-aspect and two world interpretations of the transcendental idealism, In V. Gerhardt, R. Horstmann & R. Schumacher (ed.) *Kant and the Berlin Enlightenment*, Vol.9, pp.679-790, Boston: De Gruyter