THE PROBLEM OF SOURCEHOOD IN THE LIBERTARIAN ACCOUNTS OF JOHN SEARLE AND ROBERT KANE

KERİM CAN KIRAÇ

BOĞAZİÇİ UNİVERSITY

2019

THE PROBLEM OF SOURCEHOOD IN THE LIBERTARIAN ACCOUNTS OF JOHN SEARLE AND ROBERT KANE

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Philosophy

by

Kerim Can Kıraç

Boğaziçi University

2019

DECLARATION OF ORIGINALITY

I, Kerim Can Kıraç, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature K. C. Kıraç

Date 10/11/2019

ABSTRACT

The Problem of Sourcehood in the Libertarian Accounts of

John Searle and Robert Kane

Free will is one of the most hotly debated issues in the history of philosophy, which also caused a lot of confusions in the minds of people who are arguing about it. In this thesis, I will mainly focus on the sourcehood problem in John Searle and Robert Kane's ideas concerning the nature of free will. First, the sourcehood principle which we may consider as a simplistic version of agency concepts becomes a quite problematic issue in the light of a) reductive physicalism and b) causal determinism. In Chapter 1, I will touch upon the basic concepts and discussions on free will which are relevant to our main inquiry. Secondly, when we start discovering the basic premises of Searle and Kane, we will be able to see how their ontological assumptions pose serious problems with regard to their libertarian intuitions which they are trying to defend. I will try to show in Chapter 2 that Searle's biological naturalism (BN) entails epiphenomenalism, and in Chapter 3 that Robert Kane's Self-Forming Actions fails to provide a sense of agency that exercises any more control than its compatibilist counterparts. I hold that free will, at least, entails top-down causation; and it is very hardly to be met by either of the two accounts that I will be discussing throughout this thesis. In Chapter 4, I'll examine the problems that are brought about by physicalism with regard to any metaphysically persistent sense of selfhood, which I consider to be crucial for libertarianism.

ÖZET

John Searle ve Robert Kane'in Liberteryan Görüşlerindeki Asliyet Problemi

Özgür irade, felsefe tarihinde en çok tartışma konusu olmuş ve üzerinde düşünen insanlarda belki en fazla kafa karışıklığına sebep olmuş meselelerden biridir. Bu tezde John Searle ve Robert Kane'in görüşlerinde özgür iradenin doğasına dair asliyet problemlerine odaklanacağım. Birincisi, faillik (agency) konseptlerinin basitleştirilmiş bir versiyonu olarak anlayabileceğimiz asliyet (sourcehood) prensibi a) indirgemeci fizikalizm ve b) nedensel determinizm ışığında oldukça bariz bir problem olarak öne çıkmaktadır. Birinci kısımda, özgür irade tartışmalarının tezimizin temel argümanlarını ilgilendiren kısımlarını genel olarak aydınlatmaya çalışacağız. İkincisi, Searle ve Kane'in temel varsayımlarını incelemeye başladığımızda, kendilerinin ontolojik varsayımlarının yine kendi libertaryen sezgileri karşısında nasıl ciddi sorunlara yol açtığını görmeye başlayacağız. İkinci kısımda Searle'ün Biyolojik Naturalizm'inin nasıl epifenomenalizme götürdüğünü ve üçüncü kısımda Robert Kane'in kendini-gerçekleştiren eylemlerinin (self-forming actions) uyumculuk (compatibilism) tezini savunanlara nazaran daha fazla control sahibi bir fail görüşü sunmakta başarısız olduğunu göstermeye çalışacağım. Görüşüm o dur ki özgür irade, en azından tepeden aşağı (top-down) bir sebebiyet zincirini öngörür ve bu tezde baştan sona üzerinde tartışacağım bu iki libertaryen tezde bu koşulun sağlanabilmesi pek muhtemel görünmemektedir. Dördüncü kısmı fizikalizmin metafizik açıdan stabil ve bölünemez vasıflar içeren herhangi bir benlik görüşünü nasıl derinden sarsabildiğini göreceğiz ki bu tür bir benlik kavramının libertaryen sezgilerimiz açısından pek de vazgeçilebilir olduğu düşüncesinde değiliz.

In Gratitude to
Chryssi Sidiropoulou and Stephen Voss

TABLE OF CONTENTS

INTRODUCTION

Free will is one of the most complicated problems in the history of philosophy, which can be traced back to Epicureans in Antiquity. It has been discussed in a variety of contexts such as theology, ethics, determinism and even in the light of quantum theory in modern times. Especially the problem of free will and causal determinism carries much weight in the modern discussion of philosophy. In this paper, I will try to address the problem in the context of a simplified criterion which can be named “the Sourcehood Principle”. There are two basic formulations of the principle:

First formulation of the sourcehood principle (concerning determinism): For a subject S to be the source/author of the action A, the causal relata of A or the set of reasons that justify doing A should not go beyond the causal autonomy of S. In other words, the causal relata of A should not be sufficiently constituted by the conditions over which S has/had no control.

Second formulation of the sourcehood principle (concerning reductive physicalism): For a subject S to be the source/author of the action A, the agential powers of S through which S brings about A should not sufficiently be constituted by impersonal physical forces over which S has/had no control.

If the antecedent conditions are causally sufficient for the resulting actions, I argue in a similar way with Searle and Kane that free will and moral responsibility are not possible at all. A full-blown determinism might not be true, given the implications of quantum physics; however, the same threat can persist if the quantum effects are negligible in brain-sized objects, which may imply that quantum fluctuations cannot make any difference at all. This is an open empirical question. On

the other hand, it is not clear how quantum indeterminism contributes any more control to the system even if our brains can genuinely amplify quantum indeterminism. This question is discussed in Chapter 3.

I find the second formulation much more interesting than the free will – determinism dichotomy. Searle's Biological Naturalism (BN) allows for the causal reduction of agential powers. Though Searle is a libertarian philosopher, I will argue in Chapter 2 that BN leads to some version of epiphenomenalism on the grounds that the causal powers of consciousness (not the ontological realm of consciousness itself) are rendered an observer-relative construction; and secondly, reason-explanations are rendered redundant concepts if BN is true; because the actions of a subject can ultimately be explained by underlying physical forces. We will try to clarify the distinction between strong and weak emergentism; the latter one, I argue, is the true description of BN. On the other side, we will leave it an open question whether strong emergentism is desirable given its radical implications such as that it requires postulation new fundamental patterns/laws in the world.

In Chapter 3, I'll try to sort out the basic notions and assumptions embedded in Kane's libertarianism. Kane postulates Self-forming Actions (SFAs) as the primary free actions which resolve the regress problem in libertarianism and they also function as character-developing actions in the Aristotelian sense. I'll try to show that if event-causal libertarianism allows for the causal reduction of agency into mere mental events, then this kind of libertarianism is not justified in its claim that libertarianism philosophically provides more control than compatibilism.
In Chapter 4, I will propose a version of the Brain Replacement Argument in order to address the problems of physicalism with regard to selfhood. I will try to shed light on two basic points: a) if physicalism is true and the mental states are supervenient

on underlying physical causes, then it is at best possible to account for a Humean bundle with regard to selfhood; b) if physicalism is true, our notion of death, which is certainly conceptually substantive for self-conscious beings does not fit into metaphysically substantive reality; which would not, in fact, be a great surprise. Metaphysical versus conceptual substantivity is originally put forward by Theodore Sider, and I find it quite useful in thinking about some futuristic scenarios of transhumanism and brain-replacement. There are indeed certain concepts that preoccupy our mental and social lives such as freedom, morality, death etc. However, a conception as serious as death might only be dependent upon various incompatible criteria that are put forward by different linguistic communities, without having any appeal to what is metaphysically fundamental in the world.

CHAPTER 1

PHILOSOPHICAL REMARKS ON THE PROBLEM OF SOURCEHOOD

Roger Scruton contends that we are grounded on biological roots as human animals but on the other side, we are on the edge of the nature as human persons. Being at the edge of the nature is as socially as described by Buber (1970) as I-Thou relations or as "reactive attitudes" by P. F. Strawson (1962). For instance, I am to be treated as a biological being, a human animal, while a medical doctor examines the sickness in my physiology; on the other hand, however; 'the human person' is the right description in case I act as a moral being with moral reasons in a social environment. These two levels of explanation correspond to the distinction between 'in the world' and 'of the world' (Scruton, 2018); I'm a moral being 'in the world', yet my moral/free character is not 'of the (natural) world' as if this human person is capable of producing new causal chains over and above the natural causes that constitute the human animal. Embodiment refers to 'being in the world' whereas 'personhood' signifies the transcendental character of morality, something that reaches beyond the physical closure of the world. The concept of embodied person is indispensable in our study of free will and agency. Scruton's assertions on the subject will be mentioned throughout the thesis as a subsidiary position: While we mention "transcendence", we are not necessarily referring to metaphysical reality; in Scruton's view, even if the reductive physicalism is true, self-conscious persons like us will nevertheless continue to be bounded by "reactive attitudes" and "moral reasons" in their social lives. Thinkers like Scruton and Fischer do not consider the metaphysical implications of determinism or reductionism as threats to our substantive concepts. The logical structure of human agency in libertarian versions

should be recognized as an indispensable background assumption. John Searle's analysis of that logical structure will play a central role in defending the irreducibility of genuine personhood and sourcehood as opposed to reductive physicalist tendencies that emerge even in the libertarian camp. Unlike Scruton, I will also try to make a careful analysis of the metaphysical implications in Searle and Kane's accounts.

Free will and moral responsibility have been quite hotly debated topics in the intellectual history. On every level of these discussions, there are incommensurable approaches as to what we mean by free will, what it takes to be morally responsible for our actions, what it takes for our actions to be "our actions", and the list goes on. The basic division generally appears between those who argue that causal determinism is compatible with free will and those argue on the contrary. The first group is called "compatibilists" whereas the latter is called "incompatibilists" which are the ones who defend a more traditional concept of free will. There is also a third group called "revisionists" who claim that even though our basic intuitions about the nature of moral responsibility are incompatibilist, we can nevertheless abide by the compatibilist stance with regard to metaphysics of free will.

Having focused on the compatibility of determinism with free will, I suppose the fundamental problem posed by causal determinism is the diminishing of agency. Peter van Inwagen's famous "Consequence Argument" implies that if determinism is true, then no one has any choice about how the world unfolds. Compatibilism tends to examine free will on the basis of the actual sequence of actions rather than alternate possibilities, and they argue for the sufficiency of reason-responsiveness in one's deliberative process. Frankfurt-controller arguments pose a powerful counter-argument to the incompatibilist camp in such cases. I will not argue that

compatibilism – incompatibilism paradigm is out of date now, because I grant that the truth of determinism is incompatible with free will and moral responsibility. I want to focus on the sourcehood aspect that is embedded in these discussions and examine the question of what the genuine conditions of sourcehood are.

Let us clarify van Inwagen's Consequence Argument first. Suppose that there was a judge who could have prevented the execution of a death penalty at a certain time only by raising his hand which would – let us say – conventionally signify a special pardon for the criminal (Van Inwagen, 1983, pp. 68-69). A normal person feels that a certain action was up to him and he could have done otherwise in ordinary cases on the supposition that he was not compelled in any sense. Inwagen tries to show that the truth of determinism is incompatible with any sense of causal gap that one needs to act upon in traditional sense of freedom. Question is whether the judge (J) could have prevented this particular execution at time T.

(P0: Total description of the universe in a remote past.

L: Laws of nature

P: Total description of the universe at T)

> (1) If determinism is true, then the conjunction of P0 and L entails P
> (2) It is not possible that J have raised his hand at T and P be true.
> (3) If (2) is true, then if J could have raised his hand at T, J could have rendered P false
> (4) If J could have rendered P false, and if the conjunction of P0. and L entails P, then J could have rendered the conjunction of P0and L false
> (5) If J could have rendered the conjunction of P0 and L false, then J could have rendered L false J could not have rendered L false
> (6) If determinism is true, J could not have raised his hand at T (Inwagen, 1983, p.70).

Another example: Richard Nixon was the president of United States at the time of the Watergate Scandal. He had to resign from his office after it turned out that he was involved in the scandal. The new president Gerald Ford gave pardon to Nixon, and Nixon was exempted from convictions for crimes that he was held

responsible for beforehand. Normally we tend to assert that Gerald Ford, in fact, had the choice not to pardon Nixon at that time. If determinism is true, the total description of the universe in a remote past in conjunction with the laws of nature entails that there is only one possible future. If determinism is true, we do not have or never had any choice about whether it is true or not: We do not have any choice about how the course of events unfolds. We do not have and never had any choice about the total description of universe in the remote past (a time that none of us was born) (as cited in Deery & Russell, 2013).

If Gerald Ford had the choice not to pardon Nixon at time t, then he had the choice to make the statement "Ford gave pardon to Nixon at time t" false. If the antecedent conditions entail the consequent, then it must be the case that if Ford had the choice to make the consequent false, then he must have had the choice to render the antecedent false, too. The antecedent is the total description of the universe in a remote past in conjunction with the laws of nature. In order to render the antecedent false, Ford must be able to render either the total description of universe in a remote past or the laws of nature false. As we have already seen, no one has or ever had any choice about the conditions of universe in a remote past and no one has or ever had any choice about the laws of nature. Therefore, if determinism is true, Ford has not ever had any choice about giving pardon to Nixon at that time. There was only one possible future that Nixon would have received pardon from Ford.

As opposed to this clarification of the absent alternate possibilities, for instance, Harry Frankfurt developed a strong counter-argument to show that it is not the fundamental problem whether we could have done otherwise at a particular instance but rather the reason-responsiveness of our deliberative process in the actual sequence that matters. There are many versions of the same argument that we will try

to clarify, most of which are called “Frankfurt controllers”. Frankfurt controllers ensure that an agent will perform certain action without any intervention in the actual sequence:

> Suppose someone—Black let us say—wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So, he waits until Jones is about to make up his mind . . . and he does nothing unless it is clear to him . . . that Jones is going to decide to do something *other* than what he [Black] wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones . . . does what he [Black] wants (Frankfurt, 1969, p. 835).

In some versions of this thought experiment, we suppose that Jones deliberates about killing a person – say Jim. Black, an evil neurosurgeon who implanted a control device in Jones’ brain, waits and watches the confrontation between Jones and Jim in a remote area. If Black realizes that Jones will refrain from killing Jim at any point, he will intervene in the neural processes and make sure that he won’t refrain. If Black intervenes, we would not normally think that Jones – as neurally manipulated person – is genuinely responsible for killing Jim; however – good news for Black – Jones decides to kill Jim on his own without any need for intervention. The interesting part of the story is that Black has never intervened, Jones has done everything on his own, yet Jones could not have done otherwise because of the presence of Black. Frankfurt argues that Jones is morally culpable even though he could not have done otherwise, and the so-called necessity of alternate possibilities is unwarranted (Frankfurt, 1969).

Daniel Dennett makes a similar point about Martin Luther. Luther once said to Catholic Church supporters: “Here I stand, and I cannot do otherwise." Dennett claims that even if Luther himself admits that he cannot do otherwise, he does not mean to refrain from responsibility of his position at that time, to the contrary, he takes the full responsibility (2007, p.117). Robert Kane (1996) asserts that although

most of our actions are determined, we need to be held ultimately responsible for some of our actions and for the building of our character up to the point that we may find ourselves unable to do otherwise; as Aristotle held that a wicked person is responsible for becoming one in the first place, which we will examine in the Chapter 3.

It would be to oversimplify the subject if we were to explain it all within the context of Inwagen and Frankfurt's basic arguments. Our aim here is just to set the stage for the context of our discussion which can either be titled as the sourcehood or agency problem in some libertarian accounts such as the ones that are formed by John Searle and Robert Kane. Both compatibilists and incompatibilists provide effective arguments to account for our basic intuitions about sourcehood or in other words, moral responsibility in real life cases. Although we will argue for the incompatibilist position, however, it is not even close to being self-evident whether we can consistently form a sense of sourcehood either in a deterministic context or not. Inwagen holds that we should stick to the incompatibilism unless disproved by scientific facts, but there are still grave concerns as to how we will deal with the problem of luck within the incompatibilist position (1983, p. 223).

Some keywords are widely used in free will discussions such as authorship, agency, sourcehood etc. Are they inherently incompatible with determinism, or is the so-called incompatibility just a legacy of the traditional dogma (Dennett might use the term)? It deserves a careful analysis in order for us to get clear about the concept of free will that is worth wanting. We should start by clarifying the sourcehood principle before we dwell upon a detailed examination of the problem in above-mentioned libertarian accounts. As a basic starting point, we conceive of ourselves as agents who act upon reasons and bring about changes in the world. One of the

obvious facts that ground this intuition is the fact that we are self-conscious beings. The truth of existence is predicated upon the fact that we are immediately aware of our mental states. We may simply refer to as the Cogito Argument. Regardless of determinism, we may plausibly put forward that a subject is the author of his/her actions if the relevant causal relata of these actions are properly located within the subject:

Sourcehood Principle (SP): A subject S is the source/author of his/her action A if and only if the relevant causal relata (C) of A are properly located within S; it was intentional of S to do A; S was not compelled to do A, S's mental states were reason-responsive up to moment of decision.

Similarly, Neal Tognazzini has put it in the following way:

"My choices should depend on facts about *me*!
I'm the one who made it happen.
It's in virtue of *my* activity that the world is changed." (as cited in Timpe, 2008, p. 75).

The main concern about the sourcehood concept is to secure a variety of conditions that would be adequate for moral responsibility. It immediately gives rise to the question whether it can be compatible with determinism in such a case that is mentioned by Inwagen. If determinism is true, whatever the sourcehood concept that one has in mind, the course of events unfolds in only one possible way, there is only one possible future which is the actual one. We intuitively tend to resist the idea that the judge could not have refrained from the simplest action that one can imagine as raising one's hand, yet the logic dictates upon us the truth of the proposition that the judge refraining from raising his hand at T had already been established in a distant past when the judge was not even born. We do not either want to say that the judge

would render the laws of nature different than they are at T. Therefore, as it is also widely held by many other philosophers that no one has ever had any choice about the truth value of the proposition we have mentioned in van Inwagen's argument that the judge refrains from raising his hand at T and the death penalty had to be executed, because we do not have any choice of the natural laws. This is certainly a deep concern regarding determinism as many philosophers have agreed upon in the ages-long discussions.

The second threat to the sourcehood concept is brought about by reductionism that is embedded in the physicalist understanding of the world. If physicalism does not necessarily signify reductionism, then the reader might not consider this as a real concern since there are many philosophers who claim to be nonreductive physicalists. If determinism is true, then not only the ordinary physical structure of nature has to be what it is but also the total description of the universe in a past when there was no conscious being in conjunction with the laws of nature entails only one possible future description of universe which involves the description of all the mental states of conscious beings as well. It means that the physical conditions of the universe are sufficient to bring about any kind of mental states that one has ever had so far. Of course, this logical consequence would not be of any surprise for those who adopt the mainstream physicalist stance.

Jaegwon Kim developed an argument against non-reductive physicalism in which he argues that it is inconsistent and the mental states have to be causally inert provided the causal closure of the universe (Kim, 2008, pp. 13-22). For the most part, it is no less compelling than the Consequence Argument because it is quite difficult to find the way out of epiphenomenalism without finding oneself at pains to reject the causal sufficiency of neural states or causal closure of the universe, to which we

will come back in later chapters. Tim Crane, interestingly enough, argued that emergentism and nonreductive physicalism are metaphysically indistinguishable and if the nonreductive physicalist is to remain a physicalist at all, he should adopt reductionism on the epistemological basis in order to distinguish himself from the emergentist (as cited in Gillet & Loewer, 2001, pp. 207-224). The nonreductive physicalist is motivated to secure the high-level causal efficacy at same time holding the truth of causal closure principle that every physical event has a physical cause. The difficulty arises from the fact that from the totality of the base properties which do not involve mental properties, all the interesting mental states do unfold, yet we still want to talk about genuine high-level features of the universe such as conscious agents with novel causal properties who can act upon various reasons.

Consciousness is a very exciting problem for science since we do try hard to unravel our own cognitive abilities rather than anything outside ourselves, and contemporary cognitive science is a very promising field that might possibly unravel the causal mystery beyond consciousness. We have to develop new insights to account for free will and consciousness, given that philosophers cannot just get away with their conceptual analysis without having any reference to empirically relevant existence problem of free will which Kane discusses in *The Significance of Free Will* on a large scale. Free will matters not only for its implications about our freedom of action, but also, more importantly, it is highly relevant to our character building, namely our "mode of being", which also makes our feelings and creativity considerably worthwhile, intrinsic, and ultimate.

I hold that the faculty of consciousness which distinguishes us from ordinary physical objects cannot even in principle be ontologically reduced to neural states. Furthermore, the irreducibility of consciousness is a necessary condition of

sourcehood provided that it is one of the properties that ensures the explanation of our actions are not constituted by the matters of fact that lie beyond ourselves. The faculty of consciousness is so widely confused with other arbitrary, observer-relative high-level features of physical objects or artefacts in many analogies and thought experiments that are used by philosophers. It is hard to resist the seemingly obvious fact that consciousness is just a by-product of highly complex biological machine that is called “the brain”; however, the hard question of how it arises from the brain still remains one of the most compelling mysteries about the world.

We will not cover all varieties of compatibilism and incompatibilism in this paper. For instance, even if there is only one possible future, it does not mean that determinism is true in a physicalist framework. The reader will recall that some philosophers hold that God’s foreknowledge entails predestination. Provided the omniscience of God, if God has the knowledge of the future facts as necessary truths, then it is necessary for these future facts to unfold exactly in accordance with God’s knowledge because He cannot possibly be wrong about these facts. This is another serious problem in theology which has not anything to do with ordinary determinism – free will discussions.

The incompatibilist is generally amazed by the compatibilist’s insistence upon the idea that one can be regarded as free even if he could not have done otherwise ever whereas the compatibilist has very much difficulty in making sense of how indeterminism provides any more control than the scope of control which already consists in reason-sensitive, uncompelled agents in a deterministic world. My observation is that the difference between these two camps goes way beyond some simple premises and it does not in any sense stem from a verbal debate. They not

only rest upon different premises like "uncompelled, reason-sensitive versus alternate possibilities" but also, they consist in quite different worldviews in general.

There seems to be two kinds of frame of meaning through which we justify our concepts of free will as worth wanting: one of which may be labelled as "the Spinozistic Worldview" and the other as "the Theistic Worldview". The reader might consider these categories as an oversimplification of quite complex positions, and I do not either claim that these categories may apply to all sorts of philosophical positions on free will, except the ones that we will be discussing in this paper such as Searle, Kane and Dennett's distinctive positions.

According to Ilham Dilman's interpretation of Spinoza on the question of how one can be free in a deterministic world:

> Spinoza's answer is that we shall be free by understanding and acceptance – understanding that we are part of a bigger whole and seeing that, as such, nothing that happens to any one of us could have fallen otherwise, given the state of the whole from which it arises. Once we see this clearly, we shall stop fretting and we shall come free from the cycle of ego-centric, reactive transactions in which we are puppets on a string (1999, p. 129).

Spinozistic/Deterministic worldview seemingly rests upon the central theme of aesthetics. The frame of meaning is oriented in the entirety of the universe, not in the standpoint of individual. In such a world, an agent is like a stage actor who is following a certain scenario. Although the scenario of the play is totally out of one's control, nevertheless the actor can engage in the flow of it and make sense of his existence as a part of this aesthetic tragedy. Of course, most part of the story will involve reactive attitudes that were explicated by Strawson in the daily encounters of life: one will genuinely have the feelings of resentment and indignation for others; however, these individualistic transactions are not to be taken seriously. What is genuinely valuable is the world in its entirety and individual moral responsibility is actually an illusion from the fundamental standpoint. Whether you were born as a

slave or a king, your individual existence does not mean anything other than the role that you partake fora short period of time. One may find the similar view in Dennett's account though I do not claim that Dennett is a Spinozist. Dennett is trying to convince us that determinism does not pose a real threat to real life cases of free action on the grounds that the nature is not an agent above us, who can manipulate our actions in accordance with any personal agenda; and therefore, rational agents have a certain kind of elbow room through which they can find their way out of real life scenarios without worrying about manipulation arguments. Although nature is pulling our strings, a rational agent is not a sphex that is devoid of genuine rational deliberations (Dennett, 1984, pp. 1-19).

On the other side, incompatibilism/libertarianism is nurtured by a quite different worldview that we may call "the Theistic Worldview". It is the traditional worldview that is embedded in the Abrahamic religions although there are various interpretations of human agency in those religions. I think the central theme of this worldview that we should focus on is the unique responsibility of the individual. It is the God who gives the individual just one chance for a short lifespan and expects him to act in accordance with the divine and moral laws and pass the divine test. Despite many challenging temptations throughout his life, the individual is thought to have genuine causal powers to change the course of events and thoughts that are under his control. Because of this genuine and unique power, it is justified to hold him responsible for his actions and non-actions that he brought about (or not) during his lifespan. The individual is the true source/author of the actions for which he is genuinely responsible, he does not have the privilege of going with the flow as in the deterministic case simply because he is capable of changing the flow and partly responsible for it. Even though there are many real-life cases out of control, we are

responsible for our own deliberations, actions, and partly for the personhood that we developed on our own. Robert Kane's Ultimate Responsibility Principle (UR) is a clear signification of this worldview, and also John Searle defends a robust agency and self that have genuine causal powers in the face of alternate possibilities.

The ultimate problem of determinism and reductive physicalism with regard to sourcehood principle is the fact that the agent himself and his actions are merely constituted by impersonal forces that lie beyond the agent's control. It may turn out determinism is true and the world has always had only one possible future, and consequently we may have to give up on genuine free will concepts that are – I think – incompatibilist concepts which we make sense of genuine responsibility by. I partly agree with Dennett on that we do not have to worry about the nature's pulling of our strings, however, the lack of manipulative, evil intentions in nature does not change the fact that our strings are always pulled by a combination of natural forces over which we do not have any control. Although I'm an incompatibilist, I also consider the Spinozistic Worldview as a viable option. Going back to the metaphysical concerns again, Derk Pereboom endeavours to show us that there is no fundamental difference between the nature's pulling our strings and ordinary manipulation instances in his famous Four-Case Manipulation Argument:
There are four scenarios in all of which Professor Plum kills White. In each scenario, Plum is reason-responsive to moral reasons, he does not act out of an irresistible desire, but nevertheless Plum's egoistic reasons weigh heavily in each case for different reasons. His second-order and first-order desires are in accordance with each other and this makes him morally culpable in Frankfurt's conditions, however, the argument tries to show that there is no fundamental difference between

manipulation scenarios in which ill-intended manipulators do the job and the logical consequences of determinism.

> Case 1: Professor Plum was created by neuroscientists, who can manipulate him directly through the use of radio-like technology, but he is as much like an ordinary human being as is possible given this history. These neuroscientists manipulate him to undertake the process of reasoning by which his desires are brought about and modified. They do this by pushing a series of buttons just before he begins to reason about his situation, thereby causing his reasoning process to be rationally egoistic. Plum does not think and act contrary to character since his reasoning process is often manipulated to be rationally egoistic. (…)
> Case 2: Plum is like an ordinary human being, except that a team of neuroscientists has programmed him at the beginning of his life to weigh reasons for action so that he is often but not exclusively rationally egoistic, with the consequence that in the circumstances in which he now finds himself, he is causally determined to undertake the reasons-responsive process of deliberation and to possess the set of first- and second-order desires that result in his killing White. (…)
> Case 3: Plum is an ordinary human being, except that he was determined by the rigorous training practices of his home and community so that he is often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1 and 2). His training occurred when he was too young to have had the ability to prevent or alter the practices that determined his character. As a result, Plum is causally determined to undertake the reasons-responsive process of deliberation and to possess the first- and second-order desires that result in his killing White. (…)
> Case 4: Physicalist determinism is true, everything in the world is completely physical, and Plum is an ordinary human being, raised in normal circumstances, who is often but not exclusively rationally egoistic (just as egoistic as in Cases 1–3). Plum's act of killing White results from his undertaking the reasons-responsive process of deliberation, and he has the specified first- and second-order desires. He also possesses the general ability to grasp, apply, and regulate his behaviour by moral reasons, but in these circumstances the egoistic reasons weigh very heavily for him, and as a result he is causally determined to murder White (Fischer, Kane, Pereboom & Vargas, 2007, pp. 94-95).

In each case, Plum engages in reason-sensitive deliberation as to whether to resist the temptation by his egoistic reasons or not, yet the conditions for bringing about the act of killing are already developed by the causal relata that lie beyond the control of him. In the context of these scenarios, the sourcehood principle is violated and Plum is not properly culpable of his act of killing although this is quite counter-intuitive in real-life scenarios. On the other hand, we should be careful about the

logical consequences of physicalist determinism. Although physicalist determinism degrades the robust sense of agency, it does not wipe it away altogether.

The causal determinism also entails that the mental states are entirely determined by physical conditions. It has been argued before that the causal closure of physical universe renders mental phenomena causally redundant, which we will discuss. Donald Symons says:

> The view that only physical events can cause other physical events, and that mind is mere epiphenomenon of brain activity, appears untenable if natural selection is the primary creative force in nature. An omnipotent deity presumably could, if it wished, create mind as a functionless epiphenomenon, but natural selection implies (at least to some scientists and philosophers) that were mind functionless it would long ago have withered away, not become elaborate as it obviously has (1981, p.39)

Whether our agential powers are merely constituted by brute physical forces or not, human rationality has undeniably been bringing about changes in the world in an evolutionary perspective as well. That is the reason why even reductive physicalists refer to causal powers of consciousness/rationality in their "Just So Stories" about the human evolution. In the course of evolutionary history, humankind seems to have gone beyond the level of sphexishness and acquired a larger range of elbow room, which Dennett has clearly explicated. Regardless of the truth of physicalist determinism, I also argue that rationality matters and makes a real difference in the world.

Furthermore, the truth of physicalist determinism does not entail every kind of imaginable possibilities of manipulation. This is a crucial point which is sometimes overlooked by the libertarians. For instance, let us suppose that an evil demon created this universe on a physicalistic determinist basis and he had certain ill-intentions on me. He had every physical means to create certain conditions on which he has the ultimate control. He creates the universe in such a way that I would

be born in 1992 exactly the same way and I would be tempted to perform very minor unethical actions starting from 2017 onward, say downloading hundreds of books in PDF form without paying anything for it. My brain was designed in such a way that even though my deliberation was reason-responsive up to the moment of decision, my selfish reasons would outweigh my ethical judgements. My actions had a certain financial cost for the publishing houses no matter how minor it was. Theoretically speaking, it was a kind of digital theft. However, it would almost be impossible for this demon to manipulate me to steal an actual hard copy from a publishing house even if I could have got away with it; in order to achieve this, he would have to have designed the conditions in such a way that my deliberative processes would break down before the moment of my decision because of (let's say) some swerves of atoms in my brain or tumour or anything that will wipe away my reason-sensitiveness. So, the reason-sensitiveness of a person provides him with a certain level of robustness and stability with regard to his character and decision-making. We should ask the question whether you could, for instance, create a person that would have a decent life and moral integrity until his 50's and then he would start a killing spree on his 51st birthday without at the same losing his reason-responsiveness; even if you were granted all the physicalist possibilities and complete determinism; I hardly think so. There are, of course, very clear evolutionary reasons for this as the reader can quite easily imagine. Partly by virtue of the temperament of yourself and your ancestors, here you are. Especially in a society that is predicated on and sustained by social/moral relations, one has to be careful about his/her attitude and abide by the rules regardless of the truth of determinism; and clearly one needs to have a healthy neurobiological structure as a

legacy of his/her ancestors, which may facilitate the subject to sustain his/her temperament and moral integrity.

Fischer et al. (2007) argues that even if we woke up some day and see the newspaper headline “Causal Determinism is True”, it would not and should not change our ordinary view of human beings as robustly morally responsible agents and deeply different from animals. There are two different aspects of the claim which should be evaluated independently. It is true that human beings are deeply different from animals by virtue of their rational capacities which provide them with linguistic abilities and much larger scope of reason-sensitiveness to act in the world, yet it does not change the fact that they are bound to act in only one way no matter how sophisticated their abilities are. Fischer, in fact, tries to point out an important distinction between the regulative and guidance control here. He establishes the distinction by reference to Locke’s famous example as follows:

> Here a man is transported into a room while he is asleep. When the man awakens, he considers leaving, but he decides to stay in the room for his own reasons. Locke says he voluntarily chooses to stay in the room and voluntarily stays in the room. Unbeknownst to the man, the door to the room is locked, and thus he could not have left the room. According to Locke, the man voluntarily stays in the room, although he does not have the power to leave the room. He exhibits a certain sort of control of his staying in the room (what I would call guidance control), even though he cannot do otherwise than stay in the room (and thus, he lacks the regulative control over staying in the room) (2007, p. 44).

In the actual sequence of his choosing to stay in the room, the man still has all the interesting abilities that are peculiar to humankind, one of which is the reason-responsiveness aspect. In a deterministic universe, a human person does not have the ability to do otherwise any more than a piece of rock does; but compatibilists are trying to make us focus on the actual sequence abilities which we would still preserve regardless of the truth of determinism. So, I agree on that we would still be radically different from a sphex even if determinism is true.

On the other hand, it is hard to say that we would still be robustly morally responsible agents. Most probably P. F. Strawson is right in arguing that metaphysical arguments would not change our reactive attitudes to each other, which are deeply embedded in our daily practices, and Dennett similarly holds that people do not really care about the metaphysical implications of the traditional arguments except philosophers who entertain themselves with the so-called traditional dogmas. However, we should still take the metaphysical implications seriously in order to find our way through the possibility of determinism. What determinism would make us devoid of is the unique responsibility and centrality of the individual as I briefly explained in previous paragraphs, but nevertheless there is still the Spinozistic Worldview that we may consider as a valuable option at the cost of giving up on genuine responsibility.

The reason behind the absence of genuine responsibility can be clarified by the Transfer of Powerlessness Principle ("Compatibilism", 2018) in the Consequence Argument which signifies the absence of any fundamental difference between my powerlessness about the conditions of the world before I was born and my powerlessness on the moral choices that I made throughout my life. Of course, I have been exerting guidance control all over them, however I did not have the regulative control, namely the power to do otherwise. This line of reasoning relies on the assumption that only with power, comes the responsibility; in other words, "should" entails "can", and Transfer of Powerlessness implies that an agent does not possess the power to do other than he has done in the actual world. This fact does not override the reason-sensitivity in the actual sequence of deliberation, yet in order for us to be able to say that the agent possessed the power to act in a different way, the

prior conditions of the world must have been otherwise, which is something totally out of our control.

Therefore, I will argue that the actual sequence control is not sufficient for the sourcehood; in Fischer's terms, one also needs the regulative control in order to be genuinely responsible for his actions. We have very similar reasons with the ones that are explicated in the Consequence Argument for this claim, and the Transfer of Powerlessness is the crucial element of this argument. Similarly, I hold that the guidance control is not sufficient for sourcehood because all of the state of affairs in a very distant past were sufficient for the constitution of one's agency, and no one had ever had any control over those states of affairs. One may object that it is quite reasonable to claim the ownership of those actions even though their initial conditions in a distant past were not under our control, that is to say we are just how the nature brought us about. It might be argued that whatever we refer to as "I" is a construction of the (impersonal) physical forces operating in our bodies, and there is no reason to exclude them from our self-identity and claim that we are the prisoners of natural forces regarding the assumption of determinism. Sam Harris has put it in a similar way: "You are not controlling the storm; you are not lost in it. You are the storm." (2012, p.7).

These objections are understandable in a naturalistic point of view and the Sourcehood Principle that I proposed might be impossible to be met. However, we need to examine the logical structure of free will; it is an indispensable assumption in our daily communications, I-Thou encounters, or reactive attitudes toward each other. We make causal explanations for natural events whereas we make reason explanations for the attitudes of human beings, as Scruton claims, there is a fundamental difference between subjects and objects in our minds (Scruton, 1996). I

think the sourcehood aspect is quite fundamental for those who hold that individuals are responsible for their actions. We rarely consider each other as simply biological beings, unless you are a surgeon, but rather we see each other as accountable and free agents who are the true source of their actions.

Robert Nozick argued "No one has ever announced that because determinism is true thermostats do not control temperature." (1981, p. 315). We will discuss the question whether there can be any fundamental difference between physical objects or artefacts and human agency in physicalistic framework. For many of us, I think, the distinction between thermostats' having a certain control over the temperature and a rational agent's having control over his significant moral actions is not just a matter of complexity. In another case, many of us might feel disturbed by the analogy between the intentional stance that we attribute to thermostats or IBM's chess computer Deeper Blue and our own intentional control on our actions. These examples are, again, about two different concerns about sourcehood: determinism on the one hand, and the reductive physicalism on the other. I briefly argued against the compatibility of determinism and our authorship of our actions in a genuine way, I also claim to the contrary of Dennett that if "the personal stance" that we attribute to rational agents has no fundamental distinction from the "physical stance", namely if it is only about my epistemic incapability to see a person as a moral agent rather than a totally physical computer that is constituted by hundreds of sub-systems like small robots (Miller, 2012) as Dennett holds, then our perception of each other as moral agents is metaphysically unwarranted, and we are not justified in acting in such a way on a metaphysical basis; but this implication does not bother all philosophers.

I will try to show that our sourcehood principle is incompatible with most sorts of physicalistic or naturalistic accounts. It seems easier to define the boundaries

of physicalism rather than defining naturalism. What I mean by physicalism is that every possible event or non-abstract entity and their entire causal structure can be deduced from the science of physics, namely atoms, subatomic particles and physical laws including the conceivable future physical laws yet to be discovered. It seems that if a naturalistic theory of mind/consciousness is not to be question-begging, then it is required for it to account for consciousness or mind in scientific terms that do not themselves postulate any term or entity over and above biology / chemistry / physics. One should know that it is exactly the point put forward by Dennett (1991) who is considered to be one of the forerunners of naturalistic view. John Searle is also quite well-known in this field, but we will see that "biological naturalism" is simply incoherent and thus his free will theory falls with his biological naturalism.

Immanuel Kant asserted that free will is an indispensable background assumption of moral life and action, although we cannot sort its nature out by rational inquiry since it is one of the key notions that supposedly belong to noumenal realm; we are not able to acquire positive knowledge about it in the same way we acquire knowledge about natural world. Kant considered the compatibilist position as 'subterfuge' because of the incompatibilist nature that his notion of free will consists in, although it is also highly controversial. One of the primary aspects of this reasoning is that human actions and natural causes have inherently different kinds of explanations. A very similar line of thought is explicated in Searle's *Freedom and Neurobiology* (2007) where he draws the distinction between these two kinds of explanations one of which relies upon a non-Humean self acting on a particular set of reasons other than the other set such as in choosing a presidential candidate A rather than B for an election (pp. 53-56). I will try to show why it is much better off holding the distinction between agent causation as opposed to compatibilist positions some of

which rely on reductionist premises; and I will particularly focus on three philosophers: John Searle, Robert Kane and Daniel Dennett, all of whom claim to be naturalists.

CHAPTER 2

THE PROBLEM OF REDUCTIONISM IN SEARLE'S ACCOUNT

2.1 The untenable discrepancy in Searle's weak emergentism

In this thesis, I will try to show that the main problem with regard to agency which libertarians are trying to preserve arises from the reductionist tendencies in some philosophers' accounts of libertarian free will. Searle is explicitly a libertarian philosopher within the analytic tradition. In his books titled *Freedom and Neurobiology* and *Rationality in Action*, he proposes several critical points concerning the logical structure of free will. Searle tries to decipher the concept of free will on the basis of certain logical components which are supposedly indispensable in order for us to make sense of genuine free will. Having agreed upon this logical structure which I will clarify in the next chapter, I will argue that his proposal of logical structure becomes untenable because of his commitment to causal reductionism.

Searle has endeavoured to find a solution to the mind-body problem for decades by forming a naturalist account of mind. His account has very peculiar aspects on the consideration of the discrepancy between the ontological irreducibility of consciousness and causal reductionism. Despite the several strengths of his analysis as to how the mind emerges out of 'mindless physical entities', his naturalist position seems more likely to postpone mind-body problem rather than illuminating it any further after all the ages-long discussions on the subject. The analogies he pointed out regarding the possible ways that we might solve the problem with further developments in neuroscience are highly problematic. Any analogy or thought experiment has its own limits in serving the purposes of philosophical arguments,

however, the most popular analogical arguments concerning mind-body problem break down right at the point where they are supposed to make any useful contribution to the subject. One of the most popular examples is the centuries-long discussion between vitalists and mechanists on the origin of life. If science has been able to show that we do not need "élan vital" (vital properties) in the explanation of how life emerges out of chemical structures, then, they claim, we do have good reasons to hold that mind-body problem will be settled by further neuroscientific developments.

For instance, when I accidentally hit my arm on table, my experience of pain can be accounted for at two levels. At the first level, the signalling of my nerves on my arm gives rise to excitation of neurons in my brain in appropriate areas and this is the causal history of my experience of pain. The second and the higher-level description is about mental state of pain, I suddenly perceive my arm hitting on the table and a rapid increase of pain shakes my unified conscious realm. According to Searle (2002, pp. 114-115), mental states such as pains, desires, beliefs are not ontologically reducible to neural correlates but in terms of causality, there is no further explanation once you find out how the nerves react to my arm's hitting on the table; that is the whole causal/explanatory story.

E. J. Lowe (1996, pp. 78-79) criticized Searle on the grounds that his account renders mental states causally inert. Searle, in fact, delays the solution of mind-body problem by carrying the burden of explanation to future neuroscience. The fact that we find the discrepancy between the above-mentioned two levels of explanation derives from our lack of knowledge about how these lower level neural states produce conscious experiences. It follows from the subjective ontology of consciousness that the causal explanation of mental states will ultimately be settled

by neuroscience despite the fact that we will never be able to ontologically eliminate them (Searle, 2007, p. 50).

Many physicalists/naturalists share the same intuition with respect to the mind-body problem that future neuroscience will settle the problem just as the nature of life has almost been settled with the developments in biochemistry. There is a considerable conviction among physicalist philosophers and neuroscientists on the problem of consciousness that it is somehow very proximate to their intuitions on the existence of phlogiston. In my view, those analogies are not helpful given the radical ontological distinctness of consciousness. Even the most powerful analogy which is the case of the origin of life does not fit into purposes of arguments simply because anything we know about the world is explainable in terms of third-person features of the world except the first-person, intentional features of mental phenomena. Interestingly enough, Searle clearly agrees upon what I have just suggested (Searle J. , 2016) however what makes his biological naturalism untenable is his commitment to causal reductionism.

At this point, we should put forward two formulations of the sourcehood principle in order for us to make a detailed analysis of the Searlean account:

First formulation of sourcehood principle (concerning determinist threat): For a subject S to be the source/author of the action A, the causal relata or the set of reasons that justify doing A should not go beyond the causal autonomy of S. In other words, the causal relata of A should not be sufficiently constituted by the conditions on which S has no control.

Second formulation of sourcehood principle (concerning reductive physicalism): For a subject S to be the source/author of the action A, the agential

powers of S that bring about A should not sufficiently be constituted by impersonal physical forces on which S has no control.

Searle claims that the hard problem of mind and body mainly originates from a conceptual turmoil nurtured by the philosophical traditions such as Cartesian Dualism and Materialism. Since the mental and physical phenomena are considered as mutually exclusive categories in all of them, they lead us to a certain philosophical dead-end from which Searle (2002, p. 9) wants to find the way out by trying at first to dispose of the basic terminology or by simply giving up taking them too seriously within a biological framework.

This new position is called "biological naturalism" (BN), and interestingly enough for me Searle is a libertarian about free will at the same time. The baffling hard problem of consciousness is formulated as the higher level or system feature of the brain; and consciousness is a biological phenomenon like digestion. Let us also clarify the basic premises of Biological Naturalism:

1) All of our mental phenomena are caused by lower level neuronal processes in the brain and are themselves realized in the brain as higher level, or system, features. (Searle, 2002b, p. 57) The form of causation is "bottom up," whereby the behaviour of lower level elements, presumably neurons and synapses, causes the higher level or system features of consciousness and intentionality. (This form of causation, by the way, is common in nature; for example, the higher-level feature of solidity is causally explained by the behaviour of the lower level elements, the molecules.)

2) Consciousness is ontologically irreducible to third-person qualities of the world. The material world is publicly accessible and is pretty much as described by physics, chemistry, and the other hard sciences; but the conscious, experiential, phenomenological world is not publicly accessible. It has a distinct private existence.

We know it with certainty from our inner, private, subjective experiences (Searle, 2002b, p. 58).

3) We are living only in one world (Monism). There are not two fundamental categories such as the mental and physical (Searle, 2002b, p. 59).

4) Consciousness is causally reducible to brain processes, because all the features of consciousness are accounted for causally by neurobiological processes going on in the brain, and consciousness has no causal powers of its own in addition to the causal powers of the underlying neurobiology. Consciousness does not name an extra entity over and above its neurobiological base, consciousness is the state that the brain is in (Searle, 2002b, p. 60).

When it comes to the ontological irreducibility of consciousness, Searle (2002b, pp. 60-61) recognizes the breakdown of analogies like solidity of a wheel as the higher-level property, "there is nothing in the car engine except molecules", or liquidity of water which does not show itself in individual atoms. So, it means that there is only one phenomenon in the universe that is ontologically special, in other words, it is only in the case of consciousness that causal reduction does not entail ontological reduction. Searle, in his famous article titled *Why I'm not a Property Dualist*, endeavours to diverge from property dualism, in other words, strong emergentism which is the idea that proposes some higher level structures in the universe give rise to novel causal structures which cannot be explainable by the base properties and those novel higher level features facilitate downward causation in a sense radically different from Searle's analogical examples like the solidity of wheel or liquidity of water.

The concept of consciousness as a simply higher-level feature of the brain raises the concerns about epiphenomenalism against which Searle does not seem to

have satisfactory counter-arguments. I do suppose that if all the causal powers of consciousness can be reduced to neural firings in the brain, then consciousness itself is simply causally inert on the grounds that we also want to avoid causal overdetermination. At this point, Searle makes an assertion that seems clearly naive to me:

> Nobody thinks that we are forced to postulate that solidity is epiphenomenal on the grounds that it has no causal powers in addition to the causal powers of the molecular structures, nor do they think that if we recognize the causal powers of solidity we are forced to postulate causal overdetermination, because now the same effect can be explained either in terms of the behavior of the molecules or the solidity of the whole structure (Searle, 2002b, pp. 3-4).

It reminds me of the famous paper titled "Epiphenomenalism and Eliminativism" by Trenton Merricks. Based on the Eleatic Principle, Merricks (2001) asserted that baseball does not exist over and above the baseball atoms. In ordinary language, there seems not to be any problem in saying that shattering of the window is caused by hitting of the baseball. Merricks holds that if we grant the truth of Eleatic Principle (To exist is to have causal powers.) in conjunction with the overdetermination principle, then baseball does not do anything to the window except the atoms composed baseball-wise. So let us keep it short for the purposes of this proposal: Applying this line of reasoning to biological naturalism, how can a causally autonomous, ontologically irreducible agent who has the genuine power of acting on reasons arise from neural processes and how is it possible for this agent to be a genuine one given the causal reducibility of him in the context of Eleatic Principle? I don't consider myself as a nihilist on the ontology, but I surely find the Eleatic Principle quite useful while reflecting on the difficult questions in philosophy of mind.

Let us focus on the Overdetermination Argument that is suggested by Merricks in detail:

> (1) The baseball—if it exists—is causally irrelevant to whether its constituent atoms, acting in concert, cause the shattering of the window.
> (2) The shattering of the window is caused by those atoms, acting in concert.
> (3) The shattering of the window is not overdetermined.
> Therefore, (4) If the baseball exists, it does not cause the shattering of the window (Merricks, 2001, p. 56).

While reflecting upon this argument, I paused for moment and consciously decided to light a cigarette. It is quite illuminating to touch upon the problem of epiphenomenalism in the framework of overdetermination. So, I will try to make an analysis of biological naturalism in this regard, in order to find out whether there is any crucial similarity between conscious actions and causal efficacy of ordinary physical objects:

1) My conscious decision to light a cigarette – if it exists – is causally irrelevant to whether my neurons, acting in concert, cause the lighting of the cigarette.
2) The lighting of the cigarette is caused by these neurons, acting in concert (in a certain structure of neural pathways).
3) The lighting of the cigarette is not overdetermined. (Given the Causal Closure Principle)

Therefore, 4) If my conscious decision to light a cigarette exists, it does not cause the lighting of the cigarette.

Premise (1) seems absurd because of the ontological irreducibility of mental states. I am immediately aware of my desire to light a cigarette even though I do not know how the mental states of mine cause my physical behaviour and bring about changes in the world. It is also unacceptable with regard to biological naturalism provided that it does not allow for extra causal factors over and above the total structure of the brain unlike property dualism/emergentism. So, whatever

consciousness is capable of, it is capable of doing it by virtue of the causal powers of the brain; in other words, consciousness does not have a causal autonomy of its own. Therefore, premise (2) fits into BN quite well, and also premise (3). I think the central motivation behind the causal reducibility of consciousness into neural states is to avoid violating the Causal Closure Principle (CCP); Searle achieves this goal, however, at pains of giving up on causal autonomy or novel causal powers of consciousness.

These unsatisfactory results give rise to the question of how to deal with the ontological status of consciousness with regard to the Eleatic Principle. It is not difficult to be an eliminativist about baseballs whereas it is impossible to do so in the case of consciousness (given the Cogito Argument). Even if you are not a nihilist on the ontological issues, I suppose it is easier for one to accept that baseballs do not exist than to accept that consciousness is an illusion. I accept the logical consequences of the Cogito Argument that is proposed by Descartes, as much as John Searle does. Provided the second premise of BN, consciousness has a first-person ontology, it is not possible to eliminate it in the way science showed that rainbows are illusions.

What is the difference between these two kinds of epiphenomenalism? In the case of baseball, it is purely constituted by physical particles all of which can be described by physical properties or in other words third-person qualities. However, it is not possible to give a total description the mental states purely in physical terms or third-person qualities. Therefore, again, we should note that it is only in the case of consciousness that causal reduction does not allow for ontological reduction. Even though one may claim that the neural states are sufficient to bring about corresponding mental states, nevertheless the first-person phenomena cannot be

reduced to the third-person. At the first glance, it is not easy to make sense of this assertion with respect to Eleatic Principle, one should ask then what does consciousness do at all and what is the distinction between these two instances?

Epiphenomenalism case for baseball: Provided that the total description of baseball can be given in third-person terms and all the causal structure of baseball itself can be reduced to its constituent elements, namely its atoms, baseball in its entirety has no causal power at all. Applying the Eleatic Principle to baseball, we may plausibly argue that it does not exist in fact.

Epiphenomenalism case for consciousness: Assuming that the causal structure of consciousness is equivalent to the causal structure of its neural correlates and there are no extra causal factors over and above the neurobiology of it; it follows that whatever consciousness brings about in the physical world, it brings it about by virtue of its third-person, underlying neural structure. It seems quite similar to the first case of epiphenomenalism above. There is one important difference that is derived from its ontological irreducibility into third person qualities. Therefore, the Eleatic Principle does not allow us to eliminate consciousness in the same way we eliminate baseball.

What I mean is in fact quite simple: Epiphenomenalism case for consciousness only implies that mental states are causally inert with respect to physical states of the body/world, given the Causal Closure Principle. However, there is another legitimate question that arises from this: Given that consciousness is causally inert with regard to any physical event in the world, then what does it change if we eliminate consciousness, leaving only brain states? The reader should recall the implications of the Zombie Argument at this point, and my answer to this question is very simple: Imagine someone is captivated as a prisoner of war and

tortured to death. You recall that he was one of your best companions whom you met at the time of war and it was him who saved your life for a couple of times. A few weeks after his captivation, your team carries out a surprise attack on the enemy camp and you find your friend's deceased and obviously tortured body lying on the ground. In the end, it turns out that the research and development team in the ministry of defence, unbeknownst to you, had succeeded to develop zombie soldiers who are functionally identical with human beings but lack mental states or at least capacity to feel pain. (I do not claim it is physically/biologically possible.) One of the firmest convictions concerning our phenomenology is that suffering matters, yet your companion did not have any accompanying mental states although he seemed to you that he was strikingly identical to you by virtue of his functional dispositions. I doubt this news would be much of a relief to you even if you are rationally convinced that he was not a real person; however, it is anyway a huge difference that he did not have any mental states.

When we compare a zombie soldier with his genuine counterpart which actually has all the accompanying mental states in addition to functional abilities, it would not make any difference as to how the course of events unfold, in other words it does not have any additional effect on the physical history of the universe. Nevertheless, Eleatic Principle does not necessarily lead us to ignore the fact that the former is a zombie with no mental states whereas the latter is a full-fledged human being. One may object that Eleatic Principle in fact implies that to exist is to have physically causal powers or to have the capacity to change something in the physical world. Perhaps that is the true formulation of the principle, however, this formulation has some undesirable results if one is willing to accept the Cogito Argument at the same time. I know I exist along with my mental states more surely than I know

anything else exists in the world. On the grounds of my firm conviction, I insist that all I can say is that my mental states exist, and they are happening in the world. If something real and which happens in the world has to be physical whatever its properties are, then I would warn the reader about such an inflationary definition of the physical. So, one might take this route and accept that mental states are physical entities despite their first-person ontology at the cost of the serious discrepancy between their essential properties (first/third-person qualities). Even if mental states were epiphenomenal with regard to physical closure of the world, they might be causally efficacious in bringing about other mental states. However, we do not know that; and I do not know how we can know what they are causing if we cannot acquire any knowledge about it from the third-person point of view. In the studies on emotions, for example, psychologists are ontologically committed to the existence of emotions and there are of course quite fruitful studies in this area; but we do not know whether those emotions can causally be reduced to neural states. If they are all functionally reducible, we may not have any further clue about what they are capable of doing on their own, not solely by virtue of their neural basis. Searle is right in the sense that we cannot doubt the existence of consciousness even if its causal powers are exhausted by underlying biological processes; in fact, Searle thinks that it is the case but doesn't acknowledge that it leads to epiphenomenalism.

Expanding the definition of the physical so as to cover everything happening in the world is philosophically unsatisfactory, because it renders the attributions of physicality uninteresting and trivial. If the physicalist claims to engage in a serious philosophical inquiry concerning consciousness, the attribution of physicality to any possible mental phenomenon has to mean something other than whatever possibly exists. So, I will repeat what I claimed in a different way in order to clarify my

concern in the framework of the Eleatic Principle. Imagine two different possible worlds; in the universe A, God creates human beings with exactly the (functionally) same dispositions however, they lack mental states whereas, God creates functionally the same human beings in the universe B, with the exactly the same biological/physical history but also equipped with mental states. Considering the physical history of the two universes, they are exactly the same however, it is only in the universe B that human beings have mental states such as pains, desires, belief etc. Leaving aside the attribution of physicality to mental states, I simply argue that the course of events unfolds differently in those universes despite their physical identity because there are extra mental events happening in the B-universe. All I can say at this point is that mental states are real and happening in the world, regardless of their exact ontological status. However, this might be tricky.

Noting that I am certainly a realist about the mental states and qualia, the assertion of mine in the previous paragraph might be objected to in the following way. Let us consider the example of the judge's raising his hand in the case of Consequence Argument. If the judge raises his hand, he would have given pardon to the convict and prevented the execution of death penalty. In this context, the judge's raising of his hand constitutes his giving pardon to the convict; in other words, his giving pardon is nothing over and above his raising of hand. Of course, there are long debates concerning the nature of event causation versus fact causation; yet my point is simple enough here. Applying the same chain of reasoning to the argument in the previous paragraph, the occurrence of mental state M is causally nothing over and above the corresponding physical/neural event P. Kim's (2008) Causal Exclusion Argument (CE) points out that provided the causal sufficiency of P to bring about M simultaneously and also the subsequent physical event P2, the mental state M does

not do anything by itself. By of virtue of the fact that mental states accompany physical states in the universe A, there is a factual cross-world difference between A and B. But the crucial point is that it is only in A, certain physical/neural states also cause the mental states, thus they are two different worlds, even if they are physically the same, one of them contains only zombies. In a physicalist universe, once mental states are caused by brain, they might not have any further causal power on their own, as I said, we might not know what they are capable of doing from a third-person point of view; even though we know that our mental states exist.

This debate could go on much longer, but let us also mention another argument at this point. Neuroscientist Mary in the black and white room is an epistemic counter-argument against epiphenomenalism. Suppose that Mary learns everything about colours and the visual system in the brain, sitting inside the room, except she does not have any qualitative experiences of colours. Once she steps outside the black and white room, it is important to ask whether she acquires any knowledge in addition to what she learned inside the room. I will argue, she does. Assuming that the brain is a just a complex computer whose causal powers can be explained on a purely syntactical basis; some thinkers may argue that the syntactical/informational basis of the colour-vision is sufficient to explain everything about the behaviours that are causally relevant to those visual experiences. Let us also suppose that Mary has serious difficulties in explaining the nature of her qualitative experiences following the days after she stepped outside. Now, Mary might start arguing against the Information-Processing Theorist (IPT) that her qualitative experiences cannot causally be explained by the underlying syntactical structure. She now argues that the fact that they are having this debate is causally traced back to the facts about her experiences that cannot themselves be sufficiently

explained by syntax; therefore, her having qualitative experiences makes a physical difference in the world, namely having an actual discussion. IPT replies that the syntactical background of her experiences – unbeknownst to her (like most of our cognitive abilities unbeknownst to us) – makes it causally sufficient that they are having that conversation at the time, by various feedback loops, entangled and self-scanning systems in the brain. So, IPT actually opposes Mary's counterfactual analysis of causation on the grounds that consciousness, being the higher-level feature of the brain, makes the counterfactual analysis irrelevant in this case.

On the other hand, being a critic of the Information-Processing Theory and epiphenomenalism[1], Searle (2007) makes a similar assertion:

> Epiphenomenalism is sometimes said to be explained by counterfactuals. Multiple causes apart, the truth of, "Even if A had not occurred then B would still have occurred", is supposed to be the test for whether A is epiphenomenal. But this test is at best misleading. Assuming that both the experiences of the gap and the final decisions are fixed at the neuronal level, then if the experiences had not occurred, the decision would not have occurred, or at least its occurrence would not have been guaranteed, because they are both caused by the same neuronal processes (p. 69).

I also grant that counterfactual analyses might be misleading on the grounds that we do not quite know whether the underlying neurobiological structure is sufficient to bring about those experiences, and also it might turn out that many of such dispositions are in fact fixed by unconscious processes.

One of the primary motivations of my thesis is to emphasize the issue of top-down causation in the broader context of the discussions concerning epiphenomenalism, subjective experiences, physicalist reductivism. There is more than one way to talk about reductivism and epiphenomenalism about mental states. The disturbing element of this discussion is clearly that if reductivism or

[1]I do not claim that all versiouns of Information-Processing Theory entails epiphenomenalism.

epiphenomenalism is true, then I do not think that any plausible version of libertarianism can survive. Because I think it is clear that without top-down causation, there is no (libertarian) free will. Let me make a distinction between two versions of reduction of mental states; one of which is methodological reduction. Let Ps stand for physical/neural correlates and Ms for mental properties in the following context. Let us also suppose that in order for a conscious behaviour to happen each P must co-occur with an M. So, the form of a conscious behaviour is MP, and Ps, all by themselves fall short of bringing about conscious behaviours. What if Ms are passive elements of the nature, which remain dormant until they are activated by a physical stimulus in the brain? In this case, we can in principle predict anything just by looking at third person/physical properties of the brain and body. Ms might still be (metaphysically) necessary conditions of a conscious behaviour, and we might still claim that they are in a way causally efficacious, but at the same time, we would have to admit that it leads to a kind of methodological reduction in predicting the conscious behaviour. In such a scenario, which may turn out to be true by further advancements in neuroscience, mental phenomena can be rendered irrelevant for scientific purposes although they can still perfectly be distinct from the physical/neural properties, and although they can still be irreducible in a metaphysical sense. This is an epiphenomenalist, metaphysically non-reductionist, but methodologically reductive scenario. It does not eliminate mental properties, it eliminates top-down causation and consequently the free will itself. Especially those who are interested in the literature of neuroscience can realize, that is the common ontological mindset for most of the neuroscientists in this area.

Going back to the premises of BN explained earlier in this chapter, BN leads to epiphenomenalism concerning consciousness though it successfully avoids

eliminativism on mental states. It is difficult to see why all the causal powers of our neurobiological structure do not unfold in the total darkness but rather they are also accompanied by consciousness, if consciousness is merely a higher-level feature of the brain like the solidity of a wheel. We have gone through two different cases of epiphenomenalism in earlier paragraphs, and I think the causal reducibility aspect embedded in BN renders it an instance of second case which is epiphenomenalism without elimination. Lowe (1996) asserted in a similar way that BN is an instance of epiphenomenalism:

> For Searle, thus, consciousness has *no* independent causal powers, a point he somewhat confusingly makes by saying that 'consciousness is *causally reducible to* . . . brain processes'. But he still regards it as (in another sense, plainly!) 'irreducible' — though not for any ontologically deep reason, since according to him 'the irreducibility of consciousness is a trivial consequence of the pragmatics of our definitional practices'. He doesn't, however, consider that this makes consciousness epiphenomenal, because on his account conscious states *do* have causal powers, albeit not independent ones. So, it seems, Searle is endorsing a species of non-reductive physicalism of the sort discussed earlier. But he doesn't (in my view) face up to the difficulties and inherent instability of such a position. And there seems to be, besides, a certain amount of confusion in his position in any case (as the foregoing quoted remarks on deducibility' rather serve to show) (pp. 79-80).

At this point, we should touch upon the distinction between the strong and weak emergence, which is still a hotly debated issue among the philosophers of mind. There is a variety of emergentist theories some of which postulate many emergent layers of reality such as physics → chemistry → biochemistry → biology → mental →social and so on. For the current purposes of the paper, I will remain in the boundaries of mind-body problem and argue that if there is only one strongly emergent phenomenon in the world, it is consciousness. Weak Emergence tells us that there are some layers of the universe which are emergent in such a way that the facts about their causal structure cannot be deduced from the base properties. This phenomenon fits into Searle's analogy of liquidity in relation to water molecules,

namely that the property of liquidity cannot be found in the individual water molecules and it is a system feature of a certain amount of water. Weak Emergence is related to even less interesting instances of emergence: The length of my pencil is 15 centimetres. The property of being 15 cm is not reducible to the constituent elements of my pencil because none of them has the property of being 15 cm-long, in conclusion, my pencil's being 15 cm-long is an emergent property in the world. For some philosophers, those properties are called "resultant properties" rather than emergent properties on the grounds that such uninteresting properties can in fact be accounted for simply by "add-up principles". Being-certain-centimetre-long is such a property that exists all over my pencil, we may certainly use an add-up method so as to deduce the so-called emergent property of being-15-cm-long.

Searle is very clear about what he means by emergent features: Some features of the universe are not merely resultant features like "being-15-cm-long", but rather they emerge out of causal interactions of the particulars, which give rise to higher level "causally emergent system features". Consciousness falls into this category as well as the features like solidity, liquidity and transparency (Searle, 2002a, p. 111). On the next page of *The Rediscovery*, Searle defines the strong emergence as an adventurous and naive position. The idea that consciousness cannot be explained by the behaviour of neurons means that it has a life of its own once it is "squirted" by the underlying processes (Searle, 2002a, p. 114). Though being sarcastic about the strong emergence, I think this is a true description of it even though I suspect whether it is possible. Lowe (1996), on the other hand, embraces this "naive and adventurous" idea completely and he constructs an analogy to clarify and sustain this position:

> After all, biological processes can and have brought into existence complex structures which are not themselves biological - the spider's web and the bee's

> honeycomb are obvious examples, which may encourage us to see the art and architecture of human beings in much the same light (…) Just because the spider created the web, we don't have to say that the things which the spider does with it are in no way to be credited to the web itself, independently of the spider (p.78).

David Chalmers (2006), in the article titled “Strong and Weak Emergence”, proposes four examples of weak emergence:

> (A) The game of life: high-level patterns and structure emerge from simple low-level rules.
> (B) Connectionist networks: high-level ‘cognitive’ behaviour emerges from simple interactions between simple threshold logic units.
> (C) The operating system: the fact that overloading occurs just around when there are thirty-five users on the system seems to be an emergent property of the system.
> (D) Evolution: intelligence and many other interesting properties emerge over the course of evolution by genetic recombination, mutation, and natural selection (pp. 250-251).

On the other hand, consciousness is an emergent in a quite different way that it facilitates us to claim that it is truly a novel property of the world. Let us make the basic assumptions clear:

1. Strongly emergent properties are irreducible to its base properties.
2. The high-level truths are not conceptually or metaphysically necessitated by low-level truths.
3. Strongly emergent phenomenon exerts downward causal powers on the lower-level processes, which cannot even in principle be deduced from the initial conditions of the lower-level phenomena (Chalmers, 2006, p. 249).
4. Unlike the weak emergence, strongly emergent properties are observer-independent.
5. If there are strongly emergent properties that are not deducible from the current basic laws of nature, it is required to postulate new fundamental laws of nature in order to account for them.

Imagine yourself as a physicalist philosopher, having a conversation with the experts who come from various fields such as molecular biology, mechanical engineering, sociology, chemistry and physics. At some point in the discussion, you propose "Why do not you think about building a huge supercomputer informed with all the fundamental physical laws of nature and start making any further calculations about chemistry, biology, sociology etc based on these fundamental laws and eventually render all the other sciences redundant?".

Of course, you would receive quite negative reactions some of which will imply your plain ignorance about sciences. Leaving aside the ontological indeterminism embedded in the quantum physics, we already know that chaotic systems (biological, chemical or aerodynamic) are epistemically indeterministic even though they might be ontologically deterministic. Therefore, such a project is practically impossible to achieve. However, it does not mean that weakly emergent layers of reality are not in principle reducible to physics. An omniscient god would be able to predict all the higher-level effects from the initial conditions of the universe, again ignoring the implications of quantum physics. Epistemic irreducibility does not prevent ontological reduction, and it leads us to another striking conclusion on this matter in Chalmers' (2006) words:

> We might instead understand weak emergence in terms of the ease of understanding one level in terms of another. Emergent properties are usually properties that are more easily understood in their own right than in terms of properties at a lower level. This suggests an important observation: weak emergence appears to be an observer-relative property. Properties are classed as 'emergent' based at least in part on (1) how interesting the high-level property at hand is to a given observer, and (2) how difficult it is for an observer to deduce the high-level property from low-level properties (p.251).

Dennett is very much clear on the idea that everything can in principle be deduced from physics but we are adopting intentional or personal stance just for the practical purposes of scientific prediction (Mirolli, 2002).

These are, indeed, generic and striking claims about the nature of emergence and causal reduction. We may need to make some further clarifications concerning above-mentioned premises of Chalmers. Premises 1 and 2 are very closely related to each other. It seems that mental states are quite different from brain states and there is no obvious metaphysical necessity between them. Seemingly, while neuroscientists are dealing with the easy questions of consciousness like the neural relata of vision or auditory experiences, they are taking the metaphysical connection of the brain and the mental for granted or a brute fact of nature as if it does not itself require any further explanation, which I think might be the case though very unlikely. The way my argument goes concerning the basics of strong emergence concept might lead the reader to have the impression that it is kind of an idea that is ultimately on the side of dualism, or panpsychism of a certain kind; however, Tim Crane suggests that emergentism might fit into naturalist purposes more than physicalism does. Interestingly enough, according to Crane, it is their epistemic attitudes that differentiate non-reductive physicalists and emergentists from each other; because the emergentist may plausibly embrace the brute fact concerning the connection of the brain and the mental, with a certain naturalistic piety whereas such an option is not available to the non-reductive physicalist. On these grounds, he concludes that if the non-reductive physicalist is to remain a physicalist at all, he had better adopt reductionism since non-reductive physicalism and emergentism are metaphysically indistinguishable (Crane, 2010, p.30).

The third premise is one of the crucial logical consequences of strong emergentism. Many people in the scientific fields and also social sciences easily take non-reduction for granted mostly because of their specializations that they tend to conceive of their subjects as something ontologically and causally special with

regard to other research areas. Downward causation signifies the emergent causal powers in higher level metaphysical aspects that cannot be reduced to or are not metaphysically necessitated by base-level properties. Do high inflation rates cause decrease in ordinary people's purchase power; was it the aesthetic properties of Michelangelo's works that made them famous classical pieces; did the idea of Protestant Christianity bring about profound changes in the European politics; I imagine many people would say obviously yes to each of these questions. On the other hand, let us take the examples of the solidity of a wheel that exerts downward causation on its parts or liquidity of water that facilitates swimming. The answer isnot so obvious when we take non-intentional phenomena into consideration. A thermostat might be said to have an intentional stance only of a lesser degree than full-fledged conscious subjects. A thermostat's controlling the heat inside might be a certain kind of structural cause which is partially analogous to information-processing in the brain. Leaving the brain aside at this point, I suppose that one would not seriously claim that thermostat's causal powers cannot be reducible to its base physical properties.

The libertarian is bothered by the idea that can be oversimplified as "fundamentally, downward causation does not really matter" position. Kim (1999) has proposed a three-level functional reduction for emergent properties which puts it in a serious philosophical trap:

> Step 1: (*An emergent property*) E must be functionalized - that is, E must be construed, or reconstrued, as a property defined by its causal/nomic relations to other properties, specifically properties in the reduction base B. (…)
> Step 2: Find realizers of E in B. If the reduction, or reductive explanation, of a particular instance of E in a given system is wanted, find the particular realizing property P in virtue of which E is instantiated on this occasion in this system; similarly, for classes of systems belonging to the same species or structure types. (…)
> Step 3: Find a theory (at the level of B) that explains how realizers of E perform the causal task that is constitutive of E (i.e., the causal role specified

> in Step 1). Such a theory may also explain other significant causal/nomic relations in which E plays a role (Italics are mine.) (pp. 10-11).

According to this model, BN does not seem to be threatened since it accepts causal reductionism in the first place and brain does not have anything over and above its neurobiological base. The problem is, again, about the threat of epiphenomenalism: Once we sufficiently define the causal powers of consciousness in terms of structural properties of the brain that are themselves reducible to its constituent elements such as neurons and microtubules, then there is no special kind of downward causation peculiar to mental properties. If Kim is right, of course on the other side, the second formulation of our sourcehood principle becomes indefensible given that the agential powers of a subject can be explained in base terms that go down toward impersonal physical forces that fundamentally do not have anything to do with us (assuming that one is not a Spinozist at this point.) At best, it leads us to hold that even though consciousness is not ontologically reducible to physical properties, nevertheless its content does not possess anything causally peculiar to its subjective and private realm.

Kim's (1999) argument rests upon the assumption of Mereological Supervenience that can hardly be underestimated: "Systems with an identical total micro-structural property have all other properties in common. Equivalently, all properties of a physical system supervene on, or are determined by, its total microstructural property." (p. 7). Similarly, Scruton (2012) proposed that even though we do not exactly know how the brain gives rise to mental states, neuroscientists are at least trying to show us that the matter while organized in a certain way gives us those results, namely mental states. The reader should note that this is one of the most striking challenges that the libertarian faces, because we are either forced to admit that agential powers are supervenient on the micro-physical

structure or we should admit that libertarianism cannot survive the challenge without postulating extra causal powers like noumenal selves, agent-causes, immaterial souls that Robert Kane is motivated to avoid in his libertarian theory.

Just for a short example, O'Connor and Wong proposed The Dynamic Emergence Model which allows for such divergences in subsequent events even if their prior base-level properties are identical; because according to O'Connor & Wong (2005), there are also holistic emergent levels over and above the micro-physical basis of events that facilitate the subject to evolve in different directions:

The prior causes, H1 and H2, manifest their difference solely at the emergent level, and the different emergent features, in the presence of the common state P, account for the difference in R1 and R2, respectively.

Premise 4 in Chalmers' argument is a kind of aspect in strong emergence that can hardly be ignored by Searle (2001) himself, considering his own distinction concerning observer-relative and observer-independent properties in Strong Artificial Intelligence discussions and also his famous Chinese Room Argument which contains important elements as to how we are supposed to address the observer-relative/independent aspects of language and intentionality (p. 54). Suppose that my computer requires a six-digit-code to give access to the desktop. It obviously means that the computer has a structural property / a system feature S that requires a matrix-based information code, say I, to realize the opening of desktop. It might be claimed the information code I is the functional realizer of the opening that can itself be multiply realized by different relata such as typing on my keyboard, using on-screen keyboard with a mouse, or even using an external gadget, say a wireless keyboard, which can be functional in inserting that information code to the computer. The

question is whether we seriously take that structural property S into account to explain the mechanism of this computer.

The confusing element in this line of reasoning is that we are discussing an ordinary daily behaviour that is mainly about our anthropomorphic purposes that have nothing to do with fundamental properties of universe. However, I think it is not difficult for one to be convinced that the opening of desktop with the code is totally identical with a certain electrical discharge process in the relevant parts of the computer. The computer's structural property S functions in a pseudo-intentional basis as if the computer keeps outsiders who do not know the password out of desktop; in other words, it only seems as if S is genuinely efficacious over and above the electrical realizers on its physical base. We do take pseudo-emergent properties into account when they are closely related with our daily habits, interests, semantic purposes or simply because of our epistemic deficiencies to explain certain phenomena in nature. I suppose no one is seriously willing to take the length of a pen into consideration as an irreducible property while dealing with emergent properties. Strong emergentism argues for the novel properties of consciousness whose downward causality is genuine and in a sense transcendental of the causal powers of the base properties. There is no easy problem that we are reflecting upon at this point. One party tries to say there is nothing over and above the electro-chemical forces in the brain whereas the emergentist wants to show something special about consciousness. Searle's BN is partly coherent in itself since it is an attempt to account for consciousness in down-to-earth conceptualizations whereas it falls short of accounting for novel properties which exert downward control over the base conditions. I think the true interpretation of BN should be very close to the assertions of Edelman such as the following:

> Causality can only be attributed to physical processes because the physical world is a closed space and there are no ghosts to activate the machine; there is no causality apart from a neurophysiological process. Consciousness in itself does not cause anything. Consciousness is knowing, not causing. Causality can only be attributed to electrical and chemical forces (as cited in Modell, 2008, p. 44).

Finally, the fifth premise requires the postulation of new fundamental laws in order to account for strong emergence. That is truly an unwanted consequence of emergentist position for me on the grounds that it is extremely difficult to deal with that under such epistemic limitations. According to Chalmers, it might be the case that consciousness might be such a fundamental aspect of reality that can itself exert control on quantum collapses, rather than quantum probabilities determining the subsequent states of consciousness. Under various emergentist positions, it is possible either to consider the emergence of mental states out of neural states as a brute fact of nature or to postulate extra-physical forces, holistic emergent forces over the micro-physical structure.

We have so far endeavoured to show the basic flaws in Searle's Biological Naturalism, and its consequences leading toward some version of epiphenomenalism. BN is an unsatisfactory account of consciousness since it allows for the causal reduction of mental states into neural states and totally violates the second formulation of the Sourcehood Principle, which renders BN an indefensible account for the libertarian. In the subsequent chapter, we will continue to analyse BN in the context of causal explanations as opposed to reason-explanations of human actions.

2.2 The problems of biological naturalism with regard to reason-explanations

Roger Scruton (2017), in his latest work titled *On Human Nature,* argues that what is fundamental concerning our moral lives is the distinction between a subject and an object (p. 33). We relate to each other within a certain paradigm that is described by

reactive attitudes or I-Thou relations that are explicated by Martin Buber. Since we are the kind of self-conscious beings, we very much differ from other animals on the grounds that we give reasons to justify our actions and expect reasons from other persons for justification of their behaviours in return. Despite the fact that Scruton is not quite bothered by the implications of physicalism or mereological supervenience that we briefly touched upon in the earlier chapter, it seems that there is a conceptually substantive (Sider, 2011)[2] element to reason-explanations that renders our moral lives or the ways we relate to each other genuinely sustainable as opposed to the biological world of animal kingdom. He adopts a sort of dual conception of the world which is quite similar to Immanuel Kant's understanding in this sense:

> When I give a scientific account of the world, however, I am describing objects only. I am describing the way things are and the causal laws that govern them. This description is given from no particular perspective. It does not contain words such as *here*, *now*, and *I*; and while it is meant to explain the way things seem, it does so by giving a theory of how they are. In short, the subject is in principle unobservable to science, not because it exists in another realm but because it is not part of the empirical world. It lies on the edge of things, like a horizon, and could never be grasped "from the other side," the side of subjectivity itself (Scruton, 2017, p.32).

One of the crucial aspects of Searle's libertarianism is the substantial distinction between causal explanations and reason explanations; the latter is supposed not to cite causally sufficient conditions as opposed to Donald Davidson's theory of action. These two sorts of explanations are quite different in their logical forms. Ordinary causal explanations involve causally sufficient conditions in the form "A caused B" whereas reason explanations do not cite causal sufficiency but rather a subject S acts on a certain set of reasons without necessitating the resulting decision (Searle, 2007, p.54). We should notice that the logical structure of reason explanations partly gets a grip on the worries regarding the problem of luck on the

[2]See also Ted Sider's distinction in *Writing the Book of the World*.

grounds that cross-world differences between agent's decisions are forced to be left unexplainable. Searle (2007) comments on the explanatory difficulties as following:

> Of course, such explanations, like all explanations, allow for further questions about why those reasons were effective and not other reasons. That is, if I say that I voted for Bush because I wanted an improvement in the educational system, there is a further question, Why did I want that improvement? And why was that reason more compelling to me than other reasons? I agree that such a demand for explanations can always be continued, but that is true of any explanation. Explanations, as Wittgenstein reminded us, have to stop somewhere, and there is nothing inadequate about saying that I voted for Bush because I wanted an improvement in the educational system. It does not show that my answer is inadequate to show that it admits of further questions (p. 56).

The reason explanations are the main components of the logical structure of free will in Searle's account. As we have examined in the previous chapter, mental states are ontologically irreducible elements of the reality, though Searle holds that they are at the same time causally reducible to third-person components. We give reasons in order to justify our actions in a society; and those reasons are also ontologically irreducible components which subjects act upon. Searle claims that there must be a substantial distinction between the third-person impersonal causal explanations versus the first-person reason explanations. And since these reason explanations are not ordinary causal explanations in their logical structure, our actions are only justified under the concept of freedom. In contrast with Dennett's theory, the attributions of personhood/moral responsibility to self-conscious beings should not be merely of an epistemic limitation; we should note again at this point there is no ontologically substantial difference between thermostat's controlling the heat and a self-conscious subject's controlling his behaviour in a social environment in Dennett's view.

As a result, if the antecedent conditions of a subject metaphysically necessitate his subsequent decisions, then free will is an illusion as Searle clearly

states. He proposes a neuro-deterministic scenario (out of the mythological story of Paris and Aphrodite) that brain states can be causally sufficient to fix all the mental states, which renders the experience of freedom completely illusory. Searle (2007) subsequently argues:

> But this result is intellectually very unsatisfying because it gives us a form of epiphenomenalism. It says that our experience of freedom plays no causal or explanatory role in our behaviour. It is a complete illusion, because our behaviour is entirely fixed by the neurobiology that determines the muscle contractions. On this view evolution played a massive trick on us. Evolution gave us the illusion of freedom, but it is nothing more than that—an illusion (p.62).

It is quite interesting that Searle has a serious problem in his mind concerning the Mereological Supervenience (MS) in a deterministic sense. So, it is a perfectly legitimate question why doesn't determinism lead to epiphenomenalism in the example of the solidity of wheel whereas it leads to epiphenomenalism if the brain states fix all the subsequent conditions of conscious realm? In fact, it is one of the basic premises of BN that consciousness does not have any extra causal powers beyond its neurobiological structure. However, though consciousness is still said to be an ontologically special phenomenon in the world, the charge of epiphenomenalism, I think, is mainly concerned with causal structure. Moreover, Searle even claims that rationality itself becomes a redundant concept if causal determinism is true(!); but we'll come back to this later on.

We have briefly touched upon evolutionary psychologist Donald Symons' view on the problem of epiphenomenalism in the first part of the thesis. Symons (1981) has an undermining attitude toward epiphenomenalism:

> The view that only physical events can cause other physical events, and that mind is mere epiphenomenon of brain activity, appears untenable if natural selection is the primary creative force in nature. An omnipotent deity presumably could, if it wished, create mind as a functionless epiphenomenon, but natural selection implies (at least to some scientists and philosophers) that

> were mind functionless it would long ago have withered away, not become elaborate as it obviously has (p. 39).

There is a serious ambiguity in the use of "epiphenomenalism" in both of these cases. As we have tried to clarify in the previous chapter, there is a scientifically credible scenario that our brains can be information-processing machines whose causal powers are entirely defined by the syntactical structure of codes embedded in neurons or neural groups in the brain; and some versions of that scenario can succeed in giving us a story that causally excludes qualitative experiences and at the same time holds all the functional aspects that are necessary to explain causally relevant aspects of human actions; though I still find it quite hard to believe.

Leaving aside the discussion on the epiphenomenalism of the experience of freedom, we need to clarify the basic concepts and premises that are embedded in Searle's own version of incompatibilism:

1) Freedom is an indispensable assumption embedded in human actions / deliberative processes. It does not really matter whether you believe hard incompatibilism when it comes to ordering a cup of tea or coffee in a restaurant. Our default assumption is incompatible with causal determinism. It is one of the basic concepts intrinsic to moral behaviours and speech acts; it is not only the case that we act in accordance with the assumption of freedom, but also, we hold each other responsible and accountable based on this conception (Searle, 2002, p. 14).
2) Three Gaps in the Structure of Volition: The Gaps are essential to understanding both rationality and freedom in human actions, because rationality can operate only in the gap (Searle, 2001, p. 16). In order to make sense of (1), we have to presuppose the existence of genuine metaphysical gaps in the causal structure of

the world. The preconditions of our actions should not be causally sufficient to bring about the subsequent results. (Similarly, the 1st formulation of the Sourcehood Principle in regard to determinism). There are three sorts of gaps: A) there is the gap of rational decision making, where you try to make up your mind what you are going to do. B) there is a gap between the decision and the action. C) There is a third gap that arises for actions and activities extended in time, a gap between the initiation of the action and its continuation to completion (p. 14-15). There is no extra causal component to fill the gap at any point; hence, it is simply the agent who makes up his mind and terminates the deliberation in favour of either options (p. 17) (choosing to study/stop studying/take a break/go back to studying because of the bothering voice of your conscience etc). (Searle (2001) points out the experimental findings in Wilder Peinfield's research in order to direct attention to the possibility that the kind of volitional structure he proposes can actually fit into scientifically credible results. Peinfield (1975) observed that his patients were capable of distinguishing between the instances of actions that are brought about by artificial stimulation of their brains and the instance of their ordinary voluntary actions like raising one's arm (pp. 76-77).

3) Irreducible Self: We cannot make sense of the gap, of reasoning, of human action or rationality without an irreducible notion of self (Searle, 2001, p. 75). This notion is closely related with the unity or binding principle of consciousness which is also articulated by Kant as "the transcendental unity of apperception". Searle (2001) revises the Humean concept of "bundle of perceptions" in accordance with Kant's argument as he put forward as follows:

> I do not just have the feeling of the shirt on my back and the taste of beer in my mouth, but I have them both as part of a single unified conscious field. Hume thought of each perception as separate and distinct, but that cannot be right; because then we could not distinguish between one consciousness

> having ten experiences —the feeling of the shirt, the taste of the beer, the sight of the sky, etc (p. 79).

In contrast with a Humean bundle, the self is capable of initiating and carrying out actions depending on reasons under the presupposition of freedom. The concept of Non-Humean Self in Searle's account is an attempt to argue in a coherent way for a genuine distinction between the events that are merely happening to a subject and the actions that a subject carries out by reasons. On the other hand, Searle's objection to Humean conception of self does not mean that he admits we cannot make sense of freedom or rationality without some sort of Cartesian Ego persisting through time. He wants to preserve some kind of non-substantive, purely naturalistic concept of personal identity through time. Therefore, the notion of irreducible self does not signify causal irreducibility in Searle's account, otherwise it would obviously be giving up on BN altogether. The irreducibility of the self is mainly related to its ontological irreducibility like any other mental state that cannot be reduced to third-person physical properties. I believe there is only one significant condition of personal identity through time in the relevant chapter where Searle (2001) makes this analysis. The continuity of the self is predicated upon the biological continuity (p. 78). I think the biological continuity is insufficient for an adequate account of self and I will try to clarify its shortcomings in the fourth chapter "Some Remarks on Physical Selfhood and Its Implications". As BN implies that even though prior experiences/memories of a subject are ontologically irreducible, they are ultimately sustained by the continuation of neural structure in which all of them are encoded. The self is totally reducible to underlying elements such as sequential conscious feelings, memories, subjective sense of being oneself etc (Searle, 2001, p. 79). The reader should note that when I decide to pick up a book to read and start doing so; the decision and the action of reading a particular book including the actish

feeling of doing so are causally reducible to lower level neural states; in other words, each of the causal components that constitute my action of reading is ultimately explained by the neural causes, to which Searle should clearly agree. This logical consequence will be our main target as to how BN renders the distinction between events and actions redundant to the almost full extent.

4) The causal gap does not entail explanatory gap: It is true that we cite sufficient causal conditions in explaining ordinary natural events whereas we refer to reasons in explaining our actions especially in a social context. In accordance with the causal gaps that are metaphysically available, a subject acquires the opportunity to pick out one set of reasons out of many and makes it effective by acting on the causal gap. It might be useful to conceptualize it in the framework of "Why" questions as opposed to "How" questions. When you are asked "Why did you choose to pursue this career?", you would start talking about many deliberative processes that you have been through along with your prior desires, hopes and fears which were causally relevant to constitute your frame of reasoning at those times, and you would be mentioning a set of purposeful/intentional reasons that you picked out and made prevail over the set of reasons which were supposedly available. On the other hand, there are "How" questions that mainly concern the unintentional events that are to be evaluated under impersonal causal conditions, for instance the question how the solar system developed in this and that way. Besides the fact that Searle doesn't himself refer to oversimplified dichotomy of Why/How questions; their examples are pretty close, and he does not seem to recognize a serious problem carefully here, which is the problem of luck. There are many philosophers, like Alfred Mele, who are deeply concerned about the explanatory gap not with regard to

justification of any particular action, but rather the cross-world difference between the two possible worlds: in accordance with Searle's own example, the difference between the actual world in which he decided to vote for Bush as opposed to the alternative possibility that he decided to vote for (say) Al Gore. Suppose that Searle has equally good reasons to vote for either of the presidential candidates prior to his decision. On the one hand, he is justified to vote for Bush on the reason that the U.S needs a serious education reform and Bush is a better candidate to achieve this goal; and on the other, Searle is also justified to vote for Al Gore on the reason that climate change has been reaching alarming levels and Al Gore is a more suitable candidate to take on necessary actions to prevent that. In either of his decision scenarios, Searle would be justified in choice no matter which one is going to be his choice. However, some might claim that the cross-world difference between the two possible choices is merely a consequence of luck on the grounds that there is no more fact to explain this difference since we are assuming exactly the same personal history and natural laws at play prior to his decision.

5) The gaps of volitions are predicated on quantum indeterminism: According to BN, everything is physical and there is not any natural phenomenon that cannot be causally reducible to physical processes. Many philosophers normally derive compatibilist results out of physicalism and reject the libertarian views. However, Searle thinks that it is incompatible with the concept of free will; and moreover, he considers free will as an indispensable notion in parallel with Kant (Proposition 1 above). Secondly, libertarian free will requires that there should be genuine alternative possibilities and volitional gaps (Proposition 2). The problem is that we do not normally observe any gaps in the world or in the field of

neuroscience. Searle seemingly sticks to only one possible solution to this problem in a physicalist framework, which is quantum indeterminism. Here is the argument:

> Premise 1: All indeterminism in nature is quantum indeterminism.
> Premise 2: Consciousness is a feature of nature that manifests indeterminism.
> Conclusion: Consciousness manifests quantum indeterminism. (Searle, 2007, pp. 74-75).

Quantum indeterminism is based on genuine randomness, and it cannot normally contribute to the libertarian idea. Searle tries to argue that quantum randomness in the micro-level does not necessarily manifest itself in the macro-level as well. If this is true, then consciousness can manifest indeterminism without randomness (I suspect this phrase is totally absurd for many philosophers / scientists). I'd also like to know whether this can be true in the light of quantum physics; philosophers' discussions on mereology and reductionism are generally held without any relation to findings of fundamental physics. I'll leave this proposition at a descriptive level here.

In the previous section, we went over the unwanted results of BN with regard to causal efficacy of mental states. We tried to show that BN is compatible with the logical consequences of reductionist Information-Processing theories, in other words, functional aspects of consciousness can give us the same results even if our qualitative states were epiphenomenal. In this part, we try to shed light on the basic aspects of the logical structure of free will that is also put forward by Searle, and we will see that there are certain discrepancies between Searle's libertarian free will and his biological naturalism. Here is my argument for this:

P1: Subject S decided to do A by acting on the reason R. (Searle voted for Bush on the reason that he wanted an improvement in the educational system.)

P2: The action A along with the actish feeling of doing A is realized by the neural

processes in the brain.

P3: If BN is true, then we will in principle be able to give a total description of causal processes purely in neural terms. (Causal reducibility of mental states)

P4: The logical structure of free will entails a substantial distinction between causal explanations and reason explanations. (In accordance with volitional gaps, there is a distinction between an event which is happening to S and an action A that is actually done by S)

P5: Conscious decision to do A on the reason R will be causally (though not ontologically) deduced from set of neural descriptions by future neuroscience.

P6: If P5 is true, then the action A will be causally reduced to third-person (impersonal) processes in the brain. (Violation of the Sourcehood Principle)

P7: Given P6, the causal properties of S are in fact grue-like properties in the universe. They do not correspond to metaphysically fundamental reality.

Conclusion 1: If BN is true, there cannot be a substantial distinction between reason explanations and causal explanations.

Conclusion 2: Given P6 and P7, the conceptual preference in our description (S decides to A on R) is arbitrary. Subjects, actions and reasons are not necessary concepts in this explanatory framework; they are causally dispensable. (And then, most probably our need to refer to agents, actions and reasons stems from our own epistemic limitations, nothing more.)

CHAPTER 3

KANE'S SELF-FORMING ACTIONS AND DISAPPEARING AGENTS

Robert Kane has produced one of the most influential libertarian theories on a naturalistic basis. Self-forming actions (SFA) are central to this theory, which are will-setting, regress-stopping and character-building actions. Kane endeavours to show that it is possible to defend traditional sense of free will within the boundaries of modern science and naturalist view of the world. There are two basic principles on which SFAs are built upon in his account, namely the Principle of Alternate Possibilities (AP) and Ultimate Responsibility (UR). For most of the libertarian accounts, the ultimate worry about determinism is that the conditions for action stretch beyond purely impersonal forces that took part in a distant past when the agent was not even born. UR represents this key worry and SFAs are supposed to account for philosophical difficulties that libertarians suffer in the light of modern science for a long while. Kane is quite ambitious about his account when he even claims that there are certain conditions proposed in the account that make his theory scientifically intelligible; for instance, if determinism is true or quantum indeterminism does not make any difference in brain processes, then Kane's account will be falsified.

Kane is motivated to remain in a certain metaphysical parsimony which he thinks makes his account advantageous over agent-causal accounts, Kantian theory, substance dualism etc since he does not postulate additional causes, substances, noumenal selves and so on. One should notice that it is a very similar line of reasoning with Strawson's accusation of 'panicky metaphysics'. Each phenomenon in exercising free will is already postulated by modern science such as mental events,

parallel processing of the brain, quantum indeterminism (though yet to be connected with neural processes). One of the leading thinkers of reductionism, Daniel Dennett, considered Kane's account as the best attempt in libertarianism (Dennett, 2003, p. 99). I hold that libertarian free will is incompatible with naturalism, it is in the very essence of traditional free will that it cannot survive without postulating an additional layer of causal autonomy above physical forces.

In Kane's (1996) account, self-forming actions are required to satisfy the principle of ultimate responsibility given that "If free will is exercised by finite agents and entails UR, then some undetermined self-forming actions, or SFAs, exist, and determinism is false" (p. 74). And secondly, "(…) SFAs are both undetermined (by virtue of U) and such that the agents willingly performed them and "could have voluntarily (or willingly) done otherwise" (by virtue of R). In other words, SFAs satisfy AP in an incompatibilist sense requiring indeterminism" (p. 75). SFAs are supposed to be undetermined and consciously done by the subjects in order to satisfy the Principle of Alternate Possibilities (AP). So basically, *the required indeterminism is sustained by quantum indeterminism in a physical world where agents exercise free will by virtue of their agent-involving mental events realized in their brains, as opposed to agent-causal libertarianism which requires the settlement of such self-forming actions/willings by agents as substances. He also adopts the Aristotelian notion of character building in terms of Self-Forming Actions somehow similar to habituation.* Thus, it is quite interesting and sophisticated view unlike many other views on free will which reduce the notion of free will to freedom of action (the paraphrase of Kane's assertion in italics, 1996, p. 61).

Let us illustrate the conditions of SFAs further as such: An action A is a basic exercise of free will and performed by the subject S on the grounds that:

1) At least one opposite choice B was open to S up until the moment of decision t. (AP: S could have done otherwise until t.)

2) A was intentional/purposeful of S.

3) S was performing dual-voluntaries up until t in such a way that while he was trying to justify A over B, at the same time he was performing an opposite effort to make the set of reasons B prevail over A. (Plural/Dual Voluntariness Principle).

4) (For AP to be satisfied in naturalistic terms) Quantum indeterminism amplified by conscious dual effort (or concurrent effort) in the parallel-processing system of brain should be fused by the effort of will. Quantum indeterminism does not precede or succeed the process of dual effort (or effort of will in general) (Kane, 1996, p. 151). (Fusion Principle)

One of the generic examples in *The Significance of Free Will* is a case about the moral dilemma of a businesswoman who witnesses an assault while catching up to an important meeting:

> Consider a woman on the way to an important sales meeting who witnesses an assault or mugging in an alley. Should she stop and call for help or press on to avoid losing a sale crucial to her career? (…) The businesswoman faces a moral conflict, the engineer a prudential one. But both are torn by conflicting motives and must make mental efforts to get their ends or purposes sorted out—to "set" their wills in one way or another. They are strongly tempted to act from self-interest, in the one case, or to relieve present tensions, in the other. But they are also committed to moral beliefs and long-term plans, and are making efforts to overcome temptations that threaten these commitments (Kane, 1996, p. 126).

On the confrontation of this assault, her prior dispositions give rise to two incommensurable sets of reasons to act on:

1) Moral reasons to call the police in order to help the victim

2) Selfish reasons to keep driving to catch up with the meeting.

Either course of action is conceivable given her past experiences and character, however they are not compatible with each other. SFAs come to stage

when a subject is truly considering who she is or what kind of a person that she wants to become. The conscious tension between these competing motives gives rise to a tension in the parallel-processing structure in the brain and the chaotic structure of the system become quite sensitive to quantum indeterminism. On the phenomenological aspect of will-setting, the agent engages in dual-voluntaries by forming concurrent efforts to justify one set of reasons to prevail over the other and on the neural counterpart of SFA, there is a tension of parallel-processing in fusion with quantum indeterminism up until the moment of decision. At the end, either moral or selfish reasons prevail and the business woman, by engaging in SFA, either strengthens her moral character for the future encounters or weakens it by appealing to her selfish concerns. According to Kane's account, a person can be morally degenerate and be almost fully determined to perform evil actions. As also Aristotle pointed out, we need to consider how far she is responsible for becoming such a wicked person. Only answer is how far she engaged in SFAs and evil actions which were once under her control degraded her moral character.

Much of debates concerning Kane's libertarian theory has been revolving about the problem of luck, and one should check Alfred Mele's *Free Will and Luck*, *Aspects of Agency* (one of his latest books) about the details of that specific debate which had far-reaching influences on free will literature. There is another concern that we will deal with in this paper, which is the regress problem that is mainly articulated by Galen Strawson, in his article titled "The Impossibility of Ultimate Responsibility". For sure, the Principle of Ultimate Responsibility is the fundamental principle that Kane's theory is predicated on. We may briefly give out an Aristotelian example which is quite in parallel with Kane's libertarianism:

Suppose while you are driving your car at night, you suddenly hit a pedestrian on the way, and consequently he dies. When the police arrive at the place, they find out that you were seriously drunken at that time. Obviously, you would be taken to the court and be punished in accordance with the penal code. It would be be naive to suggest that that wasn't a free or responsible action because the perpetrator wasn't conscious of what he was doing, and the Principle of Alternate Possibilities does not apply to him. An agent does not have to be directly responsible for an action, in order to be held culpable for it.

In the Turkish penal code (I'm sure it is also the case in the Anglo-Saxon law), there is a distinction between wilful killing and reckless killing, and the latter one is the true description of our example here. The car accident falls into the category of conscious reckless killing on the grounds that the perpetrator conducts the act despite foreseeing the results, he does not discharge the duty of care and attention. You simply neglected the precautionary steps to prevent such an event. We may reflect upon the implications of that on a broader scale and suggest, in accordance with Kane, that one is responsible for the actions that flow from one's character or current situation/defects in proportion to his responsibility in becoming the person who he/she is. So, we may only determine one's proportion of responsibility by understanding to what extent SFAs are involved in the development of one's current character.[3] It is also admitted by Kane that only an omniscient God can fully measure this, but the point is that one should have been through some basically free actions (SFAs) in order to be held culpable or praiseworthy for the determined acts which occupy most of our daily lives. Hence, there are some

[3]This is a critical point which is put forward by Daniel Dennett. He means to say that SFAs are epistemically indistinguishable from pseudo-SFAs (the ones that are perfectly compatible with determinism), hence SFAs are practically redundant concepts, which means that they are not of any use in real life (See also Dennett, 2003, p. 127).

primarily basic actions that cannot be traced back further, and that is the main reason why SFAs are regress-stoppers.

Having clearly agreed upon the significance of UR for moral responsibility, Galen Strawson (2000) does not think that it is possible at all. Here is the argument against UR by Strawson:

> (a) If you are to be UR for the way you are, you must have intentionally brought it about that you are the way you are.
> (b) You have somehow intentionally brought it about that you are the way you now are, in certain mental respects: suppose you have brought it about that you have a certain mental nature Z in such a way that you can now be said to be UR for Z.
> (c) You must already have had a mental nature Y, in the light of which you intentionally brought it about that you now have nature Z.
> (d) For it to be true that you are UR for how you now are, you must be UR for that nature, Y, in the light of which you brought it about that you now have nature Z.
> (e) You must have intentionally brought it about that you had Y.
> (f) You must have existed already with a prior nature, X, in the light of which you brought it about that you had Y, in the light of which you brought it about that you now have Z (pp. 150-151).

Although I grant that the argument is valid, it may also be an exaggeration of the concept of free will. Provided that we do not have absolute control over who we are, it does not mean that none of our actions is basically free actions. Kane is trying to tell us that there are indefinite number of moments in our lives that we find the chance to set a new direction for ourselves and overcome the impediments that are brought on us at birth/childhood. One might have been through many unfortunate cases of mistreatment in his/her childhood, which might provide one with lots of ill-intentions toward other people; however, there are in general chances to reflect upon those inner drives, habitual resentful ideas and try to overcome them by generating (say) second-order desires to be the person that would no longer have those malicious intentions toward others. Of course, we are only theoretically reflecting upon such a philosophical possibility, but nevertheless such a possibility would not

be so far from what UR is meant to achieve although the term “ultimate” in UR might be too strong word to describe it, as I think, Strawson was trying to show us. It is by no means clear whether such a notion of free will can be real. For my part, I’m less optimistic than Kane. For instance, psychopaths comprise 1% of the general population, although they are responsible for 25% of criminal offences. The pessimistic part of it is that some of the psychopaths are known to be raised by quite decent families, but nevertheless they seem to be born as manipulative and “emotionally colour-blind”. Notorious serial killer Theodore Robert Bundy’s (Ted Bundy) mother said: “He was raised in a good, loving, caring family.... We still love and care for him, but we want to know: what caused this?” (Slotjaw, 2009). In the interviews that are put together in a documentary movie titled *Conversations with a Killer: Ted Bundy Tapes*, Ted Bundy said:

> Everybody is fascinated with the notion that there is cause and effect. That we can put our finger on it and say, “Yes, his father beat him when he was a boy. We could see it when he was a kid.” (…) There is nothing in my background which would lead one to believe that I was capable of committing murder.
> [Stephen Michaud] Absolutely nothing?
> [Ted Bundy] Absolutely nothing (Bundy, 2019).

Some neuroscientists claim that psychopaths’ characteristic features are determined by a certain malfunctioning in their brains. In the near future, some neuroscientists who are interested in libertarian theories which rely on quantum indeterminism might conduct serious research on the area and conclude that SFAs are impossible (let’s say at least for psychopaths due to certain structural properties of their brains). Nevertheless, none of us would be willing to render them inculpable for their actions especially on a legislative basis. Reflecting upon the biographical data about some of those criminal minds in history, they are so broken, degenerated, lost in their moral darkness that it is hard to believe they were even trying to see what kind of a person they would like to become.

There is another counter-argument that is put forward by Dennett (2003) in his book titled *Freedom Evolves*, what we may call "the Mammalian Argument":

> (1) Every mammal has a mammal for a mother.
> (2) If there have been any mammals at all, there have been only a finite number of mammals.
> (3) But if there has been even one mammal, then by (1), there have been an infinity of mammals, which contradicts (2), so there can't have been any mammals. It's a contradiction in terms. (…)
> What should we do? We should quell our desire to draw lines. We don't need to draw lines. We can live with the quite unshocking and unmysterious fact that, you see, there were all these gradual changes that accumulated over many millions of years and eventually produced undeniable mammals (pp. 126-127).

Dennett's point is quite solid on this problem, it can even be simplified as "The problem of regress is all in one's mind". The theory of evolution has obviously changed our conception of life in a radical way. As Popper has argued before, what is peculiar about Darwin's Theory of Evolution is that he showed us that teleological systems in the world can be explained in non-teleological terms (Popper, 2006, p. 57). Despite all of our fascination with human's cognitive capabilities, moral concerns, existential turmoil etc, Dennett is always inviting his reader to recall that we are biological beings who evolved through millions of years out of micro-organisms; and we should stick to the evolutionary facts thereby avoiding loss of common sense in philosophical problems. Hence, we are supposed to give up on our so-called exaggerated concerns about being the prime movers of our lives, namely the primary free actions, SFAs, in this context.

Robert Kane's understanding of free will is not limited to moral decisions and character development; however, I'll keep on focusing on this significant part of his theory. It is not always clear to what extent agents exercise this kind of free will in their lives; nevertheless, it occupies quite significant moments which set new directions to one's life. We are basically dealing with the torn decisions in which

agents exert dual-voluntariness prior to its settling. The businesswoman who witnesses the assault finds herself in the midst of a moral turmoil in which she genuinely starts considering what kind of a person she is and what kind of a person she wants to become. The assumption is that on the neural side of this story, moral dilemma activates some kind of a chaos in the parallel-processing system of the brain, which makes the neural net sensitive to quantum fluctuations that might serve our purposes like "a background noise" which breaks the neural determinism that existed prior to the phenomenological dilemma. She starts forming two sets of reasons which are incompatible with each other; one is the set of reasons that justify ignoring the assault and go on heading to the meeting, and the other set of reasons is about the justifications of moral reasons for staying and calling for help. Both the phenomenology of moral struggle and the neural counterpart of it are accompanied by quantum fluctuations that supposedly open up genuine alternate possibilities. The decision is settled at an undetermined moment and also which of these sets of reasons will prevail is an open question up until the moment of decision. Once the struggle is resolved, then an instance of SFA is complete and after all, she exercised a primary free action of her own which contributes to her character-development. Kane's line of reasoning is generally attacked with regard to its implications of luck, to which he has tried to provide all sorts of answers in various occasions. In our critique of Searle's BN, we mostly stressed its reductionist implications. Given that Kane has a common point with Searle in terms of physicalism, we should proceed with our inquiry in the context of reductionism versus the Sourcehood Principle. The main question is whether Kane's agents who exercise SFAs are appropriately the source/author of their actions. There are two influential philosophers who are suspicious whether the kind of agency provided by SFA illustrates a genuine notion

of agency. Derk Pereboom (2012) developed the Disappearing Agency Argument (DA) as follows:

> On an event-causal libertarian picture, the relevant causal conditions antecedent the decision—agent-involving events—would leave it open whether the decision will occur, and the agent has no further causal role in determining whether it does. With the causal role of the antecedent events already given, whether the decision occurs is not settled by any causal factor involving the agent. In fact, given the causal role of all causally relevant antecedent events, nothing settles whether the decision occurs. Thus, plausibly, on the event-causal libertarian picture, agents lack the control required for moral responsibility (pp. 60-61).

I think the worry is that the decision is settled not by the agent but rather by mental events in the brain. The agent is torn between two sets of reasons and disappeared in a sense, and consequently the agent has no further causal role to bring about this or that action, as if the mental events process themselves by engaging in a quantum leap and one of them is somehow realized. This is partly similar to Dennett's (2003) objection that no one can control quantum indeterminism, thus its providing any more control to the agent is inconceivable:

> The problem is compounded for Kane, since he has to figure out some way to get the undetermined quantum event to be not just *in you* but *yours*. He wants above all for the decision to be "up to you," but if the decision is undetermined—the defining requirement of libertarianism—it isn't determined by you, whatever you are, because it isn't determined by anything. Whatever you are, you can't *influence* the undetermined event—the whole point of quantum indeterminacy is that such quantum events are not influenced by anything—so you will somehow have to co-opt it or join forces with it, putting it to use in some intimate way (…) (p. 123).

Hence, Dennett (2003) rejects the intelligibility of Fusion Principle which is one of the key premises in Kane's account. The brain's amplifying some sort of quantum indeterminism is not enough, quantum indeterminism should be located and merged with the effort of will in an appropriate way. Some event-causal libertarians locate the quantum indeterminacy somewhere prior to decision-making as a simple randomizer that should break the chain of determinism; however, for Kane, quantum

indeterminism should be present up until the moment of decision, not simply prior to the effort of will or decision. It somehow seems to me Kane is downplaying the implications of quantum indeterminism with respect to objections of luck, by appealing to the primacy of wilful efforts and dual-intentional character of them in a phenomenological sense (as cited in Fischer et al., 2007, p. 173). And Pereboom (2012) claims that such a phenomenological strategy is also compatible with determinism and equally available to the compatibilist as well. Moreover, it is hard to believe that quantum indeterminism, in such a case, can function as a safe passage-way of alternate possibilities; because there is nothing in standard physicalism to indicate that anyone can have an appropriate control over it or that an irreducible agent can persist through such a quantum passage while at the same time preserving the afore-mentioned agential controls beside/in it.

If conscious efforts of will are supervenient elements in our brains, then we should question how exactly we can account for any kind of irreducible self in the light of physicalism. Robert Kane adopts a version of non-reductivism which is not very different from Searle's BN. In one sense, it is true that brain mechanism can give rise to such wilful efforts that are at least epistemically irreducible to lower elements. In Kane's framework, not only must his non-reductive physicalism be coherent but also such an irreducible agent must be able to integrate the quantum fluctuations to oneself in order to escape from neuro-determinism. Previously, we have argued that BN is a form of weak emergentism that can hardly be helpful to account for irreducible sourcehood in an ontological sense; there is a similar problem that Kane needs to resolve at this point. In Chalmers's dichotomic analysis of weak and strong emergence, it was also clear that strong emergence required seriously courageous philosophical moves such that we may have to consider consciousness a

fundamental component of the world. Chalmers also touched upon such possible solutions elsewhere: He basically suggests that according to the Integrated Information Theory of Consciousness (IIT), consciousness might be a basic factor in the collapse of quantum fluctuations. Going back to Kane, most philosophers interpreted SFAs in such a way that SFAs in fact resolve themselves via some succession of wilful efforts in combination with quantum fluctuations, in which seemed that there is no more control than the compatibilist can also account for. In contrast, Chalmers' interpretation of ITT implies that it is the irreducible agent/consciousness who brings about the collapse/termination of quantum states. I'm not in a position to evaluate its scientific credibility but I think the contrast is very clear in this context (Chalmers, 2017).

Another objection with regard to quantum indeterminism in SFAs came from Manuel Vargas, implying that none of the quantum-based models of consciousness are welcomed by neuroscientific communities recently, such as the ones that are developed by Roger Penrose and Hameroff. Vargas argued "Hameroff 's, for example, have been widely rejected by neuroscientists, philosophers, and mathematicians. Although brain science is by no means a complete and settled science, it is clear that there are no widely accepted indeterministic models of brain activity, nor, for that matter, even an influential but contested model of indeterministic brain activity. In the words of Henrik Walter, a neuroscientist and philosopher, "to date there is no solid empirical evidence that local quantum phenomena play a role in neurons, and that there are good arguments to the contrary" (Fischer et al., 2007, pp. 143-144).

Kane, on the other hand, asserts that scientific data is far from being clear at this point and we should keep an open mind on the subject. According to Kane, the

brain does not have to implement such global quantum effects on neural pathways; so, SFAs are empirically quite less demanding than the theories that are put forward by Penrose/Hameroff. In response to Vargas, he proposed:

> Minute indeterminacies in the timings of firings of individual neurons would suffice, because the indeterminism in my view plays only an interfering role, in the form of background noise. Indeterminism does not have to "do the deed" on its own, so to speak. One does not need a downpour of indeterminism in the brain, or a thunderclap, to get free will. Just a sprinkle will do (Fischer et al., 2007, p. 183).

Kane's ontological stance is not very clear on how to achieve that by quantum background noise. Most of the time, it seems to me that Kane is trying to make sense of SFAs along with an irreducible and coherent conception of agency within the boundaries of a physicalistic/weak emergentist framework while at the same time inviting the reader's strong emergentist intuitions into the same picture. In general, agent-causal libertarians appeal to extra-causal factors which Kane avoids, and draw a fundamental distinction between event causation and substance causation. Therefore, the ontology of SFAs has a serious tension between its two implicit motivations: a) avoiding any kind of special causation or extra-causal factors or noumenal selves etc; and b) preserving an observer-independent, genuinely irreducible sense of agency. (The reader should recall the crucial difference between observer-relative and observer-independent phenomena as we explained in the previous chapter.)

Let us try to examine the ontological aspects of SFAs in the light of the Sourcehood Principle with regard to the threats that are generated by reductive physicalism:

(1) A subject S's constituents are grounded by physical aspects of the world. The causal powers of S are predicated on the physical atoms, sub-atomic particles and

fundamental components of physics like mass, velocity, volume, electro-magnetism and the compositions of those elements.

(2) The total description of the micro-states of S, call that P, also grounds the macro-level truths about S. The causal power of the mental state M (macro-level description) is in fact deducible from P. Therefore, there cannot be any change in M without any change in P. (See also Kim's Principle of Mereological Supervenience) It also entails that the physical entities whose micro-states are identical are also identical on the macro-level.

(3) Given the truth of (1) and (2), the agential powers of S (the combination of prior desires, character traits, daily habits, prejudices, deliberations, wilful efforts etc.) are constituted by impersonal elements of the world (and S's past) prior to his/her first SFA. Since the genuine self-development of character is traced back to S's first SFAs in order to avoid the regress problem, S does not exercise genuine free will prior to the first SFA; hence, S does not have any choice about who he/she is before that.

(4) The first SFA can arguably be exercised as soon as S's brain develops to a certain point so that S becomes aware of his/her moral concerns with a sense of judgement and his/her brain is now capable of implementing quantum indeterminism by being torn between certain moral choices. Of course, we do not mean that S's moral judgement clears up these concerns instantaneously or quantum indeterminism is brought into the picture all of a sudden; but nevertheless, SFAs have to start at a certain point because they are the primary free actions which conceptually ground freedom of the will on a large scale. SFAs can start with very negligible moral dilemmas, say at the age of 5 (maybe), but there are no semi-SFAs that gradually evolve along with S's cognitive abilities.

(5) Given the premises (2) and (3), the proposition P contains all the necessary conditions that brought about the mental states of S just before the first SFA at T, which implies that S's control over them can only be understood in a compatibilist sense. Just before T, S is not the source/author of who he/she is. It also entails that S is not the source of a) moral concerns on which his/her judgement is torn apart, b) the chaotic turmoil that is amplified in S' brain c) the generation of quantum noise in the neural pathways along with the chaotic turmoil, d) the start and carrying out of wilful effort as a result of (a). As we have seen before, Kane gives very special importance to wilful efforts, and (d) might be strongly objected to; however, it should be clear to us that whatever gives rise to the wilful effort prior to the first SFA, it cannot be of S' own creation. It will not be helpful to appeal to voluntariness, wilfulness aspects of the conscious effort, given that the same move is perfectly available to the compatibilist. So, it should be obvious to the reader as well, if you grant the sourcehood of the wilful effort prior to the first primary free action, then the libertarian loses the ground on which he/she objects to the compatibility of free will with determinism as well. The reader will also notice that a version of the Transfer of Powerlessness that we covered along with Inwagen's Consequence Argument also applies here, though not in the context of determinism this time.

(6) Given the premise (5), S is a developed subject just prior to T only in a compatibilist sense. In other words, S had been exerting only "guidance control" in his/her conscious decisions without any appeal to alternate possibilities. In the first instance of SFA, S will also be exerting "regulative control" in regard to quantum indeterminism. Regulative control enters into the picture when the quantum indeterminism is amplified in the neural pathways, which facilitates the whole system to be open to metaphysical alternate possibilities. The neural system makes it

so by virtue of the randomizing aspect of quantum fluctuations over whose results no one has any control. The metaphysical opening of an alternate possibility is only dependent upon the quantum fluctuations which S can invite via guidance control on moral reasons, but S cannot take any control over it. Remember that the termination of deliberative process at T is also an arbitrary timing since when the decision will be executed is also undetermined because of the ontological indeterminism. Suppose that the quantum indeterminism is amplified by your cell phone radiation signals rather than the brain itself, and those signals causally function in SFAs exactly the same way the quantum indeterminism the brain invites. You would still have exactly the same guidance control in regard to wilful efforts and dual voluntariness and the same kind of quantum indeterminism, only by an external cell phone at this time (See also Dennett, 2003, p. 132). A possible objection might be that an external object's providing the background noise is not the same on the grounds that the source of the noise is external to the agent. However, the case of the brain's giving rise to indeterminism has no causally relevant difference because the indeterminism itself has no controller at all, as Dennett correctly put forward. Another objection might be that the Fusion Principle that we were discussing might be supposed to make a qualitative difference, that is to say, quantum indeterminism cannot be causally excluded from SFAs because what matters is the qualitative aspects of the whole system rather than the causal powers of its components. For example, if you remove the ability of vision, then you do not just remove a component of sensation but also the whole range of ideas and experiences entangled with it. Couldn't quantum indeterminism assume a broader role by opening up alternate possibilities? I think not. Every other element in the structure of consciousness causally contributes to the range of control that S possesses. In contrast, quantum indeterminism has no such

role; it doesn't broaden the horizon of conscious realm, and it does not have any informative contribution like the components of consciousness. It is clearly not the case that S takes the control over the quantum states after it enters into wilful efforts no matter to what extent it fuses with them. S is still only equipped with reason-responsive and wilful efforts in the face of the moral dilemma at T, which equals the range of agential powers that the compatibilist provides us.

Conclusion 1: The S does not exert any further control than the sense of compatibilist control in the first SFA in his/her life.

Conclusion 2: The S is not responsible for his action in this instance of so-called SFA. And obviously, S cannot have any further free will if he/she doesn't exert it in the first possible SFA, given the regress problem. The SFAs fail to stop the regress of powerlessness in our determined actions for which we are held responsible: An alcoholic driver case cannot be resolved with reference to the earlier developments if the first SFA fails to be a genuine one.

I will also add some further points which will be more simplistic than the above argument. Given the truth of physicalism, even though there are high-level truths of reality, we should also notice that those high-level truths supervene on base-properties. S is constructed by those base properties; the whole existence of S supervenes on the fundamental description of universe despite the fact S occupies a very small place in it. The compatibilist abilities such as reason-responsiveness can also perfectly be functionalized and reduced to their base if we ignore our epistemic limitations, as we tried to show in the earlier chapter. It seems easier for us (libertarians) to disregard the compatibilist framework on the grounds that there is only one possible future. On the other side, the nature does not give us the quantum indeterminism as an escape door behind; we should note that the nature already

determines the probabilities embedded in quantum fluctuations. I'm no expert; however, I think there is nowhere any point that we get involved in these quantum phenomena. Once again, a) the probabilities are predetermined by prior conditions, b) no one seems to be able to intervene in the quantum process itself, and c) there is nothing one can do about which result will come out of a fluctuation.

If this is true, then it is fair to ask how it is different from "a swerve of atom" idea in Epicureanism. Robert Kane is trying to show that the compatibilist control in fusion with quantum indeterminism leads us toward new causal chains in an intentional way. Except for the quantum fusion part, there is nothing different from ordinary compatibilist control. The breaking the chain of causal determinism is brought about by an intentional, reason-sensitive deliberation in fusion with quantum indeterminism. However, the chain-breaking aspect of this process is solely dependent on the quantum indeterminism itself on which we do not have any control. All the control, wilful effort, and even the very interesting Taoist Effort (Kane, 1996, pp. 164-165) that we exert during deliberative processes are perfectly compatible with determinism. So, there is a distinction between the following propositions:

(1) The presence of quantum states makes it the case that the action that S will bring about can metaphysically unfold more than one way, in contrast with determinism.

(2) The presence of quantum states makes it the case that it is in S's power to bring about the action A or its alternative B.

Kane strongly endeavours to convince the reader that quantum indeterminism is somehow integrated into S's agential powers as in the second proposition. However, the true description of it seems to be the first one. All it entails is that the presence of quantum indeterminism brings about the state of affairs that the resultant

action is undetermined; and it is undetermined only by virtue of quantum indeterminism. Hence, Dennett's objection to the Fusion Principle is sound and it is an untenable principle in such a physicalistic framework. As we touched upon before, Chalmers' interpretation of the Integrated Information Theory allows for agential interventions in the quantum states in a genuine sense. However, Kane would have to admit that consciousness is somehow a fundamental feature of the universe and it exerts a special kind of causation; if he were to save the Fusion Principle by this way, certainly it would be a save at the cost of violating his entire initial motivation to remain in the boundaries of modern science/physicalism.

That was my argument against Kane's libertarianism; however, this may not fully do justice to this sophisticated account. My argument stands and falls with the failure of non-reductive physicalism, and it was not very clear what kind of non-reductivism his account was based upon. Therefore, I decided to ask Prof. Kane myself, about how he manages to deal with the problems in non-reductive physicalism in the face of Kim-like attacks. I was also informed that he will deal with the questions about human agency on a larger scale in his upcoming book. At this point, I would like to present some fragments of his reply:

1. Reality is describable at different levels, and the higher levels having to do with human thought and agency are among those not reducible to (though they may supervene on), lower physical levels.
2. The designation of Kane's account as "event-causal" is misleading. It turns out that when Kane started developing his ideas on free will in 1960's, event-causal libertarianism was absent in the philosophical literature. Later on, philosophers began to make another classification of those accounts which do not either fall under the category of non-causal libertarianism or agent-causal. However, Kane was never contented with the implications of event-causal concept since it reduces freedom of the will into causation by mental events. Kane's ontological position with regard to agency is different from event-causal libertarianism in the sense that agential powers cannot merely be explained by/reduced to a succession of indeterministic mental states. In a similar way, he objects to Davidsonian framework of mental causation (beliefs + desires = action).

3. Kane's response to Kim-like attacks is grounded on the same principles that are explicated by Peter Ulrich Tse in *The Neural Basis of Free Will* and the book titled *Downward Causation and the Neurobiology of Free Will* which was edited by Murphy, Ellis and O'Connor (R. Kane, personal communication, April 21, 2019).

Peter Tse's book is a quite sophisticated one as well as *Downward Causation*. Because I do not see any fundamental difference in their response to Kim-like objections, I just want to touch upon *the Neural Basis of Free Will* in this context. Tse has created a three-stage model for mental causation and free will, that signify a powerful objection to reductive physicalism and reductive versions of Information-Processing Theories of mental causation. Tse proposes a very useful term in order to define his own version of mental causation: "Criterial Causation". In this account, "that *patterns in input can be genuinely causal only if there are physical detectors, such as neurons, that respond to patterns in input and then change the physical system in which they reside if the criteria for the presence of a pattern in inputs have been met"* (Tse, 2013, p. 9). By detectors, it is meant that for instance, NMDA detectors can scan certain patterns in neurons or neuron groups and this is what determines whether the criteria for subsequent firings/excitation are met or not. In simpler terms, the brain functions with regard to quasi-immaterial patterns/sets of information/criterial codes. This is ultimately what makes the neural basis of mental causation something of a very close "approximation" (Remember: Dennett) of Kantian mind which take everything into consideration. Criterial codes that are embedded in neural pathways or individual neurons are the prime movers of the volitions, which supposedly renders this account very different from the previous IPTs. Here are the three stages of the model for Tse (2013):

> (1) new physical/informational criteria are set in a neuronal circuit on the basis of preceding physical/mental processing at time *t* 1, in part via a mechanism of rapid synaptic resetting that changes the effective inputs to a

> postsynaptic neuron; these changes can either be driven volitionally or non-volitionally, depending on the neural circuitry involved.
> (2) At time *t* 2, inherently variable inputs arrive at the postsynaptic neuron; and
> (3) at time *t* 3 physical/informational criteria are met or not met, leading to postsynaptic neural firing or not. Randomness can play a role in the first two stages, but not substantially in the third, because intracellular potential either passes the threshold for firing or it does not (p. 25).

In principle, the criterial causation is supervenient on the physical basis of brain. To keep it short, we are going back to the same discussion: Even though it seems so obvious that there are higher level causal systems in the brain, which is true in a sense; these higher-level features are weakly emergent or supervenient upon the physical complex system out of which they come into existence. The criterial causation is not something uninteresting like my pen's length being-15-cm-long, as we have admitted before, the chaotic systems cannot be explained by simple add-up compositional principles; however, it does not mean that they are not in principal totally determined by the underlying particles and forces. In some sense, the brain is an approximation of an idealized mental system that lies beyond the blind physical forces, as if there are "immaterial" information codes that consist in our mental states and are doing a real job.

So, despite my fascination in the face of this account, I will argue the same way: Criterial causation along with the other information-processing accounts is in fact causally observer-relative, the sense of substantiveness it establishes depends on our epistemic limitations to describe the chaotic systems in reductive terms. Construed in this way, it seems like a substantive solution to Kim-like attacks; however, it achieves this at the cost of rendering itself an observer-relative phenomenon. Therefore, criterial causation also fails to encounter Kim-like attacks on a genuine ontological basis. If this is the neural basis of free will, the sourcehood principle cannot be sufficiently met. Is Kane's insistence that one has to refer to "the

agent" not simply "agent-involving mental events" scientifically accountable in any version of these physicalist accounts? I cannot make a strong judgement about it. However, I have strong reasons to believe that we need to have a robust sense of agency and self that are irreducible to underlying forces. If Dennett is right, then our philosophical speculations about such a need might be just the insistence on "the traditional dogma"; the sense of robust agency might just be on a par with "gravitational centers in physics" about which Dennett (1991) asserts: "So indeed do physicists, who, if asked what a center of gravity was made of, would say, 'Nothing!'" (p. 95). I'm not sure about the potential limitations of cognitive science, but there is a good chance that they can show us that the concept of irreducible agency/self is just an entertainment of the philosopher, if this is truly the case. Once this happens, we shall all become compatibilists and limit our sense of responsibility and existential meaning in a kind of Spinozistic framework.

CHAPTER 4

SOME REMARKS ON PHYSICAL SELFHOOD AND ITS IMPLICATIONS

Both John Searle and Roger Scruton agree on the irreducibility of the notion of self. For Searle, we cannot make sense of free will and rationality without a non-Humean concept of self that extends through time. Raymond Tallis (2011) put forward a similar point that one of the indispensable concepts regarding personhood is "tensed time" which cannot be accounted for in the framework of physicalist neuroscience (p. 101). Phenomenological considerations of an agent who is trying to put his actions together involve his memories, present states of affairs and also his self-conception projected to the future. Scruton, however, tries to deal with the problem without ontological concerns like supervenience or emergentism; claiming that what neuroscience tries to show is that we can in principle bring about causal capabilities when material properties of brain are put together; which he asserts is probably true. The fundamental distinction, for Scruton, is the distinction between the subject and object; which is of more likely conceptual substantivity rather than an ontological one (conceptual – metaphysical substantivity distinction is originally proposed by Ted Sider).

As we have mentioned in the beginning, Scruton refers to reactive attitudes pointed out by Strawson which cannot supposedly be thrown away by changes in our metaphysical convictions. It is simply that whenever I am asked about my decisions or courses of actions, the appropriate way of justification is always based on reasons rather than causes or predictions which we refer to in natural events. When I'm asked why I decided to study philosophy in MA program or to work on the problem of free will, I'll be telling my story by giving out my reasons how I came to this point in

each step even if I'm convinced that causal determinism is true. Yes, however I think there are serious philosophical worries arising out of the convictions on physicalism and causal determinism combined.

Let us consider the following version of the Brain Replacement Argument ("The Definition of Death", 2017). Granted the truth of physicalism, scientists develop highly sophisticated techniques to replicate each component of human organs artificially, unfortunately however there is still no cure for brain tumours. One day, the subject John is diagnosed with brain tumour in his frontal lobes this time. Having decided to replace his frontal lobes with artificial ones, neurosurgeons scan his brain to sketch all the neural pathways that he developed so far. After a while he undergoes a very successful surgery and regains all of his abilities beyond any reasonable doubt. One neurosurgeon puts the replaced frontal lobes in a vat for further scientific research. After half a year, John starts suffering awful anomalies again and sees the doctor. It turns out that there developed a major tumour on temporal lobes this time, and John undergoes the same process. The replaced part is put in the vat. John go through the same processes for other tumours development at different parts of the brain, all of its original parts are preserved afterwards. John has now a brand-new brain but all of the necessary conditions for preserving his memories and relevant neural pathways to sustain his original decision-making processes.

After a year, John dies in a terrible car accident, and his brain is irreparably crushed. While his family is mourning at the hospital near his dead body; the neurosurgeon, who had been working on the brain tumours, comes along with the vat in which all of John's original brain components are cured and put together back. Thus, each component of his original brain was cured and reconstructed by the

original parts. John's original brain is re-implanted to his body and his heart is replaced with an artificial one or is made to function again. John wakes up after the most difficult surgery in medical history, with only some of his memories for the past 1-1,5 half year lost. In this strange Ship of Theseus case, is John still alive and under what conditions?

Let us consider some examples: If the answer is that John is not alive given the memory loss, then are the patients with Alzheimer Disease in fact dead or new persons? If John is not alive given the vast amount of change in the physical ground of his body, then we should remember that we undergo vast changes through several years both in our body structure and also neural pathways, at least John's neural pathways and body structure are preserved in many ways though it is reconstructed by artificial material. It does not even have to be synthetic material, for the sake of argument, let us imagine we will be able to produce genuine cells and neurons with 3D printers by installing whatever physical properties we want and with the same kind of material that constitute our organic cells; in such a case, only the construction is brought about in an artificial way.

Consider the possible answer that John is still alive with an affordable memory loss. Then consider also a parallel scenario while John with the artificial brain did not have any car accident, the neurosurgeon who managed to cure the tumour in a lab connected his brain in vat to a sophisticated computer system. When John's brain regains the consciousness, he would be able to talk to the doctor by a computer screen which translates John's mental states into speech and sends feedback to his brain whenever the doctor says something to him. Obviously, in this tragic scenario, John is told that his original body with an artificial brain implant has been living for more than one year and lives a decent life with his family. John's

original brain is shocked and told by the doctor that he cannot go back to his original life because his original body now belongs to someone else who officially sustains his original identity. The doctor says sorry to John and explains the legal concerns that if he had implanted his brain to an artificial body and let him go, then the doctor would end up in prison on the grounds that he cloned a person without even his consent or vice versa, the other doctors would face similar charges at pains to explain to the jury members how come two persons emerged after a succession of brain surgeries.

The problem is that we do not really care about the ways in which the body sustains itself biologically as far as the body and limbs are all functioning properly. I imagine that very few people would be concerned about personal identity problems if the doctors promised them health back again. I think most people who are brought up in theistic traditions (this is more than half of the world population) are quite convinced that their selfhood in fact depends on something transcendental of biological ground, like a noumenal self. On such a conviction, one can only see the biological processes as a means for the connection of soul with the earth. If this is right, you may simply allow for a very wide range of replacements in your body without worrying about selfhood. In the social sphere, we treat each other as individual souls or substances, bounded by rationality and moral reasons regardless of its possibility in a physicalistic framework.

The above thought experiment concerning the persistence of a person in a physicalistic framework partly shows that even the death becomes metaphysically arbitrary in such a picture. Theodore Sider makes a distinction between metaphysical and conceptual substantivity, the latter of which signifies our concepts do not match the fundamental joints of the world. Personal identity is given a variety of definitions

and criteria by numerous philosophers before. If there is no fundamental answer to which of these criteria or definitions are more credible than the others in a metaphysical sense, then we should accept that those problems are not metaphysically fundamental and we might only arbitrarily stick with a philosophical position about it.

There are similar ideas that are explicated by P. F. Strawson in his famous article titled *Freedom and Resentment* that favours the evaluation of moral responsibility within the boundaries of reactive attitudes. On the other hand, if I'm right about Scruton's position, there is no real threat posed by physicalism as far as the distinction between subject and object is preserved properly. In parallel lines with Kant, there are two ways through which we perceive the world: as appearances and things in themselves. We perceive the empirical objects of inquiry and also the morally responsible agents ("as if" their choices emanate from their noumenal selves); however, the dichotomy in our perception is not a matter of choice. I can only treat someone with self-consciousness as a responsible agent, not as an ordinary object on my table. Hence, Scruton tries to show that it is the conceptually substantive aspects that preserves our basic intuitions about freedom and responsibility, not the metaphysical subjects whose level of complexity goes far beyond the understanding of millions of people who exercise their freedom and reactive attitudes on a daily basis.

If this is true, then interestingly enough, BN cannot pose a serious threat either. Searle has made a clear distinction between ontological and causal reducibility of consciousness. Even though the causal history of one's actions go beyond the conditions over which he/she can exert control, the phenomenological actish feelings are still preserved, since the accompanying conscious states cannot be

reduced or eliminated in an ontological sense. So, no matter how much my conscious realm is a playground of external forces, I can anyway exert compatibilist control over them in contrast with non-self-conscious animals, and this would still preserve the subject-object dichotomy and also the reactive attitudes on which our sense of moral responsibility is (supposedly) grounded.

I find it quite interesting that physicalism might fall short of our basic intuitions about the concept as serious as death. It may turn out that our notion of death is not standing on a solid ground as much as we imagine it to be. The definition of death might just be going between a class of definitions that are put forward by different linguistic communities. There are different answers to be given as far as you think that the criterion of personal identity is psychological continuity or it is biological continuity; and it might turn out that none of these arguments match with what is metaphysically substantive in the world. At this point, we might need to get clear on our premises in order to prevent any confusions on the subject:

1) Personal identity and death are conceptually substantive in our phenomenology and social lives.[4]
2) In the above brain replacement case, physicalistic framework falls short of covering which person is in fact alive or dead, and which of them properly persists through time.
3) Given (2), a subject S's selfhood (let's say one's persistence through time) is an observer-relative phenomenon; it relies on the criteria put forward by linguistic communities, not the metaphysical joints in nature. (In Sider's terms)

[4] See also Searle, *Rationality in Action*, p. 2: "Death, one might say, is the horizon of human rationality; but thoughts about death and the ability to plan with death in mind would seem to be beyond the limitations of the ape's conceptual apparatus."

4) Given BN (by Searle), consciousness is not an observer-relative property (which is definitely true).
5) Given (4), S's instantaneous existence is not observer-relative; nevertheless, given (3) the criteria of S's persistence through time are observer-relative.
6) Given BN, we have a relatively stable criterion for persistence such as the biological continuity on which mental/psychological persistence supervenes. Most of the time biological persistence is not a problematic issue.
7) Given (1), (2), (3) and (5); there might come such futuristic scenarios (maybe in a recent future) that S in his/her (metaphysically) genuine momentary existence at T might be left clueless as to whether he/she will wake up at T+1 after the surgery.
8) Given the causal reducibility of conscious realm in BN, S is ultimately what his underlying biological conditions dictates; and given (6), the psychological continuity of S (the feeling that ensures S that he/she is the same S at T+1) is observer-relative in the sense that some other biological conditions can dictate the same feeling at T+1 (after the surgery) even if original S was in fact dead.
9) Given (8), there is no fundamental states of affairs to decide whether S truly survives the surgery at T+1, that is to say, we can only rely on S's sense of psychological continuity and biological persistence (which can somehow be metaphysically unwarranted). We might have the same psychological continuity and strong sense of the self even if we were no more than bundles.
10) Given (9), I think it is plausible to say, we are only left with some version of a Humean bundle in such a framework. It is not fundamentally Subject 1 or Subject 2 persisting through time, but rather there is a succession of living and thinking moments going on one after another. To put it less poetically, physical/biological

conditions fall short of individuating a subject (See also Lowe, 1996, p. 6), which leads us to conclude that personal identity and death are observer-relative issues; on the other hand, BN might still provide us with a relatively stable criterion that can cover all the intuitions in daily social lives with the exception of certain futuristic scenarios. But, again, the future is very close, and we will no longer be able to avoid such problems; and the safe zone of practical attitudes will be interrupted.

CONCLUSION

I have tried to address a variety of serious concerns that arise as a result of physicalistic implications in the libertarian accounts of John Searle and Robert Kane. Causal reductionism has as many serious results as causal determinism on the grounds that it entails that agential control becomes a redundant concept with regard to the fundamental causal structure of the universe. It does not mean to say we have the proper epistemic tools for appealing to any such serious reductive terms; however, it renders the Sourcehood Principle indefensible in a philosophical point of view. I think it is also true that our intuitions about sourcehood are largely shaped by social practice and reactive attitudes as Strawson has argued; but nevertheless, if any of the versions of physicalism that we discussed turns out to be the case, then it is quite difficult to claim the ownership of any of our actions, not in social practices but rather in a fundamental perspective.

In contrast with Kane, I'm not overreactive toward the option of compatibilism even though I also think that it falls short of covering genuine moral responsibility (Fischer et al., 2007, p. 181).[5] Our basic intuitions about moral praise and blame are still incompatibilistic and will continue to be so especially in social practices. One can still embrace (for example) a Spinozistic attitude with respect to free will, and he/she can wilfully accept the idea that he/she is what the universe makes of him/her. In a similar way, Sam Harris has put forward that we are not lost in the storm (neural chaos), or in control of it; we are (simply) the storm. Such a

[5] Kane's "Response": "If I doever read Fischer's future headline and it is true, I would give up my libertarianview and perhaps go over to one of these other views. I think empiricalevidence matters. But I don't know which of these other views I'd go to. Forsomeone with libertarian intuitions like me, it would be like being askedwhether I wanted to live in the desert, in the middle of the jungle or at theSouth Pole."

perspective requires a paradigm of morality and existential meaning quite distinct from the traditional one which I call “the Theistic View”.

John Searle has been an influential philosopher in philosophy of mind and philosophy of language. His famous position that is called “Biological Naturalism” is an attempt to account for the place of consciousness in a naturalistic outlook. Biological Naturalism successfully distinguishes itself from the emergentist accounts of mind that postulate special kinds of causation which we may classify as the strong emergentist view. Based on Chalmers’ distinction between weak and strong emergentism, we have evaluated BN on a large scale and argued that it leads to some version of epiphenomenalism in the sense that BN renders the causal powers of consciousness relational and reducible to underlying micro-physics. We did grant that reductionism is not a simple subject and with this regard we endeavoured to be clear about the kind of reductionism that we stress. It is quite clear in BN that mental states are causally reducible to electro-chemical correlates in the brain. There seems to be nothing causally special about classifying consciousness as a biological phenomenon if the biology itself is also reducible to micro-physical base properties. I think it must have been clear enough for the reader that the so-called higher-level and causally efficacious properties of consciousness are in fact so only by virtue of its micro-physical properties; and their being higher level properties is observer-relative. The causal reducibility of mental states thus becomes a clear-cut violation of the Sourcehood Principle; and it is also quite problematic with respect to the logical structure of free will that Searle defends in the chapter. It seems to us that the nature can be describable at different levels and those levels of description are equally credible. We have tried to address the mistaken part in this conviction. The indispensability of the levels of description in fact stems from our epistemic

limitations, and in Theodore Sider's words, there are those descriptions which do not carve the world at its joints. Reason-explanations have to refer to a certain class of ontologically irreducible entities in the world such as: beliefs, desires, goals, fears etc. However, in BN, whatever the causal aspects of those referents that activate physical effects in the world, they can actually be described in the micro-physical level that does not postulate any of those mental states. How these ontologically subjective entities emerge out of neural correlates is a mystery of its own; however, BN is clear about their causal reducibility and by virtue of this fact, agential powers of a subject do not have a causal autonomy of its own. It is true that biological persistence and the integrity of biological structure ensures a certain level of stability in the decision-making system of a subject; but this is not sufficient to account for a genuine sense of sourcehood/authorship.

As for Kane's libertarianism, I argued that SFAs fail to account for sourcehood in a proper way. In my argument against Kane, I tried to show that, one of the key principles, the Fusion Principle, falls short of constructing a kind of control that the compatibilist cannot afford. Dennett's criticism of SFAs is adequate in the sense that not only the quantum indeterminism must be within the agent but also the agent must take the ownership of quantum indeterminism, which does not seem plausible. There is logical difference between the amplification of quantum indeterminism making it the case that the course of action can metaphysically unfold in more than one way and the quantum indeterminism in the brain making it the case that the agential control is enhanced by virtue of the integration of quantum indeterminism. We grasp the sort of discrepancy in Kane's account that is similar with the one in BN: both of these accounts are designed to preserve an irreducible agency whereas they are highly motivated to avoid any special kind of causation.

And that is the key issue that we endeavoured to address throughout the thesis: Physicalism leads to the blurring of the key distinction between the subject and object by virtue of the kind of causal structure it dictates; so there seems not to be a fundamental difference between acting upon or be acted upon in such a causally reductive picture.

I find the Dennettian position more coherent in most of the cases because Dennett admits that, for instance, there is no deep fact to distinguish the intentional stance of a thermostat from the (so-called) underived intentional content of an agent. Dennett's strategy is grounded on pushing those with incompatibilist intuitions outside the naturalistic picture; so, he can claim the victory of compatibilism over the rest, on the criterion of scientific credibility. The ultimate advantage of this physicalist monism is that it accounts for the so-called dual aspects of subjects in the world and binds them together in a singular explanatory basis. Dennett is an ontological reductionist in one sense; however, he is trying to convince his reader that the truth of physicalist monism does not necessarily render our cherished values redundant but rather it provides a well-grounded rational substitute for them.

We have mentioned Roger Scruton's assertions as a subsidiary position throughout the paper. Scruton is not interested in metaphysical implications of physicalism, on the contrary, he embraces its implications and the third-person-oriented projects of cognitive science. Scruton is trying to remind us that our intuitions of accountability and freedom are not predicated on our metaphysical convictions and they can't change the nature of reactive attitudes on which moral responsibility is grounded. We encounter each other via the appearances and treat each other as sources of morality by looking at each other's faces by which we infer each other's motivations and integrity. If Scruton is right, we need not worry about Churchland-like projects

which are motivated to render these appearances redundant, since they are deeply embedded paradigms in our perception of each other. Our perception of the world is supposedly based upon two incommensurable levels as Kant has explicated before: one is the causal relations and second one is the moral paradigm through which we attribute freedom and sourcehood to each other. According to this view, physicalism is something to be avoided because we cannot simply choose to eradicate the second vision of the world, namely the moral perception. In other words, we do not have any choice to see each other as mere physical objects that are only causally related to one another.

On the other side, Robert Kane, John Searle and some other libertarians including myself are seeking a way to ground libertarian intuitions on a proper metaphysical basis. There is another option that is called "agent-causal libertarianism" which we have not discussed in this paper. John Searle is explicitly a reductionist about the causal relata of consciousness, which renders BN an epiphenomenalistic account with respect to Kim-like criticisms. Robert Kane's failure in accounting for a robust sense of agency, as I tried to show, rests upon similar physicalistic implications. I do not claim any physicalistic libertarianism is doomed to failure, but that is the case with regard to Searle and Kane. An influential philosopher in this related area, Derk Pereboom, has seen the weaknesses of event-causal accounts (in this case, even if Kane is not an event-causalist, I assume he is very close to it), and decided to find a way in agent-causal theories. This move is not really a desirable one because it pushes you out of the scientific plausibility even if plausibility has vague boundaries. In one of his latest books titled "Living Without Free Will", Pereboom gave up on libertarianism altogether on the grounds that he decided agent-causal view is impossible. I do not know whether it is impossible; and

I don’t know whether we really need to have a metaphysical ground to protect our libertarian intuitions in real-life scenarios, either. As one of the Oscar Wilde’s (2004) characters says: “The truth is rarely pure and never simple. Modern life would be very tedious if it were either, and modern literature a complete impossibility!” (p. 54).

REFERENCES

Blackmore, S. (Ed.). (2005). *Conversations on consciousness: What the best minds think about the brain, free will, and what it means to be human.* Oxford: Oxford University Press.

Block, N., Flanagan, O., & Güzeldere, G. (Eds.). (1997). *The nature of consciousness: Philosophical debates.* Cambridge, MA: MIT Press.

Buber, M. (1970). *I and thou*. New York: Scribner.

Bundy, T. (2019, 01 24). The Ted Bundy tapes. *Conversations with a killer.* (S. Michaud, Interviewer) Netflix. Retrieved from https://www.netflix.com/watch/80226550?trackId=14170289&tctx

Chalmers, D. (2006). Strong and weak emergence. In P. Clayton, & P. Davies (Eds.), *The re-emergence of emergence: The emergentist hypothesis from science to religion* (pp. 244-254). New York: Oxford University Press.

Chalmers, D. (2017, August 28). David Chalmers - Physics of consciousness. (R. L. Kuhn, Interviewer) Closer To Truth. Retrieved from https://www.youtube.com/watch?v=CIooLxk2jxs

Churchland, P. S. (2002). *Brain-wise: Studies in neurophilosophy.* Cambridge, MA: MIT Press.

Clovis Meroveus. (2018, July 1). Retrieved from YouTube: https://www.youtube.com/watch?v=aQVEsBMUVQE

Crane, T. (2001). The significance of emergence. In C. Gillet, & B. Loewer, *Physicalism and its discontents* (pp. 207-224). New York: Cambridge University Press.

Crane, T. (2010). Cosmic hermeneutics vs. emergence: The challenge of the explanatory gap. In C. Mcdonald, & G. Mcdonald (Eds.), *Emergence in mind* (pp. 22-34). New York: Oxford University Press.

Crawford, M. B. (2008). *The limits of neuro-talk.* Retrieved October 19, 2018, from The New Atlantis: https://www.thenewatlantis.com/publications/the-limits-of-neuro-talk

DeGrazia, D. (2017, March 21). The definition of death. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved May 10, 2019, from https://plato.stanford.edu/archives/spr2017/entries/death-definition/

Dennett, D. (2012, September 26). The atheism tapes - Daniel Dennett [6/6]. (J. Miller, Interviewer) The Miller Tapes. Retrieved from https://www.youtube.com/watch?v=fvG-q7VrFPg

Dennett, D. C. (1984). *Elbow room: Varieties of free will worth wanting.* Oxford: Clarendon Press.

Dennett, D. C. (1991). *Consciousness explained.* New York: Back Bay Books.

Dennett, D. C. (2003). *Freedom evolves.* New York: Penguin Group.

Dennett, D. C., Searle, J. R., Bennett, M., & Hacker, P. (2007). *Neuroscience and philosophy: Brain, mind and language.* New York: Columbia University Press.

Dilman, I. (1999). *Free will: An historical and philosophical introduction.* New York: Routledge.

Frankfurt, H. (1969, December 4). Alternative possibilities and moral responsibility. *Journal of Philosophy, 66*, 823-839. doi:10.2307/2023833

Frie, R. (Ed.). (2008). *Psychological agency: Theory, practice, and culture.* Massachusetts: MIT Press.

Glannon, W. (Ed.). (2018). *Free will and the brain: Neuroscientific, philosophical, and legal perspectives*. Cambridge: Cambridge University Press.

IanRamseyCentre. (2012, August 2). Retrieved from Youtube: https://www.youtube.com/watch?v=249MmChENhs&t=2693s

Inwagen, P. v. (1986). *An essay on free will.* Oxford: Clarendon Press.

Kane, R. (1996). *The significance of free will.* New York: Oxford University Press.

Kim, J. (1999, August). Making sense of emergence. *Philosophical studies: An international journal for philosophy in the analytic tradition, 95*(1/2), 3-36. Retrieved from https://www.jstor.org/stable/4320946

Lowe, E. J. (1996). *Subjects of experience.* Cambridge: Cambridge Univeristy Press.

McKenna, M. D., & Coates, J. D. (2018, December 21). Compatibilism. In E. D. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved May 10, 2019, from https://plato.stanford.edu/archives/win2018/entries/compatibilism/

Mele, A. (2006). *Free will and luck.* New York: Oxford University Press.

Mele, A. R. (2017). *Aspects of agency: Decisions, abilities, explanations, and free will*. New York: Oxford University Press.

Merricks, T. (2001). Epiphenomenalism and eliminativism. In *Objects and persons* (pp. 56-84). Oxford: Clarendon Press.

Merricks, T. (2001). *Objects and persons.* Oxford: Clarendon Press.

Mirolli, M. (2002). A naturalistic perspective on intentionality. Interview with Daniel Dennett. *Mind & Society, 3*, 1-12.

Modell, A. (2008). The agency of the self and the brain's illusions. In R. Frie, *Psychological agency: Theory, practice, and culture* (pp. 35-49). Cambridge, MA: MIT Press.

Murphy, N., & Brown, W. S. (Eds.). (2007). *Did my neurons make me do it?: Philosophical and neurobiological perspectives on moral responsibility and free will.* New York: Oxford University Press.

Murphy, N., Ellis, G. F., & O'Connor, T. (Eds.). (2009). *Downward causation and neurobiology of free will.* Berlin: Springer. doi:10.1007/978-3-642-03205-9

Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.

O'Connor, T. (2000). *Persons & causes: The metaphysics of free will.* New York: Oxford University Press.

O'Connor, T., & Wong, H. Y. (2005, December). The metaphysics of emergence. *Noûs, 39*, 658-678. Retrieved from https://www.jstor.org/stable/3506115

Parfit, D. (1984). *Reasons and persons.* Oxford: Clarendon Press.

Peinfield, W. (1975). *Mystery of the mind: A critical study of consciousness and the human brain.* New Jersey: Princeton University Press.

Pereboom, D. (2004). Is our conception of agent-causation coherent? *Philosophical Topics Vol. 32, No. 1/2, Agency*, 275-286.

Pereboom, D. (2012, March). The disappearing agent objection to event-causal. *Philosophical Studies*, 56-69. doi:10.1007/s11098-012-9899-2

Pereboom, D., Fischer, J. M., Kane, R., & Vargas, M. (2007). *Four views on free will.* Oxford: Blackwell Publishing.

Popper, K. (2006). *Hayat problem çözmektir: Bilgi, tarih ve politika üzerine.* (A. Nalbant, Trans.) Istanbul: Yapı Kredi Yayınları.

Russell, P., & Deery, O. (Eds.). (2013). *The philosophy of free will: Essential readings from the contemporary debates*. New York: Oxford University Press.

Scruton, R. (2017). *On human nature.* Oxford: Princeton University Press.

Searle, J. (2002). *The rediscovery of mind* (9 ed.). Cambridge, MA: MIT Press.

Searle, J. (2002). Why I am not a property dualist. *Journal of Consciousness Studies*, 9, 54-64.

Searle, J. (2016, 01 28). John Searle - Can Brain Explain Mind? (R. L. Kuhn, Interviewer) Retrieved January 11, 2019, from https://www.youtube.com/watch?v=ehdZAY0Zr6A

Searle, J. R. (2001). *Rationality in action.* Massachusetts: MIT Press.

Searle, J. R. (2007). *Freedom and neurobiology: Reflections on free will,language, and political power.* New York: Columbia University Press.

Sider, T. (2011). *Writing the book of the world.* Oxford: Clarendon Press.

Slotjaw. (2009, September 10). Retrieved from Youtube: https://www.youtube.com/watch?v=wBM4-QZMfX4

Strawson, G. (2000, January). The unhelpfulness of indeterminism. *Philosophy and Phenomenological Research, 60*, 149-155. Retrieved from https://www.jstor.org/stable/2653435

Strawson, P. (1962). Freedom and resentment. In G. Watson (ed.), *Proceedings of the British Academy*, Volume 48: 1962. Oup Oxford. pp. 1-25.

Symons, D. (1981). *The evolution of human sexuality.* Oxford: Oxford University Press.

Tallis, R. (2011). *Aping mankind: Neuromania, darwinitis and the misrepresentatıon of humanity.* Durham: Acumen.

Timpe, K. (2008). *Free will: Sourcehood and its alternatives.* New York: Continuum.

Tse, P. U. (2013). *Neural basis of free will: Criterial causation.* Massachusetts: MIT Press.

Vargas, M. (2007). Revisionism. In J. M. Fischer, R. Kane, D. Pereboom, & M. Vargas, *Four views on free will* (pp. 126-165). Oxford: Blackwell Publishing.

Wilde, O. (2004). *The importance of being earnest.* (R. Jackson, Ed.) London: A & C Black.