# FREE WILL AND DETERMINISM:
# ARE THEY EVEN RELEVANT TO EACH OTHER?

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

HASAN ÇAĞATAY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF PHILOSOPHY

FEBRUARY 2012

Approval of the Graduate School of Social Sciences

_____

Prof. Dr. Meliha Altunışık
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

_____

Prof. Dr. Ahmet İnam
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

_____

Prof. Dr. Teo Grünberg
Supervisor

**Examining Committee Members**

| | | |
|---|---|---|
| Assoc. Prof. Dr. Erdinç Sayan | (METU,PHIL) | _____ |
| Prof. Dr. Teo Grünberg | (METU,PHIL) | _____ |
| Prof. Dr. David Grünberg | (METU,PHIL) | _____ |
| Prof. Dr. İlhan İnan | (BOUN,PHIL) | _____ |
| Assoc. Prof. David Pierce | (MSGSU,MATH) | _____ |

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

Name, Last name : Hasan Çağatay

Signature :

# ABSTRACT

FREE WILL AND DETERMINISM:
ARE THEY EVEN RELEVANT TO EACH OTHER?

Çağatay, Hasan
Ph.D., Department of Philosophy
Supervisor: Prof. Dr. Teo Grünberg

February 2012, 132 pages

Many philosophers tend to defend the view that there is a significant relation between the problem of *determinism* / *indeterminism* and the problem of *free will*. The belief that there exists such a significant relation is supported by our intuitions; however, in this thesis, I defend just the opposite view: free will has no significant dependence on the deterministic or indeterministic character of causal relations. In the same way, I propose that the question, whether or not determinism is true, cannot be answered based on observations about the problem of free will.

I believe that the genuine question whose answer would illuminate the darkness surrounding free will is whether or not *will* supervenes on anything other than itself. Therefore, in order to decide whether or not we are free, the question we should ask is "Does will supervene upon something other than itself?" Moreover, I defend the position that no matter whether the world is deterministic or

indeterministic, if physicalism is true, i.e. if properties of free will supervene upon physical properties, then we cannot enjoy *genuine* freedom.

The position of the thesis has some important ethical implications: If we cannot be genuinely free, we cannot be genuinely responsible for our actions either. This implies that retributive and admirative desires towards other persons are rationally untenable. I defend the view that only practical attitudes like reinforcement and punishment or isolation and inclusion are rationally tenable.

Keywords: Compatibilism, Determinism, Free Will, Foreknowledge, Incompatibilism

# ÖZ


ÖZGÜR İRADE VE BELİRLENİRCİLİK:
BİRBİRLERİYLE GERÇEKTEN İLGİLİLER Mİ?

Çağatay, Hasan
Doktora, Felsefe
Tez Yöneticisi: Prof. Dr. Teo Grünberg

Şubat 2012, 132 sayfa

Birçok düşünür *özgür irade* problemiyle, *belirlenircilik/belirlenmezcilik* problemi arasında önemli bir ilişki olduğunu savunma eğiliminde. Böyle bir ilişkinin var olduğu görüşü sezgilerimizle destekleniyorsa da, bu tezde karşıt fikri savunacağım: özgür irade problemi, nedensel yapının belirlenir ya da belirlenmezliğine bağımlı değildir. Aynı biçimde, bana göre, belirlenirciliğin doğru olup olmadığı sorusuna, özgür iradeye dair gözlemlerimizden yola çıkarak cevap verilemez.

Bana göre, özgür irade kavramının çevresindeki sisi dağıtacak asıl soru, iradenin kendisinden başka bir şeyin gölgesi olup olmadığı, kendisinden başka bir şeye bağımlı olup olmadığıdır. Bu nedenle, özgür olup olmadığımızı anlamak için sormamız gereken soru "İrade kendisinden başka bir şeye bağımlı mıdır?" ya da "Fizikalizm doğru mudur?" olmalıdır. Ayrıca, fizikalizm doğruysa, belirlenirci bir dünyada yaşayıp yaşamadığımızdan bağımsız olarak, *gerçek anlamda* bir özgürlükten bahsedilemeyeceğini savunuyorum.

Bu pozisyonun bazı önemli etik imaları da var: Eğer gerçek anlamda özgür değilsek, gerçek anlamda sorumluluk taşıyabileceğimiz de söylenemez. Bu, kişilere karşı intikamcı ya da övücü duygularımızın akılcı olmadığına da işaret ediyor. Bunların yerine, ödül ve ceza ile izolasyon ve bir arada bulunma gibi pratik tutumlar savunulabilirliğine inanıyorum.

Anahtar Kelimeler: belirlenircilik, belirlenmezcilik, önbilgi, özgür irade, tahmin

To My Family

# ACKNOWLEDGMENTS

I wish to express my gratitude to my supervisor Prof. Dr. Teo Grünberg and Prof. Dr. David Grünberg for their guidance. They have always encouraged me to think in an explicit way and to develop new ideas.

I would like to thank Assoc. Prof. Dr. Erdinç Sayan for his very important contributions. He is the one who introduced me the problem of free will throughout his comprehensive courses. In these courses and in our private discussions, I had opportunity to come across with new questions and perspectives to the problem. Moreover, forth chapter of this thesis is based on the discussions between him and me.

Prof. Dr. John M. Fischer helped me to realize one of my critical mistakes in this thesis. Prof. Dr. Ted Honderich and Prof. Dr. Daniel C. Dennett encouraged me when I informed them about my thesis.

Assoc. Prof. Dr. David Pierce supported me in understanding a mathematical study related to this thesis. Besides we had very interesting discussions about the ethical implications of the thesis.

Çiğdem Çukurcu and Dr. Mahmut Özer shared their valuable opinions on the thesis. They have also supported me psychologically throughout this work. Thanks to them for being by my side.

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

*ALD*: Almost Laplace's Demon

*NDP*: No-Dependence-Position

*PAP*: Principle of Alternative Possibilities

*PPA*: Principle of Possible Action

*PPP*: Principle of Possible Prevention

*SFA*: Self-Forming Actions

*UR*: Ultimate Responsibility / Ultimately Responsible

# CHAPTER I

## INTRODUCTION

Will is some kind of power to determine one's actions. Or as Jonathan Edwards (2001) describes "the faculty of the will is the power of, or source in, the mind by which it is capable of choosing." *Choosing* or *determining one's actions* presupposes desires. Without desires and aversions, we would not care outcomes of the actions and we would have no reason to choose. Therefore, will presupposes desires. Without desires, being organic computers or robots, human beings would not enjoy freedom. Will requires desires; however, our having desires is not sufficient condition to assure that we have will. To say that we have will, we have to have a power to perform or at least to *try* to perform the actions that we desire to do. We all agree that each of us apparently has will, since we frequently satisfy our desires by determining our actions. We want to stand up, and we satisfy this will by standing up. We want to raise our hand, and we satisfy this will by raising our hand.

We also believe that this will we have is *free*. By our will's being free, most of us roughly mean that it is determined by nothing but ourselves. Even if we have an intuition that our will is free (it is not determined by anything besides us), philosophers tend to test this intuition before they accept it.

The supposition that events in our world are completely or at least partly determined has always been perceived as a threat for free will. In this thesis mainly the view that the world is completely determined by causal laws, namely *causal determinism*, is focused on. If causal determinism is true, then everything happening is depend on the strict causal laws of nature. Therefore, if causal determinism is true, then for the complete set of true propositions for each time point $t_1$ (input), there is only one possible complete set of true propositions for a future or past time point $t_2$ (outcome). For determinists, when I buy a lottery ticket, it is strictly determined whether or not I would win. Causal laws have such a deterministic character, such that even the choices of a person who would draw for the lottery are strictly predetermined.

If determinism is true, then with respect to alternative possibilities, the future is not different than the past: They are both determined or unchangeable. In other words, if determinism is true, an event at a time point *t* could not have happened or may not happen in a way different than it was or it will. Determinism seemingly implies that there is no real probabilistic event or *genuine free will.*[1] If determinism is true, even if we do not know in which way an electron, a die or a person would behave, each of these may behave in only one strictly predetermined way at a particular time point *t*. This implies that complete empirical data about the world for any time $t_1$ and conjunction of all the causal laws are adequate to derive complete empirical data belonging to a past or future time point $t_2$, which can schematically expressed as follows:

[1] By "genuine free will," I mean the referent of the concept of free will used mostly by incompatibilists, which demands conditions which are hard to be satisfied like *alternative possibilities* or *self origination*: I name this kind of concepts of free will as "*strong free will*" later in this section. See also van Inwagen, 1983, 1989; Kane, 1989, 1996, 1999.

$Et_1$ (Complete empirical data for a time point $t_1$)

$L$ (Conjunction of laws of nature)

$Et_2$ (Complete empirical data for a time point $t_2$)

One of the most fundamental questions about free will comes up here: If every event, including our decisions, are completely dependent on the causal laws and facts belonging to a time when we had not been born, then how can our wills or anything in our world can be free? And since we are responsible for neither the state of world belonging a time when we had not been born, nor for the laws of nature, how can we be responsible for our actions, which are predetermined by these two factors?[2]

Resting on various arguments, some thinkers believe that we cannot be free, if determinism is true. Some others asks, "Why not?". There are four fundamental opinions about the dependency of *freedom of will* on determinism: (1) Firstly, *hard incompatibilists* (or *pessimistic incompatibilists*), like Derk Pereboom (1995), Galen Strawson (2000), and Janet Radcliffe Richards (2000) believe that we may not have free will, no matter whether determinism is true or false. (2) The other incompatibilists, *libertarian incompatibilists*, *agnostic incompatibilists*[3] and *hard determinists* like Peter Van Inwagen (1983, pp. 93-105; 1989, pp. 404-405), Robert Kane (1989, 1996, 1999) and Paul Holbach (1957) believe that especially determinism is incompatible with free will, because of its distinctive features. For them, if we have free will, then our world has to be indeterministic. (3) There are

[2] Peter van Inwagen rejects compatibilism using a more systematical version of this argument, *the consequence argument* (1983, pp. 93-105). I will return to this argument and deal with it in the section "Another Interpretation of PAP," in detail.

[3] Thinkers who believe that free will and determinism are incompatible but are agnostic about whether or not we have free will and whether or not our world is deterministic.

also some compatibilists who believe that if we have free will, this means that our world *has to be* deterministic (Hobart, 1934). Since they do not believe that free will and indeterminism are compatible, it can be proper to call them "*pessimistic*[4] *compatibilists*." (4) Some others philosophers, like Harry Frankfurt (1969, pp. 838-839)[5] and Daniel C. Dennett (1984a, 1984b) argue that free will and determinism are compatible, like the former group of compatibilists. However, these philosophers also accept that free will and indeterminism are compatible, unlike the former compatibilists. This is why it can be accurate to call them "*optimistic compatibilists*." Even if pessimistic incompatibilists and optimistic compatibilists seem to be positioned at two poles of the discussions, they have an opinion in common: These two schools are likely to hold the position that the problem of free will and the problem of determinism / indeterminism are *not* actually dependent on each other (at least not in the sense that the other thinkers believe).

The position of the first chapter is quite close to that of pessimistic incompatibilists and optimistic compatibilists. Just like them, I believe that the question whether or not we have free will is not dependent on the question whether or not our world is deterministic. I share some intuitions defended by each of these schools: Even if my strong pessimistic intuition pushes me to believe that we do not have free will, I also have a weaker optimistic intuition which says that we have free will, whether or not the world is deterministic. These seemingly opposing intuitions do not necessarily contradict each other. There are two basic types of concepts of free will[6]. In the first chapter, I argue for the view

---

[4] By naming philosophical positions as "pessimistic" or "optimistic," I do not intend to make value judgments about having freedom in any sense. I am just following the convention.

[5] He also defines (1971, 1987) and defends a new compatibilist concept of free will based on higher-order desires.

[6] Obviously, there are more than two concepts of free will in the literature; however, for the sake of simplicity, in this thesis, these various concepts will be classified as concepts of "*weak free will*" or concepts of "*strong free will*". To illustrate, Frankfurt, Gary Watson and Susan Wolf

that we have one of them, which is traditionally called "*compatibilist free will"* and which I call "*weak free will*," and that we do not have the other, which is called "*libertarian free will*" and which I call "*strong free will*".[7] In the second chapter, I test this claim by considering some mainstream sets of criteria for free will.

The third chapter is more practical than the first two: A few thought experiments which are somehow related to Laplace's demon[8] are discussed. These thought experiments may seem (wrongly in my opinion) to some thinkers to be demonstrations of the *significant* relation between problem of free will and problem of determinism / indeterminism. Throughout that chapter, various candidates of solutions for the puzzling thought experiments are considered and it is defended that these thought experiments do not really threaten the position of the first two chapters. Moreover, in that chapter, logical and nomological possibilities of foreknowledge is discussed.

Determinism is perceived as a threat not only to freedom, but also to responsibility. If every event has to happen in the way it does, how can we be genuinely responsible for what we do? In the last chapter, it is argued that we may not have genuine responsibility. Some of the ethical and practical implications of our having not *strong* free will, but only *weak* free will is discussed in that chapter.

---

suggest various concepts of weak free will, and Kane and Strawson suggest various concepts of strong free will.

[7] Ted Honderich (1996, p. 856) defines two parallel concepts. See also, Kane, 1996.

[8] The intelligence Pierre-Simon Laplace (1902, p. 4) describes.

# CHAPTER II

# FREE WILL AND DETERMINISM: ARE THEY EVEN RELEVANT TO EACH OTHER?

## 2.1 Two Concepts of Free Will

In his 1996 (p. 856) paper, Ted Honderich proposes that "we have two equally important ideas of freedom" and he claims "that for this reason both Incompatibilism and Compatibilism are mistaken[…]."

For him, behaviors which satisfy only voluntariness condition are free in the weaker sense and voluntariness is perfectly compatible with determinism. To describe the other, hard-to-satisfy idea of freedom he first defines *origination*: "An action is *originated* if it is within the control of the agent but is not the effect of a certain causal sequence" (Honderich, 1996, p. 855). To have the freedom in the stronger sense, our actions should be both voluntary and originated and Honderich believes that this idea of freedom is not compatible with determinism (Honderich, 1996, p. 856). I think this border between two ideas is reasonable, and Honderich's claim that only one of these ideas of freedom is compatible with determinism is true. However, he also implies that strong freedom is compatible

with indeterminism by emphasizing determinism's (*only* determinism's) being incompatible with free will in the stronger sense. This is the point where we begin to disagree. I discuss our disagreements with Honderich in the section *Origination.* In this section, the distinction between the *weak* and the *strong* free will, to which the rest of the dissertation will have numerous references, is elucidated.

Concepts of weak free will are simpler to deal with. The Humean concept of free will can be considered to be one of the most fundamental versions of the concept of weak free will: Hume claims that if you have some *volitional* actions, in which your desires seem to have a causal role, you have (*weak*) free will (Hume, 1975, p. 95).[9] You have weak free will, if and only if you *could (would) do otherwise, if you wanted to do so*. I will concentrate on the concepts of weak free will mostly in the sections named "*I, Puppet*," "*Compatibilist PAP*" and "*Free to Will What We Want to Will*."

Weak freedom is easy to talk about and easy to ascribe to our volitional actions. Strong free will, on the other hand, is too slippery to handle. It cannot be satisfied merely by the existence of seemingly volitional actions. To say that we have strong free will, we have to satisfy harder criteria like *ultimate responsibility* in Kane's sense (1989, 1996, 1999), *self-origination* (Strawson, 2000), or *principle of alternative possibilities*.

The position of this thesis is dependent on these two concepts of free will: I defend the view that independently of the problem of determinism, we *do* have *weak* free will and we do *not* have *strong* free will. Philosophers seem to agree in that we do have weak free will. Therefore, I will take the problem of free will as the problem of strong free will. In the rest of the work, I will use "free will" to refer to "strong free will", unless stated otherwise.

---

[9] See also 6th chapter of Moore, 1912 and McIntyre, 1994.

Taking “free will” as “strong free will” makes the position of this thesis quite close to hard incompatibilism. However, since I claim that, just like the problem of strong free will, the problem of weak free will is not dependent on the problem of determinism / indeterminism, it can be more self-explanatory to refer to my position as “*no-dependence position*” (*NDP*).

## 2.2 I, Puppet

There is a critical monist assumption this dissertation is committed to: I assume that physicalism is true, like most of the contemporary philosophers do implicitly when they are dealing with free will. If physicalism is true, our world supervenes upon the physical. Everything in our world is dependent on physical entities. This assumption may seem to be a little pessimistic since it implies that we are just puppets of the physical entities, which act according to the laws of physics. We are complex puppets with feelings, wills, ideas, and so on; and the physical is the puppeteer. Since we achieve our goals harmonious with our desires, we, puppets, possess a belief that we have “free” will.

First, let us suppose that the puppeteer (the physical world) behaves deterministically; it animates us according to a set of deterministic rules, such that for each animation (behavior, feeling, thought, etc.) of the puppets (us) at $t_i$, the next animation to be performed at $t_{i+1}$ is strictly determined. Now, no matter how complex and *seemingly* free these puppets are, is not it counter-intuitive to claim that these puppets (us) have free will? Their wills and behaviors are strictly determined by the puppeteer. Knowing that their wills are in the hands of a puppeteer, should we still call them “free?” As far as my intuitions are concerned, if we are living in a physicalistic deterministic world, we cannot have free will.

We may have a belief in free will, and we may be capable of achieving goals which are coherent with our will, but since our will and behaviors supervene upon physical entities, we cannot call them "free," I believe, without weakening the meaning of "free."

Probably, some compatibilists would not agree with me. However, this is not due to our conflicting beliefs about the empirical facts. They would not agree with me, because their concept of *weak* free will is more embracing than the concepts of *strong* free will under consideration. They would say, "We, puppets, have complex structures like desires and wills anyways. Even if these complex structures are strictly determined by the physical world, *still*, goals coherent with them are somehow reached. Therefore, we, the puppets, should be considered to be free. "After all," they would say, "the concept of (*weak*) free will does not require being independent of the physical world (the puppeteer)."

As long as we have the same opinion that living in a classical physicalistic world makes us, puppets, in the control of the physical world, I do not think it is crucial to discuss whether or not we should call ourselves "free". Let compatibilists say that we are free, having the concept of *weak* free will in their minds, and let me say that we are not free, having the concept of *strong* free will in my mind.

The implications of a deterministic physicalistic world having been presented, now the other physicalist possibility, the possibility of our living in an indeterministic physicalistic world, will be discussed: This time, the puppeteer (physical world) animates puppets (us) in an indeterministic way. Now it animates us either randomly or with probabilistic laws (of physics). If this is the case, the audience of the puppet show cannot, in principle, be sure of the puppet's next action by looking at the previous series of actions. Does it make any difference? Does being unexpected make us freer? Could these indeterministic puppets be free from the puppeteer? My answer to these questions are "no;" because in an indeterministic physicalistic world, even if the laws of physics are probabilistic,

there is no possibility of our doing something other than what the physical world imposes on us.

Probably, this claim needs a clarification: An ambivalent driver's decision whether or not she will stop when the red traffic light is on can be dependent on a particular electron's state in her nervous system. Let us say that she will decide to stop if the electron in her nervous system is in *spin up state*, and she will pass the red light if the electron is in the *spin down state*. To simplify, let us assume that since there is no electromagnetic effect on this electron, it can be in spin up state or spin down state with probabilities of 1/2 each. Now, whether or not this electron will be in spin up state may not be determined; however, it is strictly determined that she decides to stop, if the electron is in spin up state. In this case, can it be said that the puppets, whose perfect strings are in the hands of a puppeteer, have free will, just because the puppeteer decides how to animate the puppet with probabilistic rules? Can a puppet be free to choose which action to perform? Indeterminacy imposed by the puppeteer does not make the puppets free, since probability involved in this indeterminacy is not in the control of the puppets. In the same way, indeterminacy about the physical does not make us free, since probability involved in the indeterminacy is not in our control.

If the analogy between puppets in the control of a puppeteer and persons in a physicalistic world is a proper one, it seems inevitable to conclude living in a physicalistic world does not (or would not) allow us to be free, independently of the problem of determinism. In other words, if no puppets can be considered to be free, then no decision or action in a physicalistic world can be considered to be free. As I stated previously, with a different conception of (*weak*) free will, I could also reach the opposite conclusion: If we are living in a physicalistic world, no matter if it is deterministic or indeterministic, we *have* weak free will, as long as the correlation between how we want to act and how we act is preserved at least to a degree. What is important here is that the problem of free will is not

actually dependent on the problem of determinism / indeterminism. The point in the analogy is the puppets' dependency on something external, the puppeteer; in other words, it is about the supervenience of puppets' properties upon the puppeteer (something external) as our properties supervene upon the physical (something external) properties in a physicalistic world.

## 2.3. Is the Analogy Proper?

The argument presented in the last section rests on the analogy between *puppets / puppeteer* and *persons / physical world*. If this analogy is not a proper one, last section may simply be an "intuition pump" that does not rest on a reliable ground. The questions of this section are, "Are we really analogous to the puppets, in the sense required for the soundness of the argument?", and "Is the physical world really analogous to the puppeteer (in the sense required for the soundness of the argument)?". Obviously, my answers to these questions are "yes," although there is an important asymmetry between the two relata of the analogy: Physical entities constitute us, whereas the puppeteer does not constitute the puppets. That is, there is a widely shared intuition that we need our body to exist, when puppets may exist without the puppeteer. For identity theorists the problem is even more crucial. For them, we are the same thing as our physical constituents, when the puppets are indeed not the same thing as the puppeteers. Moreover, the relationship between the puppets and the puppeteer is causal, when the relationship between us and physical entities is the relation of supervenience.

So I have to accept that the analogy between *puppets / puppeteer* and *persons / physical world* is far from being a perfect one, not surprisingly. "Not surprisingly," because not being perfect is the *nature* of analogies, even if it is a

disadvantage. However, I believe that these imperfections are not significant for the subject under discussion; even if there are differences between puppet-world-relations and physicalistic-world-relations, in both of the worlds the supervenience relation holds: Just as puppets supervene on the puppeteer, everything in the physicalistic world supervenes on the physical entities. Just as given that two physical worlds are qualitatively identical, persons in these worlds have to be qualitatively identical with the assumption that physicalism is true; given that two puppeteers are qualitatively identical, the puppets have to be qualitatively identical.

My sympathy towards the puppet analogy arises from our pessimistic intuitions about the possibility of a puppet's having free will. A more accurate analogy could be constructed between *robots / physical world* and *persons / physical world*; and the argument suggested in the last chapter would work for this analogy too. In the next section, I discuss the robot analogy and show that the arguments for *NDP* smoothly work for this analogy. Just like the illustration about a driver deciding whether or not to stop at a red light, the robot illustration is not vulnerable to the objections based on the properness of the analogy.

## 2.4. I, Robot

It can be claimed that to assume that physicalism is true, does not make us complex puppets. Maybe the assumption that physicalism is true makes us complex robots instead. Complex robot analogy is free from some of the problems puppet analogy suffers from: First of all, robots' parts constitute the robot, just like our body constitutes us. For an identity theorist, robots are the same thing as compounds of their parts, just like we are the same thing as our physical

constituents. And the relationship between the robots' parts and the robots is based on supervenience, just like the relationship between us and our physical parts. So robot analogy seems me to be less vulnerable to criticisms based on the improperness of the analogy.

As it is stated in the last section, and for the reasons stated in the last section, here it will be tried to be shown that arguments presented in *I, Puppet* works also for the robot analogy.

If physicalism is true, we can be considered to be complex robots with feelings, wills, ideas, and so on.[10] Since we achieve our goals harmonious with our desires, we, robots, constitute a belief that we have free will.

First, let us suppose that our (robots') parts and the physical world causally connected to our parts behaves deterministically; they determine events related to us with respect to a set deterministic laws of physics, such that for each action (behavior, feeling, thought, etc.) of the robots (us) at $t_i$, the next action to be performed at $t_{i+1}$ is strictly determined. Now, no matter how complex and *seemingly* free these robots are, isn't it counter-intuitive to claim that these robots (us) have free will? Every response of robots is dependent on physical laws and entities which are external to the robots.

And the other side of the thought experiment: Robots are *living* in an indeterministic world. In an indeterministic world, robots would behave more *adventurously*. At least, probably, they would seem to us be so. Does their behaving in accordance with probabilistic rules governing the physical world, make them free? Could these robots behave in a way other than the physical world imposes?

---

[10] In this section I will occasionally, reuse some expressions in the section "I, Puppet" by only replacing "puppet" with "robot" and "puppeteer" with "physical world." This is meant to support that the puppet analogy is as proper as the robot analogy for the purpose of the argument. If the reader agrees with me in that the puppet analogy was not misleading, it can be proper to skip this section.

## 2.5. Argument of No-Dependence-Position

It can be claimed that if we have only analogies we have nothing.

(1) If physicalism is true, then everything is determined by laws of physics, probability involved in them and initial state of the world. (Premise)

(2) We have no control over laws of physics, probability involved in them and initial state of the world. (Premise)

(3) If physicalism is true, our actions are determined by laws of physics, probability involved in them and initial state of the world. (Universal instantiation on 1)

(4) If we are free for an action, then we have control over this actions. (Premise)

(5) If physicalism is true, then (no matter whether determinism is true or false) our actions are not free. (From 2, 3 and 4)

(6) If physicalism is true, the problem of free will has no dependency on the problem of determinism / indeterminism. (From 5)

Proposition (1) and (2) are scientifically well supported premises. (3) obviously follows from (1) by universal instantiation. (5) and (6) logically follows from first 4 steps.

However (4) is questionable, since there are numerous criteria of free will, most of which refer to the concept of control. Because there is no agreement on the criteria for a will to qualify as free, *NDP* must face with various mainstream

conceptions of free will. Could valid arguments of *NDP* be constructed for different reasonable criteria of free will? The second chapter of the thesis is devoted to test the argument of *NDP* with mainstream conceptions of free will suggested by the compatibilist and incompatibilist schools.

# CHAPTER III

# ON THE VARIOUS CRITERIA FOR FREEDOM OF THE WILL

## 3.1. Principle of Alternative Possibilities

### 3.1.1. Two Kinds of Principle of Alternative Possibilities

Most philosophers who work on the problem of free will, base their concepts of *free will* and *responsibility* on each other. This is intuitive: If we are not free in our actions, we are not responsible for them. Most incompatibilists believe that someone can be considered to be responsible for an action, only if she could have avoided performing this action (Van Inwagen, 1978; Ginet, 1996; Kane, 1989, 1996, 1999). That is, someone can be considered to be responsible for doing something, only if she could have done something else (otherwise). Likewise, thinkers favoring this principle believe that an action performed is a mark of a free will, only if the performer might have avoided performing this action. That is to

say, a thief's action of stealing a car is free and he is blameworthy, only if he could have chosen not to steal the car. Or a hero's saving the world is free and she deserves praise only if she could have avoided her saving the world. This kind of principles in general are the *principles of alternative possibilities* (*PAP*).

Most thinkers excluding Frankfurt and Dennett agree that freedom requires *being able to do otherwise*; however, it is mostly not clear what is meant by "being able to do otherwise." There are different forms of *PAP* and through the rest of the chapter, some of these *PAP*s are examined. To begin with, a thinker may mean (1) "being able to do otherwise, *given that the past is the same*" (*PAP concerning determinacy*), or (2) "being able to do otherwise *in exactly the same (current) circumstances*" (*PAP concerning supervenience*), by *PAP*. Even if these two versions of *PAP* seem very similar, they should not be used interchangeably, since they have dramatically different implications.

In the case of the ambivalent driver, given that past is exactly the same, the driver could have done otherwise, since the state of the electron in her nervous system had not been determined before the time she decided. However, given that circumstances are exactly the same, which involves the final state of the electron in the driver's nervous system, she could *not* have done otherwise. I believe that the *PAP* concerning determinacy, which is related to determinacy, is not a necessary or sufficient condition for free will and responsibility. On the other hand, *PAP* concerning supervenience, which is related to relation of supervenience, I claim, is a necessary condition for strong free will and genuine responsibility.

Therefore, with respect to the conceptualization formulated above, given that our physicalistic world is indeterministic, we *may not* do otherwise in exactly the same circumstances; while we *may* do otherwise, given that the past is the same. It seems more appropriate to me to use "*may* do" instead of "*can* do," because we do not have any control over indeterministic physical events, which inevitably assign

our choices and actions. And since it does not give us control over our actions, I defend the claim that satisfying this *PAP*, which is related to determinacy, does not give us freedom required for responsibility.

I defend the view that the *PAP* concerning determinacy is irrelevant to the problem of freedom; however, most incompatibilists use *PAP* concerning determinacy to decide if we have free will in deterministic or indeterministic worlds. In his *An Essay on Free Will*, Van Inwagen puts forward three arguments, in order to show that free will and determinism are incompatible. His third argument, *the consequence argument*, is based on the *PAP* concerning determinacy, resting upon the sameness of the past.

First, he defines an operator "N" that attaches to sentences such that "N *p*" means "*p* and no one has, or ever had, any choice about whether *p*." Then he claims that two intuitively attractive rules namely Rule ($\alpha$) and Rule ($\beta$) are true (Van Inwagen, 1983, pp 93-94).

Rule ($\alpha$) $\Box p \vdash Np$

Rule ($\beta$) $N(p \supset q), Np \vdash Nq$.

His argument, which is based on these principles and premises takes 7 steps:

$P_0$: State of the world at a distant past time point.

L: Conjunction of all the laws of nature.

P: Any true proposition.

If determinism is true, then it follows that

(1) $\Box (P_0 \,\&\, L \supset P)$

is true. From (1) we may deduce

(2) $\Box (P_0 \supset (L \supset P))$

by elementary modal and sentential logic. Applying rule ($\alpha$) to (2), we have:

(3) N ($P_0 \supset$ (L $\supset$ P)).

We now introduce a premiss:

(4) N $P_0$.

From (3) and (4) we have by Rule ($\beta$):

(5) N (L $\supset$ P).

We introduce a second premiss:

(6) N L.

Then, from (5) and (6) by ($\beta$):

(7) N P (Van Inwagen, 1983, pp 94-95).

I am suspicious about the truth of the rule ($\alpha$). According to rule ($\alpha$), if *p* is necessarily true, then no one has or ever had any choice whether *p*. On the other hand, what if someone necessarily chooses something, i.e. someone *necessarily* chooses *p* to be true. Such cases are logically possible: God necessarily chooses what is good or if determinism true, then everybody "*chooses*" what physical laws necessitates.

It may be argued that if something is determined, we should not call it "choice;" however, I do not agree because of the reasons I present in the next section. I believe that Van Inwagen's intuition in deriving this argument is based on his commitment to *PAP* concerning determinacy and this *PAP* is not an acceptable criteria for free will in any sense. I propose that a less problematic version of the consequence argument could be set up in the following way

(1) If determinism is true, for every complete set of true propositions belonging to a time point, there is only one possible complete set of true propositions for each later time point. (From the definition of determinism)

(2) There is only one complete true set of propositions for every past time point. (Premise)

(3) There was a time at which there were no living beings. (Scientifically well supported premise)

(4) If determinism is true, there is only one possible complete set of true propositions for each time point at which there are living beings. (From 1, 2 and 3)

(5) If determinism is true, we cannot do otherwise (with respect to *PAP* concerning determinacy). (From 4)

The arguments in this form show only that *PAP* concerning determinacy can be satisfied, only if indeterminism is true. They show *only* that given that the past is the same, we *may* do otherwise, only if the world is indeterministic. And it is not only us; the same is true for even lifeless objects: A die might have come up some value other than it did, only in an indeterministic world. However, this does not lead us to believe that a die is free in an indeterministic world. The reason why we believe a die satisfying this *PAP* is not free is simply that a die has no control over which value it comes up. Do we have any control over the behaviors of physical entities, which determines our choices in a deterministic physicalistic world? No, we do not. On the contrary, we are governed by physical laws. And even if our world is indeterministic, as long as physicalism is true, we do not have any control over physical world, which governs us. To sum up, I think, any argument based on *PAP* concerning determinacy does not prove any proposition favoring incompatibilism.

### 3.1.2. Ms. Determined

*PAP* concerning determinacy has some other problems to be dealt with: Let me introduce you Ms. Determined. She is a poker star who is very good at math and whose memory is perfect for playing poker. By considering all the relevant information in a game, she makes decisions with best pay off.

When you are watching her in a tournament, an incompatibilist comes by and tells you that Ms. Determined does not have free will.

You ask: “Why do you think so?”

“I am watching her for a long time,” the incompatibilist replies. “Whenever she has a pair of aces, she is trying to raise the bet, hiding strength of her hand.”

“Isn’t it reasonable? Shouldn’t you raise the bet when you have a pair of aces?”

The incompatibilist is tired of not being understood: “Of course, you should! But you don’t understand! Since she is so good at mathematics and psychology, since she has such a good memory, and since she has such an enormous desire to play optimally, there is no way she would not try to raise the bet when she has a pair of aces.”

“So?”

“So she could not have done otherwise and this means she does not enjoy freedom, at least not when she is playing poker.”

“And me, as a rookie poker player,” you say, “make inconsistent decisions. I may not raise the bet when I should. This *means* (!) that I am *freer* than she is on the poker table. Am I right?”

This kind of thought experiment makes me believe that indeterminacy is not something a genuine choice *must* conceptually involve, contrary to the incompatibilist position. It may be claimed that I caricaturized the defenders of the *PAP* concerning determinacy. It is also possible that a libertarian incompatibilist would interpret the case of Ms. Determined in the following way: "Maybe there is no way for Ms. Determined not to try raising the bet when she has a pair of aces given her passion and ability to win; but if she wanted not to raise–if she wanted to lose– Ms. Determined could prevent herself from raising." However, such a defense should not satisfy an incompatibilist, since this conception of free will (a version of weak free will) belongs to the compatibilist view and it implies that compatibilism is true. Ms. Determined could have prevented herself from raising, if she had wanted to do so; not only in an indeterministic world, but also in a deterministic world. This would probably upset incompatibilists, wouldn't it?

*PAP* concerning determinacy makes rookie players freer than perfect ones. In the same way, according to *PAP* concerning determinacy, *perfectly* good willed angels are not free and praiseworthy unlike *fairly* good persons, since only the latter ones could do bad things.[11] Note that *PAP* concerning supervenience does not suffer from these problems, because in a physicalistic world neither rookie nor perfect poker players could do otherwise in exactly the same circumstances. Neither of them qualifies to be genuinely free or responsible. This is evident: After all, in a physicalist world, given that complete set of true propositions about physical is the same, nothing could be different than it is.

Indeed, I agree with incompatibilists in that if causal determinism is true, no one has, or ever had, any indeterminacy in their choices. However, I do not believe that this proves anything they intended. Genuine choices necessitate a *plurality*;

[11] Another one of this kind of thought experiments is examined in the last chapter, which is about responsibility.

but this is not plurality of possible actions, it is plurality of options. In other words, a genuine choice necessitates more than one options; however, it is not necessary that a genuine choice is indeterminate. A *perfect* chess player does not behave *indeterministically*, but since she has *options* and we assume that she has control over her actions we take her to be free.

### 3.1.3. *NDP* and *PAP* Concerning Supervenience

And the main problem: Is *NDP* coherent with the concept of free will based on the *PAP* concerning supervenience? First, it is obvious that *PAP* concerning supervenience cannot be met in a deterministic, physicalistic world. Is it, on the other hand, possible that in a physicalistic and indeterministic world, we can or could do otherwise in exactly the same circumstances? Let me return to the puppet analogy: In a physicalistic and indeterministic world, the puppeteer decides what we will do by following probabilistic rules. But still, for each state of the puppeteer, which is assigned in an indeterministic way, the puppets have only one possible state. This should conceptually follow, since we assume that physicalism is true and everything supervenes on the physical world.

I have already tried to clarify this idea with the thought experiment in which an electron belonging to an ambivalent driver's nervous system determines whether or not she will stop when the red traffic light goes on. We can revisit the idea using the puppet analogy. Let us suppose that the puppeteer chooses its next movement by flipping an onticly probabilistic coin. If the coin comes up heads, it raises its hands; and if the coin comes up tails, it does not raise its hands. And when the puppeteer raises its hands, let us say the puppet jumps. In this case, the

puppet's role seems to be analogous with our role in an indeterministic physicalistic world. Now would it not be absurd to say that the puppet has free will in this case? It would, I believe. And a working version of the principle of alternative possibilities, which I call "*PAP* concerning supervenience", implies the same, since the puppets cannot do otherwise, in exactly the same circumstances. It may not be determined whether or not the puppeteer raises its hands before the time it does so, but it is strictly determined that the puppet jumps when the puppeteer raises its hands.

That is, not only in a deterministic world, but also in an indeterministic and physicalistic world, we cannot do otherwise in exactly the same conditions. Or in other words, *PAP* concerning supervenience implies that in a physicalistic world we cannot have free will, and this is true for both deterministic and indeterministic physicalistic worlds. A version of the argument of *NDP* works for *PAP* concerning supervenience.

I suspect that some incompatibilists would not accept this *supervenience* interpretation of *PAP*. They would say that even if there is only one state of will corresponding to each possible condition of the physical world, this outcome is not determined before the time it emerges. They may accuse me of misinterpretation, and insist on *PAP* concerning determinacy. In that case, my defense could be simply that their interpretation of the requirement of *PAP* implies that puppets *do* have free will in an indeterministic world.

However, this defense works only if libertarians think that *PAP* concerning determinacy is necessary and *sufficient* condition for free will. On the other hand, most libertarians thinks that *PAP* is a necessary, but not a sufficient condition for free will (Kane, 1996, p. 33). They suggest some additional criteria for a will to qualify as free. I deal with these additional conditions throughout the rest of the chapter.

### 3.1.4 Arguments Against and for Indeterminacy PAP

*PAP* concerning determinacy has been the dominant *PAP* (I think misleadingly) in the contemporary incompatibilism. That is why throughout this section, I will refer to *PAP* concerning determinacy as "*PAP*." In spite of problems I articulated in the last section, *PAP* concerning determinacy is perceived as one of the strongest castles for incompatibilist accounts. However, this castle of incompatibilism, which is strengthened by the walls of intuition, has not been immune to criticisms at all. Harry G. Frankfurt complains about the tendency to take *PAP* (*PAP* concerning determinacy) as if it is an *a priori* truth and he claims that this principle is mistaken. He suggests a counter-example to support his opinion:

> Suppose someone–Black, let us say–wants $Jones_4$ 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 $Jones_4$ is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that $Jones_4$ is going to decide to do something *other* than what he wants him to do. If it does become clear that $Jones_4$ is going to decide to do something else, Black takes effective steps to ensure that $Jones_4$ decides to do, and that he does do, what he wants him to do. Whatever $Jones_4$'s initial preferences and inclinations, then, Black will have his way (Frankfurt, 1969, p. 835).

Frankfurt assumes that Black has enough power to make Jones perform the action (*B*) that Black wants him to do: Threatening, manipulating, using a neurological device to determine Jones's decisions… Anything imaginable. Then he assumes that Jones decides to perform *B* by his own will. In this case, it seems obvious that Jones would be responsible for his action, since Black's conspiracy did not have a

role in determining Jones's action *B*. He would have done *B*, even if Black had not even existed. On the other hand, according to *PAP*, although Jones does what he wants to do without any causal effect from Black's conspiracy, since Jones could not have done otherwise, he is not responsible and his action is not free. The thought experiment seems to threaten *PAP*.

However, not everyone grants consistency of Frankfurt's counter-example: David Widerker (2000)[12], defines a sign which he calls "*S1*".[13] Observing if *S1* arises by a time earlier than Jones's potential action under consideration, Black understands if Jones will perform the action without his interruption and decides whether or not to show his hands. Indeed, Widerker is right in that such a sign is necessary to make Frankfurt's thought experiment complete. Then, Widerker argues that if *S1* is such a perfect predictor of Jones's decision, then Jones's decision should be predetermined. Contrary to Frankfurt's presupposition, Jones is not responsible for what he did, since he could not have done otherwise. And if *S1* is not a perfect predictor, then it is not true that Jones could not have done otherwise and this is why Jones is responsible for what he has done. In other words, by assuming that a free action can be perfectly predicted in his thought experiment, Frankfurt preassumed that *PAP* is false.

Accepting that Frankfurt's thought experiment was not enough to disprove *PAP*, Alfred R. Mele and David Robb attack *PAP*, by modifying the original scenario. Before beginning their thought experiment, they guarantee that their world is not deterministic and there is no sign like *S1* in their thought experiment. However, they also remind us that the world's being indeterministic does not mean that there is no causally determined event in this world.

---

[12] He puts his remark forward first in 1995.

[13] Actually, Widerker defines S1 for his own version of a Frankfurt-style counter-example. In order not cause unnecessary confusions, I will apply his argument to Frankfurt's original counter-example.

This time, Black is trying to make Bob steal a car and he uses some kind of conditional process *P* to guarantee this. *P* assures that Bob decides to steal the car at a certain time $t_2$, in the following way: If Bob does not decide to steal the car until $t_2$, then *P* will cause Bob to decide to steal the car. Consequently, like in all the similar counter-example scenarios, Bob decides to steal the car by his own indeterministic decision before $t_2$ and *P* plays no role in this decision (Mele and Robb, 1998).

This new scenario, does not necessitate a sign, and it seems that it disproves *PAP*. However, Widerker does not agree. He believes that this new scenario is inconsistent too. The first question Widerker asks is this:

> Given the presence of the deterministic process P, how is it that P does not cause Bob's decision to steal the car, but rather it is Bob who makes that decision on his own (Widerker, 2000, p. 183)?

He claims that this is possible only if when two processes are *about* to cause Bob's decision, *P* is preempted by Bob's decision process. Let us denote the time when Bob's decision would preempt *P*, by "$t_3$." Obviously, $t_3$ is earlier than $t_2$. And Widerker asks his critical question: How can we know that Bob would not change his decision within the time interval [$t_3$, $t_2$]?

If the preemption does not take place at the last time point before the decision takes place, Widerker's criticism seems to be acceptable and the existence of such a *very last* time point before a time point is controversial.

Eleonore Stump (1996) tries to solve this problem without making the preemption take place at the last time point before the decision takes place. Instead, she gives a neuroscope to Black[14] which lets him observe and manipulate neural firings in

[14] Again for the simplicity's sake, I will apply Stump's counter-example to the original thought experiment articulated by Frankfurt.

Jones's nervous system. Black observes by his neuroscope that every time Jones decides to do *B*, a sequence of neurons, which always contains *a*, *b*, *c* near its beginning, fires; and every time Jones decides *not to do B*, another sequence of neurons, which always contains *x*, *y*, *z* near its beginning, fires. Then Black adjusts his neuroscope such that whenever the neuroscope detects the firing of *x*, *y*, *z*, it inhibits the neural sequence and activates a coercive neurological mechanism resulting in the firings of *a*, *b*, *c* which would cause Jones to decide to do *B*. Otherwise, the neuroscope does not intervene Jones's nervous system and lets Jones do *B* by himself. Then Stump assumes that Jones performs *B* without any neural intervention by the neuroscope.

The dodge in Stump's scenario is that there is a time gap between the beginning of the natural mechanism causing Jones's decision and his decision itself. In this scenario, Black has time to stop Jones from deciding to do *B*. However, Stump sacrifices the indeterminacy lasting until the decision. Widerker objects to Stump's scenario, by arguing that in this story, will is causally determined by the sequence of firings of *a*, *b*, *c* and Jones could not have done otherwise in the sense *PAP* requires.

Considering his objections to Mele, Robb and Stump, it maybe that for Widerker, *PAP* is not just that a person is responsible and free only if she could do otherwise, but more strongly, a person is responsible and free only if she could do otherwise *at the very time point she decides*. However, indeterminism alone does not guarantee that the indeterminacy involved in an event lasts until the very last time point it takes place. Widerker's argument implies that satisfaction of *PAP* requires more than a simple indeterminacy. Why do indeterminacies in our choices have to last until the last moment the decision takes place? Widerker may answer this question simply by saying "In order to make these choices free." Even if I do not believe that classical *PAP* concerning determinacy is a necessary or sufficient condition for freedom or responsibility, especially for the reasons I

present in the rest of the section, I believe that Widerker indeed has reasons to criticize Frankfurt's work.

Another very strong argument Widerker puts forward against Frankfurt's counter-examples is based on his summary of Frankfurt's argument defending that *PAP* is false:

> 1. The fact that in an IRR-situation[15] an agent could not have avoided performing a certain act plays no role in the explanation of why the agent performed the act.
>
> 2. If a fact is irrelevant to the explanation of why the agent performed a certain act, then this fact has no bearing on the agent's moral responsibility for the act.
>
> 3. Hence, in the absence of any other excusing factors, M [The agent is responsible for his/her act.] is the case.
>
> 4. Therefore, PAP is false (Widerker, 2000, p. 189).

Widerker and apparently Kane believes that (2) is false. And Widerker exposes a counter-example to (2) from an unpublished manuscript written by Kane:

> A hurricane is approaching a coastal city. Smith, a resident of the city, mistakenly believes that by chanting an elaborate incantation [he][16] could cause the hurricane to veer off to sea and never strike the coast. Smith deliberates whether or not to mutter the incantation and (as [he] believes) thereby to prevent the impending disaster. But, [he] decides not to do so (as cited in Widerker, 2000, p. 189).

---

[15] Cases in which although a performer has to perform action A (the performer could not have done otherwise) because of the set of circumstances C, C has no role in determining the performer's doing A (or simply cases Frankfurt counter-examples describe).

[16] Modifications belong to Widerker.

Smith's choice not to save the coast is irrelevant to the fact that the muttering would not stop the hurricane. He would not save the coast even if he could. However, because of the fact that the muttering would not stop the hurricane, he is not responsible for the hurricane. This case can be a proper counter-example for (2). In other words, maybe this counter-example shows that Frankfurt could not successfully find why *PAP* was false. However, he can still be right in that *PAP* is false, unless Frankfurt-style counter-examples are overcome.

Up to now, Frankfurt's critique, new versions of Frankfurt's scenario and Widerker's defense for *PAP* are presented. To be honest, I find all of these discussions too confusing. There is a simpler and convincing answer to Frankfurt's 1969, I believe, and this response was presented in 1978. In my opinion, Peter Van Inwagen had overcome the difficulties, long before these discussions Widerker primarily referring to were suggested.

## 3.2. Van Inwagen's PAP Immune to Frankfurt's Counter-Examples

As an incompatibilist, Peter Van Inwagen asserts that *principles of alternative possibilities* which are immune to Frankfurt's counter-examples can be generated and he suggests a set of rules composed of three versions of *PAP* corresponding to three different types of entities:

Entity: Acts

*PPA* (the Principle of Possible Action): "A person is morally responsible for failing to perform a given act only if he could have performed that act" (Van Inwagen, 1978, p. 201).

Entity: Particular events

*PPP1* (the Principle of Possible Prevention): "A person is morally responsible for a certain event (particular) only if he could have prevented it" (Van Inwagen, 1978, p. 206).

Entity: Universals

*PPP2* (the Principle of Possible Prevention′): "A person is morally responsible for a certain state of affairs only if (the state of affairs obtains and) he could have prevented it from obtaining" (Van Inwagen, 1978, p. 210).

Van Inwagen's reply to Frankfurt is convincing: Each individual event is only identical to itself. Therefore, Jones's performing the action that Black wants *without any causal effect* of Black (*E*) is not the same particular event as Jones's performing the action that Black wants *under the causal effect* of Black (*E′*). Or the action corresponding to *E*, is not the same particular action as the one corresponding to *E′*. So Jones could have prevented occurrence of *E*. This means he was responsible for *E*'s occurrence (by *PPP1*). He was also responsible for trying to perform the action that Black wants him to do, since he could have prevented himself from trying it (by *PPA*). But he was not responsible for the state of affairs caused by this action (in the universal sense), since he could not have prevented this state of affairs to be realized due to Black's conspiracy. The distinctive applications of *principle of alternative possibilities* for different kind of entities seem to solve the problem Frankfurt articulated.

On the other hand, Van Inwagen does not say explicitly that Frankfurt's counter-examples failed to disprove *PAP*. This is why almost two decades later, Carl Ginet (1996, pp. 406-407) wrote his paper "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing." Ginet concentrates on the Frankfurt's counter-examples intensively. First he makes some temporal specifications in Jones's case: He defines $t_1$ as the time when Jones would do the action Black wants (*B*) without any causal effect of Black's conspiracy. Black decides whether or not Jones would have done *B*, at

time $t_2$, which should obviously be later than $t_1$. And if Jones has not performed the action desired by Black at $t_1$, the conspiracy causes him to perform this action at $t_3$. However, like in all of the cases, in the end, Jones does *B* by his own choice. Now, Kane claims that Jones is responsible for his doing *B*, by $t_1$; however, he is not responsible for his performing *B* by $t_3$ (Ginet, 1996, pp. 406-407).

These new versions of *PAP* concerning determinacy are improved versions of the old one, but arguments of *NDP* can still be applied to them: Van Inwagen and Ginet claim that Frankfurt made a mistake by not specifying events as much as he should; and I agree with them. However, a similar specification should be applied to the other side of the coin: circumstances. Just as Jones's performing *B* because of the conspiracy of Black is not the same event as Jones's performing *B* by his own choice, the circumstance which leads Jones to perform *B* because of the conspiracy of Black is not the same circumstance as the circumstance which leads Jones to perform *B* by his own choice. Therefore, if Van Inwagen and Ginet want to show that somebody could do otherwise in exactly the same circumstances, they have to show that this person can choose two different things in one single particular circumstance with one single state of her nervous system, or more precisely, with one single state of the physical. Note that modifying *PAP* concerning determinacy in this way transforms it into *PAP* concerning supervenience.

As I claimed before, *PAP* concerning supervenience is impossible to satisfy in a physicalistic world. No matter if the puppeteer is deterministic or indeterministic, no matter if we are talking about particular or universal events, no matter how complex the puppets are, the puppets could not have done otherwise in exactly the same circumstances; and their being able to do otherwise given that all the facts about the past are the same does not give them strong freedom required for genuine responsibility. What puppets desire is strictly determined by the strings in the hands of a puppeteer. They cannot be free, they cannot be responsible. Our

desires are in the hands of the physical; we cannot be free, we cannot be responsible.

## 3.3. Origination

Ted Honderich claims that we have two important concepts of free will and he contends that for this reason both incompatibilists and compatibilists are wrong. To have free will in the weaker sense, it is enough to have voluntary actions. And to have the free will in the stronger sense, our actions have to be both voluntary and originated (*causa sui*) (Honderich, 1996, p. 856).

Honderich argues like most thinkers that voluntariness is compatible with determinism. He also states like most incompatibilists that origination is not compatible with determinism. I agree with him up to this point. However, he does not maintain *NDP*. He believes that the problem of strong freedom has dependencies on the problem of determinism (Honderich, 1996, p. 856).

In my opinion, neither voluntariness (indeterminacy free will) nor origination (strong free will) has a significant relationship with the problem of determinism / indeterminism. Voluntariness requires only a regularity between our desires and fulfillment of these desires. If we regularly satisfy some of our desires, then we are free in the weaker sense. It is not controversial that we can do what we want. We may act voluntarily in both deterministic and indeterministic worlds. Only in the possible worlds in which our desires have no positive correlation with our actions, could we not act voluntarily.

On the other hand, origination, as I see it, is impossible to satisfy in any physicalistic world. In a physicalistic world, *will*, which is assumed to be non-

physical, cannot be originated; *will* is just a *shadow* of physical. Therefore, no matter whether our world is deterministic or indeterministic, we do not have Honderich's stronger free will, which requires origination. In other words, the problem of free will has no dependency on the problem of determinism / indeterminism.

## 3.4. Ultimate Responsibility or Self-Origination

Another criterion of free will is shaped by the concept of *ultimate responsibility* (*UR*), by Kane (1989, 1996, 1999). Kane defines three conditions for free action and he believes that these conditions can only be satisfied in an indeterministic world:

> (1) (The Production Condition) is the intentional termination of an effort of will that is the agent's effort of will, and
>
> (2) (The Rationality Condition) the agent ($r_1$) has reasons for doing so (whichever occurs), ($r_2$) does it *for* those reasons, ($r_3$) does not choose (for those reasons) compulsively, and ($r_4$) believes at the time of choice that the reasons for which it is made are in some sense the weightier reasons, more worth acting upon than their alternatives, and
>
> (3) (The Ultimacy Condition) given the facts of the situation, no other explanation (other than the conjunction of (1) and (2)) for the agent's choosing A or choosing otherwise (whichever occurs) is possible, unless that explanation can in turn be explained by the conjunction of (1) and (2) itself. (I.e. the explanation provided by (1) and (2) is "ultimate" or "final.") In particular, any explanation of the agent's making the effort of will in (1) and of the agent's having the character and the reasons or motives for choosing in (2) will not also explain the choice, even though (1) and (2) will explain the choice (Kane, 1989, p. 232).

I do not have much to say about the production condition. Just a clarification: What Kane means by termination of an effort's being "*intentional*" is simply that its being done *knowingly* and *purposefully* by the performer. Kane believes that this condition is compatible both with determinism and indeterminism. So do I, as I understand the condition. The production condition intuitively seems to be satisfied, and it is not one of the problems between compatibilists and incompatibilists.

Among sub-conditions of the rationality condition I want to concentrate on $r_3$ which is the only one which has difficulties in my opinion. The difficulty is simply about the ambiguity in the meaning of "compulsion" and Kane intends to solve it by referring to Frankfurt's analysis of compulsion based on higher-order desires: He uses Frankfurt's (1971, p.9) example of addiction: Somebody having a first-order desire to use a drug qualifies as a person, only if she also has a second-order desire directed towards or away from the drug. Now, Frankfurt and Kane claim that if a person follows her first-order desire and uses the drug despite her conflicting second-order desire (or I would say "*highest*-order desire"), she chooses to use the drug *compulsively*. And if she follows her second-order desire and uses or does not use the drug, she does it *freely*. Even if Frankfurt's analysis based on higher-order desires has problems that I will concentrate on later in this dissertation, I will ignore them for now. In any case, fulfillment of the rationality condition is not dependent on the problem of determinism.

And the ultimacy condition, the most important one of all three for the problem of free will and determinism: The explanation provided by (1) and (2) should be ultimate. Kane believes that this is impossible in a deterministic world. He can be right in that free will's being involved in the ultimate explanation for an event is impossible in a physicalistic world which is deterministic. But if we are living in a physicalistic world, no matter whether it is deterministic or not, I believe that free will's being involved in the ultimate explanation for an event is impossible. It

seems to me that, the ultimate explanations for a physicalist world should be a physical explanation. On the other hand, if we are living in a dualistic or idealistic worlds, in which the will supervenes on nothing besides itself, the ultimacy condition can be satisfied without any dependency on the problem of determinism / indeterminism. Besides, I believe that the concept of *explanation* is too anthropocentric, too subjective to be the best candidate to base free will on. No need to add that, like many of the concepts in philosophy, *explanation* is not free of problems: There is a huge literature on the conflicting accounts for explanation.

Galen Strawson attacks the same condition with a different argumentation: In his view, *UR* can only be satisfied if we have *will*s which are *self-originated*:

> … *what is the argument that UR requires some kind of radical self-creation?*
>
> It has two steps.
>
> (1) When you act, you do what you do, in the situation in which you find yourself, because of the way you are.
>
> (2) So if you are to be UR for what you do, you must somehow or other be UR for the way you are–at least in certain crucial mental respects (Strawson, 2000, p.151).

The first step is intuitive and second one naturally follows from it. Therefore, if we are *UR* for an action, we have to be *UR* for our being the way we are. However, Strawson claims (rightly, in my opinion) that being *UR* for the way we are is impossible. His argument to prove his point is the following:

> *But why, exactly, can't you be UR for the way you are?*
>
> Well, it seems clear that
>
> (3) 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.
>
> And the problem is then this. Suppose

> (4) 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.
>
> For this to be true
>
> (5) 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. (If you didn't already have a mental nature then you didn't have any intentions or preferences and can't be UR for the way you now are, even if you have changed.)
>
> But then
>
> (6) 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.
>
> So
>
> (7) You must have intentionally brought it about that you had Y.
>
> But then
>
> (8) 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(Strawson, 2000, p.151).

Kane's *UR* requires (*ultimate*) *self-origination* and it seems that Strawson shows that it is impossible to have a self-originated will. He also notes that his argument is independent of any view whether or not determinism is true. So his argument supports *NDP*.

As I have already stated, I believe that in order to satisfy the condition of self-origination, our actions must not be determined by something out of our control. Strawson points out that our character which has a role in our actions is *out of our control*. And this implies that we cannot be *UR* for our actions. My point already presented several times is relevant but different: It seems to me that this condition can only be satisfied in a dualistic or idealistic world, satisfying certain characteristics; because only in such worlds, will's state can in principle be *in our control*. And this is contrary to my assumption that physicalism is true. In a

physicalistic world, the physical renders truth of every proposition inevitably, none of our actions can be (ultimately) self-originated, and truth value of indeterminism, which does not have any implication about physicalism, does not change this. Therefore, no matter whether our world is deterministic or indeterministic, for this criterion of free will, we do not have free will; i. e., the problem of free will has no dependency on the problem of determinism / indeterminism. Again, *NDP* faces no problem in dealing with this concept of free will, which is based on ultimate responsibility or self-origination.

## 3.5. Compatibilist PAP

Another version of the principle of alternative possibilities, which is used by compatibilists, has implications which are *apparently* conflicting with indeterminism and *PAP* concerning supervenience's implications discussed in the *Principle of Alternative Possibilities*. For these compatibilists, an action is free, if and only if this action could have been avoided by the performer, *if the performer had wanted to do so* (Hume, 1975; Moore, 1912; McIntyre, 1994). I will call this principle "*compatibilist PAP*". Compatibilist PAP is much easier to satisfy than any *PAP*, since it is not a criterion for *strong* free will. Actually, these compatibilists are dealing with *weak* free will. No matter whether we are living in a deterministic or an indeterministic world, as long as we are concerned with a volitional action, we believe that we could have done otherwise if we had wanted to do so. This statement seems to be *conceptually* true: "Volitional actions" are actions which the performer could have vetoed if she had wanted to do so. It leaves no doubt that we have (*weak*) free will. No matter whether we are living in

a deterministic, indeterministic, dualistic or monistic world, we have *some kind of* free will.

However, weak free will seems too embracing: It is so embracing that complex puppets discussed in "I, Puppet" qualify to be free. Recalling and detailing the nature of the puppets can be useful here: These puppets have some complex mechanisms like wills, desires, feelings, thoughts, and so on. However, of course the strings of these mechanisms are in the hands of the puppeteer. Even if the puppets want things to happen, what they want is strictly determined by the puppeteer. Let us say: When the puppeteer decides that the puppet should jump, it first makes the puppet want to jump by using the strings and then makes the puppet jump, again by using the strings. There is no causal relationship between wanting to jump and jumping. The puppeteer is the common cause of the wanting to jump and jumping itself. On the other hand, it is still true that the puppets could do otherwise if they wanted to do so; since the puppeteer always makes the puppet want an action before it makes the puppet perform this action. Compatibilist PAP, this embracing interpretation of free will, implies that these puppets have (*weak*) free will. What about the philosophical intuitions? Can we say that an action is free, even if will has no causal effect on the action? Should the philosophical problem of free will deal with this concept of (*weak*) free will? As I stated earlier, the claim that we have weak free will is not controversial.

No matter what the answer to this question is, the argument of *NDP* still persists. Again, for this concept of free will, the problem of free will is not dependent on the problem of determinism / indeterminism. We have free will either way, according to compatibilist *PAP*.

## 3.6. Frankfurt's Concepts of Free Will

### 3.6.1. Frankfurt's *PAP*

Frankfurt does not only show that the *PAP* concerning determinacy is misleading without Inwagen-style-interpretations. He also suggests a new version of *PAP*, nine years earlier than Van Inwagen's version of *PAP*: "a person is not morally responsible for what he has done if he did it only because he could not have done otherwise" (Van Inwagen, 1969, p. 838). According to this principle, a volitional action imposes responsibility on its performer, if it is really *intended* to be performed by the performer. Since moral responsibility and free will is widely accepted to be akin, I will derive the following principle based on Frankfurt's: A volitional action we performed is *free* if and only if it is not the case that we did it only because we could not have done otherwise. In other words, a volitional action we perform is *free* if and only if either (1) we desire to perform this action although we cannot veto this action or (2) we can veto this action if we desire to do so. This condition of free will does not seem very different from compatibilist *PAP* with respect to its relation with *NDP*. It seems Frankfurt is another thinker who accepts our puppets have free will, and just like compatibilist *PAP*, Frankfurt's solution does not threaten *NDP*.

### 2.6.2. Free to Will What We Want to Will

There is another, a more famous criterion of free will that Frankfurt (1971, p. 15) suggests. Frankfurt (1971, p. 14) discriminates *freedom of action*, which a variety of complex species enjoy, from *freedom of will*, which only persons enjoy. He claims that compatibilist *PAP* can be considered to be a criterion only for freedom of action, which is a different concept than the freedom of will. For him, someone has free will if and only if she is "free to will what she wants to will" (Frankfurt, 1971, p. 15). Frankfurt's concept of free will is linked to his concept of *person*, which requires *second-order desires*. He suggests that only ones who have the capacity of willing what they want to will can qualify as persons.

Frankfurt's point seems important in scientific and everyday perspectives: Suppose that there are three people finding wallets which contain some money and an ID card identifying the rightful owner of it.

Mr. Giveitup

First-order desire: Wants to keep the money.

Second-order desire: He is okay with his first-order desire.

Action: Keeps the money.

Mr. Superego

First-order desire: Wants to keep the money.

Second-order desire: He is not okay with his first-order desire.

Action: Finds the owner of the wallet and gives it back.

Mr. Kleptomaniac

First-order desire: Wants to keep the money.

Second-order desire: He does not have a sufficiently strong second-order desire for the time being, since he is kleptomaniac.

Action: Keeps the money.

In such a case, I believe, most of the people, in accordance with modern justice systems, would intuitively perceive Mr. Giveitup and Mr. Superego to be responsible for their actions and see them as praiseworthy or blameworthy; however, they would not take Mr. Kleptomaniac as responsible for and free in his action. Therefore, Frankfurt's point is important in scientific and daily perspective.

On the other hand, the *philosophical* adequacy of his conceptualization is questionable, I believe. Why are second-order (or $n^{th}$-order) wills or desires or volitions necessary to be person or to have free will? Why is it not necessary to have *third*-order (or $n+1^{st}$-order) wills or desires or volitions to be a person or to have free will? Besides, if second-order desire makes first-order desires free and if $n+1^{st}$-order desires makes $n^{th}$-order desires free, what makes the highest-order desires free? Just like Gary Watson (1975, pp. 217-219; 1987, pp. 147-149), I could not find satisfactory explanations for these questions in Frankfurt's (1971, 1987) relevant articles.

Even if I do not find this interpretation of free will intuitive, I will still test *NDP* here. At first glance, it seems that Frankfurt's interpretation implies that we cannot be free in a physicalistic world. After all, in neither deterministic nor indeterministic physical world, we (the puppets) are free to will what we want to

will because the strings of our wills are in the hand of the puppeteer. However, since we have complex structures like second-order wills, desires, volitions and so on, even though the strings of these latter are in the hands of the puppeteer (physical), we qualify to be free, no matter whether the physical behaves deterministically or indeterministically.

Therefore, in either way, for this interpretation, if our world is physicalistic, there is no relation between the problem of free will and the problem of determinism / indeterminism.

## 3.7. Self-Forming Actions

In his 1999, Kane contends that there are free actions which are predetermined by our characters and motives. Our wills maybe determined in such cases, he claims; however, they are determined by nothing external but our own past free choices which are not predetermined. He calls these earlier undetermined actions "*self-forming actions*" (*SFA*). To sum up, self-forming actions, which are *undetermined*, determines the future free actions, by forming the self. And for Kane, thanks to the indeterminacy in the earlier *SFA*s, these determined actions are free.

Indeed, he does address the question "How being undetermined makes *SFA*s free?". He introduces a principle that opponents of libertarians use against his view:

> [Luck Principle] If an action is *undetermined* at a time *t*, then its happening rather than not happening at *t* would be a matter of *chance* or *luck*, and so it could not be a *free* and *responsible* action (Kane, 1999, p. 217).

Luck Principle implies that undetermined actions including *SFA*s are not free, because it takes undetermined actions to be matters of luck. Kane claims that luck principle is not true. In an indeterministic world, events do not have to take place by chance or luck. Even if indeterminism implies that events may not be strictly determined by the preceding state of the world, it does not follow that events are not affected by the preceding state of the world. Clearly, Kane is right in that luck principle is false. Then he describes several cases in which an action is not determined but the performer is still intuitively responsible:

> Consider an assassin who is trying to kill the prime minister, but might miss because of some undetermined events in his nervous system which might lead to a jerking or wavering of this arm. If he does hit his target, can he be held responsible (Kane, 1999, p. 227)?

Not surprisingly, following his intuitions, he answers "yes." By employing thought experiments like the one quoted above, he concludes that we can be free and responsible in an indeterministic world. As I did in the thought experiment about Ms. Determined, this type of arguments can be used against libertarian incompatibilism too: Consider the assassin with a perfect shooting ability. No jerking, no wavering, no distraction… Intuitively, would this assassin be less responsible? I do not think that Kane would think that only imperfect snipers can be responsible for killing people. So this kind of defense does not discriminate incompatibilism from compatibilism. Actually, it favors *NDP*.

Kane should do something to favor libertarian school against the others. He has to point out what makes indeterminism more free-will-friendly than determinism and he tries to do it:

> There is a tension and uncertainty in our minds at such times of inner conflict [while a choice resulting in an *SFA* is taking place] which are reflected in appropriate regions of our brains by movement away from thermodynamic equilibrium–in short, a kind of stirring up of chaos in the brains that makes it sensitive to micro-indeterminacies at the neural level. As a result, the uncertainty and inner tension we feel at such soul-searching moments of self-formation is reflected in the indeterminacy of our neural processes themselves. What is experienced phenomenologically as uncertainty corresponds physically to the opening of a window of opportunity that temporarily screens off complete determination by the past (Kane, 1999, p. 224-225).

Clearly, such "soul-searching moments" or such "windows of opportunities" is not something a deterministic world could offer. How about an indeterministic world? How do the micro-indeterminacies provide such soul searching moments? Before answering these questions, is Kane's world of indeterminacies a world that is colored with mystical powers, dual substances and magical powers of *agents*? He clearly states that he does not favor agent-causal libertarianism or any other kind of "extra factor" strategies (Kane, 2007). Still, in the following quote, Kane gives the impression ,to me at least, that he is favoring some kind of agent-causal libertarianism:

> When we do decide under such conditions of uncertainty, the outcome is not determined because of the preceding indeterminacy–and yet it can be willed (and hence rational and voluntary) either way owing to the fact that in such self-formation, the agents' prior wills are divided by conflicting motives. If we overcome temptation, it will be the result of our effort; and if we fail, it will be because we did not *allow* our effort to succeed (Kane, 1999, p. 225).

"[T]he outcome is not determined because of the preceding indeterminacy–and yet it can be willed …" But who is willing? Who is rational? And who is voluntary? An agent who does not obey physical laws? Remembering that Kane is an event-causal libertarian, it must not be a *rebellious* agent who wills. So what is

the one that wills, for Kane? Micro-indeterminacies? Why does a rational choice have to be dependent on a set of indeterminacies? Is not indeterminacy something threatening rationality, on the contrary? Maybe there is a window of opportunities before an indeterministic physical event takes place. However, this window of opportunities is just for the physical, not for what we call "will." Will simply obeys what physical entities *decide*. Therefore, *SFA*s are formed by the physical which behaves probabilistically. Wills *themselves* are not free of the external; they cannot make future actions free of the external.

And I believe, falsifying luck principle does not change this. It is not controversial that even in an indeterministic world, there is more than probability (luck) which determines outcomes (events): The input events, laws of physics. On the other hand, these factors are not in our control. And any factor out of our control is not less dangerous than luck, for free will.

Kane does not agree; he rejects the claim that the indeterminacy in the events is out of our control. For him, when somebody chooses not to steal a car in an indeterministic world, we should not say that "He was *lucky* not to steal the car. " or "He did not steal the car, by *chance*." Instead, he proposes to say that the person under consideration succeeded not stealing the car despite the probability of failure. For Kane, indeterminacy is a property of will's effort to choose, and this effort has no dependency on anything external (Kane, 1999, p. 232).

> The complex recurrent neural network that realizes the effort in the brain is circulating impulses in feedback loops and there is some indeterminacy in these circulating impulses. But the whole process is her effort of will and it persists right up to the moment when the choice is made. There is no point at which the effort stops and chance "takes over" (Kane, 1999, p. 232).

There is real indeterminacy in the effort of will; however, since this indeterminacy is not something external, Kane claims, it is will's effort that determines the

outcome action. Here, I am losing Kane: Does will's indeterminacy have anything to do with the laws of physics? If he believes that the indeterminacy is completely or partially free from the physical laws, this would make Kane an agent-causal libertarian defending that agents have powers to act independent of the laws of physics. On the other hand, if he believes that indeterminacy is completely dependent on the physical laws (and probability involved in them), this would imply that our actions are determined by physical laws. If the latter is the case, "self-forming actions" are physically or physically-and-probabilistically-formed. Because, just like luck, physical is not in our control; physically or physically-and-probabilistically-formed actions (*SFA*s) cannot be free.

Therefore, I believe that Kane is not successful in solving the problem of will's dependence on desires and aversions, resorting to possible indeterminacy involved in our world. Here arises a question: If I am right, if *SFA*s are not free contrary to Kane's claim, how can the problem of our wills' dependence on characters and motives be solved? Desires and aversions determine character and motives, and they determine will. How can a will simply satisfying its desires be free? It seems we are not free of our desires; we simply do what our desires command. When a volitional action is considered, can we do something that we do not want to do most? Yes, we can prefer to do what we *normally* would not want most among our choices, but only if we *want* to perform the action that we *would* not want to perform *normally*. Or simply, no, we cannot try to perform an action that we do not want most among the alternatives. An action which we want most among our choices is the same thing as what we try to do among these choices. So will is the slave of our desires by definition. As I express in the *Introduction*, for me, desires and aversions are conceptual prerequisites to will.

Then how can a slave be free? Only if it chooses its commander, Kane implies. After all, for him *SFA*s are free and other volitional actions result from our characters and motives shaped and constituted by *SFA*s. However, for me, in a

physical world *SFA*s cannot be chosen freely, because they are physically-and-probabilistically-formed, and hence they are not in our control. Besides, apparently, desires and motives precede choices. Therefore, the very first will should be shaped by a set of desires. This means that there is no will free from desires and aversions. Besides, as I claimed above, a will free from desires and aversions is a conceptual impossibility. Again, how can a slave be free? In my opinion, it simply cannot. A will freed from desires and aversions is not something that we could enjoy. If we still want to talk about freedom of will; we should take desires and aversions as something internal to will. I believe, this conceptualization is what would clear away Kane's need for self-forming actions.

Taking this conceptualization, to decide if we are free, we should seek for the source of wills, desires and aversions. Only if the source is not something external, may we have strong free will, which is not a possibility in a physicalistic world.

## 3.8. Agent-Causal Libertarianism

Libertarians believe roughly that some actions can be considered to be free, since they are not determined by factors beside the agent. *Event-causal* libertarians do not try to clarify what is meant by "agent" and they do not treat agent as a special entity which possesses special powers on the rest of the world. Up to now, in this thesis, mainly *event-causal* libertarian arguments were focused on. On the other hand, there are some other libertarians which diverge from *event-causal*

libertarians radically: *agent-causal* libertarians.[17] Agent-causal libertarians explain free will by reference to agents who can choose their action not only because of the state they are in, but also their *originated* independent choices.[18] They do not completely follow physical laws. For agent-causal libertarians, these agents cause events without being determined to do so. Indeed, they may have some tendencies which are determined by the circumstances they are in; however, they still can genuinely *choose* to do what they prefer.

I believe agent-causal libertarianism is very intuition-friendly: Most people intuitively believe that they are agents who make decisions without being determined to do so. However, is the agent-causal libertarianism consistent in itself, and is it consistent with the rest of our intuitions? Is it philosophically plausible? Is it scientifically plausible? Derk Pereboom asks these questions in his *Living Without Free Will*.

At first glance, agent-causal libertarianism does not seem to be scientifically tenable, as much as it is intuitive. It requires agents which have causal powers which are not determined by the physical. Science in general presupposes that events including choices are rendered by means of laws of physics (which indeed can be probabilistic) and previous states of the world. Agent-causal libertarians need to explain how agents could be undetermined. What of agents makes them unique? So unique that they are not causally determined. So unique that they are not fully explainable even in theory without referring to themselves.

To overcome the problems expressed above Derk Pereboom, put agent-causal libertarian strategies into three categories (Pareboom, 2001, pp. 69-70):

---

[17] For the distinction between agent-causal libertarianism and event-causal libertarianism see also Pereboom, 2001, pp. 69-88.

[18] See also Pareboom, 2001, p. xv.

1) Non-physicalist strategy: Defending the position that agent-causal powers are not wholly physical. And this is why they cannot be explained by laws of physics.

2) Ordinary non-reductive strategy: This strategy maintains that agent-causal powers are higher-level powers which are wholly constituted by complex lower-level physical organizations. And these higher level organizations are wholly governed by micro-level physical laws. However, agent-causal powers are not reducible to the physical.

3) Strong emergentist non-reductive strategy: Also for this strategy, it is true that agent-causal powers are higher-level powers wholly constituted by complex lower-level physical organizations. However, since they are something *more* than their constituents, they are not governed by physical laws.

I want to begin evaluating these strategies from the second one. In my opinion, the ordinary non-reductive strategy does not support agent-causal libertarianism. The problem whether or not everything is reducible to physics is an epistemological one. Irreducibility of will to the physical does not necessarily imply that will is independent of physical. To illustrate, it is consistent that chemical properties supervene upon physical ones, even if there is no way to reduce chemical properties to physical ones. Ignoring this gap between subjective reducibility and objective dependency, ordinary non-reductive strategy tries to base our objective liberty on subjective irreducibility. Pereboom claims that irreducibility does not provide any ground for libertarianism:

> Consider first the case in which the physical laws are deterministic. If everything is wholly constituted of microphysical entities, and these entities are governed by deterministic laws, then the complete microphysical state of the universe five million years ago, together with these deterministic laws,

> renders inevitable every subsequent complete microphysical state of the universe (Pareboom, 2001, p. 71).

Which means, our decisions are not free, given that determinism is true. And if our world is *not* deterministic

> …but is instead governed by … laws of quantum mechanics, then all of our decisions … will be constituted of events that are alien-deterministic, or truly random (those not produced by anything at all), or partially random (those for which factors beyond agent's control contribute to their production but do not determine them, while there is nothing that supplements the contribution of the factors to produce the events). To simplify, the causal history of all of the constituents of any of our decisions will be exhausted by the contribution made by factors beyond the agent's control, and nothing else (Pareboom, 2001, pp. 71-72).

And this is exactly what I am trying to support by the puppet analogy: If everything is wholly constituted and determined simultaneously by a puppeteer, it is not important whether or not our actions are reducible to the properties of the puppeteer, it is not important whether or not the puppeteer acts deterministically or indeterministically, our decisions are out of our control and we are not free.

Strong emergentist non-reductive strategy is another attempt to explain how we can be free without rejecting physicalism. This strategy maintains that even if everything is constituted by the physical, free decisions are not constrained by physical laws together with physical state. Indeed, it is stronger than the ordinary non-reductive theory, as its name suggests; however it also seems less plausible to me. Consider the case of deciding to buy a lottery ticket. Jones decides to buy a lottery ticket after a serious consideration of the odds. Now, there is a physical state (maybe a neural state) which constituted the psychological state of decision. Let us denote this physical state by "$S_p$" and the psychological state by "$S_h$." If every physical state is caused by another physical state, which a physicalist would

suppose, then there were some physical set of causes $C_p$, which caused $S_p$. Now, if the strong emergentist strategy is true, it should be possible that $S_h$ is not caused (or rendered) by $C_p$ together with physical laws. However, it was true that $S_p$ constituted $S_h$ and $S_p$ is caused by $C_p$. If $C_p$ caused $S_p$ and $S_p$ constituted $S_h$, how can we say that $C_p$ together with physical laws have not caused, determined or rendered $S_h$? How could it be possible that $S_h$ is not constrained by $C_p$ and physical laws. I do not find premises of emergentist non-reductive strategy consistent.

Pereboom has some criticisms against both the strong emergentist non-reductive and the non-physicalist strategies. He asks how these strategies would set agents be free in the sense required for moral responsibility. And he answers: If these strategies set the agents be free, then they must give agents the capability of producing *deviations* from the physical laws. Decisions of agents must not be perfectly predictable using laws of physics and information about physical states. Even if this is logically possible, Pereboom claims, it is not scientifically credible. There is no scientific evidence of such deviations for the time being. So he claims that the non-physicalist strategy and the strong emergentist strategy need to find some evidence of such deviations in order to be more than a logical possibility. And he divides objections to his claim into two groups:

1) *Reconciliation*: Accepts that there are no such deviations; however, there is still a possibility for the freedom required for moral responsibility.
2) O*verriding*: Insists that there are such deviations and there is also scientific evidence to believe this claim.

He thinks that Kant's strategy belongs to the first group. Kant claims that to have transcendental freedom which is a necessary condition for moral responsibility, agents must be able to cause something without being caused to do so. He also believes that the sensible world is deterministic: He says "… it is … necessary that everything that takes place should be fixedly determined according to laws of

nature" (Kant, 1949, p. 72). In spite of the sensible world's being deterministic, he does not think that we cannot be free. He believes that we are free in *ourselves*; and that is why the sensible world of phenomena's being deterministic does not threaten the possibility of freedom (Kant, 1998, A 541/ B 569). Pereboom accepts that Kant's attempt to prove that there is such a logical possibility is successful. However, he also believes that this *logical possibility* is not *credible* given the evidence we have:

> … Kant needs a much more substantial proposal, and this one fares differently. It is that all transcendentally free choices should be for just those possible actions whose physical components are causally determined to occur, and that none of these choices be for the alternatives. Aside from highly dubious idealistic attempts to explain how this might be, the wild coincidences implied by this proposal make it incredible (Pareboom, 2000, p. 80).

Pereboom also rejects the trials explaining free decisions based on the quantum theory. He reminds that even the quantum indeterminism has numerous regularities about the frequencies of events. So even if probability gives a space for freedom, this space is too regular to give the ability to choose freely. Moreover, this indeterminism does not give a large space for complex events like decision making. He notices that although there is a possibility that a soda can be displaced one inch spontaneously without any *macro event* causing it to do so, this probability is extremely small. In the same way, he thinks, the probability of a decision's being otherwise is extremely small, since a decision probably involves a large number of micro-physical entities.

Pareboom denies the possibility that probability in quantum mechanics is sufficient to make us free. He is right, in my opinion. However, there is more to say: Even if there were scientific evidence of significant

probabilities' being involved in our decisions, I do not think they would help us to be free. Again and again, why would I be freer, if I could do otherwise given that the complete set of true propositions belonging to the past is the same? Satisfaction of *PAP* concerning determinacy is not enough to make us free, as I argued throughout this chapter. Why would I, as a rookie poker player, am freer than Ms. Determined?

## 3.9. Conclusions

I believe our intuition that the problem of free will and the problem of determinism / indeterminism are strictly connected is misleading. The real essence of the problem of freedom is independent of the problem of determinism. Instead, it depends on the relation of supervenience. The questions we should actually ask are, "Does free will supervene on something other than itself?" and "Are free will and physicalism (world of puppets) compatible?".

To be more precise, there are both deterministic and indeterministic possible worlds with and without weak free will. After all, volitional actions do not necessitate indeterminacy or determinacy. They only require positive correlations between some of the desires and actions. It is obvious that there are deterministic and indeterministic possible worlds without weak free will: in a world in which there are no living beings, there is no free will. Indeed, there are both deterministic and indeterministic possible worlds without living beings.

In the same way, there are deterministic and indeterministic possible worlds *without* strong free will. However, there is no physicalistic possible world *with* strong free will. Strong free will requires origination, *PAP* concerning

supervenience or ultimate responsibility; and these cannot be satisfied in a physicalistic world.

Most libertarian incompatibilists take a version of *PAP* concerning determinacy as the criterion for free will. I agree with their insistence that this principle can only be satisfied in an indeterministic world; however, in my opinion it is not a sufficient condition for strong free will and not a necessary condition for weak free will. Therefore, it is not a proper criterion for free will in *any* sense. Freedom is not about indeterminacy; it is about the origination and only a *PAP* concerning supervenience may test origination. In my opinion, this criterion cannot be satisfied in any physicalistic world.

Agent-causal libertarianism, on the other hand, is not scientifically reliable. Besides, I believe that the truth of agent-causal libertarianism is not relevant to the problem of determinism either. It necessitates a substance which is independent of the physical and this requires *not indeterminism* but *non-physicalism*.

Table 1 encapsulates the answers of this chapter to the question "Are we free?". This summary supports the claim that free will has no dependency on the problem of determinism / indeterminism. In my opinion, the connection between the problem of free will and the problem of determinism / indeterminism is "nothing better than a long-lived philosophical illusion".[19]

[19] Dennett (1984a, p. 565) uses this expression for PAP.

**Table 1:** Summary of conclusions of this chapter.

| | **Deterministic World** | **Indeterministic World** |
|---|---|---|
| ***PAP* Concerning Supervenience (Strong Free Will)** | Are we free?<br>No. | Are we free?<br>No. |
| **Origination (Strong Free Will)** | Are we free?<br>No. | Are we free?<br>No. |
| **Ultimate Responsibility (Strong Free Will )** | Are we free?<br>No. | Are we free?<br>No. |
| **Compatibilist *PAP* (Weak Free Will)** | Are we free?<br>Yes. | Are we free?<br>Yes. |
| **Frankfurt's *PAP* (Weak Free Will)** | Are we free?<br>Yes. | Are we free?<br>Yes. |
| **Free to Will What We Want to Will (Weak Free Will)** | Are we free?<br>Yes. | Are we free?<br>Yes. |
| **Self-Forming Actions (Strong Free Will)** | Are we free?<br>No. | Are we free?<br>No. |

# CHAPTER IV

# TWO THOUGHT EXPERIMENTS BASED ON THE LAPLACE'S DEMON[20]

## 4.1. A Thought Experiment with False Presupposition(s)[21]

1) Assume that our world is deterministic.
2) Then assume that we have access to all the laws of nature and all the empirical data belonging to a time $t_1$.
3) Then it is possible that we make a computer[22] which is an instance of Laplace's demon based on the data and the laws of nature.
4) Suppose I want to consult this computer to know when I am going to die. And I learn that I am going to die in a particular plane crash when I am flying to İstanbul.

[20] This chapter is based on the discussions between Dr. Erdinç Sayan and me.

[21] For another thought experiment with many common characteristics, see Smedes, 2003, p. 969.

[22] A computer that predicts future perfectly using the precise empirical data and laws of nature which are fed to it.

5) The idea of death scares me. I do not want to die. And because of the physical mechanisms behind my free will and my fear of death, I would prefer not to go to İstanbul at least on this particular plane which is going to have the accident.

6) Either I would board the plane or I would not. My boarding seems problematic because of the fifth supposition. So I would avoid boarding the plane. And this is not consistent with the third supposition.

This puzzling thought experiment is based on Pierre-Simon de Laplace's work, *A Philosophical Essay on Probabilities*. In this work, he describes an intelligence which could calculate all the truths about the past and the future of the world using empirical data for a time point and laws of nature, in his famous quote:

> We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it an intelligence sufficiently vast to submit these data to analysis it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes (Laplace, 1902, p. 4).

Laplace's demon intuitively may seem consistent; however, it is *not*: Karl R. Popper (1950), D. M. McKay (1960) and Taede A. Smedes (2003) are some of the philosophers who articulated problems with Laplace's demon.

Finally, in 2008, in his influential paper "Physical Limits of Inference," David Wolpert has proven using formal mathematical language that this dream of making Laplace's demon could not come true. Wolpert first defines mathematical entities that perfectly predict future. He calls these entities "inference devices." Each of the inference devices are composed of a pair of functions: ($X$, $Y$) over a

set of world lines[23], *U*. Here, *X* is called the "setup function," which represents how the device is configured. And *Y* is the "conclusion function" representing what the inference device predicts. One more definition: An inference device, $C_1$, *strongly infers* another one, $C_2$, if and only if $C_1$ has the ability to infer the whole conclusion function of $C_2$. Then Wolpert proves the following proposition:

"*Let $C_1$ be a inference device over U.*

*…There is a device $C_2$ such that $C_1 \not\gg$* [does not strongly infer] $C_2$" (Wolpert, 2008, p. 1266).

This proposition implies that for every inference device $C_1$, there is at least one true proposition which is not inferable by $C_1$, but inferable by another inference device $C_2$. And this means that for every inference device, in any worlds, there is a limit of success. By proving this proposition, Wolpert shows that even if our world is deterministic and even if we knew all the physical laws and empirical data belonging to a time point, it would be impossible for any inference device to predict the future of our world perfectly. In other words, Laplace's demon cannot exist. This shows that my third presupposition in the thought experiment was false and the thought experiment was inconsistent.

Wolpert might have shown that Laplace's demon was not a consistent thought experiment; however, nothing in his paper implies that a very strong computer may not predict the future, *almost* perfectly.[24] Let us call this computer *ALD* (Almost Laplace's Demon). I believe that, in order to understand implications of determinism, to grasp the problems with Laplace's demon, and implications of

[23] A world line of an object is the sequence of four dimensional (space and time) events corresponding to the history of the object. So each point of a single world line is a position in four dimensions. And the world line of a physicalistic universe includes all the empirical data on the universe.

[24] *Cf.* Binder, 2008.

Wolpert's paper, a thought experiment based on *ALD* can still be fruitful. That is why I will present one, in this chapter.

An *ALD* is a relative of Laplace's demon whose existence does not imply any contradiction. Indeed, we have no apparent reason to believe that it is inconsistent; however, with the assumption in the thought experiment above, is there any reason to believe that an *ALD* which deserves its name may not exist? How strong can an *ALD* be? Can it predict far future? Or would *sensitive dependence on initial conditions* (butterfly effect) sabotage its predictions? Would it make false predictions or would it just be incomplete? Can something have all the empirical data including the empirical data about itself? To answer at least some of these questions, I will modify the original thought experiment by putting an *ALD* instead of Laplace's demon.

Before modifying the thought experiment, some clarifications: In this chapter, I refer to a version of weak free will by "free will," contrary to the other chapters. Now, we have free will as a product of our body and its interaction with the physical environment. When we make a choice, actually some neurological events happen in our nervous system. Free will is not the genuine cause of the choices; the physical mechanism behind free will is the real cause. Even if this approach does not make free will *genuinely free* or *causa sui*, it still explains how we can have a belief that we have free will in a physically determined world.

## 4.2. The Thought Experiment about ALDs

Changing Laplace's demon with *ALD*, the puzzle may seem to be eliminated:

1) Assume that our world is deterministic.

2) Then assume that we have access to all the laws of nature and all the empirical data belonging to a time $t_1$.

3) Then it is possible that we make a computer which is an instance of *ALD* based on the data and the laws of nature.

4) Suppose that I want to consult this *ALD* to know when I will die. And I learn that I am will die in a particular plane crash when I will be flying to İstanbul.

5) The idea of death scares me. I do not want to die. And because of the physical mechanisms behind my free will and my fear of death, I would prefer not to go to İstanbul at least on that particular plane which would have the accident.

6) Either I would board the plane or I would not. My boarding seems problematic because of the fifth supposition. So I would avoid boarding the plane. And this is *not* logically impossible, since an *ALD* is not a perfect, but a very strong predictor.

In this thought experiment, there is no straightforward contradiction with my not boarding the plane. However, there should be something wrong with the thought experiment, since it leads to implausible conclusions: *ALD*'s failing once in a while can be considered to be plausible. Or despite my knowing how and when an unwanted event would occur, my not being able to prevent this event may also be plausible. However, if there is nothing wrong with the thought experiment, each time *ALD* makes a prediction about an unwanted event, either it fails, or my effort to avoid this event fails. Does *ALD* fail to predict the future in all of the similar cases which involve an unwanted and seemingly preventable event? Then what makes an *ALD* an "*Almost* Laplace's Demon"? It does not seem plausible to me that whenever we learn an unwanted and seemingly preventable event from an *ALD*, we fail to avoid these events. I do not have any intuition implying that we

would begin to fail controlling (in the weaker sense) simple events about our future after we consult an *ALD*.

## 4.3. Questioning the Possibility of Determinism

It may be argued that such puzzles may not be solved by showing that nothing can predict everything about future perfectly. Even an *ALD* which predicts future imperfectly is a problem with the assumption that determinism is true. Therefore, a libertarian may argue that such thought experiments and puzzles coming with them shows that we are not living in a deterministic world. And because we are living in an indeterministic world, an *ALD*, which calculates most facts for the past and future time points, is impossible.

This claim is coherent with quantum theory. We know that, at least for now, we are not able to determine behaviors of some sub-atomic particles, which may seem necessary to make an *ALD*. And some scientists and thinkers believe that the empirical success of quantum theory supports the claim that the world is not deterministic. So maybe this thought experiment shows us that these indeterminists are right.

On the other hand, I believe the indeterministic candidate for a solution is not the one that would solve the puzzle. *ALD* may perfectly predict that its telling me that I will board the plane will cause me not to board the plane and its telling that I do something else will cause me not to board the plane. No matter how good the prediction is, whatever *ALD* predicts, I would do just the opposite of its prediction. That is why I believe that not the probability involved in causal relations but the directions of these causal relations makes it impossible for an *ALD* to foretell future correctly.

Moreover, even if our world is not deterministic, the puzzle would remain: Suppose that our world is indeterministic. Even if determinism is false it is obvious that our world is not *wildly* indeterministic; there are only minor events that are ruled by indeterministic laws. Apparently, even in this indeterministic world, an *ALD* making *imperfect* predictions with very high accuracy could exist. So an *ALD* in this world should be successful with a very high frequency in its predictions, about cases considered. However, the thought experiment we consider implies just the opposite: Apparently, almost every time I will learn how and why I will die, I would avoid it and fail *ALD*. Therefore, at least a mild indeterminism does not solve the puzzle. And what if our world is wildly indeterministic? Apparently, probabilities involved in our world are not wild. We do not expect airplanes to fall because of the recent indeterministic behaviors of a micro-entity without any macro-cause or we do not expect the presidents to activate a nuclear bomb because of indeterminacy caused by micro-particles in his/her nervous system.

Besides, an undoubtedly deterministic computational model of the thought experiments and puzzles can be constructed, as I explain below.

## 4.4. The Deterministic Computational Model of the Thought Experiment

First a character, *X*, which moves deterministically is defined.

Rule 1: *X* goes one step right, one step left, unless it knows it will die. (Figure 1)

Here *X* represents me, who moves deterministically in accordance with the laws of nature.

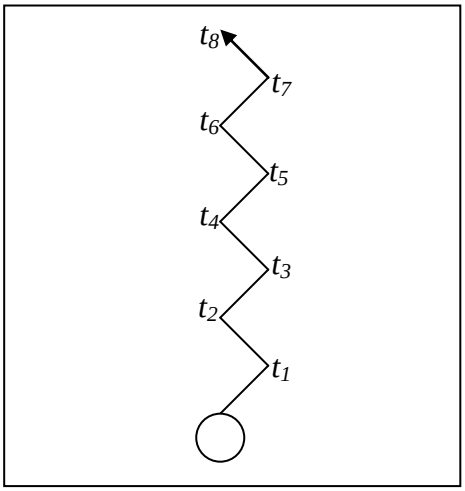


**Figure 1:** Outcome of rule 1.

Now *D*, representing death, is to be defined in the model. Suppose it is at $t_3$ on the right side (Figure 2). *X* does not *know* where *D* is, just like we do not know when we will die.

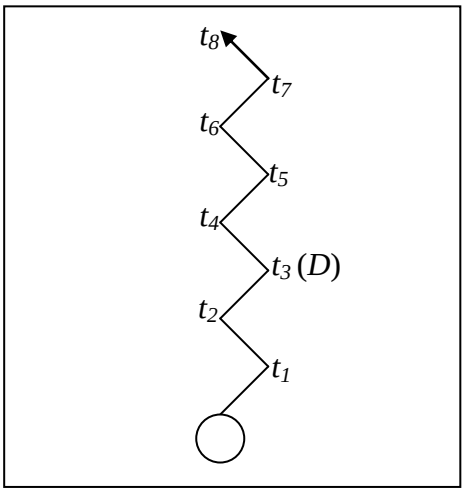


**Figure 2:** Outcome of rule 1 with *D*

To make a working analogy, another object *C*, which knows both the rules of *X* (the rule 2 is below) and where *D* is placed, is needed. Indeed, *C* knows everything about this *software world* just like the *ALD* aims to do in the parallel thought experiment. Before each step of *X*, *C* sends a message telling if *X* will meet *D* in its next step.

There remains one more rule to make the analogy complete. Rule 2 represents the volitional actions:

Rule 2: If *X* gets the message that in the next step it will meet *D* following rule 1, then for one step, it moves in the reverse direction of what rule 1 suggests.

In this model, it is obvious that the world is deterministic. There is no ambiguity or probability in the laws of the software world. *C* knows all the facts affecting *X*. So every component of the thought experiment seems to be modeled in a deterministic world.

Such an object, *X*, would move according to rule 1, until it reaches $t_2$. If *X* goes on its movement according to rule 1, it would meet *D* at $t_3$ and the dilemma arises here: What will *C* do now? It will either send a message telling *X* will meet *D* at $t_3$ or it will send a "No danger" message. If *C* sends the message saying that *X* will meet *D* at $t_3$, *X* will move according to rule 2 and will not meet *D*. So *C* will fail to predict the future. On the other hand, if *C* sends a no-danger-message to *X*, *X* will go on with respect to rule 1 and will meet *D*. And because *C* did not predict that *X* will meet *D*, *C* will fail again.

Apparently, it is possible to construct a deterministic computational model of the thought experiment. This implies that the problem about the thought experiments is not the deterministic assumption. If the reason why there are problems with the original thought experiment is the false assumption that the world is deterministic, what is the reason why there are problems with this computational model of the thought experiment, which is obviously deterministic?

## 4.5. Is Free Will a Delusion?

Another possible source of the inconsistency in the thought experiment is our supposition that human beings have some kind of *free* will. Maybe in a deterministic world people cannot have free will and my compatibilist assumption is the proposition that makes the thought experiment inconsistent. Such a conclusion would support the incompatibilist school and question *NDP*.

On the other hand, I am very skeptical about this candidate for solution, because actually in the thought experiment, we have not assumed that human beings have incompatibilist strong free will, which is hard to satisfy. What we pre-supposed was a compatibilist weak free will, which exists dependently on a physical mechanism. A compatibilist free will, which can be satisfied merely by our volitional actions. Would our volitional actions disappear if we learn about our future almost perfectly in a deterministic world? Would I board the plane that will carry me to my death, in spite of or in accordance with my will? Why would I board this plane even if I believe that the plane will have a crash? I do not think that this solution is tenable.

Moreover, this solution cannot be applied to the computational model either. Since in the computational model, there are no complex entities like feelings, consciousness, strong free will, and so on, it is reasonable to expect that the central problem about the puzzling thought experiment is not about these complex entities. This means, I believe, new candidates for solutions should be considered.

## 4.6. Unavoidable Unwanted Events Solution

If neither of the candidates for solution discussed up to this point is successful, then maybe the problem is somewhere else. In this and the following sections, I try to change my attitude towards the puzzle. Maybe the *ALD* would not tell me that I die in the plane crash when I am going to İstanbul; because if it tells me, I would avoid my death and this would not support the other assumptions. Here is a new scenario: When I am on my way to ask about my time and manner of death to *ALD*, for some reason, a contract killer follows me. And when I sit down in front of the *ALD*, the killer points his gun on me from behind. I want the *ALD* to show my death and the *ALD* shows me dying in a few seconds, in front of the *ALD*, looking at my killer. Maybe I cannot avoid looking backwards to the killer because of the shock. And I die as just as the *ALD* predicts.

This approach saves determinism, foreknowledge and free will by changing the structure of the deaths. It worked for one way of dying. But can we find such scenarios for *almost* all of the expected and unwanted events? Perhaps yes, but we have no reason to believe that all the unwanted events will become unavoidable after the *ALD* is built. What physical effect of the *ALD* will make all unwanted events unavoidable? When we know more because of the *ALD*, why do we die more in unavoidable ways? Knowing more, should we not expect being more successful in our decisions, instead of being the subject to unavoidable unwanted events? To save this candidate of solution, a satisfactory explanation for this explanatory gap has to be suggested.

## 4.7. Preferable Events Solution

When we know more, thanks to the *ALD*, maybe we should be more successful in our lives. Our decisions should be more accurate. That is what we are used to: The more knowledge, the more optimal decisions. Then the *ALD* would not inform me that I would die in a plane crash; but it would not tell me that I would die in an unavoidable way either. Perhaps, it would inform me that I would have a preferable death after a preferable life, such that I would not want to change my future at all. Maybe on the screen of the *ALD*, I would see that I die lying in a hammock, in a paradise island, after a very long and pleasant life. Or I would simply see that I die because the science and engineering find a way to let people live forever. This would also solve the puzzle, since in such a case, I would not try to change my preferable future and I would not falsify the *ALD*.

*ALD*s' changing our lives in a positive manner seems more reasonable than its changing our lives in a way that we have many unavoidable unwanted events in our lives. But this candidate of solution shares a problem with the unavoidable unwanted conclusions solution: What is the physical explanation of this change in our lives? Which of the *ALD*'s physical properties would change our future positively? I cannot find an explanation for this question.

If what improves the future is not the *ALD*, maybe *ALD* and the future's changing positively have a common cause: Perhaps, not *ALD*, but the technological and scientific improvements that make *ALD* possible also improve the future. To construct an *ALD*, we have to know (almost) everything about the physics: At least almost all the laws of nature, almost all the empirical facts. This would yield an amazing level of technology. Medical doctors would have the best diagnosis and treatment methods, security organizations would be more successful, technology and knowledge would improve quality of life dramatically. We might

even become immortal. This can be the explanation of the preferable events solution. It seems that this candidate of solution has fewer problems than the last one. Now I know why the *ALD* would not report that I would die in a plane crash: In a world with the scientific and technological knowledge necessary for *ALD*, there would not be any crash.

On the other hand, this explanation has an important problem. Changing the thought experiment a little bit, problems reappear. Suppose that an alien which knows all the laws of nature and has all the empirical data builds an *ALD*, brings it to our world and leaves it there. In this case, neither the level of medicine, nor the security organizations should be any more competent than they are today. In such a case, what would be the explanation of the positive changes in our lives? It seems that preferable events solution fails to be convincing either.

And besides, neither of last two solutions has a projection on the computational model of the thought experiment. In the computational model, rule 2 makes it impossible for *X* to meet *D*, if it receives the danger-message from *C*. This means unavoidable unwanted events solution is not applicable to the computational model. But there is no rule to guarantee that *X* would not meet *D*. This implies that preferable events solution is not applicable to the computational model either.

## 4.8. Can an ALD Have All The Empirical Data?

This approach denies neither the assumption that we live in a deterministic world, nor the supposition that we have free will in the weaker sense. Maybe in fact we live in a deterministic world, and in this world we have weak free will.

After trying to solve the puzzle about the *ALD* with other candidates of solution desperately, from now on impossibility of foreknowledge will be questioned: candidates of solution that reject possibility of an "*Almost* Laplace's Demon" deserving its name. Maybe it is not really possible to design such an *ALD* that knows (calculates) almost everything (all the history and future of the world). Such an *ALD* has to know (or have in memory) everything about our world including itself to calculate the past and the future. It can be argued that nothing / no one may "know" (have in memory) every bit of empirical knowledge about a system including itself, because there is infinitely large knowledge to be known about self. Even if we only take one proposition *p* which is true, the *ALD* has to know infinitely many true propositions. Obviously *ALD* will have to know at least one proposition *p*. And it is likely that for each *ALD* deserving its name, there exists a true proposition *p*, such that the *ALD* knows every proposition referring to *p*. But if the *ALD* knows every proposition referring to *p*, then it has to know that it knows that *p*. Let us denote the proposition "The *ALD* knows that *p* is true" as *K(p).* But the *ALD* has to know also that it knows that *K(p)* or *K(K(p))*, and so on. Then an *ALD* which knows every proposition referring to *p* should have infinitely large data even about a single proposition: {*p*, *K(p)*, *K(K(p))*, *K(K(K(p)))*...}. But because the memory of any *ALD* will have finite capacity, it is impossible for any *ALD* to know (have in memory) infinitely many data.

On the other hand, our physicalist presupposition does not support this solution. For Laplace and contemporary physicalists, *ALD*s do not deal with statements like *K(p).* They only have and they only process the physical data and laws of nature corresponding to *K(p).* If the physical events are the sources of all the states and if there are finitely many basic particles of which our world (including the *ALD*) is composed, then an *ALD* should be able to have all the necessary data for its predictions. It may not be able to interpret physical knowledge, but surely people who observe these past and future events in physical form, would be able to

interpret them. So I think if this solution is the one that we are seeking for, it has to be improved and clarified.

I believe that this strategy contending that nothing may have all the empirical data about a system including itself can be adapted to our physicalistic presupposition. In a physicalist world, we represent physical data in physical form. So, we represent physical properties by means of physical properties. Therefore, an *ALD* needs to possess at least as many physical properties as the world has, to be able to represent all the physical data to be represented. However, this is not possible, since an *ALD* itself is part of the world. That is to say, nothing can represent all the empirical data of a system of which it is a proper subset. And this implies that the only (*A*)*LD* which has all the empirical data can be the world itself. Note that at any time point, this *ALD*, the world itself, cannot have any additional knowledge like laws of physics. So it cannot predict the future, before it reaches the future.[25] This is one of the solutions I favor: For no *ALD*, is it possible to possess all the empirical data.

## 4.9. Can an ALD Predict the Future of a System which Includes Itself?

One of the simplest ways to solve the puzzle is to take the *ALD* out of the system and not let people communicate with it. In this situation, the puzzle would be eliminated. In this case, since I am not going to learn that I will die in the plane crash, I go to İstanbul on the plane and I die in the plane crash. Therefore, since

[25] Popper (1950a, p. 130) reaches a similar conclusion when he discusses the computation problem. I will discuss the computation problem in the next section. For the discussion whether or not the physical universe is just a computer, see also Wolpert, 2008.

the *ALD* knows the future events just as they would happen, the puzzle would be evaded.

So maybe we should simply accept that nothing may know *almost* everything in a system in which it is included. But why? Suppose that I, as an *ALD*, am trying to guess what dream one of my friends will have tonight. And assume that I have all the necessary knowledge.

1) I calculate that she will see that she is flying in her dream.
2) But wait!.. That was before I say to her that she would fly in her dream. If I tell her my prediction, maybe she will dream of something else… I should calculate again. If I say that she will see her flying in her dream, she will see me and herself flying in her dream… Then I should say that she will see us flying in her dream.
3) But that was before I say to her that we will fly in her dream…

So perhaps nothing can guarantee to predict how the future state of a system will be, unless it stops communicating with the people in the system. Here *communication* is not a magical action. Indeed, it is a special case of *causation*. Therefore, it would be more appropriate to conclude that nothing can guarantee to predict an outcome in a way dependent on the effects of its prediction, since it does not know what it will predict yet. Nothing may know (have in memory) the result of its calculation before it calculates. And if the progress of calculation affects the future, it cannot infer the future.

Smedes (2003, pp. 969-970) formulates the same problem as infinite regress of computational levels, for Laplace's demon: In order to compute the future, a computer has to use all the data related the world. To do that, it has to consider its own physical structure which is a part of the world. However, it cannot do that at the same level ($L_0$). Therefore, it has to make calculations at a higher level $L_1$. However, it has to make calculations about the physical structure behind $L_1$, since

this structure is also a part of the world interacting with the rest of the world directly or indirectly. Therefore, it has to make calculations about the structure behind the $L_1$, at a higher level $L_2$, and so on.

On the other hand, if the computer does not compute at hierarchical computational levels, Smedes claims, it will face another problem: To make a calculation, the computer has to have complete data about $d_0$ itself at a time point $t_0$. However, after it records $d_0$, its state and complete data about itself changes to $d_1$ at a later time point $t_1$. Then it has to record this new data and change its state again. Another infinite regress making Laplace's demon impossible.

Popper also states that a machine $C$ cannot predict its own future, even if another machine with the same capabilities $C^+$ can. And he also clarifies the reason why $C$ fails in the task which $C^+$ with the same capacity succeeds accomplishing:

> If we furnish $C$ and $C^+$ with the same information (e.g. about $C$) and set them the same task, they will of course produce precisely the same predictions. Now we assume we have furnished only $C^+$ with this accurate and full information about $C$, and that $C^+$ alone produced predictions about $C$'s $t_5$ state. No doubt these predictions will be true, and $C$ will be at $t_5$ in the state predicted. But if $C$ also is given that information, this will mean that we have interfered with $C$, and that its $t_5$ will be different from the previous case (Popper, 1950b, p. 179).

However, claiming that it is possible for $C^+$'s to predict C's future states, Popper ignores some problems that he is absolutely aware of: $C^+$ cannot guarantee to predict $C$'s future state, since it cannot predict its own state and effects of its own state on $C$. So $C^+$ needs a $C^{++}$ and it needs a $C^{+++}$, and so on.

There is one more reason why $C^+$ cannot predict the future of C, which Popper does not mention: If $C^+$ is not physically more complex than $C$, $C^+$ can barely guarantee to model $C$, as it is argued in the last section. Since no additional data

can be represented by $C^+$, it cannot guarantee to have the rules for the calculation of the future state of *C*.

It seems quite clear that these problems of calculation and memory will not let Laplace's demon be realizable even in principle. What about *ALD* which is less demanding? In my opinion, problem about impossibility of perfect prediction would be enough to fail an *ALD*, regardless of the fact that it is not as demanding as Laplace's demon. Especially for cases in which a person learns the predictions of an *ALD* about herself, her seeing the predicted future may yield totally different conclusions than *ALD*'s prediction, because of the processes in the nervous system of the person. This simple divergence produced by the person's learning her predicted future can be considered to be an instance of "the butterfly effect," since human nervous system is very sensitively dependent on the initial conditions.

Popper also holds that position: For him, every predictor may *strongly interfere* with the system it predicts, after a sufficiently long period of time. The interference may be negligible at first, but this negligible interference probably results in a strong interference because of the butterfly effect (Popper, 1950a, p. 129).

This means that it is impossible to construct an *ALD* (deserving its name) even in a deterministic world, in which it is included. This is the second candidate of solution I favor.

## 4.10. Questioning Truth of Determinism Again

Following almost the same path, Popper reaches a conclusion which I did not:

> [W]e can assert that classical mechanics is not deterministic, but must admit the existence of unpredictable events.
>
> If this is correct, then Laplace's determinism, and that of others who were influenced by the *prima facie* deterministic character of classical mechanics, is based upon a misinterpretation (Popper, 1950b, p. 193).

I agree with Popper about most of his claims, which are about the impossibility of a perfect predictor, in both parts of his *Indeterminism Quantum Physics and in Classical Physics*; however, concluding that our world is not deterministic, only because it is impossible to construct Laplace's demon, is not consistent with the contemporary concept of determinism. After all, determinism is not an epistemological concept and being *deterministic* is not the same thing as being *perfectly predictable*. In the section *The Computational Model of the Thought Experiment*, a deterministic system having the problem of predictability is illustrated already. It is not reasonable to conclude that a computational system which does not involve any chaotic or probabilistic relations is not deterministic, only because it has the problem of perfect prediction *within the system*.

As I said, it does not seem acceptable to me to use an epistemological criterion for determinism; however, if determining a criterion of determinism using the concept of predictability is insisted on, the following would be the only one acceptable I can think of: "Determinism is true if and only if it is *logically possible* that a perfect predictor exists." By showing that it is nomologically impossible to predict future, we cannot conclude that determinism is false. And in

no studies I have encountered, it has been shown that it is logically impossible to predict future. In each work trying to show impossibility of Laplace's-demon-like beings, there were empirical assumptions, like Laplace's demon's being physical, being causally efficient, being available for human beings, being finite, etc.

After all, it is logically possible that there exist a perfect instance of Laplace's demon causally isolated from our world, although a Laplace's demon intervening with the world has problems. Therefore, even if we take an acceptable epistemological criterion for truth of determinism, prediction problems mentioned in this chapter do not imply that determinism is false.

## 4.11. Is It Possible that a Secretive ALD exists?

Up to now, *ALD*s which are very user-friendly, *ALD*s which do not keep any secret from the users have been considered. Is it, on the other hand, possible that there exists an *ALD* which does not reveal some of its predictions to the users? Perhaps such a computer would not tell me that I would die in the plane crash, since it would predict that such an action would make itself fail in its prediction. It would not give me at least the complete details of its prediction. It could tell me only the future facts which would not make me change them, in order to assure its success in its prediction. To illustrate, instead of giving complete details of my death, it could just tell me the time when I would die. Is it plausible to believe that such a secretive *ALD* may exist?

First of all, an *ALD*'s being secretive would not change the fact that it would not be able to hold all the empirical data about the world including itself because of the reasons presented in "*Can an ALD Have All The Empirical Data?*". However, since it is not a perfect Laplace's demon, maybe it does not have to have all the

empirical data anyway. Now the question to be asked is "Can an *ALD* have enough data about the world, such that it predicts as many facts as an *ALD* deserving its name should predict?". Since an *ALD* can only perfectly represent the data about itself, and since there is enormous amount of other data to be represented, I am not optimistic about *ALD*s' ability to keep large enough data for its purpose.

Moreover, if the causal effects of the process having role in an *ALD*'s predictions is significant, being secretive would not help the *ALD*. Consider an *ALD* which makes calculations using some electronic mechanism. These electronic mechanisms would produce some electromagnetic waves affecting on the rest of the world. For example, these electromagnetic waves could result changes in people's nervous systems and these changes would have the potential to affect on people's choices and to fail the *ALD*'s predictions. Or these electromagnetic waves could produce some other physical changes and these changes would yield a "butterfly effect" resulting in significant or even catastrophic events, making *ALD*'s predictions fail. Since an *ALD* does not know what calculation process is required for its prediction before this process ends, it would be impossible for it to calculate how this process would affect the rest of the world. That is to say the problem of *ALD*s presented in "*Can an ALD Predict the Future of a System which Includes Itself?*" works for secretive *ALD*s which do have significant causal interaction with the rest of the world too.

On the other hand, if the *ALD* were very secretive and the effects of the process carrying out the *ALD*'s predictions were negligible, this would mean that it is practically causally isolated from the rest of the world. But it is not interesting to conclude that an *ALD* causally isolated from the world is possible. In addition to these, an *ALD* may not know when its insignificant interventions on the rest of the world would result in a significant intervention. Therefore, I conclude that even a secretive *ALD* has important problems.

## 4.12. Foreknowledge by Computation vs. Foreknowledge by Precognition: The Puzzle Awakens

Up to here, the possibility of foreknowledge by computation has been discussed. An *ALD* or a Laplace's demon is supposed to calculate facts about a time point $t_1$, using the facts about another time point $t_0$ and laws of physics. How about foreknowledge by precognition? Is it nomologically or at least logically possible? Could we have an ability that lets us *see* the future without making any kind of computation based on causal laws? Could we see the future just like God is supposed to do? The assumption that some entity which does not causally interact with the rest of the world could see the future and be right in all of its predictions seems logically consistent to me. Having a meta-temporal perception ability would let an entity perceive the future and the past without affecting them. This gift demands neither an infinite database, nor the infinite computational ability that Laplace's demon needs. It seems possible that right now a genuine psychic located outside of our world knows everything about our future and past. This is not very surprising, since a version of Laplace's demon which is isolated from our world is also possible, as I have argued.

However, there are indeed problems with psychics who are causally interacting with the world they examine. Let me call these imaginary physics "Magnificent Psychics" following Dr. Erdinç Sayan. Here is a new version of the thought experiment which is presented at the beginning of the chapter:

1) Assume that there is a magnificent psychic, who can perfectly foresee the future.

2) It is possible that someone asks the magnificent psychic wants to see when she would die. She may learn that she will die in a particular plane crash when she is flying to İstanbul.

3) The idea of death would probably scare her. She may not want to die. And she would prefer not to ever go to İstanbul by plane.

4) Either she would board this plane or she would not. Her boarding seems problematic because of the third supposition. So she will *not* board this plane. And this is not consistent with the first assumption.

This time not the causal determinism, but only the logical determinism (fatalism) is committed to by the thought experiment. In other words, for this thought experiment to be consistent, the world does not need to be governed by deterministic causal laws, even if it is necessary that everything that happens has to happen in the way it does. In logical determinism, there is only one possible future, yet this future cannot be computed using causal laws even in principle.

If logical determinism is false, it is impossible that there exist a magnificent psychic: As Sayan pointed out in one of our discussions, if logical determinism is not true, then either there is no future before we reach it, or there are multiple possible futures. If there is no future to see before we reach it, obviously nothing would be able to perceive it by precognition. If, on the other hand, there are multiple possible futures, each time a (so-called) magnificent psychic tries to see the future, she would see a different one and fail in at least some of her meta-temporal perceptions.

For the thought experiment about *ALD*s, I have argued that a solution which rejects determinism is not favorable, since apparently the problem is not about indeterminacy involved in causal relations but the directions of them. I hold a similar position regarding the thought experiment about the magnificent psychic, which only presumes logical determinism to be true. Even if logical determinism

is false, there are a lot of regularities in our world because of causal connections. The truth value of "I will not be abducted by a *UFO* by midnight" may not have a truth value before midnight; however, we have good reason to believe it is true. Our living in a world with lots of regularities is enough to lead us to the puzzling point in the thought experiment.

In other words, even if logical determinism is false and the magnificent psychic's ability to see future fails once in a while, the problem cannot be overcome. In this case, the magnificent psychic should have both a very good accuracy in her predictions and ability to avoid unwanted avoidable events frequently, since she has volitional actions. However, her ability to predict future and her ability to succeed in avoiding unwanted events cannot be strong enough together: If her precognition ability is strong enough, her volitional actions should be sacrificed, and if efficiency of her volitional actions is not bizarrely weak, her ability of precognition cannot be strong enough.

Besides, we can construct a logically deterministic computational model of the thought experiment, in which each time *C* tries to guess how *X* will move it fails, unless *X* fails to satisfy rule 2.[26] Moreover, if it is true that indeterministic solution does not solve the puzzle about *ALD*s, it is reasonable to expect it does not solve puzzle about the magnificent psychic either, since the two thought experiments have much in common.

Because the thought experiment presumes not strong free will, but only volitional actions, denial of free will does not seem to be an option for solving the puzzle either.[27] It seems very strange to think that if we had the ability of a magnificent psychic, we would not try to change our destiny.

---

[26] For details, see the section "The Computational Model of the Thought Experiment" in this thesis.

[27] See the section "Is Free Will a Delusion?" in this thesis.

For the magnificent psychic thought experiment, *unavoidable unwanted events solution* and *preferable events solution* does not get rid of the problems mentioned in the relevant sections either. Why would unavoidable unwanted events begin to be much more frequent, if I got the ability of a magnificent psychic today? Or why would I see preferable events in my future before even trying to change them, by getting the ability of precognition?

A magnificent psychic does not have to have a model of the world, or does not have to have a computational ability, like an *ALD*. This means, solutions based on computation and memory incapacity of *ALD*s are not applicable to the case of the magnificent psychic.

The puzzle of the magnificent psychic seems very similar to the puzzle of *ALD*s, and intuitively, I believe, so should their solutions be. For the case of *ALD*s, besides claiming that it is impossible to predict the future because of the absurd conclusions perfect prediction implies, more detailed explanations for this impossibility has been presented in 4.8, 4.9 and 4.11. However, for the case of the magnificent psychic, I have no suggestion for a detailed explanation of why for a magnificent psychic it is impossible to foresee future by precognition. It seems to me that this is one of the weaknesses of the arguments presented in this chapter. For now, I see no way of filling this explanatory gap.

## 4.13. Two Thought Experiments on the Infallible Foreknowledge

### 4.13.1. Schlesinger's Solution for Newcomb's Paradox: It is Impossible That a Perfect Predictor Exists

If the claims of these chapter are correct, all the paradoxes which presume a perfect ability of prediction or precognition have at least one premise which cannot be true.

So does Newcomb's Paradox: In the paradox, a player is playing a game with a perfect predictor. There are two boxes in the game: Box I and box II. The player is trying to get as much money as he can, by taking only box II or both boxes. In the box I, there is always $1,000. On the other hand, the predictor may or may not put $1,000,000 in box II depending on its prediction of the player's choice. If the perfect predictor predicts that the player is going to take both boxes, it does not put $1,000,000 in box II; and if it predicts that the player is going to get only box II, it puts $1,000,000 in it. Assume that the predictor made its prediction and it put either $1,000,000 or nothing in the box II. What should the player do to increase its payout? Should she take only box II or both boxes?

First Argument: Some say that the player should take only Box II, since it is guaranteed by a *perfect* predictor that if the player were to take both boxes (or equivalently, if the player takes both boxes), there will not be any money in Box II. This position obviously makes sense.

Second Argument: On the other hand, some others argue that the player should take both boxes. Their argument also seems alright: Regardless of perfect predictor's prediction, if the player takes both boxes, she will get $1,000 more. If the perfect predictor predicted that the player would get two boxes, the player

would get $1,000 instead of nothing and if otherwise she gets $1,001,000 instead of $1,000,000.

To decide his position, Schlesinger (1974, p. 210) modifies the problem. Instead of a perfect predictor, he puts an observer in the game. Again, there is $1,000 in box I and the observer gives an additional $1,000,000 to the player *after* her decision, if and only if she takes only box II. Indeed, the second argument presented above is not rational for this case, since it is obvious that there is no possibility that the player takes nothing or $1,001,000 for this case. Schlesinger claims that taking the assumption about the capabilities of the perfect predictor's *seriously,* the original problem is not really different than this modified one. Therefore, as far as Schlesinger concerned, if the problem is consistent, the player should take only box II. However, he does *not* think that the problem is consistent.

He makes another modification in the problem to point out the inconsistency. This time we stick to the original scenario in that the perfect predictor puts the money beforehand. However, there is a guest, Jones, a very good friend of the player who gets the opportunity to see if there is $1,000,000 in box II. To clarify, let us say that box II is transparent and visible from Jones's position, however, it is hidden from the player's view. Obviously in both of the cases Jones would advise the player to take both boxes to ensure the player get an extra $1,000. This new modification does not harm the core of the original problem either. So the second argument is right too.

Considering these two modified problems, Schlesinger argues that both of the arguments, which contradict one another, are derivable from the same set of premises. Therefore, he concludes that the problem itself must be inconsistent. More specifically, he claims that the inconsistency about the problem is that it is not possible that a perfect predictor exists.

I find this solution to be one of the most convincing ones for Newcomb's paradox, and the consistency of Schlesinger's conclusion with those of the other arguments

presented in this chapter supports the position that even in a deterministic world perfect foreknowledge is impossible.

### 4.13.2. Prediction Paradox and Free Will

Prediction paradox is another problem having some common characteristics with the thought experiments discussed in this section. Again perfect prediction is at the center of the problem. This is another paradox which may seem to support the thesis that we do not have free will; however, Ardon Lyon argues just the opposite:

He uses an examination version of the paradox (Lyon, 1959, p. 510): A headmaster ensures his pupils that an examination will be given on an *unexpected* day of the next term. A clever pupil, apparently rightly, claims that the exam cannot be given on the very last day of the term, because if it were to be given on the very last day, the pupils would know when the exam would be given on the eve of the examination day. And this would violate the rule that the exam would be unexpected. Moreover, if the exam cannot be given on the last day, it follows that the exam cannot be given on the day before the last day either. Otherwise, the exam would be expected on the eve of this day, in the same way, since the exam cannot be given on the very last day of the term. Following this reasoning the clever pupil concludes that this exam will never be given. The headmaster listens the reasoning without commenting and gives the exam one week later. And obviously this exam surprises the pupils and satisfies both rules of the headmaster.

Lyon analyzes the two rules the headmaster sets up:

> Rl An examination will take place on one day of next term.
>
> R2* The examination will be unexpected in the sense that it will take place on such a day that on the previous evening it will not be possible for the pupils to deduce from Rules 1 and 2* that the examination will take place on the morrow (Lyon, 1959, p. 512).

Here what is meant by *R2** is very critical and may change the conclusions to be derived from the paradox. Lyon simplifies the case, to analyze it. He convincingly claims that when *A* holds two cards (one black and one white) and says that *B* cannot predict which card he will lay, what he means is that *B* cannot predict the color of the first card, not the second one after the first card is laid. In the same way, Lyon claims, when the headmaster says that the exam will not be predicted, he does not mean that it cannot be predicted whether or not the exam be given on the very last day of the term. Obviously, on the eve of the very last day, pupils are going to know whether or not there will be an exam on the last day, like *B* knows the color of the second card after the first is laid. What the headmaster actually means is simply that the pupils cannot logically deduce the examination day, if it is not to be given on the very last day of the term.

For Lyon, there are two ways to interpret *R2**:

> S1 The examination will be unexpected in the sense that ... it will not be possible for the pupils to deduce from Rules R1 and S1 that the examination will take place on the morrow, unless it takes place on the last day.
>
> S2 The examination will be unexpected in the sense that ... it will not be possible for the pupils to deduce from Rules R1 and S2 that the examination will take place on the morrow, even if it takes place on the last day (Lyon, 1959, pp. 512-513).

If what the headmaster meant by *R2** is *S1* (probably it is), Lyon argues, the headmaster's rules were logically consistent in the first place and the clever pupil misinterpreted the *R2**. And this is the reason why the exam given after one week

was unexpected. Lyon seems to be right, since if *S1* is the intended interpretation, then it is not true that the headmaster cannot give the exam on the last day and the whole argument of the clever pupil rests on this false proposition. The clever pupil did not interpret the rules as the headmaster intended. If *S1* is the intended interpretation, then the exam can be given on the very last day without violating his rules.

And if the headmaster means *S2*, obviously his rules are logically inconsistent and cannot be satisfied. If this is the case, the pupil's proving any proposition is not surprising at all, since the headmaster suggests an inconsistent set of premises. Even if in reality, the exam is not given on the very last day and the headmaster succeeds in surprising the pupils, Lyon claims, it is still false that if the exam were to be given on the very last day, it would be unexpected. I find Lyon's analysis to be clear and convincing.

Then he makes a very interesting claim: For him, even if there is a perfect predictor, and the world is deterministic, if a perfect predictor informs you what you will do, you can do otherwise. So he claims that a perfect predictor failing because of our freedom is logically consistent (Lyon, 1959, p. 515). Apparently, this means that determinism and free will are consistent. However, his conclusion is based on a problematic approach: He assumes that the world is deterministic and the predictor is perfect; nevertheless, he lets the perfect predictor fail. If the predictor is perfect, as he assumed, then we cannot fail the prediction even if we get to know the perfect prediction. And if we fail the predictor, it cannot be a perfect predictor in a deterministic world.[28]

Even if it seems that his argument about the compatibility of free will and determinism is inconsistent, his thought experiments have a core which is tenable

[28] For a detailed criticism, see Canfield, 1961.

and coherent with a kind of compatibilism. His criterion for freedom is some kind of *PAP*.

Lyon's *PAP*: An action is free, if and only if the performer can do otherwise, if she gets to know that a perfect predictor had predicted that she would do this action.

Changing the material conditional with counterfactual conditional, makes a working criterion for weak free will discarded from the problem mentioned.

Lyon's *PAP* Revised: An action is free, if and only if the performer could do otherwise, even if she got to know that a perfect predictor predicted that she would do this action.

Lyon's *PAP* implies that the thought experiments focused on in this chapter, which are related to precognition or perfect prediction, are not a threat for the compatibility of free will and determinism. On the contrary, according to this principle, the thought experiments under consideration show that determinism and free will are compatible. After all, no matter whether we are living in a deterministic or indeterministic world, no matter whether the predictor is fallible or infallible, if we were told that we would voluntarily do something in a future time by a predictor, we could do otherwise (at least if the prediction is not the most preferable choice for us among the realizable ones). What Lyon (1959, p.517) concludes supports compatibilism. He defends that we can be free in a deterministic and physicalistic world. And this is consistent with the conclusions of this thesis. Lyon's revised *PAP* practically is not different than the compatibilist *PAP*.

Lyon's *PAP* Revised: An action is free, if and only if the performer could do otherwise*, if she got to know that a perfect predictor predicted that she would do this action.*

Compatibilist *PAP*: An action is free, if and only if the performer could do otherwise, *if the performer wanted to do so*.

Apparently, I could do something other than what the perfect predictor predicted (and I did), only if I *wanted* to fail the perfect predictor for some reason. So Lyon's *PAP* is satisfied, only if compatibilist *PAP* is satisfied. Therefore, just like compatibilist *PAP*, Lyon's *PAP* implies that weak free will is compatible with determinism, which is among the conclusions of this thesis.

## 4.14. Conclusions

After stating the *ALD* thought experiment, I have tried to solve in six ways. Only two candidates of the solution seems acceptable to me. Each solution serves as an explanation why it is impossible that an *ALD* exists.

To sum up:

1) It is possible to build a Laplace's demon, which would infallibly know the future in a causally deterministic world, *only if* it is causally isolated from the world.
2) It is impossible to build an *ALD* which guarantees to continue its success for two reasons:
    a. No object may know everything about a system of which it is a part.
    b. No object can guarantee to predict an outcome dependent on the effects of its prediction, since it does not know what it will predict before it makes the prediction.

3) It is impossible for a magnificent psychic to foresee future by precognition, if it is not isolated from the world.

These imply that the thought experiments which involve infallible precognition or prediction has at least one false premise. The relation of the prediction paradox and Newcomb's paradox with the thought experiments about the *ALD* and the magnificent psychic is discussed in two sections of this chapter.

The thought experiments discussed in this chapter may seem a threat for compatibilism. However, a few reasons why candidates of solution which reject compatibility of free will are not credible are presented in several sections. If these arguments are reasonable, as they seem to me, there is nothing which implies that there is a significant connection between free will and determinism in these thought experiments. *NDP* seems me to be one of the consistent and perhaps most plausible position in the problem of free will and determinism.

# CHAPTER V

# RESPONSIBILITY AND DETERMINISM: ARE THEY EVEN RELEVANT TO EACH OTHER?

## 5.1. The Reason Why We are not Responsible for What We Do

I defend the view that if physicalism is true, no matter whether the world is deterministic or not, we have weak free will and we do not have strong free will. This raises an important ethical question: Is weak free will enough to constitute a reliable foundation for responsibility? If we do not have strong free will and if weak free will is not enough to constitute a foundation for responsibility, then most of the ethical paradigms and contemporary justice systems may be questionable. If somebody is not responsible for her actions, how can we blame or praise her?

Epictetus (1888) thinks that our desires are in our power. He claims that our opinions, intentional actions and even desires and aversions are in our power. For him, this is the reason why we should take the responsibility of those. It seems that Frankfurt (1971) also believes that our desires are in our power by means of

higher-order desires. For Frankfurt, we can choose what to want by our second-order desires: "I don't want me desiring to abuse drugs." Now, *persons* who abuse drugs are responsible for wanting to use drugs, since they could avoid wanting to use drugs, by means of their higher-order desires.

Obviously, if Frankfurt is right, then we should be responsible for our intentional actions. If he is right, responsibility, which is the foundation of contemporary justice and ethics is safe. However, I believe that his concept of higher-order desires is unhelpful for saving responsibility just like it is unhelpful to construct a complete and consistent compatibilist concept of free will or personhood.

Following Frankfurt's path does not help responsibility: We are responsible for our first-order desires, because we have second-order desires confirming our first-order desires. And we are responsible for our second-order desires, because we have third-order desires confirming our second-order desires. But this chain cannot continue infinitely. How about our highest-order desires? Why are we responsible for our highest-order desires? In order to be convincing, Frankfurt has to show that at least the highest-order desires are not caused by something *external*, which seems counterintuitive. "To have significance," claims Gary Watson, "…the hierarchy [of multiple level of desires] must be grounded in something else that precludes externality" (Watson, 1987, p. 149).

I think we cannot be genuinely responsible. First, our desires and aversions are not in our control. Indeed, we could do otherwise, if we wanted to do so; but we could not have wanted to do something other than what we actually wanted to do. Or in Frankfurt's conceptualization, and contrary to his view, even if we have freedom of *action*, we do not have freedom of *will*. In short, we cannot choose

what to want. We cannot, because our highest-order desires, aversions and will supervenes upon the physical, something external. Since we cannot be responsible for deterministic or indeterministic laws of physics and the previous state of the world, we cannot be responsible for our desires, aversions, wills or actions, which are determined by these laws and states.

What I suggest is rationally quite clear even if it is not intuitively so. In order to gain an intuition supporting the arguments defended in this chapter, I suggest following a path of thought experiments. First assume that a blind woman bumps into a person when she was trying to walk on the street. Obviously, one would not think that the woman is responsible for hitting the person in this case, since not bumping into hit the person was not in the blind woman's power. Epictetus or Frankfurt would also interpret this situation in the same way and he would advise one not to take the blind woman as responsible for hitting the person.

Now, let us change the case a little bit. A man who is schizoid hits a person because he is wrongly sure that this person will kill a child, if he does not do so. Should we consider this man responsible for what he has done? Despite of his good will, should we punish him because he hit an innocent person, when it is clear that his action is a symptom of his illness? Probably again in this case, one would not think that the schizoid man is responsible for what he has done. And probably Epictetus or Frankfurt would not object to this either. The second case is different than the first one only in that inability to do otherwise is caused by not a *physical* handicap but a *psychological* (which is in its roots physical too, if physicalism is true) one.

And finally, let us suppose that a man who is rude hits someone intentionally just to annoy this person. It is obvious that such actions are mostly influenced by prior psycho-physical history and genetic factors. Suppose, for example that the man is mean to others, because he has a unconscious belief formed by his experiences that everybody is his enemy. And moreover his genetic makeup makes him less

happy than other people, *via* determining how his serotonin system works. Now, can we say that this rude man is responsible for what he has done? Can we say that he has no psycho-physical disorder, unlike the schizoid man; or can we say that despite his genetic makeup and experiences, it was in his power not to hit the person, unlike the blind woman? Can we say that he is responsible for what he has done? Is he really responsible for his not being rational enough? Is he really responsible for the neural activity in his brain, which caused him to hit the person? I defend that he is not, since this case is different than the last one only in that the handicap resulting in inability to do otherwise is not *psychotic* but *neurotic*.

Perhaps, Epictetus would think that the rude man is responsible for what he has done, since Epictetus believes that even our desires and aversions are in our control. However, it is hard to hold this position, unless contemporary philosophy and science are ignored. Indeed, Epictetus is not responsible for not having heard of the experiment of Benjamin Libet, in which Libet (1999) measures *readiness potential*, an electrical change in the brain which regularly precedes the conscious experience of decision. Or he is not responsible for not having heard of modern medicines which change our desires without consuting to our "free" will. I believe, since the case of the rude man is not qualitatively different than the case of the blind woman, or the schizoid man, he is not responsible for his action. Since no action in a physicalistic world is qualitatively different than any of these actions with respect to *PAP* concerning supervenience, we are not genuinely responsible for anything we do.

We have only weak free will and it is *not* enough to constitute responsibility. I believe that we are *not* responsible for our actions. This implies that our justice system and ethical beliefs do not have the ground of responsibility they need. If physicalism is true, we are just complex deterministic or indeterministic machines

and machines just do what the physical imposes; they cannot own genuine responsibility.

> Traditionally, it has been assumed that moral responsibility requires us to have some type of free will in producing our actions, and hence we assume that human beings, but not machines, have this sort of free will. At the same time, there are reasons for regarding human beings as more like machines than we ordinarily suppose. These reasons stem from various sources: most prominently, from scientific views that consider human beings to be parts of nature and therefore governed by natural laws, and from theological concerns that require everything that happens to be causally determined by God (Pareboom, 2001, p. xiv).

Not the incompatibilists' *PAP* concerning determinacy, but *PAP* concerning supervenience is a necessary condition for *genuine* freedom and responsibility. If we cannot do otherwise in exactly the same circumstances, we cannot be responsible for our actions. We *cannot* do otherwise in a physicalistic world, and this is why we are *not* responsible for our actions.

Dennett also rejects that *PAP* concerning determinacy's being a necessary condition for responsibility: He calls *PAP* "nothing better than a long-lived philosophical illusion" (Pareboom, 2001, p. 565). He claims that the reason why we ask the question "Could the performer have done otherwise [given that *complete set of true propositions belonging to the past is the same*]?" after somebody commits a crime has nothing to do with free will or responsibility. According to Dennett, when we ask "Could the performer have done otherwise?" what we are actually concerned about is practical conclusions to be drawn from this event (Pareboom, 2001, p. 562). To illustrate, we want to predict whether or not the performer will do something similar again. I agree with Dennett on that. Common function of asking questions of *PAP* concerning determinacy is not finding out if somebody is genuinely responsible, but making practical conclusions for a more convenient future.

Another problem with the mainstream libertarian incompatibilist position reminds me of the case of Ms. Determined. Changing the case a little bit leads it to demonstrate *PAP* concerning determinacy's problems in being a criterion for responsibility. Ms. Determined is not a poker player in this case. She is a angel with perfect good will. She is so nice that she can never do anything bad. She can never stop being nice. If *PAP* is a necessary condition for responsibility, then she can never be praiseworthy for her good actions, since she could not have done otherwise. And as a person with good and bad sides, when I do something good, I deserve praise, thanks to my "ability" (!) (in my opinion, the right word is "possibility") to do otherwise, given that the complete set of true propositions belonging to the past is the same.

*PAP* concerning supervenience, which requires being able to do otherwise in exactly the same circumstances, is free of this problem, because neither in deterministic nor in indeterministic physicalistic worlds, can we satisfy it. I believe, neither me, nor Ms. Determined is praiseworthy for our actions. Even if I have good or bad sides, if I am controlled by the physical, I cannot do any action other than what I do, just like Ms. Determined cannot.

Therefore, I agree with Dennett in that libertarian incompatibilist interpretation of *PAP* has nothing to do with responsibility; however, I also believe that a proper *PAP* (*PAP* concerning supervenience) could be an acceptable criterion for genuine responsibility.

## 5.2. Practical Implications of Lack of Strong Free Will

Perhaps, some practical questions should be answered to clarify what is meant by our not having genuine responsibility for anything we do. It is a common intuition

that the reason why we should try to change our environment by acting is that we have control over our actions. If we do not have genuine control over our actions, i.e. if we are not genuinely free, should not we try to make our lives better?

First of all, our not having genuine power to determine what we do, does not imply that we have no will or ability to act voluntarily. We have weak free will which is determined by external causes which have active roles in the causal chain. Even if these roles are determined by previous states of the world and laws of physics, both of which seem to be out of our control, because of the regularities originating from laws of physics, what we do regularly corresponds to what we want to do when uncoerced volitional actions are concerned. So when one wants to raise her hand, she has good reason to believe that she will be able raise her hand. The important thing is that we have no control over what we want. I do not know whether or not we should try to change our lives, but I am sure that we *will* do so. We have evolved in this way. And our decisions which are caused externally may make our lives "*better*", or "*worse*" even if we are not *genuinely* responsible for them.

Another practical question possibility of our not being genuinely responsible reminds is about the role of reward and punishment in our lives: If we are not responsible for anything we do, should not we praise and blame people. Should not we reward and punish people for their actions? I.e. does our not having responsibility for anything we do threats the truth value (if there are such truth values, in the first place) of the moral judgments? I believe not only threats but also eliminates the possibility of the moral judgments. There is no reason to admire or blame people. Admiring and blaming wholeheartedly requires commitment to genuine responsibility, which I claim we do not have. On the other hand, there are reasons to reward or punish people to change structure of their actions, as we shall argue below. It seems to me that *expressing* that you admire

or blame persons are *only* useful as psychological rewards and punishments. Beyond that they are groundless, irrational and useless.

Now what happens to contemporary justice system? I think we should face a truth: We do not punish people since they deserve punishment. Instead, punishment serves for a pragmatic end: We punish "criminals" to change the structure of the actions in the way majority or some powerful persons deserve.[29] Thus, even if people do not deserve punishment, punishment has a function from the perspective of majority. We isolate "criminals" to stay away from the dangers. So our lack of genuine responsibility requires almost no change in our actual justice systems. We just should, I believe face its foundations and confess: Persons do not deserve reward or punishment, we just use reward and punishment for our, majority's and powerful persons' ends.

## 5.3. Conclusions

Honderich describes two ways of disapproving a vicious politician:

> […] [1] we may hold a vicious politician responsible, which is to say we disapprove of her morally for an action, where this particular disapproval involves a retributive desire–a desire to subject her to discomfiture or worse–and where the disapproval issues in certain behavior. [2] We may also disapprove of her morally for her action in another way. This attitude involves desires, but it does not involve a retributive desire, and it issues in distinct behavior. It follows that the attitudes have different contents, and in particular that the first takes the politician's action to be both voluntary and originated, and the second only takes it to be voluntary (Honderich, 1996, p. 856).

[29] *Cf.* Dennett, 1984a, p.562.

I believe, if we do not have genuine responsibility (which I defend), only the second disapproval is rational. If we do not carry any responsibility, there is no way to legitimately rationalize retributive or admirative desires. Only avoidance of bad consequences and pursuit of good ones is rational.

Pereboom lengthily discusses the implications of hard incompatibilism on the management of criminal behaviors. He defends that if hard incompatibilism is true, since nobody is responsible for any action, nobody deserves to be punished (or to be rewarded); and that is why retributive punishment is not a way of managing criminal behaviors which hard incompatibilism could rationalize. However, even if we are not responsible, there are legitimate punishments he claims: Punishments for the purpose of education is acceptable even if hard incompatibilism is true. Deterrence, isolation, rehabilitation is not more problematic in a world with lack of responsibility, than it is in a world with responsibility.[30] Therefore, being committed to that we are not genuinely responsible would not dramatically change *practices* of our justice system.

However, it would change foundation and principles of justice system dramatically: Instead of punishing people because we believe that they deserve punishment, we would punish people to reconstruct their personality in a way that is more adaptable to the society. Instead of holding people responsible which may causes *hatred*, we would rehabilitate, isolate them, which sounds more like *problem solving*.

It seems to me that human being are getting more and more convinced with these no-genuine-responsibility-positions, consciously or unconsciously, based on their reasoning or based on their evolving intuitions. That is why death penalties practiced, punishment of mentally disabled people, and retributive torture is

[30] For a more comprehensive discussion, Pareboom, 2001, pp. 158-86.

getting rarer, while rehabilitation, and isolation is getting more and more central in the contemporary justice system.

# CHAPTER VI

# CONCLUSIONS

One of the widely shared intuition is that indeterminism is more free-will-friendly than determinism. Even some compatibilists like Honderich take this proposition granted. In this thesis, I tried to weaken this intuition like Hume, Frankfurt, Dennett, Pareboom and G. Strawson do. Hume, Frankfurt and Dennett defend that weak free will is compatible with determinism, just like it is compatible with indeterminism. Pareboom and Strawson, on the other hand, defend that strong free will is incompatible with indeterminism, just like it is incompatible with determinism. I think both of these schools, "optimistic compatibilists" and "pessimistic incompatibilists" hold a tenable positions. And I try to combine these two positions in *NDP*. If physicalism is true, both determinism and indeterminism is compatible with weak free will and incompatible with strong free will. On the other hand, if physicalism is false, I believe both weak and strong free will may or may not exist, independent of the problem of determinism. If we have wills independent of everything external, it does not matter whether or not our wills follow a deterministic or indeterministic pattern, we are free in the strong sense. Therefore, a perfectly (perfectly determined by goodness) good angel is not less free than a fairly good person, or a perfect chess player (perfectly determined by

math and logic) is not less free than a rookie chess player independent of the problems of physicalism and determinism.

**Table 2:** Implications of this thesis about the relationship between freedom and physicalism.

| | Weak Free Will | Strong Free Will |
|---|---|---|
| Physicalistic World | Possible | Impossible |
| Non-Physicalistic World | Possible | Possible |

I believe, if physicalism is true, we are simply machines. We do not get mad at machines, and we do not admire them wholeheartedly. Instead, we use them for our purpose. I know that this sounds mean; however this is not something I prefer, this is something, I defend, we cannot escape from. To clarify, how should my most respectable professors take the *Acknowledgements* section of this thesis? I do not reject that we have a room for sympathy, feeling of admiration, and so on. What I claim is that these feelings do not have a rational ground like responsibility or freedom, which they are assumed to have. Personally, I feel lucky for having a bit of these rationally groundless, evolutionary helpful feelings.

The reason why P. F. Strawson (1962) proposes that contemporary problem of free will is over-intellectualized is its not corresponding to the daily experiences of feelings and *reactive attitudes*. We have feelings and reactive attitudes, which are shaped evolutionary and culturally. We do not have genuine responsibility and genuine freedom. Our appreciation towards "good" people, and resentment towards "bad" people are feelings and reactive attitudes, which does not have to have a rational ground. However, *responsibility* and *strong freedom* are two

concepts which can and should be rationally grounded. Therefore, I agree with Strawson in that we over-intellectualize philosophy of freedom by inventing concepts like *responsibility* and *freedom of will*, which perhaps does not have any referent in our world. Perhaps, we perceived our weak freedom (or volitional actions), over-intellectualized it and created these stronger concepts of freedom and genuine responsibility, which does not have any referent. So maybe we should just ignore these strong concepts.

We are machines in one sense; however machines are not as predictable as we tend to believe. Even if determinism is true, even if we know all the laws of physics, we cannot predict future of our own system. Contrary to Popper's claim, obviously unpredictability does not show that our world is not deterministic. Unpredictability does not prove or disprove that we have free will either. I think this is simply irrelevant to problem of compatibilism / incompatibilism. There are contradictions followed by existence of a Laplace's demon or a magnificent psychic and these contradictions, I defend, are irrelevant to problem of determinism and free will. For Laplace's demon, some of the problems are impossibility of registering complete data about oneself, and impossibility of calculating the future of a system including oneself. For magnificent psychics, on the other hand, I have no explanation. It seems to me that Chapter 3 is just a start for the enquiries about possibility of perfect foreknowledge. Much more complete, systematical and simpler explanations for the problematic thought experiments could be presented, I believe.

The relation between desires and will is another point which should not be missed. Since a will free from desires is not a possibility, if we still want to talk about freedom of will, we should take desires as something internal to will. A being acting against its strongest physical or mental desire would not seem us to be freer than a drunk person, who is having trouble even with walking straightforward.

# REFERENCES


Binder, P. M. (2008). Theories of almost everything. *Nature*, *455* (18), 884-5.

Canfield, J. (July 1961). Determinism, free will and the ace predictor. *Mind*, *70* (279), 412-6.

Dennett, D. C. (1984a). I could not have done otherwise--so what? *The Journal of Philosophy*, *81*(10), 553-65.

Dennett, D. C. (1984b). *Elbow room*. Cambridge: MIT.

Edwards, J. (2001). *Freedom of the will*. Tennessee: Lightning Source Inc.

Epictetus, (1888). *The enchiridion, or manual*. (G. Long, Trans.). Retrieved November 04 2010 from the World Wide Web: http://www.ptypes.com/enchiridion.html

Frankfurt, H. (1969). Alternate possibilities and moral responsibility. *The Journal of Philosophy*, *66* (23), 829-39.

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, *68* (1) 5-20.

Frankfurt, H. (1987). Identification and wholeheartedness. In F. Schoeman, (Eds.), *Responsibility, character, and the emotions: New essays in moral psychology* (pp. 27-45). New York: Cambridge.

Ginet, C. (1996). In defense of the principle of alternative possibilities: Why I don't find Frankfurt's argument convincing. *Philosophical Perspectives, 10,* 403-17.

Hobart, R. E. (1934). Free will as involving determinism and inconceivable without it. *Mind, 43* (169), 1-27.

Holbach, P. (1957). The illusion of free will. In P. Edwards & A. Pap (Eds.). *A modern introduction to philosophy: Readings from classical and contemporary sources* (pp. 317-326). New York: The Free Press of Glencoe.

Honderich, T. (1996). Compatibilism, incompatibilism, and the smart aleck. *Philosophy and Phenomenological Research, 56* (4), 855-62.

Hume, D. (1975). *An enquiry concerning human understanding*. L. A. Selby-Bigge, ed., third edition revised by P. H. Nidditch, New York: Oxford.

Kane, R. (1989). Two kinds of incompatibilism. *Philosophy and Phenomenological Research, 50* (2) 219-54.

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

Kane, R. (1999). Responsibility, luck, and chance: Reflections on free will and indeterminism. *The Journal of Philosophy, 96* (5), 217-40.

Kane, R. (2007). Libertarianism. In J. M. Fischer, R. Kane, D. Pereboom, and M. Vargas (Eds.), *Four views of free will* (pp. 5-43). MA: Blackwell Publishing.

Kant, I. (1949). *Fundamental principles of the metaphysics of morals*. (T. K. Abbott, Trans.). New York: The Bobbs-Merrill Company, Inc.

Kant, I. (1998). *Critique of pure reason*. The Cambridge Edition, (P. Guyer & A. W. Wood, Trans.). New York: Cambridge University Press.

Laplace P. S. (1902). *A philosophical essay on probabilities*. (F. W. Truscott & F. L. Emory, Trans.). New York: John Wiley & Sons; London: Chapman & Hall Limited.

Libet B. (1999). Do we have free will? *Journal of Consciousness Studies*, *6* (8-9), 47-57.

Lyon, A. (1959). The prediction paradox. *Mind*, *68* (272), 510-7.

McIntyre, A. (1994). Compatibilists could have done otherwise: Responsibility and negative agency. *The Philosophical Review*, *103* (3), 453-88.

McKay, D. M. (1960). On the logical indeterminacy of a free choice. *Mind*, *69* (273), 31-40.

Mele A. R. & Robb, D. (1998). Rescuing Frankfurt-style cases. *The Philosophical Review*, *107* (1), 97-112.

Moore, G. E. (1912). *Ethics*. Oxford: Oxford University Press.

Pereboom, D. (1995). Determinism al dente. *Noûs*, *29* (1), 21-45.

Pereboom, D. (2001). *Living without free will*. New York: Cambridge University Press.

Popper, K. R. (1950a). Indeterminism in quantum physics and in classical physics. Part I. *The British Journal for the Philosophy of Science*, *1* (2), 117-33.

Popper, K. R. (1950b). Indeterminism in quantum physics and in classical physics. Part II. *The British Journal for the Philosophy of Science, 1* (3), 173-95. P. 179.

Richards, J. R. (2000). *Human nature after Darwin: A philosophical introduction*. London: Routledge.

Schlesinger, G. (1974). The unpredictability of free choices. *The British Journal for the Philosophy of Science, 25* (3), 209-21.

Smedes, T. A. (2003). Is our universe deterministic? Philosophical and theological reflections on an elusive topic. *Zygon, 38* (4), 955-79.

Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy , 48*, 1-25.

Strawson, G. (2000). The unhelpfulness of indeterminism. *Philosophy and Phenomenological Research, 60* (1), 149-55.

Stump E. (1996). Libertarian freedom and the principle of alternative possibilities. In D. Howard-Snyder and J. Jordan (Eds.), *Faith, freedom and rationality* (pp. 73-88). Totowa, N.J.: Rowman and Littlefield.

Van Inwagen, Peter. (1978). Ability and responsibility. *The Philosophical Review, 87* (2), 201-24.

Van Inwagen, P. (1983). *An essay on free will*. Oxford: Oxford University Press, Clarendon Press.

Van Inwagen, P. (1989). When is the will free? *Philosophical Perspectives, 3,* 399-422.

Watson, G. (1975). Free agency. *The Journal of Philosophy, 72* (8) 205-20.

Watson, G. (1987). Free action and free will. *Mind, 96* (382), 145-72.

Widerker, D. (1995). Libertarianism and Frankfurt's attack on the principle of alternative possibilities. *The Philosophical Review*, *104* (2), 247-61.

Widerker, D. (2000). Frankfurt's attack on the principle of alternative possibilities: A further look. *Philosophical Perspectives, 14*, 181-201.

Wolpert, D. (2008). Physical limits of inference. *Physica D, 237*, 1257-81.

# APPENDIX A

## TEZ FOTOKOPİSİ İZİN FORMU

**ENSTİTÜ**

Fen Bilimleri Enstitüsü ☐

Sosyal Bilimler Enstitüsü ☐

Uygulamalı Matematik Enstitüsü ☐

Enformatik Enstitüsü ☐

Deniz Bilimleri Enstitüsü ☐

**YAZARIN**

Soyadı : Çağatay
Adı : Hasan
Bölümü : Felsefe Bölümü

**TEZİN ADI**: Free Will and Determinism: Are They Even Relevant To Each Other?

**TEZİN TÜRÜ** : Yüksek Lisans ☐ Doktora ☐

1. Tezimin tamamından kaynak gösterilmek şartıyla fotokopi alınabilir. ☐

2. Tezimin içindekiler sayfası, özet, indeks sayfalarından ve/veya bir bölümünden kaynak gösterilmek şartıyla fotokopi alınabilir. ☐

3. Tezimden bir bir (1) yıl süreyle fotokopi alınamaz. ☐

**TEZİN KÜTÜPHANEYE TESLİM TARİHİ**:

# APPENDIX B

## CURRICULUM VITAE

### PERSONAL INFORMATION

Surname, Name: Çağatay, Hasan
Nationality: Turkish (TC)
Date and Place of Birth: 25 July 1981 , Malatya
Marital Status: Single
Phone: +90 554 497 50 68
email: e111690@metu.edu.tr

### EDUCATION

| Degree | Institution | Year of Graduation |
|---|---|---|
| PHD | METU, Philosophy | 2012 |
| MA | METU, Philosophy | 2007 |
| BS | METU, Mathematics Edu. | 2002 |

### WORK EXPERIENCE

| Year | Place | Enrollment |
|---|---|---|
| 2007- 2009 | SEBIT | Educational Designer |
| 2002-2003 | MEB Beypazarı Lycee | Mathematics Teacher |

### FOREIGN LANGUAGES

Advanced English

### PUBLICATIONS

1. Çağatay H. and Ekemen C. "Mary'nin Odası ve Fizikalizm", Felsefe Tartışmaları, 47(47), 225-228 (2012)

### HOBBIES

Writing, Chess, Movies, Coding.

# APPENDIX C

## TURKISH SUMMARY

## ÖZGÜR İRADE VE DETERMİNİZM: BİRBİRLERİYLE GERÇEKTEN İLGİLİLER Mİ?

### İki Tür Özgür İrade Kavramı

*İrade* insanın eylemlerini belirlemesine yarayan bir çeşit güçtür. Bu belirlemeye "seçim yapmak" diyoruz. Seçim yapmak kavramı pozitif ve negatif istekleri (istek ve rahatsızlık) varsayıyor. Eğer isteklerimiz olmasaydı, dünya umurumuzda olmazdı ve seçim yapmak için hiçbir itki ve sebebimiz olmazdı. Öyleyse, irade istekleri kavramsal olarak varsaymaktadır. İsteklerimiz olmasaydı, organik bilgisayarlar veya robotlardan farksız olurduk ve özgürlükten de bahsedemezdik. İstekler olmadan iradeden bahsedemeyiz; ancak isteklerimizin olması irademizin olduğunu tek başına göstermez: İrademiz olduğunu söyleyebilmemiz için, yapmak istediğimiz şeyleri yapabilmemizi sağlayacak, ya da en azından yapmayı denememizi sağlayacak bir güce ihtiyacımız var. İsteklerimizi sıklıkla gerçekleştirebildiğimiz için, görünüşte irade sahibi olduğumuzda hemfikir olmak

eğilimindeyiz. Ayağa kalkmak istiyoruz ve bu isteği ayağa kalkarak tatmin ediyoruz. Elimizi kaldırmak istiyoruz ve bunu kolaylıkla gerçekleştiriyoruz.

Aynı zamanda sahip olduğumuza inandığımız bu iradenin *özgür* olduğunu da düşünüyoruz. İrademizin özgür olmasıyla, genellikle onu kendi kendimize, dışarıdan müdahale olmaksızın belirlediğimizi kastediyoruz. İrademizin özgür olduğu yönünde genel bir sezgimiz varsa da, filozoflar bu sezgiyi sorgulamadan kabul etmiyor. Dünyadaki nedensellik zincirinin tamamen ya da kısmen deterministik olduğu düşüncesi, genellikle özgür irademize bir tehdit olarak algılanıyor. Determinizmin farklı türleri olsa da, bu çalışmada genel olarak *nedensel determinizm*, yani dünyanın tümüyle nedensellik kurallarıyla belirlendiğini iddia eden doktrine odaklanılacak. Eğer nedensel determinizm doğruysa, tüm olgular doğanın değişmez ve şaşmaz olan kurallarına bağımlı olmalı. Bu durumda, nedensel determinizm doğruysa, herhangi bir $t_1$ anındaki eksiksiz doğru önerme kümesi (girdi) için, daha sonraki herhangi bir $t_2$ anı için sadece bir tane eksiksiz doğru önerme kümesi (çıktı) olasıdır. Deterministlere göre, bir piyango bileti aldığımda, çekiliş yapıldıktan sonra ikramiyeyi kazanıp kazanamayacağım bellidir. Nedensellik kuralları çoklu olanaklara izin vermeyecek biçimde olduğundan, çekilişi yapacak kişinin seçimleri bile bu kurallarla tek olanaklı biçimde önceden belirlenmiştir.

Determinizm doğruysa, alternatif olanaklılık açısından, gelecek geçmişten farklı değildir. Her ikisi de belirlenmiş ve değiştirilemezdir. Bir başka deyişle, determinizm doğruysa, herhangi bir anda olmuş ya da olacak olan bir olay, olduğundan farklı biçimde olamazdı ya da olamaz. İlk bakışta, deterministlere göre gerçek anlamda olasılıktan ya da gerçek anlamda özgür iradeden bahsetmek mümkün değil gibi görünüyor. Determinizm doğruysa, herhangi bir *t* anında bir kişinin ya da bir elektronun nasıl davranacağını bilmiyor olsak bile, önceden belirlenmiş bir tek biçimde davranmak zorunda olması gerekir. Bundan herhangi

bir $t_1$ anına ait eksiksiz ampirik bilgiyle, nedensellik kurallarını birlikte kullanarak herhangi bir $t_2$ ana ait eksiksiz ampirik datayı çıkarsayabileceğimiz sonucuna ulaşılabilir. Bu şematik olarak aşağıdaki gibi ifade edilebilir:

$Et_1$ ($t_1$ anına ait eksiksiz ampirik bilgi)

*L* (Doğa kanunları)

---

$Et_2$ ($t_2$ anına ait eksiksiz ampirik bilgi)

Özgür iradeye dair en temel sorulardan biri işte burada akla geliyor: Eğer bizim seçimlerimiz ve diğer tüm olgular daha doğmadığımız bir zamana ait olgulardan ve nedensellik kuralları tarafından önceden belirlenmişse, irademiz ya da dünyamızdaki herhangi bir şey nasıl özgür olabilir? Ve doğa kanunlarından ya da doğmadığımız bir zamandaki olgulardan sorumlu olmadığımıza göre, bu iki faktör tarafından belirlenmiş olan davranışlarımızdan nasıl sorumlu olabiliriz?

Bazı düşünürler, çeşitli argümanlara dayanarak, determinizm doğruysa özgür olamayacağımızı iddia ederken, bazı diğerleri de bu argümanların yanlış olduğunu savunuyor: Özgür iradenin determinizm doktrininin doğruluk değerine bağımlılığı meselesine ilişkin dört temel görüş var: (1) Öncelikle, Derk Pereboom (1995), Galen Strawson (2000), ve Janet Radcliffe Richards (2000) gibi *ikili uyuşmazcılar* (*hard incompatibilists*) ya da *kötümser uyuşmazcılar* (*pessimistic incompatibilists*) deteminizm doğru da olsa yanlış da, özgür irademiz olamayacağını düşünüyorlar. (2) Peter Van Inwagen (1983, s. 93-105; 1989, s. 404-405), Robert Kane (1989, 1996, 1999) ve Paul Holbach'in (1957) aralarında bulunduğu diğer uyuşmazcılar, *özgürlükçü uyuşmazcılar* (*libertarian incompatibilists*), *agnostik uyuşmazcılar*[31]

[31] Determinizmin doğru olup olmadığını konusunda taraf seçmeyip, özgür iradeyle uyuşmadığını savunanlar.

ve *uyuşmazcı deterministler (hard determinists)*, özellikle determinizmin, indeterminizmden ayırıcı niteliklerinden dolayı özgür iradeyle uyuşmaz olduğunu savunan düşünürler. (3) Eğer özgürsek bunun determinizmin doğru olduğunu gösterdiğini savunan bir cephe de bulunmaktadır (Hobart, 1934). Onlar özgür irade ve indeterminizmin uyumlu olmadığını düşündüklerinden, onları "*kötümser uyuşurcular*" diye adlandırmak uygun olabilir. (4) Harry Frankfurt (1969)[32] ve Daniel C. Dennett'ın (1984a, 1984b) aralarında bulunduğu bir diğer cephe de, özgür irade ve determinizmin uyumlu olduğunu savunmaktadır. Bu felsefecilerin öncekilerden farkı olarak, özgür iradenin indeterminizmle de uyumlu olduğunu düşünmeleridir. Bu nedenle bu ekol de "*iyimser uyuşurcular*" diye adlandırılabilir. Her ne kadar kötümser uyuşmazcılar ve iyimser uyuşurcular tartışmanın iki kutbunda konumlanmış gibi görünseler de çok önemli bir ortak tezleri bulunmaktadır: Bu iki ekol de determinizm / indeterminizm problemiyle, özgür irade probleminin birbirine (en azından diğer felsefecilerin iddia ettiği türden) bağımlı olmadığını savunmaya yatkındırlar.

Ben de, tıpkı iyimser uyuşurcular ve kötümser uyuşmazcılar gibi, özgür irademizin olup olmadığı sorusunun, dünyamızın deterministik olup olmadığı sorusuna bağımlı olmadığını düşünüyorum. Her iki ekolün argümanlarında da büyük oranda katılıyorum. Özgür irade probleminin determinizmin doğruluk değerinden bağımsız olduğunu kabul ettikleri sürece, özgür olduğumuzu destekleyen ve olmadığımıza işaret eden iki güçlü sezginin de temeli olduğunu düşünüyorum. Bu görünüşte birbirine karşı olması beklenen sezgiler çelişmek zorunda değil; çünkü özgür irade tartışmalarında iki tür özgür irade kavramına referans veriliyor: Geleneksel olarak "*uyuşurcu özgür irade*" ("*compatibilist free will*") ve "*uyuşmazcı özgür irade*" ("*incompatibilist free will*") diye adlandırılan

[32] Frankfurt (1971, 1987) ayrıca, yüksek-düzer-istekler (higher-order desires) kavramını temel alarak yeni bir özgür irade kavramı da tanımlamaktadır.

bu kavramlara bu çalışma boyunca, sırasıyla "*güçlü özgür irade*" ve "*zayıf özgür irade*" diye anacağım.

Ted Honderich (1996) de "aynı önemde iki özgürlük kavramımız var" diyor ve bu nedenle hem uyuşurcuların hem de uyuşmazcıların yanıldığını savunuyor. Ona göre, sadece istemlilik şartını sağlayan eylemlerimiz zayıf anlamda özgür ve istemlilik determinizmle tamamen uyumlu. Diğer zor sağlanır özgürlük kavramını tanımlamak içinse *kişiden kaynaklanma* (*origination*) kavramını tanımlıyor: "Bir eylem, eğer bir nedensellik zincirinin sonucu değil de, kişinin kontrolündeyse, bu eylem kişiden kaynaklıdır" (Honderich,1996, s. 855). Bir eylemin güçlü anlamda özgür olması için, hem istemli, hem de kişiden kaynaklı olması gerekir ve Honderich'e göre, bu şart determinizm doğru ise sağlanamaz (Honderich,1996, s. 856).

Bana göre iki kavram arasında çizilmiş bu ayrım akla yakın ve bu iki kavramdan sadece birisinin determinizmle uyumlu olduğu da doğru. Ancak Honderich determinizmin (*sadece* determinizmin) güçlü özgür iradeyle uyumsuz olduğuna vurgu yaparak, indeterminizmin güçlü özgür iradeyle uyumlu olabileceğini ima ediyor. İşte bu noktada onunla hemfikir değilim. Çalışmanın ilerleyen kısımlarında neden Honderich'le aynı fikirde olmadığımı açılayacağım; ancak öncelikle zayıf ve güçlü özgür iradeyi biraz daha netleştirmek istiyorum:

Zayıf özgür irade kavramları daha kolay anlaşılır ve daha kolay test edilir kavramlardır. Hume'un özgür irade kavramı bu türün en temel örneklerinden biri sayılabilir. Hume göre, eğer isteklerimizin nedensellik zincirinde yer bulduğu istemli eylemler gerçekleştiriyorsak, özgürüz. Bir başka deyişle, *eğer istediğimizde davrandığımızdan farklı davranabiliyorsak, zayıf özgür irademiz vardır*. Zayıf özgürlük hakkında tartışılması ve istemli davranışlara atfedilmesi kolay bir kavram.

Öte yandan, güçlü özgür irade kavramı, kaygan, kavranması zor. Sadece görünüşte istemli olan davranışlarla temellendirilmesi mümkün değil. Güçlü özgür irademizin olduğunu söyleyebilmek için eylemlerin Kane'in (1989, 1996, 1999) kastettiği anlamda *nihai sorumlusu* (*ultimately responsible*) olmalıyız ya da bu eylemler *kişiden kaynaklı* olmalı (Strawson, 2000). *Tam olarak aynı koşullar altında, birden çok eylem yapabilme olanağını taşıyor* olmalıyız.

Bu çalışmanın pozisyonu bu iki kavramla yakından ilgili: Savunduğum şey, determinizm doktrininin doğruluğundan bağımsız olarak, zayıf özgür iradeye sahipken, güçlü özgür iradeye sahip olmadığımızdır. Filozoflar zayıf özgür irademizin varlığında hemfikir gibi görünüyor. Bu nedenle özgür irade problemini, *güçlü* özgür irade problemi olarak ele alacağım. Çalışmanın devamında aksini belirtilmediği sürece "özgür irade" ifadesini "güçlü özgür irade" anlamında kullanacağım.

"Özgür irade" "güçlü özgür irade" olarak ele alındığında, bu çalışmanın pozisyonu kötümser uyuşmazcılığa oldukça yakın hale getiriyor. Ancak tıpkı güçlü özgür irade gibi, zayıf özgür iradenin de determinizminin doğruluk değerinden bağımsız olduğunu iddia ettiğim için, pozisyonumu "*ilgisizci pozisyon*" diye isimlendirmek daha doğru olabilir.

## Kukla Olmak

Bu çalışmanın kritik bir monist varsayımı var: Birçok çağdaş felsefecinin özgür irade problemiyle ilgilenirken yaptığı gibi, ben de fizikalizmin doğru olduğunu

varsayacağım. Eğer fizikalizm doğruysa, dünyamız fiziksel özellikleri *takip etmektedir* (onlardan doğmaktadır, onlara dayalıdır, *supervene upon*). Dünyamızdaki her şey fiziksel varlıklara bağımlıdır. Bu biraz kötümser bir varsayımmış gibi görülebilir; çünkü bu durumda biz fizik kanunlarına göre davranan fiziksel kanunlarının yönetimindeki kompleks kuklalarız. Bizler hisleri, iradeleri, düşünceleri, vb. olan kompleks kuklalarız ve fiziksel varlıklar da kuklacı. İsteklerimizle uyumlu amaçlarımıza ulaştığımız için, biz kuklalar, "özgür" irademiz olduğu inancını oluşturuyoruz.

Öncelikle, kuklacının (fiziksel dünya) deterministik davrandığını düşünelim. Bizi her $t_i$ anındaki davranış, duygu ve düşüncelerimiz için $t_{i+1}$ anındaki davranış, duygu ve düşüncelerimizi kesin biçimde bilineceği şekilde, bir deterministik kurallar bütününe göre hareket ettiriyor. Şimdi, bu kuklalar (bizler) ne kadar kompleks ve *görünüşte* özgür olursa olsun, özgür irade sahibi olduklarını iddia etmek sezgilerimizle uyumsuz değil mi? İrade ve davranışları kuklacı tarafından kesin biçimde belirleniyor. İradelerinin bir kuklacının ipleriyle alternatife yer vermeyecek biçimde yönlendirildiğini bile bile, onlara "özgür" demeli miyiz? Benim sezgilerime göre, fizikalistik deterministik bir dünyada yaşıyorsak, özgür olamayız. Özgür olduğumuz inancını geliştirebilir, ya da isteklerimizle uyumlu amaçlarımıza ulaşabiliriz; ama irademiz ve davranışlarımız fiziksel nesnelere tamamen bağımlı olduğundan, davranışlarımıza özgür diyemeyiz, "özgür"ün anlamını değiştirmeden.

Muhtemelen bazı uyuşurcular benimle hemfikir olmayacak. Ancak bu ampirik gerçeklerle ilgili uyuşmuyor oluşumuzdan kaynaklanmıyor. Benimle hemfikir olmayabilirler; çünkü onların akıllarındaki zayıf özgür irade kavramı, benim şu an üzerine düşündüğüm güçlü özgür irade kavramından daha kucaklayıcı. Söz gelimi şöyle diyebilirlerdi: "Her ne olursa olsun, biz kuklaların, arzu, irade gibi kompleks yapıları olduğunda aynı görüşteyiz. Bu kompleks yapılar fiziksel varlıklar

tarafından kesin biçimde belirleniyorsa bile, onlarla uyumlu amaçlara ulaşabiliyoruz. Bu nedenle biz kuklalar özgür sayılmalıyız. (*Zayıf*) Özgür irade kavramı fiziksel dünyadan (kuklacı) bağımsız olmayı gerektirmiyor."

Fizikalistik bir dünyada yaşıyor olmamızın bizi fiziksel dünyanın kontrolünde kuklalar yaptığında anlaştığımız sürece, kendimizi "özgür" addedip addetmeyeceğimizin öncelikli bir tartışma konusu olduğunu düşünmüyorum. Bırakalım uyuşurcular akıllarında zayıf özgür irade kavramlarıyla özgür olmadığımızı iddia etsinler, bense aklımda güçlü özgür irade kavramlarıyla özgür olduğumuzu iddia edeyim.

Deterministik fizikalistik dünyadan sonra, şimdi de diğer fizikalistik olasılık, dünyanın indeterministik ve fizikalistik olduğu olasılığı üzerine tartışalım: Bu kez kuklacı (fiziksel dünya) kuklaları (bizi) indeterministik bir biçimde kontrol ediyor. Şimdi, kuklacı bizi ya rastgele, ya da olasılık içeren kurallarla (fizik kanunları) yönetiyor. Eğer durum buysa, prensipte kukla gösterisinin seyircileri, kuklaların daha önceki davranışlarını gözlemleyerek, daha sonraki davranışlarını çıkarsayamaz. Bu bir şeyi değiştirir mi? Tahmin edilemez olmamız bizi daha özgür yapar mı? Bu indeterministik kuklalar özgür olabilir mi? Benim bu soruya cevabım hayır; çünkü indeterministik fizikalistik dünyada, fizik kuralları olasılık içerse de, fiziksel dünyanın bizi zorladığından farklı davranamayız.

Muhtemelen bu iddia bir açıklama gerektiriyor: Kararsız bir sürücünün kırmızı ışıkta durup durmayacağı, sinir sistemindeki tek bir elektronun durumuna bağımlı olabilir. Söz gelimi, sürücünün sinir sistemindeki elektron *üst spin durumdaysa* (*spin up state*) sürücü kırmızıda duracak, elektron *alt spin durumundaysa* (*spin down state*) da sürücü kırmızıda geçecek olsun. Durumu basitleştirmek için elektronun üzerinde herhangi bir elektromanyetik etki olmadığını ve alt spin durumda ve üst spin durumda olma olasılığının 1/2 olduğunu varsayalım. Şimdi bu elektronun üst spin durumunda mı alt spin durumunda mı olacağı belirsiz

olabilir; ancak sürücünün elektronun üst spin durumunda olması halinde duracağı kesin. Bu durumda, mükemmel iplerle kuklacı tarafından yönetilen bu kuklanın, sadece kuklacı nasıl hareket edeceğine olasılık içeren kurallara göre karar veriyor diye özgür demek mümkün mü? Bir kukla hangi eylemi seçeceğinde özgür olabilir mi? Kuklacı tarafından empoze edilmiş bu belirsizlik (indeterministik olaylar), kuklayı özgür yapmaz; çünkü bu belirsizliklerde içerilen olasılıklar kuklanın kontrolünde değildir. Aynı şekilde fiziksel varlıklar tarafından empoze edilmiş belirsiz davranışlar, bizi özgür yapmaz; çünkü bu belirsizliklerde içerilen olasılıklar bizim kontrolünde değildir.

Eğer kuklacının kontrolündeki kuklalar ve fiziksel dünyanın kontrolündeki kişiler analojisi problem içermiyorsa, fizikalizmin bizim özgür olmamıza iznin vermediği (ya da vermeyeceği) sonucuna ulaşmak kaçınılmazmış gibi görünüyor. Bir başka deyişle, eğer hiçbir kukla özgür sayılamayacaksa, fizikalistik bir dünyadaki hiçbir eylem ya da seçim de özgür sayılamaz. Daha önce de belirttiğim gibi, farklı (*zayıf*) bir özgür irade kavramlaştırılmasıyla bunun aksine de ulaşılabilir: İster deterministik ister indeterministik bir dünyada yaşıyor olalım, nasıl davranmak istediğimizle nasıl davrandığımız arasındaki pozitif ilişki bir dereceye kadar korunduğu sürece özgürüzdür. Burada asıl önemli olan, ister güçlü ister zayıf bir özgür irade kavramından bahsedelim, özgürlüğümüzün determinizm / indeterminizm problemine bağımlı olmaması. Analojinin temeli kuklaların dışsal bir şeye bağımlı olması; kuklaların özelliklerinin, kuklacının özelliklerine tıpkı bizim özelliklerimizin fiziksel özellikle takip ettiği gibi takip etmesi.

Analojilerden ve düşünce deneylerinden uzaklaşmak adına ilgisizci pozisyonun ardındaki argüman aşağıdaki adımlarla ortaya konulabilir:

(1) Eğer fizikalizm doğru ise, her özellik fiziksel özelliklere takip eder. (Fizikalizmin tanımından)
(2) Eğer fizikalizm doğru ise, her özellik, fiziksel özelliklere tamamen bağımlıdır. (1'den)
(3) Eğer fizikalizm doğru ise, irade ve eylemlerimizin tüm özellikleri, fiziksel özelliklere tümüyle bağımlıdır. (2'den)
(4) Eğer *A B*'ye tümüyle bağımlıysa, *A* özgür olamaz. (Öncül)
(5) Eğer fizikalizm doğru ise, (determinizm doğru da olsa yanlış da) iradelerimiz özgür değildir. (3 ve 4'ten)
(6) Eğer fizikalizm doğru ise, özgür irade problemi, determinizm / indeterminizm problemine bağımlı değildir. (5'ten)

## İki Tür Alternatif Olanaklar Prensibi (Principle of Alternative Possibilities)

Özgür irade üzerine çalışan birçok felsefeci, *özgür irade* ve *sorumluluk* kavramlarını birbirinin üzerine inşa ediyorlar. Bu sezgilerimize uygun: Öyle ya özgür olmadığımız bir eylemden sorumlu sayılamayız. Çoğu uyuşmazcıya göre bir davranışımızdan sorumlu olabilmek için, bu davranışı göstermeme olanağımızın da bulunması gerekir(Van Inwagen, 1978; Ginet, 1996; Kane, 1989, 1996, 1999). Aynı şekilde bu prensibe inanan düşünürlere göre, bir davranışımızda özgür olmamız için de, o davranışı göstermeme olanağımızın bulunması gerekir. Öyleyse, bir hırsızın araba çalmaktan sorumlu ve araba çalarken özgür olabilmesi için, o arabayı çalmama olanağının da elinde bulunması gerekiyor. Ya da bir kahramanın dünyayı kurtarmakta özgür ve kurtardığı için takdire şayan olması için, dünyayı kurtarmama olanağını da elinde bulundurması

gerekiyor. Bu türden prensipler genel olarak *alternatif olanaklar prensipleri* (*AOP*) olarak anılıyor.

Frankfurt ve Dennett'ın aralarında *bulunmadığı* çoğu düşünüre göre, özgürlük *başka türlü davranabilme olanağını* gerektiriyor; ancak genellikle "başka türlü davranabilme olanağı" ile ne kastedildiği net değil. Literatürde birbirinden farklı *AOP*'lerle karşılaşılıyor: Bir düşünür *AOP* ile (1) "*geçmişteki geçmişin aynı olmasına rağmen*, birden fazla eylemin gerçekleştirilebilmesini" (determinizmle ilişkili *AOP* [*DİAOP*]), ya da (2) "*tam olarak aynı durumda* birden fazla eylemin gerçekleştirilebilmesini" (*fizikalizmle ilişkili AOP* [*FİAOP*]) kastediyor olabilir. Bu iki *AOP* birbirine çok benziyor görünse de, bizi birbirlerinden dramatik olarak ayrılan sonuçlara götürebileceklerinden, birbiri yerine kullanılmamalılardır.

Kararsız sürücü ile ilgili durumda, *geçmiş aynı olsa* bile sürücü birden fazla davranış gösterebilirdi (kırmızıda durmak, ya da geçmek); çünkü sürücünün sinir sistemindeki elektronun durumu kararından önce belirlenmiş değildi. Öte yandan, elektronun son durumunu da sabitleyen *tam olarak aynı durumda* olma koşulunda, sürücünün birbirinden farklı davranışları gösterme şansı yoktur. Bana göre determinizmle ilgili olan *DİAOP* özgür irade veya sorumluluğun gerekli ya da yeterli şartı değil. Öte yandan, fizikalizmi ilgilendiren *FİAOP*, güçlü özgür iradenin ve gerçek anlamda sorumluluğun gerekli şartıdır.

Bu durumda, yukarıdaki kavramlaştırmaya göre, dünyamızın indeterministik ve fizikalistik olduğu varsayımıyla, tam olarak aynı durumda alternatif olanaklarımız yoktur; ancak geçmişin tam olarak aynı olması durumunda, alternatif eylemler gerçekleştirebilmemiz mümkündür. Mümkündür, ama bizim elimizde değildir; çünkü edimlerimizi kesin biçimde belirleyen indeterministik fiziksel olgular üzerinde hiçbir kontrolümüz bulunmamaktadır. *DİAOP*'yi sağlamamız bize edimlerimiz üzerinde hiçbir kontrol vermediğinden de, *DİAOP*'nin bize sorumluluk için gerekli özgürlüğü vermediğini savunuyorum.

Öte yandan, birçok uyuşmazcı, *DİAOP*'yi deterministik ya da indeterministik dünyalarda özgür olup olamayacağımıza karar vermek için kullanıyor: *An Essay on Free Will*'de, Van Inwagen, özgür iradenin determinizmle çeliştiğini ortaya koymak için üç argüman ortaya koyuyor. Üçüncü argüman, *sonuç argümanı* (*the consequence argument*) geçmişin aynılığına dayanan *DİAOP*'yi temel alıyor:

Önce önermelere bağlanan "N" operatörünü tanımlıyor. "N *p*.", "*p* doğrudur ve kimsenin hiçbir zaman *p*'nin doğruluk değeri konusunda bir seçme şansı olmamıştır." anlamına geliyor. Daha sonra sezgisel olarak çekici olan ($\alpha$) Kuralı ve ($\beta$) Kuralının doğru olduğunu iddia ediyor (Van Inwagen, 1983, s. 93-94).

($\alpha$) Kuralı $\Box p \vdash Np$

($\beta$) Kuralı $N(p \supset q), Np \vdash Nq$.

Bu tanım ve öncüllere dayanan argümanı 7 adımdan oluşuyor:

$P_0$: Dünyanın eski bir zamandaki durumu.

L: Tüm doğa kanunlarının birlikte (conjunction) doğruluğu.

Determinizm doğru varsayımını aşağıdaki adımlar takip ediyor:

(1) $\Box (P_0 \,\&\, L \supset P)$

Doğrudur. (1)den kipsel mantıkla (modal logic), şu çıkarsanabilir:

(2) $\Box (P_0 \supset (L \supset P))$

($\alpha$) kuralını (2)ye uyguladığımızda aşağıdaki ifadeyi elde ediyoruz:

(3) N ($P_0 \supset$ (L $\supset$ P)).

Şimdi öncüllerimiz:

(4) N $P_0$.

(3), (4) ve ($\beta$) kuralından:

(5) N (L $\supset$ P).

İkinci öncül:

(6) N L.

(5) ve (6)'ya ($\beta$) kuralını uygularsak:

(7) N P (Van Inwagen, 1983, s. 94-95).

($\alpha$) kuralının doğruluğundan şüphem var. ($\alpha$) kuralına göre, eğer *p gerekli olarak doğru* (*necessarily true*) ise, hiç kimsenin hiçbir zamanda *p*'nin doğruluk değeri hakkında bir seçme şansı olmamıştır. Öte yandan, ya *p*'nin doğruluk değeri, *p*'yi zorunlu olarak doğru kılacak biri tarafından seçilmişse? Böyle durumlar mantıkça mümkündür: Tanrı zorunlu olarak iyiyi seçer, ya da determinizm doğruysa herkes fizik kanunlarının gerektirdiğini "*seçer*".

Bana göre Van Inwagen'i sonuç argümanına iten sezgi *DİAOP*'den temel alıyor ve bu *AOP* özgür iradenin kriteri olmak için uygun değil. Sonuç argümanının daha az problemli bir versiyonunun aşağıdaki gibi kurulabileceğini düşünüyorum:

(1) Eğer determinizm doğruysa, herhangi bir zamana ait eksiksiz doğru önermeler küsmesi için, daha sonraki herhangi bir zamana ait sadece bir eksiksiz doğru önermeler kümesi vardır. (Determinizmin tanımından)
(2) Geçmişteki her zaman noktası için sadece bir eksiksiz doğru önermeler kümesi vardır. (Öncül)
(3) Öyle bir zaman vardı ki hiçbir canlı yoktu. (Bilimsel olarak iyi desteklenmiş öncül)
(4) Eğer determinizm doğruysa, canlıların yaşadığı her zaman noktası için sadece bir tane eksiksiz doğru önermeler kümesi vardır. (1, 2 ve 3'ten)
(5) Eğer determinizm doğru ise (*DİAOP*'ye göre) yaptığımızın aksini yapamayız. (4'ten)

Bu türden argümanlar sadece *DİAOP*'nin sağlanması için indeterminizmin doğru olması gerektiğini gösteriyor. Geçmiş aynı olduğu sürece, yaptığımızdan başka şey yapabilme ihtimalimizin bulunması için indeterminizmin doğru olması gerektiğinden öte bir şey göstermiyor. Ve bu sadece bizim için değil, cansız objeler için de doğru: Atılan bir zarın aldığı değerden başka bir değer alabilmesi için indeterminizmin doğru olması gerekiyor. Öte yandan, bu bizi bir zarın indeterministik bir dünyada özgür olabileceği düşüncesine yönlendirmiyor. *AOP*'yi sağlayan zarların özgür olmadığını düşünmemizin sebebi zarın hangi değeri göstereceğinin kontrolünü elinde tutmaması. Peki, fizikalistik bir dünyada, biz davranışlarımızı belirleyen fiziksel varlıklar üzerinde kontrole sahip miyiz? Eğer fizikalizm doğru ise böyle bir kontrol gücümüz yoktur gibi görünüyor. Aksine, biz fizik kuralları tarafından yönetiliyoruz. Ve indeterminizm doğru olsa bile, fizikalistik bir dünyada yaşadığımız sürece, bizi yöneten fiziksel varlıklar üzerinde kontrolümüz yok. Kısacası, bana göre *DİAOP*'ye dayanan hiçbir argüman, uyuşmazcılığı uyuşurculuğa karşı destekleyen hiçbir önermeyi ispatlamıyor.

## Bayan Belirlenmiş

*DİAOP*'nin bazı başka problemleri de var: Bayan Belirlenmiş matematiği çok iyi olan ve hafızası poker oynamaya mükemmel biçimde uyumlu olan bir poker yıldızı olsun. Pokerde sıra ne zaman kendisine gelse, oyundaki bütün ilgili bilgiyi kullanarak getirisi en iyi olan kararı veriyor.

Siz onun bir turnuvasını izlerken, yanınıza bir uyuşmazcı geliyor ve size Bayan Belirlenmiş'in özgür iradesinin olmadığını söylüyor.

Siz de uyuşmazcıya neden böyle düşündüğünü soruyorsunuz.

"Onu uzun süredir izliyorum." diye cevap veriyor uyuşmazcı. "Ne zaman eline iki as gelse, elinin gücünü gizleyerek, bahsi arttırmaya çalışıyor."

"Bu mantıklı değil mi? Elinize iki as geldiği zaman bahsi yükseltmeniz gerekmez mi?"

Uyuşmazcı anlaşılamamaktan yorulmuş: "Tabi yükseltmelisin! Ama anlamıyorsun! Matematikte ve psikolojide bu kadar iyi olduğu için, bu kadar iyi bir belleği olduğu için, ve optimal oyunu oynamak için bu kadar büyük bir arzusu olduğu için, eline iki as geldiğinde bahsi yükseltmemesi imkansız."

"Yani?"

“Yani yaptığının aksini yapamaz ve bu da en azından poker oynarken özgür olmadığını gösteriyor.”

Siz de “Ve ben acemi bir poker oyuncusu olarak,” diye tamamlıyorsunuz. “…tutarsız kararlar veriyorum. Bu da benim poker masasında ondan daha özgür olduğumu gösteriyor(!). Haklı mıyım?..”

Bu türden düşünce deneyleri, uyuşmazcı pozisyona katılmamama ve belirlenmemişliğin gerçek seçimlerin kavramsal olarak içermesi gerekmediğine inanmama neden oluyor. Düşünce deneyinde, *DİAOP* savunucularını karikatürize ettiğim iddia edilebilir. Özgürlükçü uyuşmazcıların Bayan Belirlenmiş’in durumunu şöyle yorumlayacağı da iddia edilebilir: “Belki Bayan Belirlenmiş’in yetenekleri ve optimal oynama ihtirası göz önünde bulundurulduğunda, elinde bir çift ası olduğunda bahsi arttırmaması mümkün olmayabilir; ancak bahsi yükseltmek istemeseydi -kaybetmek isteseydi- Bayan Belirlenmiş kendisini bahsi yükseltmekten alıkoyabilirdi.” Öte yandan, bu özgür irade kavramsallaştırması uyuşurculara ait olduğundan, bizi uyuşurculuğun doğru olduğuna götüreceğinden, böyle bir savunma bir uyuşmazcıyı tatmin etmemeli. Bayan Belirlenmiş, indeterministik dünyalarda olduğu gibi, deterministik dünyalarda da kendisini yükseltmekten alıkoyabilirdi. Bu durum, muhtemelen uyuşmacıların hoşuna gitmezdi.

*DİAOP*, acemi poker oyuncularını, mükemmel olanlardan daha özgür yapıyor. Aynı şekilde, *DİAOP*, *mutlak* iyi melekleri, *oldukça* iyi insanların aksine özgür ve övgüye değer bulmuyor; çünkü mutlak iyi melekler, oldukça iyi insanların aksine yaptığından başka şey yapamaz. Fizikalist bir dünyada ne acemi ne de mükemmel oyuncuların tam olarak aynı koşullar altında farklı davranamayacağını düşündüğümüzde, *FİAOP*’nin aynı sorunlardan muzdarip olmadığını görüyoruz. Her ikisi de gerçekten özgür ve yaptıklarından gerçekten sorumlu olmuyor. Bu

açık: Her şeyden önce, fizikalist bir dünyada, fiziksel özelliklerle ilgili eksiksiz doğru önermeler kümesi aynı olduğunda , hiçbir şey olduğundan farklı olamaz.

Uyuşmazcılarla nedensel determinizm doğru ise, hiç kimsenin hiçbir seçiminin en ufak bir belirlenmemişlik barındırmadığında tabii ki hemfikirim. Ancak bunun onların istediği herhangi bir şeyi kanıtladığını düşünmüyorum. Gerçek karar bir çeşit *çokluk* gerektiriyor. Ancak bu olası davranışların çokluğu değil, seçeneklerin çokluğudur. Bir başka deyişle, gerçek seçim birden çok seçeneği gerektiriyor; ancak seçimin belirlenmemiş olması gerekmiyor. *Mükemmel* bir satranç oyuncusu indeterministik olarak satranç oynamaz; ama birden çok seçeneği olduğu ve oyunu üzerinde kontrolü olduğunu varsaydığımız için ona özgür deriz.

## İlgisizci Pozisyon ve Fizikalizmle İlişkili Alternatif Olanaklar Prensibi

Tekrar ana probleme dönersek, acaba ilgisizci pozisyon, *FİAOP*'ye dayanan özgür irade kavramıyla uyumlu mu? İlk olarak, açık ki *FİAOP* deterministik fizikalist bir dünyada sağlanamaz. Öte yandan, indeterministik fizikalist bir dünyada sağlanabilir mi acaba? Tam olarak aynı koşullar altında yaptığımızdan başka bir şey yapmamız mümkün mü? Tekrar kukla analojisine dönersek, indeterministik fizikalist bir dünyada kuklacı ne yapacağına olasılık içeren kurallara göre karar veriyor. Fakat kuklacının indeterministik olarak belirlenmiş olan her durumu için, kuklaların sadece bir olası durumu var. Bu kavramsal olarak şart olmalı; çünkü fizikalizmin doğru olduğunu ve bu nedenle her şeyin fiziksel özelliklere takip ettiğini varsayıyoruz. Bu durumda *FİAOP* indeterministik fizikalistik bir dünyada da sağlanamıyor olmalı.

Bu düşünce, sinir sistemindeki bir elektronun kararsız bir sürücünün kırmızı ışıkta durup durmayacağını belirlediği düşünce deneyinde zaten açıklandı. Yine de aynı düşünce kukla analojisiyle de tekrar ele alınabilir: Kuklacı bir sonraki hareketinin nasıl olacağını *ontikçe olasılık içeren* (*onticly probabilistic*) bir parayı atarak belirliyor olsun. Eğer tura gelirse elini kaldırıyor ve yazı gelirse kaldırmıyor. Ve diyelim ki, kuklacı elini kaldırdığında kukla sıçrıyor. Bu durumda, kuklaların rolü, bizim indeterministik fizikalist bir dünyadaki rolümüze analojiktir. Şimdi bu durumda kuklaya özgür demek akıl almaz olmaz mıydı? Benim kanım olacağı yönünde. Ve benim "*fizikalizmle ilişkili alternatif olanaklar prensibi*" dediğim, özgür irade probleminde işe yarar olan tek *AOP* de aynı yönü işaret ediyor; çünkü kuklalar tam olarak aynı koşullar altında yaptıklarının aksini yapamazlar. Kuklacının elini kaldırmadan elini kaldırıp kaldırmayacağı belirsiz olabilir; ama kuklacı elini kaldırırsa kuklanın sıçrayacağı kesindir.

Öyleyse, biz sadece deterministik fizikalist dünyalarda değil, indeterministik fizikalist dünyalarda da, tam olarak aynı koşullar altında yaptığımızdan başka bir şey yapamayız. Bir başka deyişle *FİAOP*'ye göre fizikalist bir dünyada özgür olamayız ve bu hem deterministik hem de indeterministik dünyalar için doğru. İlgisizci pozisyonun bir biçimi *FİAOP* ile uyuşuyor.

Bazı uyuşmazcıların benim *takip etmeye* dayanan *AOP* yorumumu kabul edeceklerinden şüpheliyim. Her ne kadar fiziksel her durum için iradenin tek durumu olsa da, bu durumun kendisini göstereceğinin o duruma geçilmeden önce belirsiz olduğunu iddia edebilirler. Beni yanlış yorumlamakla suçlayıp, *DİAOP*'de ısrar edebilirler. Bu durumda benim savunmam onların *AOP*'lerinin indeterministik dünyada kuklaları özgür yaptığı olur.

## Deterministik Sistemlerin Tahmin Edilemezliği

Biz bir anlamda makineleriz; ancak makineler düşünüldüğü kadar tahmin edilebilir değiller. Determinizm doğru da olsa, bütün fizik kurallarını bilsek de, kendimizin (sistemimizin) geleceğini tahmin edemeyiz. Aşağıdaki düşünce deneyinde karşılaşılan çelişki, bizi deterministik sistemlerin en azından başı şartlar altında tahmin edilemez olduğu sonucuna iter.

1) Determinizm doğru olsun.
2) Şimdi de bütün doğa kanunlarına ve dünyamızın herhangi bir $t_1$ anına ilişkin tüm ampirik bilgiye sahip olduğumuzu varsayalım.
3) Bu durumda, sahip olduğumuz bu bilgi ve doğa kanunlarını kullanarak Laplace'ın cininin[33] (Laplace's demon) örneği olan bir bilgisayar yapmamız mümkündür.
4) Öyleyse, benim ne zaman öleceğimle ilgili bu bilgisayara danışmak istememem mümkündür. Ve İstanbul'a belirli bir uçak seferiyle giderken uçak kazası sonucu öleceğimi öğrenmem olasıdır.
5) Ölüm fikri muhtemelen beni korkuturdu. Ölmek istememem olasıdır. Ve özgür irademin ve ölüm korkumun ardındaki fiziksel mekanizmadan dolayı, İstanbul'a en azından bu kaza geçirecek uçakla gitmemeyi tercih etmem mümkündür.
6) Ya uçağa bineceğim ya da binmeyeceğim. Uçağa binmem 5. önerme yüzünden sorunlu görünüyor. Bu durumda uçağa binmeyeceğim. Bu da 3. önermemle çelişiyor.

[33] Pierre-Simon Laplace'ın tanımladığı üstün hesaplama ve değerlendirme yetisine sahip, hesaplamalarına sayanarak geleceği kusursuz olarak tahmin eden varlık (Laplace, 1902, s. 4).

Bu çelişki barındıran düşünce deneyi, ilk bakışta, determinizmin doğru olmadığı ya da zayıf özgür irademizin bile var olmadığı sonucunu çıkarmamıza sebep olabilir. Öte yandan, bana göre buradaki problemler, daha önce bazı David Wolpert'in (2008) matematiksel olarak ispatladığı gibi determinizm doğru olsa bile hiçbir şeyin kendisini içeren bir sistemin geleceğini hesaplayamayacağı gerçeğine dayanıyor. Bu türden düşünce deneyleri ne zayıf özgür irademizin olmadığını ne de determinizmin doğru olmadığını göstermiyor. Tez boyunca bu türden düşünce deneyleri ile ilgili aşağıdaki sonuçlara ulaşıldı:

1) Nedensel olarak deterministik bir dünyanın geleceğini hatasız bir biçimde bilen bir Laplace'ın cini örneği ancak dünyadan nedensel olarak izole biçimde var olabilir.
2) Başarısını sürdürmeyi garanti eden Laplace'ın cini benzeri bir varlık yapmak da iki sebeple mümkün değildir:
    a. Hiçbir nesne kendisinin kapsayan olduğu bir sistemle ilgili her şeyi bilemez.
    b. Öngören şey öngörüsünü öngörmeden bilmeyeceğinden, öngörüsüne bağımlı bir çıktıyı bilemez.
3) Geleceği hesaplamak yerine önceden gören bir medyumun var olması için de, dünyadan nedensel olarak izole edilmiş olması gerekir.

Bu sonuçlar yukarıdaki düşünce deneyinin en az bir yanlış öncül içerdiğini gösteriyor. Eğer bu iddialar (bana göründükleri gibi) doğruysa, düşünce deneyinden özgür irade ve determinizm arasındaki bir ilişki olduğuna dair bir sonuç çıkarılması için bir sebep kalmaz.

## Sonuçlar

İstekler ve irade arasındaki ilişki başlangıçta atlanmaması gereken bir nokta diye düşünüyorum. İsteklerden bağımsız bir irade imkansız olduğundan, hala özgür iradeden bahsetmek istiyorsak, iradeyi istekleri içeren bir yapı olarak almalıyız. En güçlü fiziksel ya da entelektüel zevkine göre hareket etmeyen bir varlık, bize doğru dürüst yürümekte zorluk çeken sarhoş bir adamdan daha özgür görünmez.

Özgür irade problemi incelenirken en çok paylaşılan sezilerden birisi, indeterminizmin özgür iradeyle determinizmden daha uyuşur olduğu. Honderich gibi bazı uyuşurcular bile bu seziyi sorgulamaktan geri durabiliyor. Bu tezde, Hume, Frankfurt, Dennett, Pareboom ve G. Strawson gibi, bu seziyi, zayıflatmaya çalıştım. Hume, Frankfurt ve Dennett zayıf özgür iradenin indeterminizmle olduğu gibi determinizmle de uyumlu olduğunu savunuyorlar. Diğer yandan, Pareboom ve Strawson ise güçlü özgür iradenin indeterminizmle, tıpkı determinizmle olduğu gibi uyumsuz olduğunu savunuyorlar. Bana göre her iki ekol de -"iyimser uyuşurcular" ve "kötümser uyuşmazcılar"- rasyonel pozisyonları savunuyorlar. Ben de bu iki ekolü ilgisizci pozisyonda buluşturmaya çalıştım. Eğer fizikalizm doğru ise, hem determinizm hem de indeterminizm zayıf özgür iradeyle uyuşurken, güçlü özgür irade ile uyuşmaz. Eğer fizikalizm yanlışsa, hem güçlü hem de zayıf özgür irade determinizm / indeterminizm probleminden bağımsız olarak, var olabilir de, olmayabilir de.

**Tablo 1:** Bu tezin özgür irade ve fizikalizm arasındaki ilişkisine dair sonuçları.

| | Zayıf Özgür İrade | Güçlü Özgür İrade |
|---|---|---|
| Fizikalistik Dünya | Mümkün | Mümkün değil |
| Fizikalistik Olmayan Dünya | Mümkün | Mümkün |

Eğer dışsal olandan bağımsız bir irademiz varsa, deterministik veya indeterministik bir şablon takip etmesinden bağımsız olarak güçlü anlamda özgürüz demektir. Bu nedenle, mükemmel biçimde iyi olan (iyilik tarafından tamamen belirlenmiş) bir melek, oldukça iyi bir insandan daha az özgür değildir; ya da mükemmel bir satranç oyuncusu (matematik ve mantıkla tamamen belirlenmiş) acemi bir satranç oyuncusundan daha az özgür değildir.

Fizikalizm doğru ise, sadece birer makine olduğumuzu düşünüyorum. Makinelere içtenlikle kızmaz ya da takdir etmeyiz. Bunun yerine, onları kendi çıkarlarımıza kullanırız. Bunun sevimli görünmediğinin farkındayım; ama bu benim tercih ettiğim değil, kaçınamayacağımızı savunduğum bir durum. Daha açmak gerekirse, kendilerine büyük saygı duyduğum öğretim üyeleri bu tezin *Teşekkürler* (*Acknowledgements*) bölümünü nasıl değerlendirmeli? Sempati, takdir duygusu ve benzerlerinin var olmadığını savunmuyorum. İddia ettiğim şey, bu duyguların varsayılanın aksine sorumluluk, özgürlük gibi rasyonel bir temeli olmadığı. Kişisel olarak bu rasyonel olarak temelsiz, evrimsel olarak faydalı duygulardan nasibimi aldığım için kendimi şanslı sayıyorum.

P. F. Strawson'un (1962) modern özgür irade probleminin aşırı-entelektüel kılınmış olduğunu (over-intellectualized) düşünmesinin sebebi, problemin gündelik duygu ve *reaktif tutumları* (*reactive attitudes*) karşılamıyor olmasıdır.

Evrimsel ve kültürel olarak şekillenmiş, duygu ve reaktif tutumlarımız var. Gerçek anlamda bir özgür değiliz ve gerçek anlamda yaptıklarımızın sorumlusu değiliz. “İyi” insanları takdir etmemiz ve “kötü” insanları beğenmeyişimiz, rasyonel bir temeli olmayan reaktif tutumlara dayanıyor. Öte yandan, *sorumluluk* ve *güçlü özgürlük*, rasyonel bir temeli olabilecek ve olması gereken iki kavram. Bu nedenle, aşırı-entelektüel kılınmış özgür irade probleminin, dünyada muhtemelen karşılığı olmayan sorumluluk ve özgürlük kavramları oluşturduğunda Strawson’a katılıyorum. Muhtemelen, zayıf özgürlüğümüzü (ya da istemli davranış) algılayıp, onu fazla entelektüel kıldık ve, dünyada karşılığı olmayan güçlü özgür irade ve gerçek sorumluluk kavramlarını oluşturduk. Bu nedenle, belki de en doğrusu bu kavramları öylece yok saymak olacaktır.

Popper’ın iddiasının aksine, tahmin edilemezlik, dünyamızın deterministik olmadığını göstermiyor. Tahmin edilemezlik özgür irademiz olduğunu ya da olmadığını da ispatlamıyor. Bana göre, tahmin edilemezlik de uyuşurculuk / uyuşmazcılık problemi ile ilişkisiz. Geleceği mükemmel biçimde tahmin etmeye aday bir Laplace’ın cini fikri bazı çelişkiler içeriyor ve bana göre bu çelişkiler de özgür iradenin determinizmle uyuşurluğu probleminden tümüyle bağımsız. Laplace’ın cininin sorunlarından bazıları, kendi hakkındaki tüm bilgiyi depolamanın imkansızlığı, ve kendini içeren bir sistemin geleceğini hesaplamanın imkansızlığıdır.