

UNIVERSITY OF RIJEKA

FACULTY OF HUMANITIES AND SOCIAL SCIENCES

David Grčki

**Planning, Reasons and Rationality over Time**

DOCTORAL THESIS

Supervisor: dr. sc. Nenad Smokrović, professor emeritus

Rijeka, 2020.



SVEUČILIŠTE U RIJECI

FILOZOFSKI FAKULTET

David Grčki

# **Planiranje, razlozi i racionalnost kroz vrijeme**

DOKTORSKI RAD

Mentor: dr. sc. Nenad Smokrović, prof. emeritus

Rijeka, 2020

## *Acknowledgments*

First and foremost, I would like to thank my supervisor professor emeritus Nenad Smokrović. Without his knowledge, guidance and patience this dissertation would not exist in this form. For pulling me from the depths of continental philosophy and gently pushing me towards the realm of analytic philosophy I have to say thank you to my M.A. supervisor professor Nenad Mišćević.

Although, we have rarely agreed on pretty much anything philosophically related I have to nevertheless thank professor Majda Trobok for feisty discussions regarding my thesis.

For her never ending support and encouragement I have to thank Martina Blečić.

For the warm welcome and continuous cooperation, I have to thank Philosophy Department at University of Rijeka.

Also, I have to thank audiences from various conferences over the years for insightful comments and criticism of my thesis. My dissertation is significantly better because of it.

I have to thank my mother Anita and father Miodrag for their continuous support through my philosophical journey.

For showing me that impossible obstacles do not exist in this world I have to thank my brother Sandro.

Lastly, to my son Patrik. You are my inspiration.

## *Abstract*

The main goal of my doctoral dissertation is to investigate and explain the rationality of diachronic agency. Firstly, I will answer the following questions:

- *What is a diachronic agency?* Diachronic agency is a specific capacity of adult humans to engage into activities that are spread across time.

- *What kind of activities are those?* Prime examples of those activities are plans and resolutions.

We all make plans in our lives from benign plans like what to have for lunch tomorrow to more substantial plans like planning to buy a house. We also have resolutions. The most well know resolutions are New Year's resolutions. Examples of these kind of resolutions are the following: *I'm going to stop smoking this year. I'm going to read at least ten books this year. I'm going vegetarian this year.*

Secondly, I will investigate if there are in fact any rationality norms that should guide this type of action.

Thirdly, I will explore how and under which conditions can agents rationality change their minds. The fact is that agents do change their minds. We break promises, dishonor commitments and abandon our plans. The questions are; are we irrational in doing so and is there a way to rationally change our minds? I will answer these question by arguing my thesis *Diachronic agency can be rationally assessed in the way in which synchronic rationality is assessed*, which can be seen as a defense of the notion of diachronic rationality.

**Key words:** rationality, instrumental rationality, diachronic rationality, time-slice rationality, reasons, practical reasons, plans, commitment, intention, options

## *Prošireni sažetak*

Cilj ove doktorske radnje jest istražiti pojam dijakronijske racionalnosti. Dijakronijska racionalnost nije pojam prema kojemu imamo neke snažne intuitivne stavove. Pojam *dijakronija* dolazi od grčke riječi *dia* što znači „kroz“ i *kronos* što znači vrijeme. Dijakronija se definira kao odvijanje događaja kroz vrijeme ili odvijanje pojava jednih iza drugih u vremenskom slijedu. Pojam dijakronije se uglavnom koristi u lingvistici za proučavanje jezika kroz njegov povijesni razvoj. Nasuprot tome, sinkronijskim proučavanjem jezika, sagledavamo jezik u ovom trenutku (figurativno govoreći), odnosno ne uzimajući u obzir vremensku dimenziju. No, kakve to veze ima sa racionalnošću? Čini se da kao ljudska bića imamo sposobnost da djelujemo ne samo ovdje i sada, nego smo sposobni djelovati s obzirom na naše odluke koje smo donijeli u prošlost i ideje o tome što ćemo činiti u budućnosti. Kao ljudska bića sposobni smo djelovati u okviru vlastitih planova, osobnih projekata, obaveza, novogodišnjih odluka, čak i obećanja. Ove aktivnosti definiramo kao dijakronijske aktivnosti. Razlog tomu jest što smo kad djelujemo na takav način u nekom smislu „prožeti u vremenu“. Takve su aktivnosti specifične za ljudska bića (barem tako tvrde autori koji argumentiranju u prilog dijakronijske racionalnosti). Životinje, čak i viši primati, mogu iskazivati iznimnu razinu inteligencije, razumijevanja i emocija, no ono što ne mogu jest povezivati svoja prošla djelovanja sa budućim stanjima svijeta na način na koji to mi (ljudi) možemo. Drugim riječima, životinje ne mogu ulaziti u dijakronijske aktivnosti (ne mogu imati plan, obavezu ili drugu vrstu dijakronijske odluke). Kao što možemo formirati planove, obvezati se, donijeti novogodišnju odluku ili dati obećanje, isto tako možemo odustati od planova, ne ispuniti svoje obaveze, te prekršiti obećanje. Nameće se slijedeće pitanje; postoje li neke norme racionalnosti koje vode naše dijakronijsko djelovanje? Odnosno; jesmo li racionalno obvezani držati se svojih planova? Argumentirati ću da doista jesmo racionalno odgovorni za svoja dijakronijska djelovanja. Moja teza jest da je dijakronijsko djelovanje podliježe racionalnoj prosudbi.

Doktorska disertacija sastoji od četiri poglavlja. U prvom poglavlju postavljam okvir u kojemu se nalazi tema moje doktorske disertacije. Taj okvir jest filozofija instrumentalne racionalnosti i filozofija radnje. Zatim uvodim i objašnjavam pojam dijakronijske racionalnosti. Naposljetku nudim pregled područja dijakronijske racionalnosti po dvije osi: povijesni razvoj i suvremena diskusija. U drugom poglavlju predstavljam ono što je, po mojemu sudu, najbolja teorija dijakronijske racionalnosti. U trećem poglavlju istražujem najbolju teoriju racionalnosti koja negira postojanje dijakronijskih normi. Iz ovog dijalektičkog pristupa izranja, četvrto i

posljednje, poglavlje doktorske radnje. U njemu direktno branim svoju tezu, te naposljetku nudim uvide u moguća daljnja istraživanja dijakronijske racionalnosti.

***Ključne riječi:*** racionalnost, instrumentalna racionalnost, dijakronijska racionalnost, racionalnost vremenskih odsječaka, razlozi, praktični razlozi, planovi, obaveza, namjera, opcije

# *Contents*

|   |           |
|---|-----------|
| <b>Introduction.....</b>  | <b>1</b>  |
| <b>1 Rationality: From a synchronic idea to a variety of diachronic accounts<br/>.....</b>  | <b>10</b> |
| 1.1 Tackling the concepts of theoretical and instrumental rationality<br>.....  | 10        |
| 1.2 Instrumental rationality: An overview<br>.....  | 11        |
| 1.3 From Davidson's account of action and rationality to diachronic notion of<br>agency .....   | 16        |
| 1.4 Diachronic rationality: Definition and examples<br>.....  | 20        |
| 1.5 Charting the field of diachronic rationality<br>.....   | 23        |
| 1.5.1 Historical overview .....   | 23        |
| 1.5.2 The first phase: Diachronic rationality as the problem of dynamic<br>(sequential) choice .....  | 24        |
| 1.5.3 The second phase: Diachronic rationality as the problem of understanding<br>the nature of intentions and future-directed attitudes<br>..... | 26        |
| 1.5.4 The third phase: Diachronic rationality as the problem of the nature and<br>the existence of diachronic norms<br>.....                      | 27        |
| 1.6 Contemporary accounts of diachronic rationality: Instrumentalists, Kantians<br>and antirealists.....  | 28        |
| 1.6.1 The instrumentalist account of diachronic rationality<br>.....  | 30        |
| 1.6.2 The Kantian account of diachronic rationality<br>.....  | 34        |
| 1.6.2.1 Ferrero's account of diachronic rationality<br>.....  | 36        |

|  |           |
|--|-----------|
| 1.6.3 The antirealist account of diachronic rationality  | 41        |
| 1.6.3.1 Time-slice Rationality   | 44        |
| 1.7 Chapter summary  | 46        |
| <b>2 A case for diachronic rationality: Intentions, Plans and Self-governance</b>  | <b>47</b> |
| 2.1 The planning theory of intention   | 50        |
| 2.1.1 Planning capacities and examples   | 50        |
| 2.1.2 Plans  | 54        |
| 2.3 The normative side of intention  | 55        |
| 2.4 Some problems with Bratman’s planning theory of intention: The bootstrapping problem, The Toxin Puzzle, and The myth-theoretical challenge | 61        |
| 2.4.1 The bootstrapping problem  | 61        |
| 2.4.2 The Toxin Puzzle   | 62        |
| 2.4.3 The myth-theoretic challenge   | 69        |
| 2.5 Self-governance as Diachronic Plan Rationality   | 72        |
| 2.5.1 A place to stand   | 75        |
| 2.5.2 Synchronic and diachronic self-governance  | 76        |
| 2.5.3 The absence of self-governance   | 78        |
| 2.5.4 Rational non-reconsideration model   | 80        |
| 2.5.5 The strategy of self-governance  | 84        |

|  |            |
|--|------------|
| 2.7 Chapter Summary .....  | 89         |
| <b>3 A case against diachronic rationality: Tragedy, Options and Time-slice Rationality.....</b> | <b>91</b>  |
| 3.1 Theoretical background and the main idea of Time-slice Rationality .....                     | 91         |
| 3.2 The rationality of Time-slice Rationality .....  | 93         |
| 3.3 Motivations for Time-slice Rationality .....   | 95         |
| 3.4 Arguments for the Time-slice Rationality account .....                                       | 102        |
| 3.4.1 The Diachronic Tragedy Argument .....  | 102        |
| 3.4.2 The argument from options .....  | 106        |
| 3.4.3 The two arguments recap.. .....  | 111        |
| 3.5 Chapter summary .....  | 112        |
| <b>4 The last stand: In defense of the diachronic notion of rationality .....</b>                | <b>114</b> |
| 4.1 General overview and framing of the problem .....  | 114        |
| 4.2 The tragedy of Time-slice Rationality .....  | 116        |
| 4.2.1 The norms of Time-slice Rationality .....  | 117        |
| 4.3 The critique of the Time-slice Rationality account .....                                     | 119        |
| 4.3.1 Internalism is compatible with diachronic and synchronic accounts of rationality .....     | 120        |

|  |            |
|--|------------|
| 4.3.2 The problems with personal identity are equally problematic for diachronic and synchronic accounts of rationality .....  | 121        |
| 4.3.3 The notion of rationality should (at least in some sense) reference real-life agents.....  | 123        |
| 4.3.4 The underlying assumption behind the <i>Diachronic Tragedy Argument</i> and <i>The argument from options</i> is a formal decision theory (or the correct form of the expected utility theory) .. | 125        |
| 4.3.5 Time-slice Rationality fails to account for clear and uncontroversial cases of irrationality (reasonless or arbitrary change of mind and severe cases of forgetfulness) .....                    | 129        |
| 4.3.5.1 A simple change of heart or a case of severe capriciousness .....  | 129        |
| 4.3.5.2 Remembering to be rational .....   | 131        |
| 4.4 The clash of diachronic rationality accounts .....   | 134        |
| 4.5 Bratman’s Diachronic Plan Rationality account emerges victorious .....   | 138        |
| 4.6 New frontiers: challenges for diachronic rationality and thoughts on further investigations .....  | 140        |
| <b>5 Conclusion.....</b>   | <b>144</b> |
| <b>6 References .....</b>  | <b>145</b> |
| List of tables and figures .....   | 157        |

# Introduction

*Man plans, god laughs.*

Yiddish proverb

The main aim of this dissertation is to investigate the notion of diachronic rationality. Discussions of rationality are almost always carried out within the synchronic framework. When we discuss what rationality requires from an agent to believe (theoretical rationality) or do (practical rationality), there is a hidden presupposition that there is no time flow or change involved in these assessments. The question of what an agent should believe or do simply amounts to what an agent should believe or do at that particular moment. We assess whether agents respond to reasons, acknowledge evidence, or have consistent sets of beliefs or intentions (the list is not exclusive and the notion of rationality can be understood in various ways). When we conceptualize rationality in this way, we are operating within the synchronic framework of rationality. This means that we assess whether agents are rational at a particular point in time and this has been the dominant way of viewing rationality throughout the history of philosophy. Recently, authors have been attempting to address the notion of rationality in a different way. They are exploring the notion of diachronic rationality. Different authors take various positions regarding the issue. While some authors have wholeheartedly argued in favor of diachronic rationality (Bratman 2018), others have fervently argued against the notion of diachronic rationality (Hedden 2015a). Before we continue any further, I should explain the meaning of the term *diachronic*. Diachronic simply implies the agency stretched or spread over some period of time. It is most commonly used in linguistic analysis when authors describe the evolution of language over time as opposed to the analysis of language today (synchronic analysis). In the philosophy of action and the philosophy of practical rationality (framework in which I conduct my research), the term diachronic refers to *diachronic agency* in which human beings can engage. Humans can engage in activities in which they take into account their past commitments and future prospects. Examples of these activities include plans, resolutions and promises (among others). These are fundamentally diachronic activities and it would be strange to look at these activities in a synchronic manner. Let us take a look at the concept of promise. It would be a bit childish to say the following: I know I promised you yesterday that I would do X today but I have no reason to do it because it was yesterday and not today. We are able to grasp that there is some meaningful connection between these two points in time and that viewing the concept of promise as two isolated points in time which have nothing to do with

each other makes little to no sense. Another way to easily grasp the concept of diachronic and synchronic agency is by looking at Aesop's fable, *The Ant and the Grasshopper*.

*One bright day in late autumn a family of Ants were bustling about in the warm sunshine, drying out the grain they had stored up during the summer, when a starving Grasshopper, his fiddle under his arm, came up and humbly begged for a bite to eat. "What!" cried the Ants in surprise, "haven't you stored anything away for the winter? What in the world were you doing all last summer?" "I didn't have time to store up any food," whined the Grasshopper; "I was so busy making music that before I knew it the summer was gone." The Ants shrugged their shoulders in disgust. "Making music, were you?" they cried. "Very well; now dance!" And they turned their backs on the Grasshopper and went on with their work. (Aesop/Winter: 1919)*

The main point that I would like to illuminate with this fable is the importance and relevance of diachronic agency. The Ants are engaged in some sort of diachronic activity. We can say that they are engaged in a plan to gather and store food during the summer for winter months. Conversely, the Grasshopper is living in the moment – an extremely synchronic activity. We are inclined to believe that the behavior of the Ants in this fable is in some sense correct and the behavior of the Grasshopper is in some sense incorrect (this does not imply that every synchronic activity is necessarily irrational). It would seem that the Ants are engaged in an activity spread across time and they have some constraints on their behavior and decision making, while the Grasshopper brings upon himself his own downfall. Those constraints on an agent's activity spread across time are diachronic constraints or norms and they are a part of the domain of diachronic rationality. My thesis states: ***Diachronic agency can be rationally assessed in the way in which synchronic agency is assessed.*** It can also be seen as a defense of the existence of diachronic norms of rationality.

My dissertation contains four chapters. Each of the chapters serves the same purpose in a different way – exploring and arguing in favor of the notion of diachronic rationality.

The first chapter serves two main purposes: it gives an introduction to the subject matter and offers my systemic overview of the notion of diachronic rationality. The second chapter also serves two main purposes: it provides an introduction to Bratman's planning theory of intention and sets forth a presentation of the most robust account of diachronic rationality, namely Diachronic Plan Rationality. The third chapter presents an exploration into the greatest opposition to the idea of diachronic rationality, namely Time-slice Rationality. The fourth and the final chapter constitutes a direct defense of my thesis. In the first part, I list five reasons

why Time-slice Rationality fails as an account of practical rationality. In the second part, I endorse Bratman's Diachronic Plan Rationality as the best account of diachronic rationality. In the last part, I offer my thoughts on future investigations regarding diachronic rationality.

At the beginning of the first chapter, I set the stage by introducing the key concepts and the framework for discussion. The first is the concept of rationality. Rationality is an extremely elusive concept tackled from many different angles in philosophy. The concept of rationality is tied to the notions of reasons, reasoning, consistency and coherence (the list is, of course, not exhaustive). Traditionally speaking, the problems of rationality are divided into two categories: theoretical rationality and instrumental or practical rationality. The two approaches overlap to a certain extent but this classification has been useful throughout the history of philosophy. In theoretical rationality, we assess agent's beliefs, desires, and other propositional attitudes. In instrumental or practical rationality, we assess agent's intentions, actions, and behaviors. Beginning with Hume (1739/1975) and Davidson (1963), I ground my investigation on the notion of practical rationality. At this point, I introduce the key notions of my dissertation: the concept of diachronic agency, the concept of diachronic norms and the concept of intention. The concept of intention explains the connection between thought and action in a fairly unique way. If an agent intends to do something and nothing intervenes, then she should do it. Otherwise, she never intended to do it in the first place (Armstrong and Malcolm 1984). This sets the stage for the diachronic approach to our decision making and action. Diachronic agency is simply an agency spread over time. In other words, an agent is engaged in temporally extended agency (diachronic agency) if she at her current point in time takes into consideration her past commitments and future prospects. Diachronic norms of rationality are the norms governing that kind of agency.

The second part of the first chapter consists of mapping the domain of diachronic rationality. I propose two overviews of the diachronic rationality domain: historical overview and contemporary overview. In the historical overview, I propose three phases in the investigation of diachronic rationality. The first phase is: *Diachronic rationality as the problem of dynamic (sequential) choice*. The second phase is: *Diachronic rationality as the problem of understanding the nature of intentions and future-directed attitudes*. Finally, the third phase is: *Diachronic rationality as the problem of the nature and the existence of diachronic norms*. Each of these phases is relevant in the contemporary discussion about diachronic rationality. The first phase, although failing to address the concept of diachronic agency properly, is the foundation for the contemporary opponents of the diachronic rationality accounts, most notably

Brian Hedden's Time-slice Rationality account (Hedden 2015a, 2015b, 2015c). The second phase, by establishing the key components of diachronic rationality and the concepts of intention and future-directed attitudes, forms the basis for the contemporary investigation into diachronic rationality. The third and final phase, by introducing the opposition to diachronic rationality (Hedden 2015a, 2015b, 2015c; Moss 2015), sets the stage for the contemporary discussion about the nature of diachronic rationality.

In the contemporary overview, I propose regarding the current state of discussion about diachronic rationality as the conflict between three camps: instrumentalist accounts of diachronic rationality, Kantian accounts of diachronic rationality and antirealist accounts of diachronic rationality. Instrumentalist accounts of diachronic rationality state that diachronic agency is explained by the fact that intentions are diachronic in nature and the norms proposed are minimalistic ones, such as *persistence of intention* or *fill-in plans* (Broome 2013, Snedegar 2017). Kantian accounts of diachronic rationality state that in order to adequately explain the nature of our ability to engage in temporally extended agency, we need to invoke the concepts of *will*, *autonomy*, *narrative* and *self-governance* (Ferrero 2009, 2012, 2012; Velleman 2000, 2007; Holton 2009; Bratman 2018). Lastly, antirealist accounts of diachronic rationality, such as Time-slice Rationality and time-slice epistemology, simply claim that there are no genuine diachronic norms of rationality (Hedden 2015a, 2015b, 2015c; Moss 2015).

In the second chapter, I explore Bratman's account of practical rationality and his account of diachronic rationality. There are several reasons why both of these accounts are important for my thesis. Firstly, Bratman's account of practical rationality – *planning theory of intention* – is a new and revolutionary way of thinking about human thought, action and decision making. By rejecting the belief-desire model (which goes back all the way to Hume) as an inefficient model of explaining human practical reasoning and action, Bratman proposes *plans* and *future-directed attitudes* as bases for our understanding of decision making and action. According to Bratman, we are future-directed beings and have the ability to formulate and execute plans. Plans are intentions which have the property of stability and inertia and allow us to coordinate with ourselves at different times and with other agents. Additionally, plans have a normative side because agents have a certain amount of commitment when they formulate and begin to execute their plans. Planning theory of intention is by no means perfect (as I present some of the major concerns of planning theory of intention, such as *Intentions as reasons (The bootstrapping problem)*, *Intention formation (The Toxin Puzzle)*, and *The myth-theoretical challenge*), but it is undeniably the foundation for every contemporary diachronic account of rationality.

The second part of the chapter is an exploration into Bratman's diachronic account of rationality, namely *Diachronic Plan Rationality*. Bratman argues that human beings value governing their own lives, more precisely, they value stability and coherence in their lives. This value is called *self-governance* (Bratman borrows the concept from Harry Frankfurt). The value of self-governance allows us to act consistently over time or have some diachronic consistency in our lives. From that point, Bratman presents diachronic norms of rationality, namely the *Diachronic Plan Rationality* norm. This chapter takes most of the heavy lifting in the defense of my thesis: ***Diachronic agency can be rationally assessed in the way in which synchronic rationality is assessed***. Firstly, planning theory of intention introduces diachronic way of thinking about practical reasoning and decision making. Secondly, planning theory of intention and Diachronic Plan Rationality both explain how we engage in temporally extended agency. Lastly, Diachronic Plan Rationality offers an extensive and explanatory account of both diachronic agency and diachronic norms, which makes it one of the strongest accounts of diachronic rationality in contemporary debate.

In the third chapter, I explore the most extensive account standing in opposition to my thesis – the Time-slice Rationality account. There are two reasons for this. Firstly, introduction of the Time-slice Rationality account has reignited the debate about diachronic rationality in recent years (Hedden 2015a). Secondly, the Time-slice Rationality account stands as a defense against the possible criticism of my thesis which would claim it to be trivial. My response to that sort of criticism would be to point out that there are practical accounts of rationality which directly contradict my thesis, namely the Time-slice Rationality account (Hedden 2015a, 2015b) and time-slice epistemology (Moss 2015). Hedden's main point in presenting the Time-slice Rationality account is to demonstrate that there are no genuine diachronic norms of rationality. His project is extremely ambitious because he covers both theoretical and practical rationality. Since my thesis is set in the domain of practical rationality, I address only the practical side of the Time-slice Rationality account (with some minor exceptions). Hedden's main line of argumentation is as follows. We should accept that the notion of rationality should fulfill three main roles: *evaluative role*, *predictive and explanatory role* and *action-guiding role*. This in turn leads us to accept a modest account of internalism regarding rationality. It also effectuates the claim that we should evaluate the rationality of beliefs and action without assuming any metaphysical claims about the nature of personhood. Finally, it directs us to conclude that only a synchronic account of rationality can adequately explain the nature of rationality and that all diachronic accounts of rationality are wrong. The synchronic account of rationality that Hedden proposes is Time-slice Rationality. The Time-slice Rationality account consists of two norms:

*Synchronicity and Impartiality.* *Synchronicity* states that all norms of rationality are synchronic and *Impartiality* states that norms of rationality should not reference agent's past or future beliefs and actions. I examine two arguments presented in favor of the Time-slice Rationality account: *Diachronic Tragedy Argument* and *The argument from options*. The first argument, the *Diachronic Tragedy Argument*, states that agents are sometimes diachronically inconsistent but not irrational and at other times they are diachronically inconsistent and in fact irrational. A simple pointing to diachronic inconsistency cannot answer the question of rationality because agents are in some cases simply tragic (their actions are suboptimal in regard to expected utility theory) and not necessarily irrational. The second argument, *The argument from options*, claims that the only thing that can be rationally assessed are the decisions available to an agent. The only decisions available to an agent are the ones she has access to at this point in time. The decisions an agent has access to at this point in time are called *options* (or preferences, in the terminology of expected utility theory). Since *options* are synchronic by definition, there can be no diachronic norms of rationality. There are a few things that I would like to point out here. Firstly, I tackle the smaller portion of Hedden's Time-slice Rationality account – the one that addresses the practical notion of rationality – which is underdeveloped in comparison to its theoretical counterpart. Secondly, Hedden's account is half a decade old. Compared to Bratman's account which counts thirty years of development and refinement, Hedden's account is extremely young. These present some of the reasons why Time-slice Rationality is under such criticism of late. Notwithstanding all these facts, Time-slice Rationality stands in strong opposition to all diachronic accounts of rationality. Hedden's account is certainly the most extensive and exhaustive account of rationality explicitly denying the existence of diachronic rationality in any shape or form and it has even attracted supporters. While I wholeheartedly disagree with his arguments and claim that Time-slice Rationality ultimately fails as an account of rationality, I do acknowledge the philosophical benefits of having a fierce opposition.

In the fourth and the final chapter, I offer a direct defense of my thesis: ***Diachronic agency can be rationally assessed in the way in which synchronic rationality is assessed.*** My defense is structured in two parts: disproving Hedden's Time-slice Rationality account and accepting Bratman's Diachronic Plan Rationality account. There are five points that I make against the Time-slice Rationality account.

My first point is that there exists no clear transition that directs us from accepting internalism or mentalist internalism (a specific account of internalism Hedden argues for) to accepting a purely synchronic account of rationality. Hedden argues, as apparently obvious, the norms of

rationality to be purely synchronic because the only thing available to an agent are her mental states and her mental states are available to her only at the present (current) moment. However, without additional argumentation this is simply a *non sequitur*. Diachronic accounts of rationality are perfectly compatible with internalism. Intentions, for example, are *available* (as Hedden emphasizes this concept) to agents at their current moments in time. Therefore, there is no reason to accept the apparent connection between internalism and the purely synchronic norms of rationality.

My second point lies in Hedden's claim (which is misguided at best and seriously flawed at worst) that rationality should be divorced from the murky problems of personal identity and that this is the reason to accept a synchronic account of rationality (specifically, Time-slice Rationality). There are three major problems with Hedden's claim. Firstly, it is not clear how exactly the norms of rationality can be separated from agents to whom those norms apply. Secondly, no account of diachronic rationality explicitly endorses any metaphysical claim about the nature of personhood. Lastly, even if we disregard the previous two reasons, the fact that the problems of personal identity are murky does not lead us to accept all rationality constraints as synchronic any more that it leads us to accept all rationality constraints as diachronic.

My third point is that the norms of rationality should, at least in some capacity, reference real-life agents. Hedden claims that the norms of rationality should be impartial and ideal without reference to real-life agents or agents at all. The problem here is that we can imagine a being who acts purely on her instinct without reasoning, reflection, or any thought for that matter. We can also imagine a being who acts purely by chance without reasoning, reflection, or any thought for that matter (Lenman 2017). This being (based on instinct or luck) always "chooses" optimally following (what Hedden calls) the *correct* expected utility theory. According to Hedden's Time-slice Rationality, that being (the one based on luck or the one based on instinct) is perfectly rational. The problem lies in this being the exact opposite of the way we have been considering the concept of rationality since Aristotle to contemporary discussions about rationality. Rationality is for the most part in some capacity connected to the concepts of reasoning, reasons, and reflection. Rational belief and action is in opposition to the behavior based purely on instinct or luck and Hedden's account fails to legitimize that fact.

My fourth point is that Hedden's arguments for practical Time-slice Rationality (*Diachronic Tragedy Argument* and *The argument from options*) contain the presupposition that the correct way to think about practical rationality is through expected utility theory. Hedden's point in the *Diachronic Tragedy Argument* is that some cases of agents' changing their minds constitute

them being simply tragic and not necessarily irrational. Agents who are tragic choose in the way suboptimal to their utilities and are aware of that fact. I agree with him completely. The problem lies in Hedden's assumption that the proponents of diachronic rationality accept expected utility theory as the foundation for practical rationality, which is not true. Some do (or did) accept it in the first phase of the investigation into diachronic rationality, which I termed *Diachronic rationality as the problem of dynamic (sequential) choice*, but there are accounts of diachronic rationality that do not rely on expected utility theory (Bratman 2018; Ferrero 2009, 2010, 2012; Holton 2009).

My last point is to present two types of behavior which are clearly and uncontroversially irrational but are not seen as such in Hedden's Time-slice Rationality account. The first is a case of arbitrary and reasonless change of mind or a severe case of capriciousness presented by Brunero and adopted by Bratman (Brunero 2012, Bratman 2012). I use their *Candice* case and my more extreme *Mandy* version to show that the cases of erratic capriciousness are the cases of irrationality in purely diachronic terms and that Hedden's purely synchronic account of rationality cannot explain such cases of irrationality. The second is a case of severe forgetfulness. Inspired by the argumentation of Doring and Eker (2017) and the movie *Memento* (Nolan 2000), I construct the case of Leonard Shelby. Leonard suffers from *anterograde amnesia* (the inability to form new memories). At one point in his life he had an accident and now he tries to manage his life one step at a time. He knows everything about himself and his life but cannot form new memories. His life is a set of unconnected time slices which are 15 minutes apart. After each 15-minute time slice, Leonard "resets" to the point in time when the accident happened. Most of the time he does not know where he is, how he got there, or why he got there. Leonard is clearly and uncontroversially irrational. However, the way in which Leonard is irrational makes it uniquely diachronic. There is nothing synchronic in his irrationality. Time-slice Rationality as a purely synchronic account of rationality cannot illuminate Leonard's irrationality and for that reason this case constitutes a counterexample to Hedden's theory.

In the second part of the last chapter, I evaluate different accounts of diachronic rationality. I do so using my own terminology established in Chapter 1. Firstly, I reject antirealist accounts on the grounds just having been presented. Secondly, I dismiss authors of the first phase, *Diachronic rationality as the problem of dynamic (sequential) choice*, because they did not address the concept of diachronic agency in a satisfactory and explanatory manner (or, sometimes, at all). They were mostly concerned with the way agents can save their dynamic consistency from one point in time to the next and had little (if anything) to say about

diachronic agency. Thirdly, I discard instrumentalists for two main reasons. The first is their lack of substance in explaining diachronic agency and the second is their inability to justify cases of erratic capriciousness, such as *Candice*. This leaves only Kantian accounts of diachronic rationality. Although I consider Ferrero's and Holton's accounts as adequate accounts of diachronic rationality, in the end I side with Bratman's Diachronic Plan Rationality account. There are two reasons for this. The first reason is that Bratman's account constitutes the most extensive and explanatory contemporary account of diachronic rationality. The second reason is that Bratman does not require the somewhat controversial concepts, such as *narrative* and *diachronic will* (Ferrero 2009, 2010), in order to explain practical rationality.

In the last part of the final chapter, I offer my thoughts on future investigations of diachronic rationality. I make the point that, when we accept some diachronic rationality account (the best one being Bratman's Diachronic Plan Rationality account), there are difficult examples requiring adequate address. One of those examples is *The runaway bride case* (which I have addressed elsewhere). The way to tackle these hard cases for diachronic rationality is to invoke the concept of different frames introduced by Tversky (1975) and Kahneman and Tversky (1979) and reintroduced recently by Bermúdez (2018, Forthcoming).

Lastly, I would like to declare that the claim I argue for in this dissertation is a modest one. It simply affirms that diachronic agency can be rationally assessed in the way in which synchronic agency is assessed. ***My claim can be seen as the middle road between stubbornness and erratic capriciousness.***

# **1 Rationality: From a synchronic idea to a variety of diachronic accounts**

## **1.1 Tackling the concepts of theoretical and instrumental rationality**

The concept of rationality means different things to different people. There are numerous ways of approaching the issue of rationality. We can find the concept of rationality in virtually every branch of philosophy – from epistemology and logic to ethics and metaphysics. There is a certain agreement amongst philosophers regarding the concept of rationality. Rationality as a whole is too broad and complex to be tackled from a single perspective. For that reason, rationality is divided into two broad categories – epistemic (theoretical) rationality and instrumental (practical) rationality. On the one hand, when we evaluate someone’s beliefs or desires, we are operating in the framework of theoretical rationality. On the other hand, when we evaluate someone’s intentions and actions, we are operating in the framework of instrumental rationality.

In the framework of theoretical rationality, we are oriented at exploring our full-blown beliefs. Traditionally, the goal of theoretical rationality is considered to be knowledge. Our beliefs represent the world and, if rational, should be true and/or justified. The most prominent way in which an agent can theoretically be irrational is to hold contradictory beliefs, but even this is sometimes contested.<sup>1</sup> Theoretical rationality usually addresses the sources of rationality. Sources of our rational or justified beliefs mostly discussed by authors are memory, perception, introspection, reasoning and testimony. We, as fallible human beings, do not have access to infallible sources of rationality, so there is much debate about how we derive beliefs or sometimes even knowledge from these sources (Audi 2004). In theoretical rationality, there are basic approaches to rationality such as foundationalism (Audi 1993, 2001), reliabilism (Goldman 1986), virtue epistemology (Sosa 1991, Zagzebski 1996, Greco 2000, 2010), contextualism (DeRose 1992), gnosticism or knowledge-first epistemology (Williamson 2000, Littlejohn 2018) and veritism (Wedgwood 2002). All of these accounts aim at answering the following question.<sup>2</sup> What beliefs are justified for a rational agent to hold? Full-blown beliefs

---

<sup>1</sup> Some solutions to the preface paradox allow holding a set of inconsistent propositions.

<sup>2</sup> This list presents some approaches in theoretical rationality and is by no means exhaustive. Theoretical rationality consists of many different branches that are connected to other vast philosophical fields such as ethics and metaphysics.

are not always the primary target of investigation of authors in theoretical rationality. Some authors talk about partial beliefs or degrees of beliefs when discussing rationality. In this framework, we usually talk about Bayesian rationality (Oaksford and Chater 2007). The focus of my research will be set within the framework of instrumental rationality.

In the framework of instrumental rationality, we are concerned with our intentions and actions. We ask ourselves what is rational for us to do. The main question is what the suitable means to achieving our goals are. First, we need to understand what the concepts of *means* and *goals* mean. When investigating *goals*, we refer to an agent's desirable state of affairs and when discussing *means*, we are concerned with the tools for achieving goals such as actions, intentions, choices, decisions, etc. Second, we need to differentiate between local and global rationality.

There are two ways in which we can fail at instrumental rationality: by having an incoherent set of attitudes and by failing to have sufficient or relevant reasons for performing an action. In the first case, we demand from an agent to be rationally coherent, i.e., to have her attitudes and actions cohere with one another. In the second case, we demand from an agent to have reasons for her actions (Kolodny, Brunero 2018).

## 1.2 Instrumental rationality: An overview

In this section, I will present relevant overview of instrumental rationality starting from Hume's influential quote from *A Treatise of Human Nature*. This overview is not supposed to exhaust the broad topic of instrumental rationality but rather simply help lay the foundations and presuppositions for the core of my research – the problem of diachronic rationality (rationality over time).

A way in which we can approach instrumental rationality is by tackling practical reasoning.<sup>3</sup> Modern foundations of practical reasoning lie in Hume, more precisely, in this quote from *A Treatise of Human Nature*:

*Where a passion is neither founded on false suppositions, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it. 'Tis not contrary to reason to prefer the destruction of the world to the scratching of my finger. 'Tis not contrary to reason for me*

---

<sup>3</sup> Different authors use different terminology to discuss the same topic. When investigating practical reasoning, they refer to instrumental rationality. For more see (Millgram 1997, 2001).

*to chuse my total ruin, to prevent the least uneasiness of an Indian or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter.* (Hume 416: 1739-40/1975).

*A Treatise of Human Nature* was written in the mid-17<sup>th</sup> century and over the course of time this passage has become one of the most influential passages that touch primarily on the concept of instrumental rationality, but rationality in general as well. Throughout the years, there has been a wide range of interpretations by various authors. Nevertheless, there is one point of consensus regarding this passage. Contrary to Kant, Hume makes a clear distinction between two concepts: the concept of rationality and the concept of morality. If it is not unreasonable or, as Hume would put it, *contrary to reason* for an agent to prefer the destruction of the world to the scratching of his finger, then we cannot reduce the question of rationality to the question of morality. This is not to say that the two concepts are not related in a meaningful way, it is just that one cannot be reduced to the other.

The most prominent interpretation of Hume's passage from *Treatise* is that the content of the passage endorses instrumentalism regarding practical reasoning. According to instrumentalism, all practical reasoning is means-end reasoning. When pondering what to do, an agent simply has to decide the optimal way in which she can achieve her goals or satisfy her desires. There are simple and clear benefits to this interpretation. I will mention a few. This account is extremely approachable from the point of view of folk psychology. Some authors do consider instrumentalism to be closely related to belief-desire psychology. Agents in the real world do have goals and some ideas of how to achieve those goals. The account has a strong explanatory dimension. Under normal circumstances, we can infer someone's desires from their actions. Lastly, the non-instrumental accounts would allow for a third party to impose to an agent what are really their own desires which is, at least *prima facie*, implausible. Hume's passage also invokes the concept of internalism regarding instrumental rationality. Although there is some contention whether the concept of internalism is related to instrumental rationality or a position in the debate about theoretical rationality, Hume's passage does offer a simple and clear insight into internalism regarding rationality. Hume points out that in order to explain an agent's action, we need to look for reasons for that action and those reasons are desires. Agent's actions are explained by her desires. An agent is motivated to an action by her desires. We regard those desires as motivating or explanatory reasons. If we presuppose that an agent's desires are only accessible to the agent, then we arrive at the internalist account of

instrumental rationality.<sup>4</sup> Internalism plays an important part in the debate about diachronic rationality. Some authors (Hedden 2015a, 2015b) claim that any form of internalism is incompatible with the notion of diachronic rationality.

If we take the instrumental interpretation to be correct, Hume's idea in his passage invokes the notion of instrumental rationality as *expected utility maximization*. Expected utility maximization is a vast and complex domain of research which has its roots in economics (Von Neumann and Morgenstern 1944, Luce and Raiffa 1957), but today it is used in various fields, ranging from psychology and philosophy to statistics and management. Viewed as a form of instrumental rationality, expected utility maximization would mean the following. Agent's goals are maximizing her expected utility, where "maximizing expected utility" simply means choosing the preferred outcome. Expected utility maximization, as a version of decision theory, is a normative, abstract mathematical model. It answers the question: what is an agent supposed to do in situations when confronted with a decision-making problem under uncertainty? Being a purely normative model of how agents should make decisions under risk<sup>5</sup>, decision theory has lost substantial amount of support as a theory of rationality. This was done most notably by Bermúdez in his book *Decision Theory and Rationality*. The relevance of decision theory in the context of diachronic rationality is reflected in the fact that the authors who first discussed the problem of diachronic rationality (without using this terminology, however) did so in the context of decision theory (Strotz 1956, Elster 1979, Machina 1989, McClennan 1990).

There are other interpretations of Hume's passage that I will go over briefly. The first interpretation is nihilism. This interpretation of the passage states there to be no correct forms of practical reasoning. There are only descriptions of how agents reason, but we cannot extract any real normative force about what agents actually should do. Although there are no prominent contemporary advocates of the nihilist interpretation, there are some relevant philosophical discussions on the matter (Hampton 1998, Millgram 2001, Korsgaard 1997). The second interpretation focuses on the agent choosing insufficient means for her ends. Proponents of this interpretation (Kolodny, Brunero 2018) argue that irrationality does not consist in intending insufficient means but in the falsehood of agent's beliefs underlying that intention. This interpretation would make Hume not an instrumentalist regarding rationality. The

---

<sup>4</sup> The concept of internalism in rationality is a highly contentious and debated one. The debate regarding internalism in rationality is long and fruitful. For introduction to the debate see (Williams 1979, Wedgwood 2002) and for contemporary contributions to the discussion see (Wedgwood 2017, Littlejohn 2012).

<sup>5</sup> For an opposite view see (Resnik 1987).

interpretation is also somewhat problematic because we can imagine an internally coherent set of attitudes that are based on false beliefs.

Leaving Hume's passage behind us, there are several key notions that we need to address when discussing instrumental rationality. Those notions include: *practical coherence*, *plans* and *defeasibility*. I have briefly touched on practical coherence at the beginning. The concept of practical coherence is a highly complex and contentious notion. We may approach the concept of practical coherence in the following manner.<sup>6</sup> There are two approaches to the use of the concept of coherence in practical reasoning: the *all-or-nothing* approach and the *gradual* approach. The all-or-nothing approach is consistent with the theoretical practice of decision and game theory. Our set of preferences over some relevant outcomes needs to satisfy the conditions for having a well-defined utility function. In this sense, our set of preferences is either coherent or incoherent, there is no middle ground. It is utterly senseless to say that a set of preferences is more or less coherent. This is what I mean by the all-or-nothing approach to the concept of coherence. The gradual approach to the concept of coherence stands in opposition to the all-or-nothing approach. When we talk about coherence in the gradual sense, we can be more or less coherent in respect to our goals, subgoals, actions or even emotions and values. In this approach, an agent chooses a subset of goals or actions that best cohere with one another. Accordingly, practical reasoning can be seen as *an inference to the most coherent plan* (Millgram 2001). There are several independent benefits of this approach to the concept of practical coherence. Firstly, there exists no problem of introspection regarding the strength of agent's desires. Secondly, there occurs no problem in agent's accessing of the relevant probabilities regarding her outcomes. Finally, there is no incommensurability problem.<sup>7</sup> The gradual concept of coherence makes most sense when applied to real-life agents. In everyday situations, real-life agents clump their goals together for the simple reason of cognitive efficiency and higher probability of achieving their goals if their plans be coherent. Having coherent plans is closely tied to the notion of being a unified agent. In what way those two concepts relate to one another will be explored in depth later.

The next notion that I will address is the concept of *plans* in practical reasoning. In this context, practical reasoning consists of adopting, filling and reconsidering plans. The idea has been introduced and developed over the years by Michael Bratman. Here are the key notions.<sup>8</sup> In the

---

<sup>6</sup> For other approaches to the concept of practical coherence see (Pollock 1974, Harman 1976, 1986).

<sup>7</sup> For the discussion on the incommensurability problem see (Raz 1986, Broome 2001, 2003, 2006).

<sup>8</sup> For more on Bratman's philosophical work see (Bratman 1983, 1987, 1999, 2000a, 2000b, 2001, 2005, 2012, 2018).

framework of practical decision making, human beings are planning agents because they have future-directed attitudes. We have two main capabilities as human beings: the capacity to act purposefully and the capacity to form and execute plans. Arguably, those two capacities (although not exclusively) are what separates us from other animals. There are two reasons why we make plans and are future-oriented: cognitive usefulness and coordination. We are creatures with limited capacities. Human beings are limited by time, space and cognitive abilities<sup>9</sup>. When engaging in practical reasoning, we are rather limited in deliberating at any particular point in time. Deliberation takes time and if we were to deliberate repeatedly on the same issue, it would be cognitively wasteful and time-consuming. This is where the concept of plans comes in. When we have a plan, there is no need to rethink our actions and intentions needlessly, we are simply guided by our plan. The second reason for adopting plans is coordination. In order to achieve complex goals, which has been established as a uniquely human capability, we need coordination between our past and present activities. There are two types of coordination: *intrapersonal* and *interpersonal*. Bratman proposes these as examples of intrapersonal coordination: *writing a lecture, picking up a book at the library, attending a committee meeting* and *picking up a child at school*; and the following as examples of interpersonal coordination: *arranging and participating in a committee with several colleagues* (Bratman 2: 1987). He concludes that we need both types of coordination because we are temporally extended and social agents.

The concept of practical coherence and the concept of plans are both indispensable when addressing the problem of diachronic rationality. One of the main questions regarding diachronic rationality states the following. Is there any rational requirement for agents to have a coherent set of attitudes over a certain period of time? And if there is such a requirement, how coherent should our attitudes be? The concept of plans is one way to ensure such coherence.

The last concept that I would like to address is *defeasibility*. The concept of defeasibility is used in a wide range of philosophical disciplines, from the philosophy of language (pragmatics and conversational implicature) and the philosophy of science (when discussing laws of nature and *ceteris paribus*) to formal epistemology (defeasible reasoning) and default logic (Etherington and Reiter 1983).<sup>10</sup> The way in which I will use defeasibility is in the context of practical (instrumental) reasoning. Instrumental reasoning is almost always defeasible

---

<sup>9</sup> Bratman adopts this approach from Herbert Simon. For more see (Simon 1972).

<sup>10</sup> This list is by no means exhaustive. For a comprehensive and contemporary overview of defeasible reasoning see (Koons 2017).

(Millgram 2001). This simply means that any appropriate reasoning can be such by adding additional premises. Let us clarify this point with the following example. If an agent's goal is to have a cheesecake at a restaurant and the suitable means for arriving at the restaurant is by taking a bus, the agent can retract her decision if she suddenly discovers there to be an outbreak of salmonella at that particular restaurant. We can construct countless everyday examples of this sort in the context of instrumental reasoning.

When discussing the problem of diachronic rationality, the concept of defeasibility is extremely useful. By its definition, diachronic rationality puts constraints on agent's attitudes and actions over time. For the reason that we cannot know what our attitudes will be in the future, it makes sense that our diachronic constraints are defeasible.

### **1.3 From Davidson's account of action and rationality to diachronic notion of agency**

After Hume, Davidson is probably the most influential philosopher in the fields of philosophy of action and instrumental rationality. Davidson approaches the concept of instrumental rationality by asking the following questions.

- What are actions?
- What are reasons?
- What is the relationship between actions and reasons?

Actions are events that are properly produced by reasons. Reasons (practical) are beliefs or desires that explain agent's actions<sup>11</sup>. To explain the connection between the notion of action and the notion of reason, Davidson uses the term *rationalization*. An agent acted the way she did because she had a reason and that reason *rationalizes* her action. Davidson presents the following example to illustrate his point.

*I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home* (Davidson 4:1980a).

---

<sup>11</sup> For contemporary discussion on practical reasons and reasoning see (McHugh and Way 2018, Smokrović 2018).

There are several questions that we need to address here. What counts as action in this example? How many actions are there in this example? What is the reason for this action? According to Davidson, there is only one action here – *flipping the switch*. We could argue that the example above consists of four actions. First, *flipping the switch*. Second, *turning the light on*. Third, *illuminating the room*. And four, *alerting the prowler*. Davidson argues this to be wrong. There is actually only one action and four different descriptions of the same action. Some descriptions make the action intentional (*flipping the switch*), while others (*alerting the prowler*) make the action unintentional. We can now ask: what is the reason for the agent's action in the example above? The reason why the agent flipped the switch is because she wanted to turn on the light. This reason, the primary reason, is the cause of the action.

Davidson argues for *causal theory of action*. Simply put, reasons are causes of actions. Agent's beliefs, desires or other propositional attitudes are the reason why the agent has taken appropriate action.

This is the main idea behind Davidson's account of instrumental rationality and action. A reason is a rational cause and an action is reasonable if it was caused by a reason. A reason explains (or *rationalizes*, in Davidson's terms) why the agent acted the way she did.

Davidson's notion of action and rationality (very briefly summarized above) is oriented primarily toward the synchronic notions of our decision making and action taking. What does this mean? Davidson's account is focused on describing and evaluating agent's beliefs, desires and actions at a single point in time. More precisely, it is synchronic. In Davidson's example, the agent's flipping of the switch (or various other descriptions such as turning on the light) is conceptually presented as a single, instantaneous action. Naturally, action takes time to complete (and Davidson is aware of this fact) but the diachronic element of human agency remains completely irrelevant for description and evaluation of an agent's beliefs and actions. This seems to leave a huge explanatory gap regarding the way human beings actually make decisions and see them through. We engage in activities that seem to be extended in time. In Bermúdez's words:

*As agents we engage in sequences of choices that are not always reducible to a series of independent, individual choices. We make choices about how we will choose, and we make choices in the light of earlier commitments to choose in certain ways. We make plans for the future and we have a degree of concern for the plans that we have made in the past. (Bermúdez 112: 2009).*

As Bermúdez puts it, the way we make decisions is not always reducible to *sequences of choices* that stand completely independent from one another. We are, as human beings, future-oriented and have the capacity to engage in activities that are extended over time. We make plans, honor our commitments, keep our promises and execute our resolutions. There is a distinct diachronic dimension in the way we structure and live our lives. The question is: how do we conceptually grasp this diachronic dimension of our agency?

There are several accounts that describe diachronic dimension of our agency and I will address them later in greater detail. Firstly, I would like to address the terminological difficulties in discussing the diachronic and the synchronic notions of human agency. There is a clear terminological difference between the synchronic accounts of action and rationality and the diachronic accounts of action and rationality. Terms such as *choice*, *preference*, *decision* and *option* usually go hand in hand with the synchronic views of human agency, while terms such as *intention*, *agency*, *resolution* and *narrative* usually go hand in hand with the diachronic views of human agency.

This is, of course, a broad generalization and does not perfectly reflect the synchronic and diachronic accounts of agency and rationality. Davidson uses the concept of intention in a fairly synchronic sense and has little connection with the diachronic aspects of human agency. For Davidson, the concept of intention is a propositional attitude in the same way as the concepts of belief and desire, there is nothing diachronic about it. On the other side of the spectrum, the concept of preference is used in the diachronic sense when discussing the notion of sequential (dynamic) choice. Additionally, the concept of agency is in some sense neutral in respect to synchrony and diachrony.

There is a couple of reasons why I believe the terminological differentiation is useful and helpful. Firstly, it reflects our pre-theoretical intuitions that choice, for example, is something we make or consider making right now (at this particular point in time), while intention, resolution or promise, for example, is something “stretched over time”. Secondly and more importantly, contemporary authors use these terms in (more or less) this fashion. Proponents of the diachronic accounts of agency and rationality use terms such as *intention* (Bratman 1987, 1999), *agency* (Velleman 1997, 2000, 2003, 2006, 2007; Ferrero 2006, 2009, 2010, 2012, 2014) and *resolution* (Holton 2009) to describe those phenomena of human agency which are not reducible to a set of individual and completely discontingent choices or decisions. Conversely, proponents of the synchronic accounts of agency and rationality use terms such as *choice*, *preference*, *decision* and *option* (Hedden 2012, 2015a, 2015b, 2019) to highlight the synchronic aspects of our agency.

Terminological difference aside, there still remains the question of how we proceed to capture the diachronic dimension of our agency. The answer to this lies in carefully examining the concept of intention and the connection between agent's intention and action.

*.... a logical connection between intending to do something and doing it. If doing it is well within the person's powers, and if he has not given up the intention for some reason or other, and if he has not forgotten his intention, and if no countervailing circumstances have arisen, and if he offers no satisfactory explanation for not fulfilling that intention, and so on – then if he doesn't do the thing, we would conclude that he does not really intend to do it. This way of judging the matter is required by the concept of intention. (Armstrong and Malcolm 88: 1984).*

The first thing that should be pointed out is that Armstrong and Malcolm are not discussing the concept of diachronic agency or rationality here. The problems that they are tackling are more complex and problematic (those being the problems of consciousness and causality). That being said, their thoughts are indicative of the way other authors (such as Bratman) will discuss the notion of diachronic agency and rationality. Armstrong and Malcolm's point is the following. There is a logical connection between *intending to do something and doing it* and this is contained in the concept of intention itself. If an agent intends to do something (there being no countervailing factors), but does not do the thing that she intended to do, then we should conclude that she never intended to do that thing in the first place.

In order to offer explanation of these phenomena, contemporary authors all agree on one thing: ***the concept of intention is fundamentally diachronic in its nature*** (Bratman 2014, 2018; Holton 2009; Broome 2013; Hedden 2015a). Proponents of the diachronic accounts of rationality (Bratman 2018, Broome 2013) and opponents of the diachronic accounts of rationality (Hedden 2015a, Moss 2015) all agree that intention as a concept in practical rationality and the philosophy of action is fundamentally diachronic. In other words, agent's plans, resolutions and promises would be quite meaningless if they referred only to this particular moment in time and nothing else. These rational capacities (plans, resolutions and promises) are diachronic in nature and the reason why they are diachronic is because the concept of intention is diachronic, i.e., ***intentions persist through time***.

There are, of course, different ways of capturing the diachronic notion of human agency and intentions are only one of them (Ferrero discusses the notion of *diachronic will* (Ferrero 2009, 2010, 2012)). Nevertheless, the concept of intention is fundamental for the idea of diachronic agency. Firstly, intentions have certain properties which are uniquely diachronic. Those

properties include *stability*, *inertia*, *commitment* (Bratman 1987, 1999) and *persistence over time* (Broome 2013). Secondly, *intentions* are conceptually very fruitful in describing agent's capacity to engage in time-extended activities, such as *plans*, *promises*, *resolutions* and other future-directed attitudes.

In the next section, we will explore in more detail the different modes of capturing human capacity for diachronic agency and the way in which we can make rational assessments of that agency.

## **1.4 Diachronic rationality: Definition and examples**

In the previous part, I have addressed the different ways of viewing the concept of rationality. First, I have established the difference between theoretical and instrumental rationality. Second, I have introduced the relevant concepts regarding instrumental rationality, those being practical reasoning (means-end reasoning), practical coherence, internalism, utility maximization, plans and defeasibility. All of these concepts are generally used in the synchronic framework. Usually, when discussing instrumental rationality or rationality in general for that matter, we do it in the synchronic framework. Synchronic rationality tells us what beliefs, intentions, attitudes, decisions and actions an agent should make/hold *at a single time*. Conversely, diachronic rationality tells us what beliefs, intentions, attitudes, decisions and actions an agent should make/hold *over a period of time* (at different points in time). In other words, diachronic rationality explains how an agent should change her beliefs, intentions, attitudes, decisions and actions over time. We may define diachronic rationality in the following way. Diachronic rationality enables us to retain a certain degree of practical coherence in order to achieve complex or distant goals. The idea is that an agent has to have some sort of practical coherence when acting over time, something that “glues” her intentions and actions across time. In other words, an agent who acts from one moment in time to the next without any connections between those moments, can potentially be charged for the failure of local rationality. Those connections can take various forms in different authors regarding diachronic rationality. The need for a genuine account of diachronic rationality was eloquently and clearly described by John Broome in his book *Rationality Through Reasoning*. His diachronic norm of choice is a minimalist one and is called *persistence of intention*.

*If you have an intention, and then you stop having it, that is irrational unless something licenses you to drop it. You must not just drop it. You could not manage your life if your intentions were liable to vanish incontinently. To bring some intertemporal coherence to our lives, we regularly decide at one time to do something at a later time. But making decisions will not actually achieve coherence unless we generally do as we decide. To decide is to form an intention, and to be effective that intention must persist until we put it into effect* (Broome 177: 2013).

There are several reasons for using Broome's account of diachronic rationality as an example of diachronic rationality in general. The first is the simplicity of the diachronic norm *persistence of intention*. It consists of two concepts: intention and persistence. Intention is a fundamental concept in the philosophy of rationality. In instrumental rationality, intention is defined as *a pro-attitude towards actions having some feature, F, along with the belief that the original action has that feature* (Davidson 5–8: 1963).<sup>12</sup> Persistence in this sense literally means existence of an object through time. These two concepts amount to the minimalistic diachronic norm of rationality known as *persistence of intention*. The second reason is intuitiveness. As Broome remarks, if an agent has an intention and suddenly drops it, there is an intuitive *prima facie* reason to conclude that something went wrong. In Broome's words, we could not manage our lives if our intentions were liable to sudden disappearance. There is an intuitive *prima facie* reason for *some intertemporal coherence to our lives*. The last reason is that the opponents of diachronic rationality, such as Hedden (2015a, 2015b), make their arguments by critiquing Broome's account of diachronic rationality.

Diachronic rationality is supposed to illuminate and offer solutions to a wide array of philosophical problems: from the weakness of will (Holton 2009), procrastination (Andreou 2007), temptation and changing one's mind (Bratman 2012, 2018; Brunero 2012) to self-destructive and addictive behavior (Frederick and Loewenstein 1999) and problems in environmental philosophy (Andreou 2010). What I am interested in are the nature of diachronic agency and the norms that we can ascribe to that agency. This is the thesis that I shall be defending:

---

<sup>12</sup> One of the core presuppositions of my thesis is the notion of intention. There is a vast and complex debate over the definition and nature of intention. For more discussion on the matter see (Davidson 1963, 1970, 1971, 1973, 1978; Anscombe 1963, 1983, 1989; Audi 1973; Bratman 1987, 1999, 2009a, 2009b, 2009c, 2018; Broome 2001; Brunero 2007; Ferrero 2010, 2012, 2014; Grice 1971; Raz 2005; Searle 1983; Setiya 2007a; Tenenbaum 2014, 2018; Velleman 1989, 2007; Wallace 1999).

***Diachronic agency can be rationally assessed in the way in which synchronic agency is assessed.***

There are two presuppositions to my thesis. The first is that we can make rational assessments in a synchronic manner and the second is that we can engage in diachronic agency. The first presupposition is an uncontroversial one. Human beings have the capacity to make rational assessments as well as have the capacity to make moral assessments. This idea can be traced all the way back to Aristotle and it has remained important and relevant till this day (we have numerous contemporary accounts of rationality). That being said, we can ask the following question: what exactly do we mean by *rational assessment*? Although rationality is a complex and difficult concept to define (as is the concept of morality), I will focus my attention on the practical or instrumental aspect of rationality following the tradition from Hume and Davidson all the way to Bratman and Broome. Rationality in this sense is connected to the notions of consistency, coherence and the appropriate response to reasons. The second presupposition is an uncontroversial one as well: human beings have the ability to engage in temporally extended or diachronic agency. Human beings have the ability to make plans, resolutions, projects and promises. We have the ability to choose at a particular time to take into account our previous decisions and future prospects. An interesting and important question arises. Are there any norms that can be applied to our plans, resolutions, projects and promises or, more specifically, to diachronic agency)? In other words, are there any genuine diachronic norms of rationality? My answer to this question is affirmative and my thesis can be viewed as a defense of diachronic rationality or the existence of diachronic norms of rationality. Now, I will present some of the approaches to diachronic rationality and how they relate to one another.<sup>13</sup>

---

<sup>13</sup> For an alternative approach to the categorization of diachronic rationality see (Doody 2019).

## 1.5 Charting the field of diachronic rationality

### 1.5.1 Historical overview

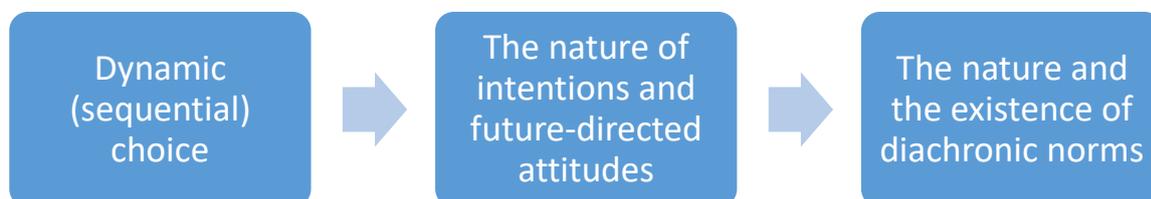
Diachronic rationality has gone through many phases over the course of the last fifty years. There are several disciplines that investigate and explain diachronic rationality: economy, psychology and philosophy. I will address it primarily from the philosophical point of view. We can identify three historical phases in the development of diachronic rationality. Those phases are:

1. Diachronic rationality as the problem of dynamic (sequential) choice (Strotz 1956; Elster 1979; Hammond 1976, 1988; Levi 1974, 1986, 1989; Machina 1989; McClennen 1990; Rabinowicz 1995),
2. Diachronic rationality as the problem of understanding the nature of intentions and future-directed attitudes (Bratman 2010, 2012; Gauthier 1997; Holton 2009; Velleman 2000),
3. Diachronic rationality as the problem of the nature and the existence of diachronic norms (Broom 2015; Carr 2015; Noody 2019; Ferrero 2009, 2012; Hedden 2015a, 2015b; Hlobil 2015; Meacham 2010b; Moss 2015; Podgorski 2016a, 2016b, 2016c 2017).<sup>14</sup>

---

<sup>14</sup> There is an approach to diachronic rationality which is not part of this taxonomy, namely *diachronic Dutch Book arguments* (Ramsey 1926; Briggs 2009; Christensen 1991, 1996; Levi 1997; Maher 1993; Schick 1986; Skyrms 1987; Teller 1976; van Fraassen 1995). The Dutch Book arguments are a series of arguments, dating back to Ramsey in his work “Truth and Probability”, which state that an agent who violates the probability axioms would be subjected to having a book made against him in which case the agent would be guaranteed to lose money. Whether or not this amounts to the failure of rationality is a matter of a vast and ongoing debate (Briggs 2009, Hedden 2013). *Diachronic Dutch Book arguments* are a series of arguments that try to address the problem of how our beliefs evolve over time, i.e., how we should update our beliefs over time. There are two most prominent diachronic norms that are put forward: *conditionalization* (Teller 1976, Lewis 1999) and *reflection* (Van Fraassen 1983). These norms address the problems of belief revision and belief retention, respectively. The reason why these arguments (*diachronic Dutch Book arguments*) and these norms (conditionalization and reflection) are not part of the taxonomy presented above is the following. By addressing the problems of the evolution of beliefs over time, belief revision and belief retention, these discussions belong in the domain of theoretical rationality and not in the domain of instrumental rationality.

**Figure 1**



### **1.5.2 The first phase: Diachronic rationality as the problem of dynamic (sequential) choice**

Diachronic rationality was first interpreted through the conceptual lenses of formal decision theory and game theory. Normative decision theory or expected utility (EU) theory ranges from (von Neumann and Morgenstern 1944) fundamental contributions to Savage’s subjectivist turn to subjective probabilities (Savage 1954) and Jeffrey’s later modifications of expected utility theory Jeffrey (1965). They are all fundamentally synchronic in their approach to decision problems or making decisions over time. The concept of making continuous and temporally extended decisions over time was conceptually incompatible with classic forms of expected utility theory. For this reason, the concept of diachronic rationality was interpreted as the problem of dynamic or sequential choice. Dynamic choice is usually defined simply as a decision problem *in which the agent is called upon to make a sequence of choices over time* (McClennen 1990) and is mostly seen as a problem that needs to be solved. The problem of dynamic choice is a problem of dynamic inconsistency. An agent finds herself dynamically inconsistent when she changes her mind regarding her previous decision, i.e., when the agent abandons her originally established plan of what to choose at a later time. The problems that arise together with the dynamic choice problem are the ones related with the second and the fourth axiom of decision theory, namely the axiom of transitivity and the axiom of independence (substitution). We are sequentially inconsistent when we break the independence axiom. The independence axiom states the following. *Axiom 5. Independence. If A, B, and C are in S,  $A \geq B$  if and only if  $(ApC) \geq (BpC)$ .* (Hastie Dawes 259-260: 2001).

According to the independence axiom, an agent may prefer A to B if and only if she prefers a complex lottery in which she prefers A with some probability to C to B with some probability to C. Whenever we have a decision problem that is “stretched”, there is always a problem of

sequential intransitivity of preferences. In his book *Decision Theory and Rationality*, Bermúdez has demonstrated that dynamic choice (or sequential choice, in his terminology) can always be shown to break the independence axiom (or the substitution axiom, in his terminology) of decision theory (Bermúdez 2009). Different authors have developed different strategies to address this issue. Some of these strategies include the resolute chooser strategy (Machina 1989; McClennen 1990, 1998) and the sophisticated chooser strategy (Strotz 1956). A brief explanation of the two aforementioned strategies is in order. Proponents of the sophisticated choice strategy (Strotz 1956, Hammond 1976, Elster 1979, Schick 1986) claim that the rational thing to do for the sophisticated choosers in the case of a dynamic choice is to foresee their own inability to carry on with their initial plans, thus realizing that they will fall into local irrationality at  $t_1$ . How are they able to do this? They have the ability to put themselves in the shoes of their future selves at  $t_2$  and anticipate the danger of falling into local irrationality. The most famous example of the sophisticated chooser in the literature regarding dynamic choice is the story of Ulysses and the Sirens (Strotz 1956, Elster 1979). As the story goes, the goddess Circe says to Ulysses:

*“First you will come to the Sirens who enchant all who come near them. If anyone unwarily draws in too close and hears the singing of the Sirens, his wife and children will never welcome him home again, for they sit in a green field and warble him to death with the sweetness of their song. There is a great heap of dead men's bones lying all around, with the flesh still rotting off them. Therefore, pass these Sirens by, and stop your men's ears with wax that none of them may hear; but if you like you can listen yourself, for you may get the men to bind you as you stand upright on a cross-piece half way up the mast, and they must lash the rope's ends to the mast itself, that you may have the pleasure of listening. If you beg and pray the men to unloose you, then they must bind you faster.”* (Samuel Butler's translation of *The Odyssey*, Book XII).

Ulysses, being the sophisticated chooser, proceeds by ordering his crew to tie him to the mast of the ship, thus avoiding the danger of local irrationality. There are also less dramatic examples of implementing the sophisticated choice strategy. For example, an agent may intend to go to a bar for only one drink because she has a lot of work tomorrow. She can also be the sophisticated chooser and realize that when she arrives at the bar and has one drink, she will then change her mind and wish to have several drinks. She can anticipate the danger of falling

into local irrationality and not go to the bar in the first place. Proponents of the resolute choice strategy (Machina 1989; McClennen 1990, 1989) claim that the rational thing to do in the case of a dynamic choice is to take into account your previously formed plans when arriving at  $t_2$ . In other words, the resolute choosers are bound by their earlier plans. In McClennen's words:

*The agent can be interpreted as resolving to act in accordance with a particular plan and then subsequently intentionally choosing to act on that resolve, that is, subsequently choosing with a view to implementing the plan originally adopted. In each such case, the plan that is judged most attractive from an ex ante perspective calls for an ex post choice that the agent would otherwise not be disposed to make, but the agent consciously makes that choice nonetheless. In doing this, the agent can be said to act on his previous decision—and in so doing to act resolutely.* (McClennen 157-8: 1990).

Proponents of the resolute choice strategy claim that the rational thing for an agent to do is to “stick to her guns” and choose in accordance with her original plan. Agents should act on their previous decisions and not diverge from their original plans.

### **1.5.3 The second phase: Diachronic rationality as the problem of understanding the nature of intentions and future-directed attitudes<sup>15</sup>**

The second interpretation of diachronic rationality engages the matter in a different light. In this incarnation, authors discuss diachronic rationality in view of understanding the nature of intentions and future-directed attitudes. This interpretation is influenced by the first, in some cases explicitly (Bratman 1999), in other cases implicitly (Velleman 2000) and in yet some other cases those two phases overlap (Gauthier 1997). In other words, authors like Bratman are reacting to the way that making decisions over time was conceptualized by the authors who view diachronic rationality as a problem in formal decision theory and their solutions which are, argues Bratman, crude, inflexible and usually unattainable by real-life agents. The difference between the second and the third phase is in substance and terminology. As stated earlier, Bratman puts focus on the nature of intention instead of consistency in the sets of

---

<sup>15</sup> Some authors think that Bratman and McClennen belong to the same category regarding the way in which they address the nature of diachronic rationality – the pragmatic-instrumentalist view (Ferrero 2009).

preferences at different points in time. By doing so, he presents new ways in which we can discuss *the nature of diachronic agency and diachronic norms* that we can ascribe to that agency. Some authors follow in Bratman's footsteps (Holton 2009, Snedegar 2017), while others, although critical of Bratman, remain heavily influenced by his account (Velleman 2000; Ferrero 2006, 2009, 2012) and vice versa. There is also a difference in terminology between the first and the second phase. In the first phase, authors use concepts like preferences, probability assignments, choices, utilities, lotteries and outcomes (this is the language of decision theory as theory of expected utility). In the second phase, authors use concepts like intentions, attitudes, norms, actions and decisions.

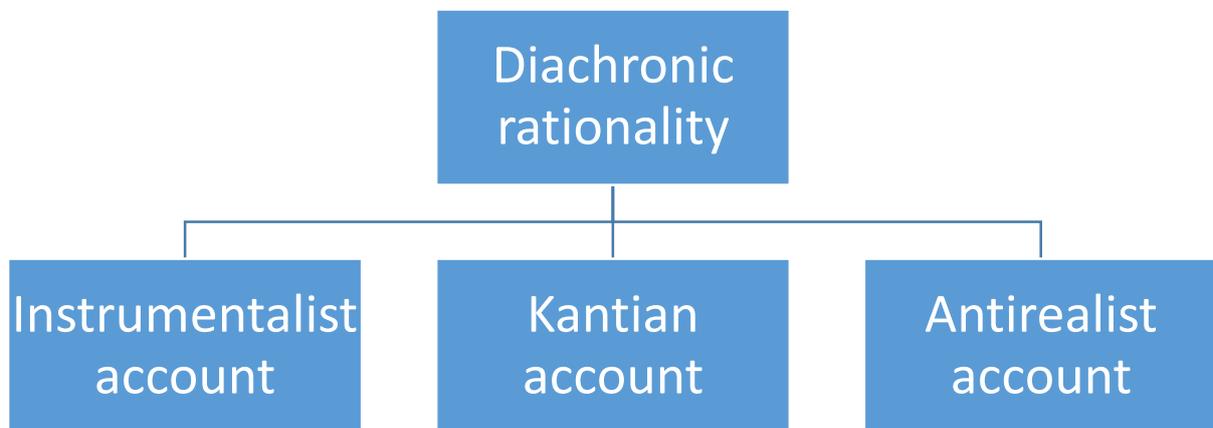
### **1.5.4 The third phase: Diachronic rationality as the problem of the nature and the existence of diachronic norms**

The third interpretation of the problem of diachronic rationality addresses the matter directly: are there any genuine nonreductive diachronic norms of rationality? There are two possible answers to this question: affirmative and negative. We can label authors who answer the former question affirmatively *synchronists* (Hedden 2015a, 2015b; Moss 2015; Doody 2019) and the ones who answer negatively *diachronists* (Ferrero 2009, 2012; Broome 2013, 2015; Bratman 2012, 2018; Podgorski 2016a, 2016b, 2017). There is the possibility of some sort of middle ground regarding the question of diachronic norms. Some authors, namely Doody (2019), can be seen as error theorists regarding diachronic norms, but for the sake of simplicity, we will regard them as synchronists. There are many different approaches in which diachronists engage diachronic rationality. Broome proposes the minimalist norm: *persistence of intention*. Bratman views *plans* as fundamental feature of human agency and proposes *diachronic self-governance* as diachronic norm of rationality. Diachronic self-governance serves as a higher-order intention which should guide our first-order intentions through time. Ferrero, influenced by Velleman, offers *diachronic will*. Diachronic will is *the exercise of rational governance from the time of decision through the time of action* (Ferrero 2007, Velleman 1997). Podgorski has a different approach to diachronic rationality. His claim is that the assessment of rationality must encompass the assessment of agent's reasoning. If we assume that reasoning is a continuous process of temporally extended causal patterns of mental states and, affirms Podgorski, we have good reasons to believe so, then that process should be governed by some

diachronic norms (Podgorski 2016a). Conversely, synchronists claim there to be no genuine nonreductive diachronic norms of rationality. Synchronists came into focus fairly recently with the rise of authors like Brian Hedden who presents a purely synchronic account of rationality which does not call for any form of diachronic norms of rationality (Hedden 2012, 2015a, 2015b).

## **1.6 Contemporary accounts of diachronic rationality: Instrumentalists, Kantians and antirealists**

**Figure 2**



In the contemporary debate regarding diachronic rationality, we can distinguish three distinct approaches: the instrumentalist approach to diachronic rationality, the Kantian approach to diachronic rationality and the antirealist approach to diachronic rationality. The instrumentalist approach to diachronic rationality states that humans have the capacity for diachronic agency and that we can ascribe some genuine diachronic norms of rationality to that agency. According to this approach, human beings are capable of making temporally extended decisions. Temporally extended decisions are those types of decisions in which an agent begins, develops, and completes temporally extended and coordinated activities or projects. These types of decisions are not reducible to a series of time-slice independent (synchronic) decisions. Temporally extended decisions have two basic properties. An agent has to know that it is she who begins, develops and completes a temporally extended decision. Also, an agent has to have backward-looking memory and the prospect for later intentional executing of her temporally extended decision. Temporally extended decisions are possible because the nature of intention

allows us to have future-directed attitudes. Diachronic norms of rationality proposed by instrumentalists are fairly minimalistic: persistence of intention (Broome 2013), partial “fill-in” plans (Snedegar 2017) and diachronic self-governance (Bratman 2018). For this reason, I call those authors instrumentalists.

Conversely, there are authors who are unsatisfied with the minimalistic approach of instrumentalists regarding diachronic rationality. Those authors claim that the instrumentalist approach to diachronic rationality is insufficient to explain human diachronic agency. I call those authors Kantians. According to them, when addressing the problem of diachronic rationality, we need to incorporate concepts such as *will*, *autonomy* and *openness to the future* (Velleman 2000; Ferrero 2006, 2009, 2012; Holton 2009). One of the concepts used to explain diachronic rationality is the concept of will. Will is agent’s ability to control her own actions and choose between alternative actions. In the debate about diachronic rationality, one of the major proponents of will is David Velleman who takes the concept of will directly from Kant. Kant describes will as “practical reason”, as something standing in opposition to pure instincts which animals have (Kant 1956). Velleman defines will as “proper object of awe” which enables us to act, but emphasizes that this is not necessarily a good thing (Velleman 2007). Without entering the discussion on diachronic rationality directly, Velleman has had a great influence on Ferrero’s and Bratman’s accounts of diachronic rationality. Inspired by Velleman, Ferrero constructed the notion of diachronic will. Diachronic will *is the faculty of intentions in its genuinely diachronic dimension and the rationale for it is that it makes possible to engage in activities with a radically novel temporal structure, activities that are not merely continuous over time, but temporally integrated and unified* (Ferrero 403-404, 2009).

Another notion is *diachronic autonomy*. Diachronic autonomy is the concept which arises when we ask the following question. How does an agent act autonomously when abiding by her previous decisions? Some authors consider the concept of diachronic autonomy to be a genuine answer to this problem (Velleman 1997, Ferrero 2010), while others remain determined to argue that there is no problem at all (Nefsky and Tenenbaum 2017). There is also the concept of *openness to the future* which is directly presented by Velleman (1989) and indirectly endorsed by Ferrero (2006, 2009, 2010, 2012). Openness to the future is a property of decisions which, *whenever we face a decision, we feel that our future is partly undetermined and thus leaves something for us to decide* (Velleman 34:1989). Having diachronic will, diachronic autonomy and openness to the future as core features of genuine diachronic agency makes these authors proponents of the Kantian approach to diachronic rationality. We should note that there

is some legitimate common ground between instrumentalists and Kantians.<sup>16</sup> They agree that we engage in diachronic agency for the reasons of scarcity in executive and deliberative resources and an unstable practical standpoint.

Lastly, there are antirealists regarding diachronic rationality. Those authors claim that we can have a comprehensive account of rationality without prescribing to any diachronic norms of rationality. They are also proponents of *Time-slice Rationality* (Hedden 2015a, 2015b; Moss 2015).

### 1.6.1 The instrumentalist account of diachronic rationality

The instrumentalist account of diachronic rationality consists of the modest account of diachronic agency and the use of minimalistic (instrumental) norms of diachronic rationality which are constructed from but not reducible to the synchronic norms of rationality. One of the simplest accounts of diachronic rationality is Broome's account of diachronic rationality. Broome's account of diachronic rationality consists of two parts: the descriptive one and the normative one. The descriptive part states that the nature of our intentions is essentially diachronic. Our intentions *persist* through time and it would be highly strange and unusual for our intentions to suddenly vanish into thin air. If our intentions were liable to disappear at any moment, it would be almost impossible for us to manage our everyday lives successfully. The diachronic feature of intentions brings the much needed intertemporal coherence to agents' lives. The second part of Broome's account is the proposed diachronic norm of rationality. That norm is *persistence of intention*. The norm states the following:

*Persistence of Intention. If  $t_1$  is earlier than  $t_2$ , rationality requires of  $N$  that, if  $N$  intends at  $t_1$  that  $p$ , and no cancelling event occurs between  $t_1$  and  $t_2$ , then either  $N$  intends at  $t_2$  that  $p$ , or  $N$  considers at  $t_2$  whether  $p$ .* (Broome 178: 2013).

*Persistence of intention* is the simplest diachronic norm in contemporary discussion. Arguably, it is derived from the diachronic nature of our intentions. The norm states the simplest form of constraint – the stability of our intentions over time. Broome's proposal rests on the fact that we need to “connect the temporal dots” between our intentions. By contrast, the norm allows us to drop our intentions at any moment of reflection or *reconsidering*. Broome's account of diachronic rationality is fully instrumental for many reasons, the main one being that his

---

<sup>16</sup> This common ground is limited to the comparison between Bratman's and Ferrero's accounts. There is no common ground between, for instance, Broome's and Velleman's accounts of diachronic rationality.

account does not incorporate any non-instrumental concepts, such as personhood, autonomy or diachronic will, which other accounts use in various degrees.

The second instrumentalist account of diachronic rationality is Snedegar's *coarse-grained plans* account. Inspiration for this account is directly taken from Bratman's *planning theory of intention*. On the basis of Bratman's planning norms (which are synchronic), Snedegar presents a diachronic planning norm which he calls the "fill-in" norm. In Snedegar's words:

*Fill In: If at  $t_1$  you have a partial plan to A at  $t_2$ , then by  $t_2$  you ought to have sufficiently filled in that plan.* (Snedegar 601: 2017).

Snedegar proposes a purely diachronic norm sensitive to our cognitive limitations and our inability to predict the future perfectly. The norm states that if an agent has a plan (a *partial plan*) at one point in time, then she *ought to sufficiently fill in that plan* by another point in the future.

The most prominent account of diachronic rationality and the best way to present the instrumentalist view is Bratman's account of diachronic rationality. Bratman's account of diachronic rationality is arguably the most extensive account of diachronic rationality in contemporary discussions about instrumental rationality. As we have acknowledged earlier, Bratman's work over the decades can be viewed in all three phases of diachronic rationality: the dynamic choice problem, the nature of intentions and future-directed attitudes and the nature and the existence of diachronic norms. However, his most prominent contributions are visible in the second phase of diachronic rationality, namely *Diachronic rationality as the problem of understanding the nature of intentions and future-directed attitudes*.

Bratman was the first author who reinterpreted the problem of diachronic rationality as the problem of temporally extended decisions. There are several crucial aspects in Bratman's philosophical opus which are relevant in addressing the problem of diachronic rationality.

Adult human activity is temporally extended (1)

Humans are planning agents (2)

Rational agents should have diachronic self-governance (3)

(Bratman 1987, 1999, 2000, 2012, 2018; numbering mine).

According to Bratman, one of the key features of human activity is that it is extended in time. This does not mean that every human activity necessarily extends through time, but rather that humans have the innate capability to engage in temporally extended decisions. We have previously established what temporally extended decisions are and now we will examine Bratman's example of a temporally extended decision.

*I see my activity of, say, writing a paper, as something I do over an extended period of time. I see myself as beginning the project, developing it over time, and (finally!) completing it. I see the agent of these various activities as one and the same agent—namely, me. In the middle of the project I see myself as the agent who began the project and (I hope) the agent who will complete it. Upon completion I take pride in the fact that I began, worked on, and completed this essay. Of course, there is a sense in which when I act I act at a particular time; but in acting I do not see myself, the agent of the act, as simply a time-slice agent. I see my action at that time as the action of the same agent as he who has acted in the past and (it is to be hoped) will act in the future. In this respect I differ importantly from those nonhuman agents who do not have the resources to understand their own agency as temporally extended (Bratman 43: 2000).*

There are three pivotal moments where Bratman addresses the concept of temporally extended decisions. Firstly, he gives an example of a temporally extended decision – *writing a paper*. There are other examples of temporally extended decisions that have been put forward by Bratman before, namely *writing a lecture, picking up a book at the library, attending a committee meeting* and *picking up a child at school* (Bratman 1987). Agents who engage in this type of agency need to begin, develop and finish their decision over some period of time. Secondly, he points out the relevance of an agent being herself throughout the duration of a temporally extended decision. There is the need for psychological unity of the agent in question for her to engage in a temporally extended decision. Thirdly, Bratman directly denies that an acting agent can be *a simple time-slice agent*. This is important because authors who deny the existence of any diachronic norms of rationality take the view that a rational agent is a time-slice agent (Hedden 2015a, 2015b; Moss 2015). Lastly, the capacity to engage in temporally extended activity is a uniquely human trait. The one that we as humans do not share with our nonhuman fellow creatures.

The second claim that Bratman presents, which has relevance to the discussion on diachronic rationality, is that humans are planning agents.<sup>17</sup> The concept of plan is a key concept in Bratman’s philosophical work. Bratman differentiates between two senses in which we can discuss the concept of plan. In one sense, we can view plan as an abstract structure that an agent

---

<sup>17</sup> There is a significant difference in the use of the concept of plan between authors such as Bratman and proponents of the resolute choice strategy such as McClennen. For McClennen, the concept of plan is simply a term to describe a commitment between two sets of preferences at different times in order to avoid dynamic (sequential) inconsistency. For Bratman, contrarily, the concept of plan is the capacity of intention fundamental for human agency as a whole. For similarities between McClennen’s and Bratman’s accounts of diachronic rationality see (Ferrero 2009).

can represent to herself. In the other sense, we can regard plan as an agent's having a plan of action, i.e., having a certain procedure for her to achieve some goal. In this sense, plans are specific mental states with some sort of commitment to action. Plans can also be defined as the features of intentions resilient to reconsideration (Bratman 1987). For Bratman, plans are the core feature of human intention and agency. Plans are the means with which we organize our own lives and our activities with other agents. Bratman's theory of intention is called *planning theory of intention*. Plans are relevant for the following aspects of human agency: *the very idea of intention, basic features of our agency, important forms of shared agency, and important forms of responsible agency* (Bratman 1: 1987).

According to Bratman, to be a human agent is to be a planning agent. In this way, Bratman departs from the game-theoretical model of diachronic agency which characterized the first interpretation of diachronic rationality (the problem of dynamic (sequential) choice). In a way, he rejects the notion of plan as an abstract representation in the sense used by decision theorists, more specifically, proponents of the resolute choice theory such as McClennen. For Bratman, plans are essential for understanding human agency, the nature of intention and future-directed attitudes. Consequently, this interpretation of diachronic rationality is named *the nature of intention and future-directed attitudes*.

Bratman's third claim addresses the problem of diachronic rationality directly. It states: *Rational agents should have diachronic self-governance*. The thesis deals with the nature and the norms of diachronic rationality. Bratman addresses diachronic rationality by introducing the concept of *self-governance*.<sup>18</sup> For Bratman, self-governance is a way in which we can govern our own lives coherently and consistently, an authority or autonomy that an agent has over her own intentions and attitudes. Self-governance can sometimes be described as policy-like commitments or self-governing policies which guide agent's actions through time. Those policies can be described as *intentions that can also be general and concern potentially recurring circumstances in the agent's life* (Bratman 1987). Some examples of these policies through time might be the following:

---

<sup>18</sup> Bratman borrows the concept of self-governance from Harry Frankfurt and Gary Watson. Although there are similarities between Frankfurt's, Watson's and Bratman's notions of self-governance, they differ in some relevant aspects. Bratman's account differs primarily in terms of use. While Frankfurt and Watson are engaged in discussing the existence of free will, Bratman is engaged in discussing the nature of intention, future-directed attitudes and rationality of action. Frankfurt defines self-governance as human capacity to reconcile second-order and first-order desires, whereas Bratman defines self-governance the norm of rationality which can be synchronic or diachronic in nature. For more see (Bratman 2012).

*Buckle up seat belts when driving a car. No more than one drink when you have to drive home. Never agree to chair more than one committee at a time. Refuse second desserts. Don't let lists get longer than seven items. Check brakes every 6000 miles. Check house insurance yearly. Change furnace filters every two months. Read some German prose every night before going to bed. Don't make important decisions at the end of a long and stressful day* (Bratman 88-89: 1999).

Bratman considers these norms our second-order intentions which help us govern our own lives. For Bratman, diachronic self-governance is a genuine diachronic norm of rationality. These norms are subject to rational requirements of consistency and coherence which are standard instrumental and synchronic norms of rationality. In this way, according to some authors, diachronic self-governance is in danger of being reduced to a set of synchronic requirements. Authors who raise those types of concerns are what I call proponents of the Kantian account of diachronic rationality.

### **1.6.2 The Kantian account of diachronic rationality**

The Kantian account of diachronic rationality has two main components: rejecting of the instrumentalist account as an insufficient account of diachronic agency and arguing for a genuine (substantive) account of diachronic rationality. These components are also reasons why this account is called the *Kantian account*. Kantians (Velleman 1997, 2000, 2003, 2006, 2007; Ferrero 2006, 2009, 2010, 2012, 2014; Holton 2009)<sup>19</sup> agree with instrumentalists (Bratman 1987, 1999, 2000, 2012, 2018; Broome 2013, 2015; Snedegar 2017) that humans have the capacity for diachronic agency. They disagree on the nature of that capacity.

Instrumentalists argue that we have the capacity for diachronic agency because we are fundamentally limited creatures. Humans are limited in many ways: by time, space, cognitive and deliberative resources. We engage in diachronic agency in order to coordinate ourselves at different times (transtemporal coordination) and in order to coordinate with other agents for

---

<sup>19</sup> There are numerous other authors who can be placed in this category, such as Heeney (2019), Hinchman (2003, 2015) and (Hardt 2018). These authors engage with the concept of diachronic rationality by invoking the concepts of *narrative understanding* (Hardt 2018), *trust* and *stable narrative* (Hinchman 2003, 2015) and *entitlement* (Heeney 2020), but here I would like to focus on authors who tackle the problem of diachronic rationality by using the concepts of *will*, *openness to the future* and *diachronic autonomy*, which are arguably most Kantian. Also, as in the previous section where I presented the instrumentalist approach through Bratman's account, I would like to present the Kantian account of diachronic rationality through one author. In this case, it is Ferrero's account of diachronic rationality.

the purpose of achieving complex and distant goals. According to instrumentalists, diachronic agency serves as a “shortcut” or “necessary evil” because of our limited human nature. This shortcut enables us to reduce the cost of ad hoc planning, dynamic or sequential inconsistencies and repetition of deliberation over time.

While some Kantians agree that there are certain benefits to the instrumentalist approach (Ferrero 2006, 2009; Holton 2009), others disagree with it completely (Velleman 2000, 2007). However, they all agree that the instrumentalist account is insufficient to explain human capacity for diachronic agency. The most influential proponent of the Kantian approach is certainly David Velleman.<sup>20</sup> His work has had a significant influence on Ferrero’s account and his critique left a great impact on Bratman’s account. He argued that the investigation on the nature of intention and future-directed attitudes requires something more than the persistence or stability of intention over time. When reflecting on the nature of intentions and future-directed attitudes, what needs to be called on is the concept of will which can meaningfully guide our actions through time. The concept of will can be traced back to Kant in his work *Groundwork for the Metaphysics of Morals (1956)* (Velleman 2007).<sup>21</sup> Velleman has also argued that in exploring diachronic agency (although he does not use that term), we need to invoke the concepts of will, openness to the future and autonomy.

### The notion of will

One of the key notions in the Kantian accounts of diachronic rationality is the concept of will (Holton 2009; Velleman 2000, 2007; Ferrero 2009, 2010, 2012). Firstly, we need to answer the following question. What is *will*? We should also acknowledge that the concept of will is traditionally used in discussions about the problem of *free will*. The exercise of *will* is standardly considered through a four-stage model.

*(i) Deliberating: Considering the options that are available, and their likely consequences; getting clear on one’s own desires, and one’s own prior plans and intentions; seeing how the options fit in with these desires and plans; establishing pros and cons.*

*(ii) Judging (deciding that): Making a judgement that a certain action is best, given the considerations raised in the process of deliberation. The upshot of the judgement is a belief.*

---

<sup>20</sup> It should be noted that Velleman does not address the concept of diachronic rationality directly. This is one of the reasons why the Kantian approach is presented through Ferrero’s account.

<sup>21</sup> Although, Velleman is mostly influenced by the work of Harry Frankfurt.

(iii) *Choosing (deciding to): Deciding to do the action that one judged was best. The upshot of this decision is an intention.*

(iv) *Acting: Acting on the intention that has been made, which involves both doing that thing, and coordinating other actions and intentions around it.* (Holton 57: 2009)

Will, as described above, is simply our capacity to achieve what we want based on our beliefs and desires. Will is the “glue” which connects our deliberation with judging, our judging with choosing and our choosing with acting. Presented in this way, there is no diachronic notion of will. This is the reason why Holton is unsatisfied with this conception of will. Holton then presents that which he calls *the willpower account*.

*Action is not determined just by the agent’s beliefs, desires and intentions. In addition, willpower plays an independent contributory role. Agents whose willpower is strong can stick by their resolutions even in the face of strong contrary desires; agents whose willpower is weak readily abandon their resolutions even when the contrary desires are relatively weak* (Holton 113: 2009)

Holton argues that the agent’s beliefs, desires and intentions are inefficient to explain her actions. In the case of temptation or weakness of will, agents have the ability to resist their current (temporary) change of preferences and stick to their initial intentions. Matter of fact being that it is rationally possible for an agent to choose contrary to her current preference. In order for her to do this, she needs her willpower.

Contrarily, Bratman has a more modest view of will. He claims that the concept of will is important in understanding intelligent action. For Bratman, “will” is simply *a complex of capacities for forming, changing, retaining, and sometimes abandoning our choices and intentions* (Bratman 97: 1994).<sup>22</sup>

### **1.6.2.1 Ferrero’s account of diachronic rationality**

Ferrero’s account of diachronic rationality is the account which can be characterized as Kantian in the full sense of the term. Ferrero rejects Bratman’s account as an insufficient way of describing human ability to engage in genuine diachronic agency. While instrumentalists

---

<sup>22</sup> Bratman uses Alan Donagan’s conception of will. For more see (Donagan 1985).

(Bratman 2012, 2018; Broome 2013, 2015) are concerned with the stability of intention over time, coherence of agent's attitudes over time and persistence of intention over time, Ferrero, mostly under the influence of Velleman's work (Velleman 1997, 2000, 2003), focuses on something he describes as genuine diachronic agency.

Ferrero claims there to be three ways in which we can conceptualize and engage with diachronic agency. In his 2006 paper, *Three ways of spilling ink tomorrow*, he argues for a genuine case of diachronic agency. Ferrero uses the following example. A philosopher has a scheduled lecture that she is presenting tomorrow at the university. Today, she intends to spill ink during tomorrow's lecture in order to illustrate the central point of her presentation.<sup>23</sup> According to Ferrero, there are three distinct ways in which the philosopher's current decision can bring about her future conduct. The first way is by a causal mechanism, the second is by precommitment and the third is by exercising diachronic autonomy.

#### Spilling the ink by a causal mechanism

The first way in which the philosopher can spill the ink tomorrow is by setting in motion a mere causal mechanism, be it external or internal to her body, which will induce her to spill the ink tomorrow. This way of spilling the ink lacks any autonomy or governance from the agent because she will be under physical constraint to spill the ink tomorrow. She has some sort of indirect control to tamper with the mechanism which has been initiated. An example of this would be Ulysses' tying himself to the mast of the ship in order not to be infatuated by the song of the sirens. Ferrero claims this to be an insufficient way of explaining human ability to engage in diachronic agency because it lacks autonomy and general appreciation for agent's attitudes across time.

#### Spilling the ink by precommitment

The second way in which the philosopher can spill the ink tomorrow is by precommitment. According to Ferrero, precommitment is *a causal device, physiological mechanism, institutional or social practice that induces the agent at the time of the action f to choose to do as she has originally decided* (Ferrero 98:2006). There are numerous ways in which an agent

---

<sup>23</sup> Ferrero's inspiration for this example is most probably John Longshaw Austin's paper titled *Three ways of spilling ink* (1966).

can “precommit” herself to act according to her original decision. An agent can precommit herself to stop smoking by taking prescription pills which in turn make smoking less satisfying. Also, an agent can decide to save money in a bank account which penalizes early withdrawals in order to compel herself to save money until retirement.<sup>24</sup> The cases of precommitments have two main characteristics. The first characteristic is that the agent has only *indirect* control over her future actions and is in some sense coerced by her past self. The second characteristic is that the agent usually needs to find some sort of *self-manipulation* tools in order to engage in precommitment. Both of those cases disqualify precommitment as genuine form of diachronic agency.<sup>25</sup>

### Spilling the ink by *diachronic autonomy*

The third way in which the philosopher can spill the ink tomorrow is by engaging in genuine diachronic agency. The way she is able to do this is by acknowledging at each relevant moment between today and tomorrow that her decision to spill the ink is a choice worthy on its own merits. She is going to spill the ink tomorrow because of her past decision to do so, but she is not bound or coerced by that decision. This makes the tomorrow spilling of the ink a genuine future-directed decision because it is not grounded in physical constraint, self-control or manipulation but, by contrast, in agent’s *diachronic autonomy*. This amounts to a particular kind of temporally extended agency which takes into account the agent’s past decisions, possible reconsideration over time and her autonomy at the time of the action. Ferrero calls this kind of genuine diachronic agency *autonomous intentional diachronic agency* or AIDA (Ferrero 2006).

### Diachronic will and narrative

Ferrero elaborates and expands his account of diachronic rationality with a view he calls the *constitutive view of diachronic rationality* (Ferrero 2009). The constitutive view can be regarded as an expansion of AIDA. There are two main components to the constitutive view. The first is the notion of *diachronic will* and the second is *narrative proneness*. Diachronic will

---

<sup>24</sup> Ferrero regards cases such as Ulysses’ being tied to the mast of the ship as cases of physical constraint and not as genuine cases of precommitment. For more see (Elster 1979).

<sup>25</sup> There are some cases in which engaging in precommitments or distal self-control can serve as a surrogate for genuine diachronic agency. For more on those specific cases see (Ferrero 2006).

is the faculty of intention in its genuine diachronic dimension. Diachronic will allows us to engage in purely diachronic activities. Those activities have an internally constituted temporal unity and

*(i) are made up of momentary actions that relate in non-local ways that span over the entire length of the activities;*

*(ii) require the agent's continuous appreciation of the structure and outcome of the extended activities taken as a whole: At each moment the agent is expected to take whatever momentary step is required by the activity's global structure rather than by her proximal concerns (Ferrero 406: 2009).*

Ferrero juxtaposes these activities to the activities such as igniting a fuse in order to cause an explosion. This is some kind of a simple diachronic activity but it is not the right kind of diachronic activity. The reasoning behind is that this kind of activity does not require any relevant future-directed intention from the agent. It only requires for the agent to presently light the fuse and passively wait for the causal chain to lead to the explosion. What is missing from this activity is the rational governance from the time of the decision making to the time of the action or, simply put, diachronic will (Ferrero 2009).

Ferrero takes the concept of diachronic will from the works of David Velleman. While Velleman addresses the concept of diachronic rationality (at least indirectly) through the cognitive view<sup>26</sup>, Ferrero emphasizes the ontological importance of diachronic will. The ontological importance of diachronic will is characterized by its enabling the agents to engage in activities that are distinctively *narrative prone*. Narrative is a way of conceptualizing a series of events as a single act which consists of distinct but interconnected parts. Narrative is not a simple recounting of events on a certain matter in chronological order. By contrast, narrative is a specific way of connecting a series of events which allows for a synoptic view of events. It allows agents to engage in narrative activities representing a dynamically coherent whole which extends both in the past and in the future. When engaging in narrative activities, agents can emphasize certain events, glance over others or entirely skip some events. The use of a narrative aims at increasing the visibility of the underlying structure of our temporally unified activities. Ferrero stresses that the best way to understand our diachronic agency is by narrative

---

<sup>26</sup> Just a few words on Velleman's cognitive view. Velleman argues that intentions are self-fulfilling predictions. This means that the agent is motivated by her desire to understand what she does and acts on the basis of such a self-fulfilling prediction in order to understand herself, i.e., in order to acquire self-knowledge. Will is not something necessarily good, but that which enables us to satisfy our desire for self-knowledge. For more on Velleman's account see (Velleman 1997, 2000, 2003, 2006, 2007).

devices. Examples of temporally extended activities best explained in narrative terms include: *engaging in conversations, making music, cultivating personal relationships, playing strategic games, storytelling, running businesses* (Ferrero 2009).<sup>27</sup>

We have examined Ferrero's account of fully fledged diachronic agency, but in order to properly address the concept of diachronic rationality, we need to explore what kind of norms or requirements Ferrero's account proposes for genuine diachronic agency. The norm presented in this account is the following:

#### *A Constraint of Temporal Unity*

(R1) *An agent A ought to (if she intends at  $tn$  ( $1 < n < 10$ ) to  $\varphi$  in the interval  $t1-10$ , then she believes at  $tn$  that—assuming she continues to have the ability and opportunity to  $\varphi$ —she will continue to autonomously appreciate and approve of her  $\varphi$ -ing until  $t10$ ).*<sup>28</sup> (Ferrero 147: 2010).

There is a lot to unpack here, so we need to take one step at a time. Firstly, we need to acknowledge that this norm is substantially different from *persistence of intention* (Broome 2013) and *diachronic self-governance* (Bratman 2012, 2014, 2018) because it presupposes the agent's ability to engage in temporally unified activities as described in Ferrero (2006, 2007, 2010). The norm presents a certain amount of rational pressure on the agent who engages in temporally unified activities, if she intends to  $\varphi$  at some point in time ( $tn$ ), to continue to  $\varphi$  through the temporally extended activity until  $t10$ . The preconditions are twofold. The agent needs to have the *ability and opportunity* to continue to  $\varphi$ . This is the instrumental requirement equivalent to Bratman's *means-end coherence* and *constancy of intention over time* which are, according to Bratman, the purely synchronic constraints on the rationality of action (Bratman 1987, 1999, 2012). What differentiates this norm from the proposed norms of Broome (2013), Bratman (2012, 2018) or other instrumentalists like Snedeger (2017) is insistence on diachronic autonomy and the agent's continued appreciation and approval for her action across time. This fact makes *A Constraint of Temporal Unity* a weak diachronic norm at best and entirely

---

<sup>27</sup> The claim about the importance of diachronic will goes even farther. Ferrero argues that even when we engage in activities that we normally consider purely static (or synchronic), such as admiring a gothic cathedral or grasping timeless abstract entities like mathematical theorems, we do so in virtue of diachronic will because we are inherently restricted and limited temporal beings. For more see (Ferrero 2009).

<sup>28</sup> I should note that Ferrero expands on constraint (R1) by introducing constraint (R2). He does so by treating intentions as *summary attitudes* and then by introducing the rational constraint on intentions as preserved summary attitudes. In the end, (R2) suffers from the same kind of problems as (R1), which is evident from the conclusion of Ferrero's paper. This is the reason why I will focus solely on the (R1) constraint. For a more detailed discussion between the (R1) and the (R2) constraint see (Ferrero 2010).

reducible to the synchronic level at worst. Ferrero explicitly takes a middle ground when arguing for this norm. He concedes that the norm, *A Constraint of Temporal Unity*, is sadly synchronic but that it has distinctive diachronic elements and serves as a constraint on diachronic practical rationality (Ferrero 2010). Ironically, this is a great segue into the third and last contemporary approach to diachronic rationality – the antirealist account – which argues all norms of rationality to be either genuinely synchronic or reducible to the synchronic level.

### **1.6.3 The antirealist account of diachronic rationality**

We can define the antirealist account of diachronic rationality in the following terms. Any account of rationality that has in its repertoire a claim that if something is a requirement of rationality, then that requirement is necessarily synchronic. According to this view, if there are some norms that appear to be diachronic norms of rationality, they are either reducible to the synchronic norms of rationality (in other words, they are not genuine diachronic norms of rationality) or are some sort of social norms (in other words, they are not genuinely norms of rationality). There are several accounts of rationality that fall under the antirealist category regarding diachronic rationality. One of them is certainly Feldman and Conee's account of *evidentialism*. According to this account of evidentialism, only one doxastic attitude towards a proposition is justified for a person *at a time*, this being the function of one's evidence (Feldman and Conee 1985, Feldman 1995, Feldman 2003). But this account does not address the concept of diachronic rationality or diachronic norms directly.

The most prominent account of diachronic rationality is *Time-slice Rationality*. Time-slice Rationality was introduced by Brian Hedden in the eponymous groundbreaking paper published in 2015. Ever since then, there has been a vigorous debate raging around the question of the existence of genuine diachronic norms of rationality. Some authors are taking Hedden's Time-slice Rationality and expanding it into the fields of epistemology (Moss 2015) and self-locating beliefs (Builes 2019), while others firmly oppose this account of diachronic rationality (Titelbaum 2015, Woodard 2019, Carr 2015, Snedegar 2017, Doring and Eker 2017, Lenman 2017). Time-slice Rationality is a complex in-depth account of how agents can be rational without referring to any diachronic norms. But before we delve into the account itself, let us start with an example as to have a clearer picture of what the proponents of Time-slice Rationality discuss. The example is about *fickle Fran*.

## A Puzzle About Fickleness

*Fickle Fran: Fran is a history buff, who has studied conflicting theories about how Amelia Earhart died. She thinks the most plausible three theories are that she died in a plane crash, as a castaway, or in Japanese captivity. On Monday, she is convinced that Earhart died in a plane crash. However, on Tuesday, she accepts the castaway theory. But she doesn't stop there: by Wednesday, she changes her mind again, now holding that she died in captivity. Fran continues to cycle through these theories, despite neither gaining nor losing—much less forgetting—any first-order evidence. Moreover, at a certain point, it's not the case that she changes her mind in virtue of noticing something new about the evidence. None the-less, she keeps changing her mind, eventually settling on one of the theories. (Woodard 2: 2019).*

Firstly, some clarifications are in order. The example originates from the paper, *Diachronic Normativity: A Puzzle About Fickleness*, in which Woodard criticizes the Time-slice Rationality account.<sup>29</sup> The reason for using this particular example is because we can clearly see the intuitions and point of view of the time-slice theorists about the nature of rationality. Fran has different attitudes regarding the same thing (namely the death of Amelia Earhart) on Monday, Tuesday, Wednesday and so on. There is no new evidence or information that Fran has acquired along her change of attitudes and there is certainly nothing that connects them diachronically. According to the time-slice theorists, Fran is perhaps fickle but nevertheless completely rational. The intuition of the time-slice theorists is that there is certainly a rationality constraint in not holding contradictory beliefs at a single time. Believing P and not P at *a single time* is certainly irrational. But there is nothing irrational about believing P and not P at *different times*. This claim is situated in the domain of theoretical rationality and the case could be made that it has little to no bearing in the discussion about the notions of instrumental rationality. But the time-slice theorists claim there to be no genuine diachronic norms regarding theoretical and instrumental rationality. Let us take a look at another example.

Sartre's pupil

---

<sup>29</sup> Similar examples can be found in Hedden (2015a, 2015b) and Tiltebaum (2015).

*You have to choose between fighting the Nazis or tending to your sick mother. There are pros and cons to each. You care about various things, and you haven't a clue as to how to weigh them off against each other. You ask your French philosophy professor for advice, but he's no help. You decide to fight the Nazis. You complete your basic training. But then you reconsider and return to your mother.* (Doody 21: 2019).

Once more, clarifications are in order. This example is taken from the paper, *If There Are No Diachronic Norms of Rationality, Why Does It Seem Like There Are?*, where Doody excepts the Time-slice Rationality account whilst offering a unique take on it. The reason for using the example above is because it presents a simple way of displaying the intuitions that the time-slice theorists have in regard to instrumental rationality. In the example, the agent has a choice either to go to war or remain home and care for his mother.<sup>30</sup> He initially decides to fight in the war against the Nazis but at some point has a change of heart and returns home to take care of his mother. We can illustrate this example as a decision problem with four possibilities.

A (Fight in the war and not care for the mother)

B (Care for the mother and not fight in the war)

C (Care for the mother after a delay)

D (Fight in the war after a delay)

We can safely put aside the problem of choosing between the mutually exclusive alternatives A and B. Although some authors consider the problem of incommensurable alternatives<sup>31</sup> to be form of a genuine diachronic problem (Andreou 2016), which the choice between A and B certainly is, we can ignore that problem for a moment and focus on choice C which the agent in our example has chosen. In scenario C, the agent chooses to fight in the war, but then changes his mind and returns home to care for his sick mother. Choice C is certainly inferior to choice B. Choice B is a scenario in which he would care for his mother from the first day. By choosing C, the agent certainly lacks stable plans and inefficiently uses his available time. Yet, the main question that the time-slice theorists will ask is the following. Does the agent behave irrationally in scenario C? The answer for the time-slice theorists is a definitive no. The agent does not behave irrationally because he is not violating any norms of rationality. What the agent

---

<sup>30</sup> In the original text, Sartre states that he is talking about a male pupil. For more see (Sartre 1965).

<sup>31</sup> For discussion on the problem of incommensurable alternatives see (Raz 1986, Broome 2000).

experiences in the example above is diachronic tragedy (Hedden 2013) or diachronic misfortune (Doody 2019). Although life is full of misfortunes and tragedies, that does not make protagonists of those tragedies irrational agents.

### 1.6.3.1 Time-slice Rationality

Time-slice Rationality is an account of theoretical and practical rationality introduced by Brian Hedden in his groundbreaking paper *Time-Slice Rationality* (2015a) and subsequent book *Reasons Without Persons* (2015b). While it draws from various authors (Stalnaker 1999, Arntzenius 2003, White 2005, Feldman 2007), it presents a genuinely new take on the concepts of theoretical and instrumental rationality. The Time-slice Rationality account consists of two theses: *Synchronicity* and *Impartiality*.

*Synchronicity: All requirements of rationality are synchronic* (Hedden 8: 2015b).

*Impartiality: In determining how you rationally ought to be at a time, your beliefs about what attitudes you have at other times play the same role as your beliefs about what attitudes other people have* (Hedden 9: 2015b).

According to Hedden, there are two main reasons why we should accept Synchronicity: internalism and the problems of personal identity. Internalism in epistemology is a view which states that what it means to be a rational agent (or have justified beliefs) is internal to the agent. There is a number of ways in which we can address the concept of internalism in epistemology, but the two most prominent ways of conceptualizing epistemic internalism include *mental internalism* and *access internalism*. Access internalism claims that agent's rationality or irrationality depends on the facts that supervene on those mental states which the agent has *access* to. Conversely, mental internalism asserts that agent's rationality or irrationality depends on the facts that supervene<sup>32</sup> on her mental states (Carr 2015). Hedden subscribes to the latter form of internalism.

---

<sup>32</sup> The etymological roots of the concept of supervenience originate from the Latin words "super" which means *above* and the verb "venire" which means *to come*. In the non-philosophical discourse, it is mostly used in the temporal context, while in the philosophical discourse, it is used non-temporally. The concept of supervenience is defined as metaphysical and/or conceptual determination-relation (Horgan 1993, 1982). The relation states that a set of properties A supervenes on a set of properties B if and only if a change in the set of properties A effects a change in the set of properties B (Bennett and McLaughlin 2018). The concept of supervenience is used

The reason why Hedden accepts mental internalism is the following. When we explore rationality, we are concerned with agent's perspective on the world or, more precisely, how agent's beliefs and actions are sensible in relation to her point of view about the world. Agent's point of view or agent's perspective on the world is one of the core features of rationality. In broad terms, Hedden claims that rationality is a matter of believing and behaving in ways that are sensible given your perspective on the world (Hedden 2015b). The second reason for accepting Synchronicity is personal identity. According to Hedden, authors who argue in favor of diachronic norms are forced to accept some ontological account of personal identity. The problems of personal identity are extremely problematic and should have no bearing on the discussion about rationality.

Impartiality is the core feature of Hedden's account of rationality. The impartiality constraint is the *time-slice* in the *Time-slice Rationality* account and is even implicitly present in the title of Hedden's book *Reasons Without Persons*. The impartiality constraint has two parts. The first part states that agents should be reduced to a set of temporally located decision-making units. The second part asserts that an agent's beliefs at other times should have the same rational force on her as other agents' beliefs at the present time. The reasons for accepting Impartiality are the same as the reasons for accepting Synchronicity: internalism and the problems of personal identity. Internalism, according to Hedden, goes hand in hand with Impartiality. Norms of rationality need to be such that the agent is able to follow them. The agent does not have the ability to "act" in the past or the future. On the contrary, the agent has only the ability to act in the present, so the norms of rationality should be applied to the agent only in the present. The second reason are the problems of the personal identity. According to Hedden, any account of diachronic rationality necessarily has to invoke an account of personal identity. In conclusion, the set of accounts of rationality which does not hinge on the problems of personal identity is superior to the set of accounts of rationality which does hinge on the problems of personal identity (Hedden 2015b).

---

in various philosophical fields of the analytic tradition, from metaethics, aesthetics and the philosophy of mind to epistemology, metaphysics and logic. For early discussions on the concept of supervenience see (Moore 1922, Hare 1984) and for contemporary discussion see (McLaughlin 2001, Bricker 2005, Moyer 2008, Bliss 2014).

## 1.7 Chapter summary

In this chapter, I established the theoretical grounds for the discussion about diachronic rationality. There are four main points that I addressed in this chapter: broader conceptions of rationality in which diachronic rationality is situated; definitions and main problems of diachronic rationality; historical development of diachronic rationality; and lastly, overview of contemporary accounts of diachronic rationality. Firstly, I positioned the scope of my research on diachronic rationality in the domain of instrumental or practical rationality. Instrumental or practical rationality is concerned with agent's intentions and actions (contrary to theoretical rationality which is concerned with agent's full-blown beliefs). Instrumental rationality is concerned with practical coherence, means-end coherence, plans and defeasibility. Secondly, I formulated my thesis: *Diachronic agency can be rationally assessed in the way in which synchronic agency is assessed*. Thirdly, I presented a novel historical overview of diachronic rationality in three phases: 1) *Diachronic rationality as the problem of dynamic (sequential) choice*; 2) *Diachronic rationality as the problem of understanding the nature of intentions and future-directed attitudes*; and 3) *Diachronic rationality as the problem of the nature and the existence of diachronic norms*. Fourthly, I offered a novel overview of the contemporary debate regarding diachronic rationality. I introduced three relevant camps in the discussion about diachronic rationality: instrumentalists, Kantians and antirealists. All of these components will serve relevant purpose for later chapters where I discuss the problems of diachronic rationality in depth.

## 2 A case for diachronic rationality: Intentions, Plans and Self-governance

In this chapter I will explore Bratman's account of diachronic rationality. This will serve two main purposes. The first purpose is to present the account that supports my thesis. The second purpose is to lay the groundwork for future investigation in the domain of diachronic rationality. This chapter will consist of two general parts: of the presentation of Bratman's *planning theory of intention* and the display of Bratman's account of *diachronic self-governance*. Both parts support my thesis; ***Diachronic agency can be rationally assessed in the way in which synchronic agency is assessed.***

The philosophical work of Michael E. Bratman can be viewed as the most extensive and comprehensive account of diachronic rationality.<sup>33</sup> As described in the previous chapter we can historically track the development of diachronic rationality through three phases: *diachronic rationality as a problem of dynamic (sequential choice)*, *diachronic rationality as a problem of understanding intentions and future directed attitudes*, and *diachronic rationality as the problem of the nature and the existence of diachronic norms*. Bratman's philosophical influence can be seen through all of these phases. According to Ferrero, Bratman and McClennen are prime examples of what he calls the *pragmatic-instrumentalist view of diachronic rationality*. They are concerned with agent's long-term satisfaction of her preferences, transtemporal coordination and dynamic inconsistencies (Ferrero 2009). Also, *commitment* is a concept that plays a large role in Bratman's account which makes his work similar to the works of Elster (1984, 1999, 2000), McClennen (1990, 1997, 1998) and Gauthier (1986, 1994).<sup>34</sup> Bratman was somewhat concerned with the problem of *intransitivity of preferences* which arises when agents fail to have consistent preferences over time. This is also consistent with the works that the authors of the first phase (Strotz 1956; Elster 1979;

---

<sup>33</sup> Bratman did not start his philosophical career with the intention to build an account of diachronic rationality. His intention was to build a new theory of intention and an account of practical rationality. Bratman has met with a lot of criticism of his work. Criticism is vast and diverse, but it can be broadly summarized as the inability of Bratman's account to offer a substantial account of rationality and its being focused on merely "mental tidiness". For more on criticism of Bratman's work see (Broome 2001, Raz 2005, Wallace 2006, Setiya 2007, Velleman 1999, 2000, 2007). In response to this criticism, Bratman has reinterpreted his philosophy from an account of practical rationality to a full-fledged account of diachronic rationality. For the full display of this transition see (Bratman 2009, 2010, 2012, 2013, 2014, 2017).

<sup>34</sup> This refers mostly to Bratman's early work. For more see (Bratman 1987).

Hammond 1976, 1988; Levi 1974, 1986, 1989; Machina 1989; McClennen 1990; Rabinowicz 1995) discussed.<sup>35</sup> However, the similarities between those authors are few and rare. For instance, Bratman was never trying to solve the problem of dynamic (sequential) choice or the problem of dynamic inconsistency which arises when agents violate the independence (substitution) axiom.<sup>36</sup> Bratman's greatest contribution to the field of diachronic rationality is visible in the second phase, named *diachronic rationality as a problem of understanding intentions and future directed attitudes*, which he, arguably, founded by introducing *the planning theory of intention*. Bratman's account of diachronic rationality is present in the third (contemporary) phase of diachronic rationality, named *diachronic rationality as a problem of the nature and the existence of diachronic norms*, in which argues for diachronic self-governance as a genuine diachronic norm of rationality. Regarding the contemporary debate, Bratman can be seen both as an instrumentalist and as a Kantian in some regard. On the one hand, Bratman is clearly an instrumentalist because of his approach to the nature of intention, action and minimalistic norms of rationality such as *Means-end coherence* and *Stability of intention*. On the other hand, in recent years he has changed his views regarding agent's intentions and actions focusing more on agent's *practical stand point, self-awareness* and *self-governance*.<sup>37</sup>

In his early work, Bratman decided to tackle some of the fundamental concepts in the philosophy of mind, philosophy of action and philosophy of instrumental (practical) rationality. Those concepts include: the concept of action (Bratman 1978), the concept of intention (Bratman 1979) and the concept of practical reasoning (Bratman 1979). At the time, the dominant views regarding these concepts were the accounts of Donald Davidson<sup>38</sup> and Elisabeth Anscombe. For Davidson, intention is a *primary reason* or a pro-attitude towards actions having some feature, F, along with the belief that the original action has that feature (Davidson 5–8: 1963). For Anscombe, intention is a concept that can be used in three different senses: *the intention with which someone acts, intentional action* and *intention for the future* (Anscombe 1963). According to Bratman, both of these accounts and any similar accounts are wrong. They are insufficient to explain the nature of intention, how we act and the process of

---

<sup>35</sup> This refers mostly to Bratman's work in *Faces of Intention: Selected Essays on Intention and Agency*. For more see (Bratman 1999).

<sup>36</sup> See Chapter 1 *Diachronic rationality as a problem of dynamic (sequential) choice*.

<sup>37</sup> This change is the result of the criticism from people like John Broome, Joseph Raz and R. Jay Wallace and the influence from people like David Velleman, Luca Ferrero and Richard Holton. For more see (Broome 2001, Raz 2005, Wallace 2006, Velleman 1997, 2000, 2003, 2006, 2007, Ferrero 2006, 2009, 2010, 2012, 2014, Holton 2009).

<sup>38</sup> Davidson was somewhat of a mentor to Bratman.

practical reasoning or deliberation. The reason why those theories are insufficient is because they reduce intentions and future-directed intentions to desires and beliefs. In Bratman's words:

*The desire-belief theory of intention in action: we understand intentional action, and action done with an intention, in terms of agent's desires and beliefs, and actions standing in appropriate relations to those desires and beliefs.* (Bratman 6: 1987).

The desire-belief theory of intention in action is regarded to be plausible and explanatorily powerful. The main point of the theory is that we can regard that agent's desires and beliefs at a certain times provide her with reasons for acting in a certain way. Agent's intentional action should be at least strongly supported by her reasons (which are based on desires and beliefs) as any of its alternatives. Reasons based on agent's desires and beliefs have direct relevance on her intentional action. There are several reasons why *desire-belief theory of intention in action* fails as a model of explanation of human behavior and action.

The first problem regarding the *desire-belief theory of intention in action* is the following. Bratman presents it in the form of a question: Why do agents bother forming intentions about the future, i.e., why do we have future-directed intentions? Why do we not, poetically speaking, simply cross our bridges when we come across them?

The second problem regarding the *desire-belief theory of intention in action* is *Buridan's ass paradox*.<sup>39</sup> We can present *Buridan's ass paradox* in the following way.

A hungry mythical donkey (ass) stands between two identical hay piles. The donkey always chooses whichever pile of hay is closest to him. Both piles, identical in every way, are exactly the same distance apart, one on his right, one on his left. Which pile of hay will the donkey choose to eat? The answer is neither because the donkey has no sufficient reason to make any choice in this scenario. The tragic consequence of the paradox is that the donkey is paralyzed by this situation and starves to death. But the tragedy extends beyond mythical donkeys into our everyday decision-making process. For example, when we are buying milk in a local supermarket we are presented with a similar dilemma.

The third issue regarding the *desire-belief theory of intention in action* is the consistency problem. According to Bratman, we can have inconsistent desires but the same does not hold for intentions. One can rationally *desire* to spend the following weekend on a hiking trip and

---

<sup>39</sup> The paradox was named after a 14th-century French philosopher Jean Buridan. The original intent of the paradox was to offer an argument in favor of the existence of free will as opposed to determinism in a free will-determinism debate. Different versions of this argument can be seen throughout the history of philosophy. For historical overview of the paradox see (Rescher 2005) and for contemporary discussion see (Weintraub 2012).

simultaneously have a *desire* to finish this chapter at the same time, but one cannot rationally *intend* to spend the following weekend on a hiking trip and simultaneously *intend* to finish this chapter at the same time. In the latter case the agent can be accused of rational inconsistency.

## 2.1 The planning theory of intention<sup>40</sup>

### 2.1.1 Planning capacities and examples

The main claim of the planning theory of intention is that human beings are planning agents. As human being we have the capacity to engage in the activity of planning or to planning ahead. Our planning activities can potentially be of any length. Here are some examples that are different in length.

*Sally, who is writing a paper, plans to take a coffee break in fifteen minutes. Amy is planning what to have for lunch today. John plans to meet with his colleague at the end of the week. Mark plans to buy a gift for his wife's birthday which is four months from now. Steven plans to buy a house in five to ten years. Donna is planning her retirement years. Zachary is making a will and planning how to pass his possessions to his family and friends.*

Our ability or capacity *to plan* is essential to us as human beings. It is this ability (among others like language capacity, reflection, higher-order cognition, etc.) that separates us from other animals. Although animals can solve some extremely complex tasks, use tools, have reasoning capacities and even have (at least limited) self-control of their actions, they cannot *make plans*. This point also stands for small children and early hominids (before the capacity for speech). Our capacity to form and execute plans stems from two general needs that we have as human beings: the need for deliberation or practical reasoning and the need for coordination. Our ability to deliberate would be of minimal use to us if we did it only moments before the time of action arrives. In order to use our deliberate capacities in their entirety, we deliberate in advance, i.e., we plan. The second need that we as human beings have is the need for coordination. We can distinguish between two types of coordination: coordination that we have with ourselves at different times (intrapersonal coordination) and coordination that we have with others (interpersonal coordination). Because we are limited creatures, both cognitively

---

<sup>40</sup> Bratman's intentions for this theory were extremely ambitious. He claimed that his theory would shed some light on problems in the philosophy of mind, philosophy of action, theory of rationality and moral philosophy. Although Bratman's planning theory of intention was revolutionary in several fields of philosophy, namely philosophy of action and philosophy of instrumental rationality, it was not very influential in other fields such as philosophy of mind.

and materially, in order to achieve complex and temporally distant goals, we equally need both types of coordination – intrapersonal and interpersonal. This is what makes us temporally extended and social agents.

Bratman offers his *planning theory of intention* as a new way of explaining connections between our mental states, intentions and actions. As we have seen at the beginning of this chapter, the *desire-belief theory of intention in action*, according to Bratman, fails to account for agent's real-life practical decision making and intentional action. Bratman presents intentions as central and distinct mental states that motivate and explain intentional action.

Intentions are specific and irreducible mental states that explain how and why agents act in the real world. They are specific because they explain why the agent does what she does and they are irreducible to desires or beliefs. This is something that the *desire-belief theory of intention in action* cannot accommodate for. We have a certain set of desires at one point in time and a certain set of desires at some other point in time. There is no reason why we should have consistent desires over time and there are no reasons why we should think about our future desires. We will either have them in the future or we will not. But how does the concept of intention help us solve this problem?

When discussing what to do, we frequently talk about the future. This is why future-directed intentions<sup>41</sup> are central for explaining how and why agents behave the way they do. This sort of intentions involves *commitment*. Bratman illustrates the point with a simple example.

*I intend now to take a United flight to Boston tomorrow. It seems that in so intending I do not merely want or desire today to take that flight tomorrow. Rather I am in some sense committed today to taking that flight tomorrow.* (Bratman 4: 1987).

Firstly, our future-directed intentions would be both useless and meaningless without some kind of commitment. Secondly, the following question arises: what does this commitment amount to? Commitments are characteristics of our intentions that facilitate persistence of intention<sup>42</sup> from one point of time to the next, i.e., over a certain period of time. These commitments are by all means revocable because circumstances can change and we cannot

---

<sup>41</sup> Bratman does acknowledge the existence of present-directed intention.

<sup>42</sup> This persistence of intention should not be conflated with Broome's persistence of intention. Bratman's persistence of intention is a property or characteristic of future-directed intentions which is a description with some normative elements. On the other hand, Broome's persistence of intention is a genuine diachronic norm of rationality. It is derived, at least in part, from the nature of intention which is, arguably, persistent over time. Both accounts share obvious similarities and that is why they are both *instrumentalist* under my classification.

expect from agents to always accurately predict the future state of affairs. Bratman's point is merely that when an agent has a future-directed intention, such as taking the United flight to Boston tomorrow, that intention should, all other things being equal, persist until the time of the action. Intentions are, as we have mentioned before, distinctive mental states that explain how agents purposely make decisions and act on them. Contrary to beliefs, intentions (and desires) are considered to be *pro-attitudes*. Pro-attitudes are, broadly speaking, a subset of attitudes that play a motivational role regarding action. In this sense, intentions and desires are pro-attitudes and beliefs are not. Intentions are, in this sense, conduct-controlling pro-attitudes, while desires are only potential influencers of action. The reason why intentions are conduct-controlling pro-attitudes and desires are not is because desires do not have a characteristic commitment to act in a certain way in the future, whereas intentions do.

Commitments are, as mentioned before, characteristics of our intentions that facilitate persistence of intention over time. There are two main dimensions of commitment: the volitional dimension and reason-centered dimension.

The volitional dimension explains the connection between intention and action. When an agent acts, she has a certain disposition to act or *a will* to act. She is motivated to act. When an agent intends to X tomorrow, and nothing relevant happens between today and tomorrow, at the time of the action she simply executes it. There is no need for a repeated formation of intention. We are in some sense bound by our previous intentions although that bound is defeasible.

The second dimension of commitment is that commitments are reason-centered. Firstly, intentions need to be deliberated and settled before the time of action. As intention holds, in the form of a commitment, between the time of formulating the intention and action, intentions are not, at least not in normal circumstances, deliberated and revisited once again. An example is in order. An agent intends to return the book to the library. She will usually not, if there are no major or relevant changes between the formation of the intention and action, reconsider her action while walking to the library. She will simply execute it at the time of the action because she has deliberated on it before and the matter is settled. Because of the reason-centered dimension of commitment, intentions resist reconsideration and they have *stability* or *inertia*. We certainly cannot predict the future since everyday life is full of twists and turns, but in order to coordinate our own lives and to coordinate with others, retention of previous intentions and non-reconsideration is, as Bratman would put it, the "default option". Secondly, commitments are reason-centered because agents reason from their means to their ends. In other words, agents reason how to achieve their practical goals. To continue on our previous example, the agent needs to reason how to get to the library. She can take a bus to the library, or drive there

by a car, or take a walk instead. Agents reason about *how*, i.e., what means to take, to complete their actions. Lastly, agents reason about the different kinds of intentions and how they are connected. Usually, we reason from more general intentions to more specific ones. Agents can have general intentions such as going to the library and more specific intentions such as how to get to the library. The level of specificity of intentions will of course vary depending on agent's willingness and the appropriate context. For example, an agent usually does not need to specify what leg she will use first when entering a bus. Nevertheless, all specific intentions need to be consistent with one another and with the general intention.<sup>43</sup>

The notion of commitment is a vital part of Bratman's theory of planning agency for several different reasons. Firstly, the notion of commitment used by Bratman is not the same as the notion of commitment used by other authors of the first phase of diachronic rationality investigation. Authors like McClennen and Elster treat commitment or pre-commitment as a norm of rationality which should keep our preferences consistent over time. Bratman does not use the notion of commitment in that sense.<sup>44</sup> Secondly, the notion of commitment has descriptive as well as normative aspects. The descriptive aspect of commitments is instantiated by the fact that agents, when having future-directed intentions, usually retain their intentions until the time of the execution of an action. The normative aspect of commitment is instantiated by the fact that the notion of commitment enables intentions to function as conduct-controlling pro-attitudes. In a sense, our future conduct is bound by our previous intentions. Thirdly, the notion of commitment will serve as a building block for Bratman's norms of practical rationality: *Means-end coherence* and *Stability of intention*.<sup>45</sup> Lastly, the notion of commitment plays an important role in Bratman's proposal of the diachronic norm of rationality. The notion of commitment, together with the concept of will, forms the concept of self-governance.<sup>46</sup> The concept of *self-governance* is a central concept of Bratman's later philosophical work and forms a genuine diachronic norm of rationality – diachronic self-governance.

---

<sup>43</sup> The claim that intentions need to be consistent is considered to be the standard view. For more on this view see (Bratman 1984, 1999, 2009, Velleman 1989, Ross 2009, Broome 2013). Conversely, there are authors who question this claim. For more see (Goldstein 2016).

<sup>44</sup> That being said, Bratman's account does struggle, at least initially, with some of the same problems that the other authors from the first phase struggle with. Primarily, with *Kavka's toxin puzzle*. More on this problem will be said later in the chapter.

<sup>45</sup> These are both synchronic norms of rationality.

<sup>46</sup> Bratman was inspired by the works of Harry Frankfurt and Gary Watson while developing the notion of self-governance.

### 2.1.2 Plans<sup>47</sup>

We have explored the two main concepts in Bratman's planning theory of intention: the concept of intention and the notion of commitment. These concepts are necessary but not sufficient for the planning theory of intention. To get the full picture we need to address the concept of *plans*. In other words, we need to ask ourselves the following question: what exactly are plans? According to Bratman, when discussing plans, we generally use the concept in two very different senses. In the first sense, we talk about plans as abstract structures that can be represented by a game theory notation and this is consistent with the first phase of diachronic rationality. In the second sense, we can talk about a plan as a *mental state* or *having a plan* for something. Bratman uses the concept of plan in the latter sense. This sense also has to exclude the notion of plan as a simple procedure or a recipe. For example, an agent can have a recipe for baking a cake but never actually have the intention to bake it. To *have a plan*, in Bratman's sense, the agent has to intend to bake the cake. The definition of plans in Bratman's own words is as follows.

*Plans, as I shall understand them, are mental states involving an appropriate form of commitment to action: I have a plan to A only if it is true of me that I plan to A. (Bratman 29: 1987).*

Plans are forms of intentions. They have the same characteristics as intentions. Plans are conduct controllers, they resist reconsiderations and provide insight in further practical reasoning. But there are severe differences between intentions and plans. While intentions are simply executed (if they are not stopped in some way) at the time of the action, plans are formed, retained, combined, constrained by other plans, filled in, modified, etc. In this sense, plans are more complicated forms of intentions that allow us to achieve complex and temporally distant goals and efficiently coordinate with other agents.

Plans have two important features: partiality and hierarchical structure. Partiality is expressed in the following way. Plans are, except in extreme circumstances, incomplete and need to be "filled in" as the time goes by. For example, an agent intends to go to the movies. She does not need to fill up every detail of her plan. She does not need to settle in advance how to get to the

---

<sup>47</sup> The word *plan* can also mean *a drawing or diagram drawn on a plane*. This is not what we are discussing here. We are using the word *plan* to mean, roughly, intention to do something in the future.

movies, what movie she wants to watch or when to buy tickets. The agent is required to have two distinct dispositions. First, she needs to be aware that her plan is incomplete and needs to intend to “fill in” the plan as the time goes by. Partiality is extremely useful and important. The reason why is because we live in an ever-changing world and some aspects of it are hard or even impossible for us to predict. Partial plans are a flexible and useful tool for us to make sense of and manage our everyday lives.

This brings us to the next feature of plans, namely their hierarchical structure. Plans come in all shapes and sizes. An agent may plan to take a glass of water that is in front of her, or plan to take a vacation next month, or have a plan as to where she wants to spend her retirement years. In order to track and make sense of our planning activities, we need to place our plans in a hierarchical structure. Specific plans need to be embedded into more general plans and plans about one’s ends need to be embedded into plans about one’s means. We can hold certain parts of our plans fixed and deliberate about other parts of our plans. In this sense, plans are hierarchical and different layers of plans need to be consistent with one another.

To summarize, plans are complex forms of intentions that enable us to manage our own activities over time, achieve complex and temporary distant goals and help us coordinate with others. They are partial, hierarchical, resist reconsideration and serve as conduct controllers. They act as a connection between deliberation and action in a strategic and systemic way. Plans need to be *internally consistent* and need to conform to the *means-end coherence*.

### **2.3 The normative side of intention**

So far, we have explored the descriptive side of Bratman’s planning theory of intention. In their everyday lives, agents do not just hold beliefs and have desires, they usually also intend to do something and then they do it. Agents have intentions which are pro-attitudes and irreducible to beliefs and desires. Those intentions are usually focused on future action, i.e., they have future-directed intentions. When having those intentions, agents have a *pro tanto* commitment to follow through on their earlier commitments. This is the “default option”. In other words, unless something substantially changes, agents retain their intention until the moment of action. Lastly, agents can plan. Plans are complex intentions that enable us to manage our own lives and the ability to make plans is one of the things that separates us from animals. This is,

roughly, the descriptive side of Bratman's planning theory of intention.<sup>48</sup> Now we will address the normative side of this theory. The normative side consists of two normative pressures: 1) Stability of intention and Non-reconsideration policy and 2) two Practical constraints of rationality, namely Means-end coherence and Consistency of intention. The difference between the two is that rational pressures are general and more loosely defined and rational constraints are specific and more clearly defined.

We shall begin with *Stability of intention*. Intentions have a specific characteristic of commitment and serve as conduct controllers for our continuous behavior. Commitment and conduct control are the two properties of intentions that make intentions stable. It is, in some sense, reasonable<sup>49</sup> for an agent to act on their previously formed intention. In other words, it is reasonable for an agent to continue on with what she is currently doing. From that we can derive some normative force. It would be almost impossible for us to manage our daily lives without stability of intention. We are not, as Bratman would put it, frictionless deliberators nor are we time-slice agents. On the contrary, we are temporally extended agents<sup>50</sup> with huge cognitive costs, especially when we make decisions over time. It would be extremely inefficient for us to reevaluate our intentions every moment between the formation of intention and the execution of action. If everything goes "according to plan", there is no need to reconsider our initial intentions. Here we have a great segue to the next rational pressure – non-reconsideration.

The second rational pressure is *Non-reconsideration policy*. We derive non-reconsideration from the stability of intention. Our intentions are stable over time and we usually do not reconsider them. Retaining our previously formed intentions is, as Bratman would put it, our

---

<sup>48</sup> Although we should point out that there are normative elements present here, namely the idea that future intentions are conduct controlling and the notion of commitment.

<sup>49</sup> The exact reason why it is reasonable for an agent to act on his previous intentions is kind of tricky to "cash out". We can argue that an agent has reasons for it, but that leads to the inevitable bootstrapping problem. Bratman is aware of this problem and addresses it in different ways. In his earlier works, Bratman first argued that intentions are not reasons (which is consistent with Broome). By doing so, Bratman avoids the bootstrapping problem but loses normative force. For more see (Bratman 1987). On another occasion, Bratman argues that future-directed intentions provide *framework reasons* for agent's previously formed intentions. Framework reasons are reasons employed in the planning framework. For more see (Bratman 1992). Later, Bratman offers a new solution to the problem by involving the concept of self-governance. For more see (Bratman 2009, 2012).

<sup>50</sup> Bratman does not use the term temporally extended agent in his early works. He is using it for the first time in *Reflection, Planning, and Temporally Extended Agency*. That being said, Bratman has since reinterpreted his philosophy in a way such that we can talk about stability of intention of temporally extended agents. For more see (Bratman 2000).

“default option”. Non-reconsiderations do not involve deliberation or reflection.<sup>51</sup> We usually do not deliberate over non-reconsiderations. It would be a waste of our time and cognitive resources. Non-reconsideration is also non-reflective and usually grounded in our habits and dispositions. Let’s take Bratman’s own example to clarify this statement.

*Earlier this year I carefully considered whether to get earthquake insurance and decided not to. Most of the time I simply do not seriously reconsider this intention of mine; I treat the matter as settled. My nonreconsideration is nonreflective, and it amounts only to **the absence of reconsideration** rather than an action of intentionally refraining from reconsidering. This is the typical case, and it is the case that is basic to my theory. Still, I do occasionally receive solicitations in the mail for such insurance, and sometimes I stop briefly to think about whether to look at them in a careful way. But then I quickly appeal to my policy of only reconsidering such matter yearly unless there is some basic change. In such cases my nonreconsideration is policy-based but not deliberative (Bratman 61-62: 1987).*

The non-reconsideration policy is in most cases the “default option”. It is simply a matter of absence of reconsideration. Reconsideration, on the other hand, is not a matter of *briefly thinking about whether* to do something. When an agent is reconsidering, she is doing the following. An agent is reconsidering her prior intention to S if and only if she seriously reopens the question whether to S. Reconsideration and non-reconsideration are not two sides of the same coin. Non-reconsideration is something that we do automatically and non-reflectively. It is a simple absence of reconsideration that enables us to continue to hold on to our previously formed intentions with as little cognitive cost as possible. Reconsideration, on the other hand, is something that we willfully and deliberately do. We reopen the question whether to S and, after reflection and deliberation, conclude whether to continue with our current intention or to abandon our current intention.<sup>52</sup>

We have addressed the two normative pressures of Bratman’s planning theory of intention and now we will deal with the two practical constraints of rationality: *Means-end coherence* and

---

<sup>51</sup> There are of course exemptions to this and Bratman goes into them, as well as into the three varieties of reconsiderations and non-reconsiderations. For more see (Bratman 1987; Chapter 5 *Reconsideration and Rationality*).

<sup>52</sup> There are subtle gray areas here which some authors (Velleman 2000) have pointed out and to which Bratman has recently responded (Bratman 2018), but those concerns do not directly address the problem of diachronic rationality.

*Intention consistency.* The consistency of intention amounts to, in Bratman's words, the following.

*Intention consistency: The following is always pro tanto irrational: intending A and intending B while believing that A and B are not co-possible.* (Bratman 413: 2009).

There are several things to dissect when addressing the *Intention consistency* norm of Bratman's planning theory of intention. We are going to start from simpler implications and work our way up to more complex implications.

Firstly, we should say that this norm is (together with *Means-end coherence*) a fundamental building block of Bratman's planning theory of intention. A rational agent cannot have contradictory intentions. In other words, if an agent has contradictory intentions, she can be accused of irrationality. In this sense, an agent is committed to one of the two possible options and cannot or should not break her commitment. This norm also solves some initial problems that belief desire models of intentional action face. Agents can rationally have contradictory desires and perhaps hold contradictory beliefs but rational agents cannot have contradictory intentions. In this sense, this norm serves as a building block for practical or instrumental rationality.

Secondly, this norm allows us, as rational agents, to have different co-possible intentions at the same time. This varies from simple future-directed intentions, through simple plans, all the way to extensive and elaborate plans whose outcomes can be located in distant future. An example is in order. An agent can simultaneously have the intention to attend a play at a local theater and have the intention how to get there.<sup>53</sup> The latter intention can change over time and be filled in as the situation demands. She can also have the intention to have dinner at a local restaurant after the play and what to order for dinner. She can have plans for an entire week, month or even a year. These intentions (or plans if they are more complex) have to abide by *Intention consistency*, i.e., they need to be co-possible. We can have multiple future-directed intentions or plans with various degrees of complexities at the same time. This makes the *intention consistency* norm an extremely simple but powerful theoretical tool when discussing instrumental rationality.

---

<sup>53</sup> This is also a necessary condition but more on that will be said when we discuss the means-end coherence norm.

Thirdly, the *intention consistency* norm takes into account one key aspect of intentional action and practical decision making – volition. When an agent intentionally A-s, this means that she is not doing it by accident, mistake, unwittingly or inadvertently, but because she wills it. She is, in the sense explained earlier, committed to such an action. According to Bratman’s account, this is more than a mere desire, it is a special pro-attitude in favor of A-ing. This special attitude is called *willing* or *volition* (Bratman 1984).

We have explained *Intention consistency* and now we are going to address the second practical constraint of rationality: *Means-end coherence*. *Means-end coherence* constitutes, in Bratman’s words, the following:

*Means-end coherence: The following is always pro tanto irrational: intending E while believing that a necessary means to E is M and that M requires that one now intend M, yet not now intending M.* (Bratman 413: 2009).

As with *Intention consistency*, there are several things to unpack here, so we are going to start with more general implications of the *means-end coherence* norm and move up to more specific ones.

Firstly, this norm serves as a minimal norm for practical rationality. If an agent intends to E, and is aware of the necessary mean to E which is M, she is rationally required to intend to M. This is the most basic norm of practical rationality and versions of this norm go back to Aristotle. We should take an example to address the matter properly.

Alice wants to turn on the light in her living room. In order to turn on the light in her living room, Alice needs to flip the switch (let us say that the only way to turn on the light is to flip the switch). Bratman’s *means-end coherence* norm tells us the following. If Alice intends to turn on the light in her living room, she needs to have another intention and that intention is to flip the switch. Failing to have the latter intention when having the former, exposes Alice to rational criticism. In other words, if Alice intends some *end*, in this case turning the light in her living room, then she is rationally obligated to intend the necessary *means* to that *end*, in this case, flipping the switch. Alice’s *ends* need to cohere with the necessary *means* to that *end* and this is what forms the *means-end coherence* norm of practical rationality.

Secondly, the *means-end coherence* norm is a suitable building block for the descriptive side of Bratman’s theory of intention. As we have established in detail before, human beings are planning creatures. We make plans and plans are multilayered and complex future-directed intentions. Our plans are, in most cases, partial, flexible and need to be “filled in”. The question

arises: what do we, as rational agents, need to “fill in” our plans with? The minimal rational requirement to fill in our plans are the necessary means. If we, as rational agents, want to achieve our plans, we need to take the necessary means to achieve them.

Both of these norms, *Intention consistency* and *Means-end coherence* serve as the normative bedrock for Bratman’s theory of planning intention. Both norms emerge from the descriptive side of Bratman’s theory (although it does not succumb to “*is versus ought fallacy*” in any meaningful way). Also, both norms have a pro tanto clause in them and the reason for that is the following. The pro tanto clause in these norms protects Bratman’s account from criticism like practical analogues of *the preface paradox*.<sup>54</sup>

Lastly, we should emphasize one important thing regarding these norms (and rational pressures that we have discussed earlier) that has relevance on the general topic at hand – the diachronic rationality. These norms, *Intention consistency* and *Means-end coherence*, and rational pressures, namely *Stability of intention* and *Non-reconsideration policy*, are all synchronic norms (or pressures) of rationality. Bratman’s theory of intention can, in some sense, serve as a model of our diachronic agency, yet be potent for thinking about diachronic norms of rationality. But Bratman’s theory of intention does not give us any genuine diachronic norms of rationality.<sup>55</sup>

---

<sup>54</sup> Although, to what extent is Bratman’s account protected from this type of criticism, remains an open question in contemporary debates in the philosophy of action and philosophy of rationality. For introduction to the preface paradox see (Makinson 1965) and for practical analogues of the preface paradox see (Goldstein 2016).

<sup>55</sup> Other authors in this field (by this field, I mean instrumental accounts of diachronic rationality) have different approaches to diachronic norms. For example, Broome’s diachronic norm *persistence of intention* is analogous to Bratman’s synchronic rational pressure *Stability of intention*. Recently, Snedegar has developed a genuine diachronic norm – “fill in plans” from Bratman’s work on the nature of intention and our planning capacities. For more on Broome’s account of diachronic rationality see (Broome 2013) and for more on Snedegar’s account of diachronic rationality see (Snedegar 2017).

## 2.4 Some problems with Bratman's planning theory of intention: The bootstrapping problem, The Toxin Puzzle, and The myth-theoretical challenge<sup>56</sup>

### 2.4.1 The bootstrapping problem

The bootstrapping problem is one of the core problems in the philosophy of action and philosophy of instrumental rationality. The problem states that we cannot view intentions as reasons when discussing instrumental rationality. The idea is simple and we should be able to understand it using a quote from Broome. In his words:

*The view that intentions are reasons is implausible. If you have no reason to do something, it is implausible that you can give yourself a reason, just by forming the intention of doing it. How could you create a reason for yourself out of nothing? Suppose, say, that you have no reason either for against doing some act, and you happen to decide to do it. Now you intend to do it. So now, if intentions are reasons, you have a reason to do it. Since you have no contrary reason not to do it, the balance of reasons is in favour of your doing it. You now actually ought to do it, therefore. But this is implausible. It is implausible that just deciding to do something can make it the case that you ought to it, when previously that was not the case (Broome 1: 2001).*

Broome argues that if we accept intentions as reasons, then we will have to accept that reasons can be brought to us *ex nihilo* and that is unacceptable. It is a form of vicious circularity (bootstrapping) and naturalistic fallacy. It is a naturalistic fallacy because we derive normative force from pure description in an arbitrary (unacceptable) manner. We can demonstrate Broome's idea in the following way.

An agent intends to do  $\phi$ . (1)

Now, an agent has a reason to do  $\phi$ . (2)

---

<sup>56</sup> This list of problems is by no means exhaustive. There are other problems that can be found in Bratman's planning theory of intention, most notably by David Velleman and Luca Ferrero. For more on them see (Velleman 2000 2007, Ferrero 2006, 2009, 2012). But these problems (Kavka's puzzle, the bootstrapping problem and non-cognitivism) are reasons why Bratman extended his theory of planning intention to a theory of diachronic rationality.

An agent should do  $\phi$ . (3)

There is, at least *prima facie*, nothing wrong with (1), it is a simple statement of an intention for a practical action. The problem lies in (2) because, if intentions are reasons, now an agent has a reason to do  $\phi$ , which in turn carries some normative weight to do  $\phi$ . Then an agent proceeds to conclude that she should do  $\phi$ . This form of practical reasoning is implausible and unacceptable. There is, more or less, a consensus in the fields of philosophy of action and instrumental rationality, from Bratman to Broome, that intentions cannot be reasons on pain of bootstrapping.

Stating that intentions are not reasons solves the bootstrapping problem but, in turn, it generates more problems. If intentions cannot be reasons, then we can reasonably ask (at least) two questions. Can we summon intentions into existence out of nothing (because we clearly cannot summon reasons into existence out of nothing)? How can we intend to do something if we have no motivation to do it (what is the motivational force behind intention formation)? In order to answer these two questions, we need to take a look at *The Toxin Puzzle*.

### 2.4.2 The Toxin Puzzle

Let us begin with *The Toxin Puzzle*. An eccentric billionaire offers you the following proposal. You need to form an intention on Monday to drink a disgusting but non-lethal toxin on Wednesday. In return, she will pay you one million dollars on Tuesday. In other words, the eccentric billionaire will pay you on Tuesday *to form an intention* on Monday to drink a non-lethal toxin on Wednesday. There are two technical caveats. The first caveat is that the intention needs to be formed clearly and honestly, i.e., without side bets. The second caveat is that the eccentric billionaire has access to advanced technology which can detect agent's intentions with great accuracy. The question of the toxin puzzle, a million-dollar question actually, is whether a rational agent can take the bet? In other words, can a rational agent form the intention at will? At first glance and intuitively speaking, if the agent desires a million dollars, she will simply form the intention to drink the disgusting but non-lethal toxin on Monday as requested by the eccentric billionaire and receive the money on Wednesday. But there is a problem with this line of thinking. On Tuesday the agent receives the money. One million dollars in her bank

account. At that moment<sup>57</sup> the agent has no reason to drink the toxin the day after (on Wednesday). The eccentric billionaire does not care if the agent drinks the toxin, she only cares about a clear and honest intention formation on Monday. As we have established, the agent's money is already in her bank account. There is no reason for her to drink the toxin tomorrow (on Wednesday) and every reason not to drink it. The real problem is: she knows all of this on Monday. Thus, the question arises: can she form an intention for something she knows she will most likely not do? This question is the true problem of Kavka's puzzle.

In his paper, Kavka argues that the toxin puzzle sheds light on two distinct problems in the philosophy of action. The first problem is the nature of intention. If intentions are "*inner performances*", or "*self-directed commands*" or *simply decisions* (i.e., mental states), then the agent should have no problem winning the million dollars. On the other hand, if intentions are reasons for or against a certain action, then there is a problem because we arrive at a contradiction – the agent has and does not have reasons to drink the toxin on Wednesday. Kavka argues that the latter is more plausible because this is how, according to him, agents make real decisions in the real world.

The second problem is divergence, as Kavka calls it, between intending and acting. On the one hand, we have rationality of intending and on the other, we have consequences of our action. In that sense, something has to give: either rational intention or rational action.

The toxin puzzle is a general problem in the philosophy of action and philosophy of practical rationality. It mostly affects instrumentalists, most notably, accounts developed by Broome and Bratman. We can see how the toxin puzzle strikes at the heart of Bratman's theory of planning intention. In Bratman's account intention formation is fundamental and primary. Agents do not need reasons to form an intention or to act, they just do it of own volition. In Bratman's view (or Broome's for that matter), intentions are not reasons. With his toxin puzzle, Kavka pushes hard against that view. He asks the following question: how can we intend to do anything if we have no motivating reason for it? Under Bratman's view, the agent will simply form the intention on Monday to drink the toxin on Wednesday. Kavka argues that the agent would try to sum up in panic the willpower she needs to form the intention to drink the toxin and she would ultimately fail to do so. The reason why she would fail at that task is simple: she knows she will not actually do it when the time comes.

---

<sup>57</sup> That moment is Monday midnight in Kavka's example. I have changed it here to Tuesday to make the case as clear as possible and it changes nothing in the example's actual content. For the original version of Kavka's puzzle see (Kavka 1983).

The toxin case strikes at the core of Bratman's planning theory of intention and he addresses the problem seriously and methodically. Firstly, he presents the notion of *the linking principle*. The linking principle amounts to, in Bratman's words, the following:

*If, on the basis of deliberation, an agent rationally settles at t1 on an intention to A at t2 if (given that) C, and if she expects that under C at t2 she will have rational control of whether or not she A's, then she will not suppose at t1 that if C at t2, then at t2 she should, rationally, abandon her intention in favour of an intention to perform an alternative to A. Call this statement of a link between rational intention formation and supposed rational intention retention **the linking principle** (Bratman 64: 1999, bold mine).*

The goal behind the linking principle is to establish the connection between intention formation and intention retention. This link serves as a normative constraint on practical rationality. We are, at least pro tanto, rationally required to hold on to our intentions from the moment we form them until the moment of the execution of action. This would mean that the agent is rationally required to drink the toxin on Wednesday, which is counterintuitive, to say the least. From the standard view, as Bratman calls it, it is perfectly rational for the agent to change her mind on Tuesday and not drink the toxin on Wednesday. Moreover, it is probably rationally required of the agent to change her mind and not drink the toxin because she has no reasons to do so.<sup>58</sup> Bratman also discovers the underlying problem of the toxin case. In this scenario the agent finds herself in a peculiar position. She is required to form an intention. In order to do that she needs a second-order intention. She needs to intend to intend on Monday to drink the toxin on Wednesday. This is not a situation in which agents usually find themselves when making decisions in the real world. Nevertheless, the objection still holds and Bratman is determined to address the problem head on.

Firstly, Bratman acknowledges that the toxin case presents a problem for his theory. The problem is multi-layered. It affects intention formation, intention retention, the nature of intention and the ability of agents to rationally change their mind. Secondly, Bratman argues that in order to solve this problem, we need to look at the larger picture. He offers the following example.

---

<sup>58</sup> The clash between these two ideas is essentially the clash between synchronic and diachronic rationality and it touches on the question whether there are any genuine diachronic norms of rationality.

## A dilemma on an airplane<sup>59</sup>

*You and I are mutually disinterested, instrumentally rational strangers about to get off an airplane. We know we will never see each other — or, indeed, the other passengers — again. We also know that we each have a pair of suitcases, and that each of us would benefit from the help of the other in getting them down from the overhead rack. We each would much prefer mutual aid to mutual nonaid. Given our seating arrangements, however, you would need to help me first, after which I could help you. You will help me only if you are confident that I would, as a result, reciprocate. But we both see that once you help me I will have received the benefit from you that I wanted. My helping you later would, let us suppose, only be a burden for me. Of course, most of us would care about the plight of the other passenger, and/or have concerns about fairness in such a case. But let us here abstract away from such concerns, for our aim here is to determine what is required solely by instrumental rationality. Let us also suppose, again artificially, that my helping you or not would have no differential long-term effects (including reputation effects) that matter to me now. Given these special assumptions, it seems I would not have reason to reciprocate after you have helped me. Seeing that, you do not help me, so we do not gain the benefits of cooperation. In such a situation I might try to assure you I would reciprocate. But suppose I am not very good at deceit and will only be convincing to you if I really intend to reciprocate if you help me. I would, then, very much like to provide a sincere assurance. Can I?<sup>60</sup> (Bratman 66-67: 1999).*

Bratman makes several points here. Firstly, a rational agent should have the option or possibility to reciprocate (or cooperate) in this example. From the perspective of the second agent (whose suitcase has been put down with the assistance of the first agent), she has no reasons to help the first agent. There is no rational obligation on the second agent's part. To make matters even worse, there is no sense in which it is rational for the second agent to help the first agent – she simply has no rational reasons to do it.<sup>61</sup> Bratman states that this is wrong. A rational agent should have the option of helping in *A dilemma on an airplane*. The question

---

<sup>59</sup> Bratman does not explicitly name this example so I do it here for clarity reasons.

<sup>60</sup> The example is a form of the prisoner's dilemma. Similar examples have been offered by David Gauthier. For more see (Gauthier 1994).

<sup>61</sup> Our intuitions, I hope, would tell us that there are reasons for the second agent to help the first agent. Those reasons are moral reasons. Although this is a perfectly valid form of reasoning, Bratman will not go down that road. The reason why is because he concerns himself with the problems of practical reason and instrumental rationality. Reducing instrumental rationality to moral reasoning has its benefits and its drawbacks. Bratman is not prepared to take the drawbacks of reducing instrumental rationality to moral reasoning.

arises: how can the second agent rationally help the first agent? Bratman claims that a rational agent must have the option of helping (or reciprocating) in *A dilemma on an airplane*. Secondly, Bratman affirms that *A dilemma on an airplane* has, when discussing instrumental rationality, the same logical structure as the toxin case. They are both instances of an agent forming an intention at t1 to act in a certain way at t2. Both cases involve, to a certain relevant degree, intention formation, intention retention and the possibility of rationally changing one's mind. But the crucial point is to find a model that will rationally allow the agent in the toxin case to form an intention and abandon the same intention at a later time, and compel the agent in *A dilemma on an airplane*, after she has formed an intention, not to drop it.

To answer this conundrum, Bratman proposes a model called the *no-regret principle*. Loosely put, the no-regret principle states that an agent should stick to her previously formed intentions if there is a reasonable expectation that she will not regret doing so in the future. Bratman presents his case with the following example.

*An evening with Ann*<sup>62</sup>

Ann is a person who enjoys spending her evenings alone. Frequently, she likes to do two things in the evening. She likes to drink beer at dinner and read a good book after dinner. Ann also knows the following. She cannot concentrate on reading if she has more than one beer at dinner. Before dinner, Ann prefers one beer at dinner and reading a good book after dinner. But at dinner, after she had one beer, she desires a second beer. So, at dinner, her preferences have changed, she now desires two beers over one beer plus reading a good book. Ann is facing a dilemma (Bratman 1999)

Bratman analyzes the example above in the following way. He presents two approaches to Ann's preferences: one before dinner and one during dinner. Before dinner Ann prefers:

*(1) one beer at dinner plus a book after dinner*

*to*

*(2) more than one beer at dinner and no book after dinner.*

*Throughout dinner, however, Ann continues to prefer*

---

<sup>62</sup> Bratman does not cite this example explicitly. I mention it here for clarity purposes. For more see (Bratman 74-77: 1999).

(3) *one beer at dinner and a book after, for all nights,*

*to*

(4) *more than one beer and no book for all nights.*

*But it is also true that during dinner Ann temporarily prefers*

(3') *more than one beer and no book, but one beer and a book on all other nights*

(Bratman 74-75: 1999).

In Bratman's analysis there are several key points. Firstly, Bratman argues that there is no conflict between (1) and (2) nor (3) and (4). The reason for this is that in (1) and (2) Ann is weighing between two specific sets of preferences that she has over the course of one night. But in (3) and (4) she is weighing between general preferences (or policies) she has decided on. From this Bratman concludes the following. We need an account of instrumental rationality that will enable agents to make general and stable plans (policies) and at the same time allow them to reasonably change their mind. This account should also properly explain and place some relevant normative force on the three examples that we have been discussing: *the toxin case*, *A dilemma on an airplane* and *An evening with Ann*. Bratman notes that there are similarities and differences between those examples. The similarities are that in all of the three cases the agent has a plan which determines her future behaviour, but when the time comes there is a conflict between intentions at the start of the plan and present intentions. There are also relevant differences. While *the toxin case* and *A dilemma on an airplane* present clear and simple examples of changing one's mind or reversal of preferences (as proponents of the first phase of diachronic rationality would call them), *An evening with Ann* is an example of a temporary reversal of preferences under the influence of her lacking willpower. This makes *An evening with Ann* an example of the diachronic weakness of will problem. Despite the differences, Bratman argues that we need an account of instrumental rationality that will adequately explain these examples.<sup>63</sup>

---

<sup>63</sup> There is one more example that Bratman uses to prove his point. That example is now famous as *The puzzle of self-torturer*. The case goes as follows. An agent has agreed for a small device to be permanently installed in his body. The device constantly generates electric current in his body at different levels from 0 (no electric current) to 1000 (very painful electric current). An agent is offered 10 000 dollars to install the device in his body and he will receive 10 000 dollars for each advance in settings. He can advance settings on the device once every week and cannot go back on previous levels of settings. The most important thing is: **the agent cannot feel any difference between any two individual settings** but he can will the difference between settings that are sufficiently far apart. The problem is the following. A rational agent would soon become the self-torturer and in

To solve this bag of problems, which includes the toxin case as well, Bratman proposes a no-regret condition account. We have mentioned it briefly before, but now we will examine it in greater detail. In Bratman's words, the no-regret principle account corresponds to the following:

*(a) If you stick with your prior intention, you will be glad you did.*

*(b) If you do not stick with your prior intention, you will wish you had.*

*So, other things being equal,*

*(c) Though you now prefer to abandon your prior intention, you should nevertheless stick with it. (Bratman 79: 1999).*

There are several key elements that need to be addressed here. Firstly, we need to establish that this, at least on the surface, functions as a norm for consistency of preferences (or, in Bratman's case, intentions) at different points in time. We should, all other things being equal, stick to our previous intentions. This line of reasoning is quite similar to other authors in the first phase of diachronic rationality, such as McClennen and Elster. They also claim that a rational agent should be resolute or sophisticated not to change her mind. In other words, an agent should stick to her previously formed intentions. But Bratman argues for "sticking to your previously formed intention" in a slightly different way. When we closely examine (a) and (b), we see that they are conditioned upon having future regret. This is apparent in the second part of (a) which states *you will be glad you did* and the second part of (b) which states *you will wish you had*. In this way, a rational agent should stick to her previously formed intention (or plan), while having the option to change her mind.

We need to formulate this more precisely. In Bratman's no-regret condition account there are three relevant time frames:

The first time frame: the time in which an agent has developed and initiated her plan for future action (t1)

The second time frame: the time of the execution of action where the agent asks herself about future regret (t2)

The third time frame: the time of future regret (or absence of regret) (t3).

---

the end find herself in excruciating pain. The problem hinges on the notion of intransitivity of slippery slope preferences and the notion of "vague projects". For more on the original problem see (Quinn 1990) and for some contemporary insights see (Tenenbaum and Raffman 2012, Andreou 2015).

With this model Bratman is able to explain how we can have general policies and sometimes not follow those same policies. The question arises: how do we know whether we should follow our policies or change our mind? Bratman's answer is: we ask ourselves at the time of action (t2) whether we will regret it in the future (at some time t3).

We need to go through the examples again. Let us start with *An evening with Ann*. What should Ann do? Stick to her policy of one beer plus reading a good book or have two beers? Bratman claims that in this example Ann should not change her mind and stick to her policy of one beer plus reading a good book in the evening. The reason for this is because at the time when Ann is having second thoughts at t2, she should ask herself the following question: will I regret this decision in the future? If we assume that the answer is yes (otherwise, she would not have the policy in the first place), then she should not change her mind and should stick to her original plan which is one beer plus reading a good book in the evening.

Let us turn to *The Toxin Puzzle*. The toxin case is arguably easy to solve with Bratman's no-regret condition account. It is apparent that an agent would regret drinking the toxin. So, on the basis of her future regret, she is allowed to change her mind and not drink it.

Lastly, let us observe *A dilemma on an airplane*. In this example, Bratman (arguably) bites the bullet. He claims that, much like in the toxin case, an agent should have the option of changing her mind and not help the first agent with the parachute, otherwise she will regret it at some point in the future.<sup>64</sup>

### 2.4.3 The myth-theoretic challenge<sup>65</sup>

The myth-theoretic challenge is an umbrella term for series of challenges presented for Bratman's planning theory of intention, all of which have one thing in common – the idea that the planning theory of intention relates to the notion of rationality which is misguided at best and false at worst. It is simply a myth that the planning theory of intention has something to do with the notion of rationality. Proponents of the myth-theoretic challenge ask the following questions:

---

<sup>64</sup> Bratman acknowledges that there is a sense in which agents may regret these decisions. More precisely, an agent would not want to be the person who does not stick to her former intentions (the toxin puzzle) or a person who does not honor her agreements (a dilemma on an airplane). But this, according to Bratman, does not seem to be the problem of instrumental rationality.

<sup>65</sup> The term myth-theory in this context originates from a fairly influential paper by Joseph Raz entitled *The myth of instrumental rationality* which was subsequently adopted by Niko Kolodny in his paper *The myth of practical consistency*. The term myth-theoretic challenge is used by Bratman in the introduction of his book *Planning, Time and Self-governance* (2018).

*Why are practical planning norms of rationality developed in Bratman's planning theory of intention norms of rationality? (Raz 2005, Kolodny 2008).*

*Can we discuss rationality without talking about reasons? (Kolodny 2008, Setiya 2014).*

*Can an agent be instrumentally rational without being theoretically rational? (Raz 2005 Harman 1999, Velleman 2007, Wallace 2006).*

These are all difficult and valid questions that put Bratman's theory of intention to the test. Firstly, we shall see Raz's take on the idea of instrumental rationality and means-end coherence.

*It appears that there is no such thing as instrumental irrationality. That is, there is no distinctive set of deliberative standards that are involved in getting us to reason correctly from ends we have to means, and that are different from those that are involved in reasoning about which ends to have. Of course, there is a difference between facilitative reasons and others. Facilitative reasons have a special kind of dependence on source reasons. But that is a difference in the content of our deliberation, not in the standards that should govern the deliberative processes. They are the same when we try to determine our will to adopt, or maintain, or abandon some ends, as when we try to determine what facilitative steps to take in pursuit of goods we take ourselves to be pursuing as our ends. (Raz 26: 2005).*

There are three important points that Raz raises here. The first point is that we cannot meaningfully discuss the normative aspects of reasoning from agent's means to her ends. In Raz's view, agents take means to achieve certain goals or ends. It is a descriptive fact and it tells us nothing about normativity at all. The second point is that when an agent is, presumably, instrumentally irrational, she fails to take the necessary means for her ends. But in reality, she simply has two contradictory goals (ends) in mind and this is what makes her irrational. The third and final point is that there is no distinctive area that can be called instrumental rationality and that instrumental rationality is simply a myth.

Secondly, there is a fundamental problem in discussing rationality without talking about reasons. Rational belief (in theoretical rationality) and rational action (in instrumental rationality) are governed by responding correctly to reasons. The question of what responding correctly to reasons entails is a different matter, but the point is that we cannot talk about rationality merely as being focused on appropriating the necessary and sufficient means for our

ends. We need reasons for those ends, otherwise we are not discussing rationality. Kolodny would say that there is no set of principles for formal coherence as such. *Instead, there is only the structure of reason, and the shadow that this structure casts, insofar as we are rational, on our beliefs and choices* (Kolodny 390: 2008).

Lastly, there is an apparent problem of irreducibility of instrumental rationality to theoretical rationality. It is not at all clear how to avoid the prospect of reducing intention to belief and subsequently instrumental rationality to theoretical rationality. If we have a theoretically rational agent, can she be instrumentally rational? Or, can an instrumentally rational agent be theoretically irrational and what exactly would that entail? There are no clear answers to this question in Bratman's theory of intention. And this is why it lacks any normative force. Reasoning in instrumental rationality, or practical reasoning, is faced with two unenviable scenarios. Either practical reasoning has nothing to do with rationality or it is reducible to theoretical reasoning. Velleman argues for the latter and claims that *practical reasoning is a kind of theoretical reasoning, and that practical conclusions, or intentions, are the corresponding theoretical conclusions, or beliefs* (Velleman 15: 1989).

We have established several issues with Bratman's planning theory of intention. Those issues are the bootstrapping problem, the toxin puzzle and the myth-theoretic challenge. The planning theory of intention has been under siege by these challenges for the last few decades. Bratman's responses to these types of criticism varied from direct defense of his position in (Bratman 1991, 1998) to slowly accepting that some of this criticism has some merits in (Bratman 2007). His defense against the toxin case, the no-future-regret account, proved to be insufficient and limited in scope. It is insufficient because it leads back to the bootstrapping problem – the idea is already present at the time of deliberation when the agent is forming her intention. It is limited in scope because there is a limited amount of scenarios in which it is internally clear to the agent whether she would indeed have future regret or not. The myth-theoretic challenge was the “the straw that broke the camel's back” when it comes to Bratman's planning theory of intention as an account of instrumental rationality. Bratman accepted some, although not all, criticism directed towards his account of rationality. In response, he has enhanced his account with the notion of self-governance and transformed it into a diachronic account of rationality.

## 2.5 Self-governance as Diachronic Plan Rationality

The criticism of Bratman's planning theory of intention, most notably by Raz and Kolodny, has landed on fertile ground. Bratman has, in some sense, abandoned the instrumentalist approach to the topic of practical rationality which goes back all the way to his line of argumentation in *Intention, Plans and Practical Reason*. But one thing is certain for Bratman: *Means-end coherence* and *Stability of intention* are insufficient to explain the concept of practical rationality and decision making. Bratman switches his focus from means to ends or goals. While in his previous work, by developing the planning theory of intention, Bratman was preoccupied with agent's ability to choose the necessary means for her ends and actively disregard the value of agent's goals in general, he now focuses his attention on agent's goals. The question that Bratman asks now is: if we settle on some goal that we deem worthy of achieving, are we rationally obligated to follow through on our goals? This new approach is developed partially as a response to Ferrero's and Velleman's criticism (authors I label as Kantians in regard to diachronic rationality). Both of those accounts argue, in substantially different ways, for the importance of diachronic activity and the goals of that activity as values in themselves. For Velleman, what we value is the autonomous and reflective diachronic agency over time which leads us to a desired goal. The term that Velleman uses to describe the autonomous and reflective diachronic agency over time is *diachronic will* (Velleman 2000). Ferrero expands on the notion of diachronic will by saying that ontological importance of diachronic will is best explained in agent's own personal narrative (Ferrero 2009, 2010). Bratman is influenced by the concepts of diachronic will and personal narrative presented in these accounts to the extent where he is "forced" to discuss the values of practical rationality. The second notion that Bratman introduces in his account is the concept of reasons. While reasons did play a substantial role in Bratman's planning theory of intention, his usage of reasons to explain practical rationality was a limited one. This is primarily because of Bratman's own claim that intentions cannot be reasons (Bratman 1987), which was later developed by Broome (2001). This view has some serious problems, most notably presented in Kavka's toxin puzzle which I have discussed in greater detail earlier in this chapter. What Kavka's toxin puzzle has shown us is that reasons must play a role in our explanation of practical rationality. Kavka focuses on the motivating side of having a reason for an action and demonstrates that agents cannot simply intend to do something if they have no reason to do it. Reasons have motivating and normative aspects, both being important when explaining agent's

practical rationality. Bratman is aware of these issues and his response includes the formulation of a new theory of practical rationality which takes into account agent's goals, values, normative reasons and diachronic agency. That theory is diachronic planning self-governance.

Bratman's work in the past ten years has been focused on the construction of an account of diachronic rationality which he calls diachronic planning self-governance. The reason why he develops this account is to answer the following questions:

1. Why be a planning agent?
2. What is the value of the planning theory of intention?

Although Bratman has developed a detailed account of the planning theory of intention, he never argued why anyone should be a planning agent, i.e., what the reason for being a planning agent is or what the value of planning consists in. We, as human beings, value governing our own lives. We value stability and coherence in our lives. But we do not value stability and coherence simply because they successfully lead us to our desired goals (that would be a regression to the instrumentalist approach), we value stability and coherence in themselves as free, reflective and autonomous agents. There is a non-instrumental value in play here, the one of governing our own lives as autonomous agents – the value of *self-governance*. This line of argumentation brings Bratman extremely close to *Kantians* regarding his view on diachronic agency and rationality. Let us remember, *Kantians* (Velleman 2000, Ferrero 2006, 2009, 2010, 2012, Holton) argue, each in their own individual way, that we cannot discuss concepts like diachronic agency and diachronic rationality purely instrumentally. We need concepts like *diachronic will* (Velleman), *personal narrative* (Ferrero) and *resolutions* (Holton) in order to properly explain agent's decision making over time. Bratman is influenced directly (he mentions them in his argumentation) by these authors and incorporates some of those ideas into his account of diachronic rationality (Holton's concept of non-reconsideration). While Velleman's proposal includes diachronic will, Ferrero's involves personal narrative and Holton's comprises a special kind of intentions which are resistant to reconsideration (*resolutions*), Bratman offers his own concept – *self-governance*.

But what exactly is the notion of self-governance? The concept of self-governance has its roots in the works of Harry Frankfurt. In his paper *Identification and Wholeheartedness* (1987), Frankfurt tackles the problem of conflicting desires of different orders, i.e., the conflict between first-order desires and second-order desires. For example, an agent may have a desire to smoke

a cigarette and have a desire not to smoke. In this example, an agent has a first-order desire *to smoke a cigarette* which conflicts with her second-order desire *not to smoke*. In order to solve this problem, Frankfurt proposes that an agent, when having conflicting desires, asks herself two questions:

1. What does it mean for an agent to identify with a certain thought or an attitude?
2. Where does the agent stand on a particular matter?

The way in which an agent should answer the first question is the following. An agent should reflect on her first-order desires until she reaches the reflective equilibrium in which she will identify wholeheartedly with a certain attitude. In the previous example, that would be a decision to smoke or not to smoke.

The second question is for an agent to ask herself where she stands *as a person* on a particular matter. In other words, how she tides her personal identity with her actions and decisions. In the aforementioned example that would be, roughly speaking, whether she considers herself to be a person who smokes or a person who does not smoke. Additionally, for an agent to have a practical standpoint is for an agent to have a coherent set of intentions, otherwise the agent does not know where she stands.

Bratman takes heavy inspiration from Frankfurt's argumentation and claims *that self-governance is a guidance by psychological structures that have agential authority* (Bratman 96: 2018). To have agential authority simply means that an agent has the ability to reflectively and autonomously make a decision on a certain practical matter. The concept of self-governance is compared by Bratman with similar concepts like *self-fidelity* (Stalzer 2004), *integrity* and *self-respect* (Dancy 2003). These concepts serve as fundamental norms of rationality by these authors and are similar in a certain sense to Bratman's notion of self-governance. The minimalist account of Bratman's notion of self-governance is, in his own words:

*"The agent's governance of his own life consists in part the guidance of his relevant thought and action by attitudes that constitutes where he stands; and important aspects of his rational functioning will be a matter of such guidance by his practical standpoint."* (Bratman 159: 2018).

There are three main components that are relevant for the notion of self-governance. Firstly, it is a value of governing our own lives coherently. Secondly, it is relevant to where the agent stands on a particular matter or domain. Thirdly, we are guided by this practical standpoint. In summary, the notion of self-governance has descriptive and normative aspects. In the descriptive sense, it is a fact that we as human beings govern our own lives. Usually, once having settled on it, we do not aimlessly wonder from one plan or project to another without finishing any. In the normative sense, we individually and collectively value personal integrity and practical coherence when it comes to our future actions.<sup>66</sup>

### 2.5.1 A place to stand<sup>67</sup>

The key notion arising from the concept of self-governance is the idea of “a place to stand”. The idea is not to simply identify a specific attitude with which an agent identifies at a particular point in time (although it is an integral part of the idea). The idea implies that there are various attitudes and other relevant features of an agent’s psychological states that come together to form a coherent standpoint. The practical point is constructed in one way by an agent’s reflection of her first-order desires and in the other way by the agent’s wholeheartedness and satisfaction which constitute her personal identity (to some extent). Wholeheartedness and satisfaction function as, more or less, *holistic properties of the agent’s psychological economy* (Bratman 2014).

A different way to understand the concept of “a place to stand” is to view it as a function – a *standpoint function*. A standpoint function is simply a function between an agent’s overall psychological profile and her relevant standpoint. Bratman presents standpoint function in the form of a synchronic norm of rationality in the following fashion:

*Synchronic Standpoint Rationality: It is defeasibly rational of an agent that (If at t her practical standpoint is committed to X, and if she knows that this commitment to X argues conclusively in favor of a course of action A at t, and she knows that her A-ing then requires her intending then to A, then at t she intends to A.)* (Bratman 160: 2018).

---

<sup>66</sup> At least this is what Bratman thinks we should all accept prima facie.

<sup>67</sup> The idea comes from Harry Frankfurt's “*Identification and Wholeheartedness*” (1987).

There is a lot of content to unpack in this proposed norm of rationality. Firstly, we should acknowledge that the norm is synchronic and says nothing about the diachronic aspect of rational decision making. Secondly, the norm is defeasible, which means that the agent is rationally allowed not to conform to the said norm in the light of new information (in a form of new evidence or reasons). Lastly, the agent has freely and deliberately (after a period of reflection and deliberation) chosen to pursue some goal and now she has a *commitment* (although a defeasible one) in favor of the action which leads to the desired goal.

There are two more aspects to the concept of “a place to stand” or *standpoint function* that I want to address. The first is the fact that standpoint function presents an essential feature to the concept of self-governance. For an agent to freely and autonomously govern her own life, she needs to have a standpoint function on a relevant matter at hand. In other words, a place to stand is a necessary but not sufficient condition for the concept of self-governance. The second is that standpoint function does not necessarily lead to diachronic agency and subsequently diachronic rationality. An agent can have different *standpoint functions* depending on her overall psychological profile. For example, an agent can have a *time-slice evaluative* standpoint function. This means that the agent makes decisions based on her current judgments and in the light of her present options without regard to past or future decisions. Contrary to this standpoint, an agent can have a *cross-temporally structured* standpoint. From the perspective of this standpoint, the agent takes into consideration her past commitments and future prospects when making decisions. The latter standpoint leads into the discussion about diachronic agency and diachronic rationality, while the former does not.

The following question emerges: What kind of a standpoint function should an agent adopt? According to Bratman, this is predicated on what kind of person a particular agent is, but a planning agent should adopt a *cross-temporally structured* standpoint and there are good reasons to do so. Briefly, those reasons are the value of governing our own lives through time and having practical coherence in our lives.

### **2.5.2 Synchronic and diachronic self-governance**

The concept of self-governance is in essence, much like its related concept of standpoint function, neutral in respect to synchronic and diachronic structures. Bratman makes a clear distinction between synchronic and diachronic self-governance. Synchronic self-governance or self-governance at the time is, in Bratman’s words, the following:

*Self-governance at the time involves guidance by a synchronic structure of attitudes that is sufficiently unified so as to constitute where the agent stands at that time, and so as to be such that that its guidance can constitute the agent's self-governance* (Bratman 81: 2012).

Not any *structure of attitudes* counts as a form of self-governance. Firstly, the structure needs to be coherent and consistent, i.e., it needs to abide by Bratman's practical norms of rationality, namely the means-end norm and intention consistency norm. Secondly, the content of attitudes needs to be relevant. This means that a consistent set of wild fantasies cannot constitute the notion of self-governance. What types of sets of attitudes do qualify to constitute self-governance? While Bratman leaves the door somewhat open for other types of sets of attitudes, he argues that we should accept planning states or planning intentions as the best candidate for the constitution of synchronic self-governance.

Now that we have defined and explained synchronic self-governance (self-governance at the time), we are ready to address the concept of diachronic self-governance. There are four conditions that need to be met in order for an agent to have diachronic self-governance (although Bratman does not present the concept of diachronic self-governance in terms of necessary and sufficient conditions).

The first condition is having synchronic self-governance. An agent cannot be self-governed over time if she is not self-governed at a particular point in time.

The second condition is a weak metaphysical claim about persistence of personal identity over time.<sup>68</sup> An agent, if she is self-governed over time, should be, more or less, the same person over that period of time.

The third condition is that an agent should have "psychological continuity" over time. She should, from the first-person perspective, be aware that she is the same person as she was, for example, yesterday.

The fourth and last condition is that there should be "a glue" which connects agent's self-governances at each time into self-governance over time. What does this "glue" consist of? Specifically, agent's self-governances at the time need to be:

*(a) semantically interconnected, and*

*(b) stable in the absence of supposed conclusive reason for a change* (Bratman 83: 2012).

---

<sup>68</sup> The claim is metaphysically weak because Bratman takes the approach, inspired by Derek Parfit, that an agent is the same person over time only if she transfers from one moment to another "what matters" of her personal identity. For more see (Parfit 1984).

Diachronic self-governance is a capability or value that human beings are apt to have or achieve. For an agent to be self-governed over time, she needs to be capable of self-governance at a particular time. Additionally, an agent needs to be metaphysically (although weakly) the same person over time and be aware that she is the same person over time (psychological continuity). Lastly, she needs to be able to connect (or “glue”) her self-governed moments in time both meaningfully and stably (in the absence of a conclusive reason for a change).

### 2.5.3 The absence of self-governance

We have established the concept of self-governance at the time (synchronic self-governance) and the concept of self-governance over time (diachronic self-governance). Now we will explore how and why the concept of self-governance is important in our everyday lives. In other words, what the function and value of self-governance consist in. We will achieve this by looking into two scenarios in which agents lack the concept of self-governance. The first scenario is *An afternoon with Candice* and the second is *An evening with Ann*.<sup>69</sup>

*An afternoon with Candice* was originally presented in Brunero’s paper *Instrumental Rationality, Symmetry, and Scope* (Brunero 2012) which discusses the problem of wide and narrow scopes of normative commitments regarding instrumental rationality. Later it was endorsed by Bratman as a fundamental problem of intending and making decisions over time in *Time, Rationality and Self-governance* (Bratman 2012) and again in his book *Planning, Time, and Self-governance* (Bratman 2018). Bratman states the problem in the following manner.

*An afternoon with Candice*

*Candice decides to go to the post office this afternoon to send out some mailings, but on the way there, she gives up on this end and decides to go buy groceries instead. But on the way to the market, she yet again trades in this end for another: going to hang out with her friend David. But on the way to David's house, she once more changes her mind and intends to spend a relaxing afternoon at home,*

---

<sup>69</sup> These terms, *An afternoon with Candice* and *An evening with Ann*, are mine and do not appear as such in (Brunero 2012) and (Bratman 1999).

*but by the time she gets home the afternoon is gone and she's accomplished nothing.* (Bratman 83: 2012).

There is a lot to analyze here. Candice, in the example above, has the intention towards some goal *a* but then she drops it, then has the intention towards some other goal *b* but then drops it, then has the intention towards some goal *c* but then drops it. Candice is changing her mind constantly and without a reason, so the question arises: what is the main problem with her conduct? Bratman's answer is: Candice lacks self-governance over time. Candice does satisfy some conditions for diachronic self-governance. Firstly, we can assume that she is metaphysically (at least weakly) the same person over the course of the afternoon. Secondly, she does have psychological continuity of her intentions and actions. What Candice lacks is "a glue" that binds her intentions at the time with her intentions over time. She lacks the last condition of diachronic self-governance which requires that her intentions be both (a) *semantically interconnected* and (b) *stable in the absence of supposed conclusive reason for a change* (Bratman 83: 2012). Candice has no conclusive reason for a change in her intentions and her intentions are not stable. Agents like Candice lead, what Bratman's calls, a "shuffle life"<sup>70</sup>. They lurk from one plan-like commitment to another without accomplishing anything in the process. Candice and *Candice-like* agents do not succeed in having diachronic self-governance, yet diachronic self-governance is something we value or at least should value in our everyday lives.

The second example of an agent possibly failing at diachronic self-governance is *An evening with Ann*. We have discussed this example in the previous chapter so I will briefly recapitulate the content of it.

*An evening with Ann*

*Ann likes spending her evenings alone drinking beer and reading a good book. She knows from experience that she cannot read a book if she has more than one beer in the evening.*

*Before the evening she prefers to spend it drinking one beer and then reading a good book. But during the evening, when she is done with the beer, her preferences change and now she prefers*

---

<sup>70</sup> Bratman takes the concept of "shuffle life" from Richard Kraut which he delineated in one of their conversations.

*to spend her evening drinking two beers over drinking one beer plus reading a good book. Ann is in a dilemma how to spend her evening* (Bratman 1999).

The case of *An evening with Ann* is a clear case of temptation (or in some sense a weakness of will). Ann has a clear policy in mind of how she wants to spend her evenings. Her policy says: one beer plus reading a good book. At the time of action, she is tempted to have a second beer. Bratman argues that there is a reason for Ann to honor her own policy, a reason to, put in these terms, stay true to herself. That reason or value is self-governance.

Both Candice and Ann, in order to govern their own lives successfully or to have self-governance, need to have a place to stand. As we have discussed earlier, there are different standpoint functions that agents can acquire. For example, *time-slice evaluative standpoint function* or *cross-temporally structured standpoint function* (the list is inexhaustible). Each of these functions is perfectly valid and legitimate for an agent to have. It seems that, in the examples presented above, Candice and Ann have some desire or idea how they want to spend the afternoon (in Candice's) or evening (in Ann's case). They have (or at least should have) a place to stand and that standpoint function should be *cross-temporally structured*. By being temporally unified, psychologically continuous agents with a place to stand, they have access to the notion of self-governance. If they (Candice and Ann) value governing their own lives, and Bratman assumes that we all do (or at least should), then Candice should stick to one of her plans how to spend the afternoon and Ann should stick to her policy of one beer plus reading a good book. Now we are ready to engage with Bratman's account of diachronic rationality.

#### **2.5.4 Rational non-reconsideration model**

Bratman's account of diachronic rationality consists of two parts: of the Rational non-reconsideration model and the strategy of self-governance. Firstly, I am going to examine the Rational non-reconsideration model. Secondly, I will present the strategy of self-governance. Lastly, I am going to explain Bratman's account of diachronic self-governance.

The idea behind the *rational non-reconsideration* model is quite simple. The model presents to an agent a rational option to choose contrary to her current preferences. The motivation for it is the fact that we as human beings often find ourselves in situations in which we choose contrary to our better judgment. These situations are, for instance, cases of temptation, weakness of will and possibly addiction. The rational non-reconsideration model presents a

way in which agents can rationally overcome these temporary changes in preferences and rationally choose contrary to their current preferences.

There are two main inspirations for the development of the rational non-reconsideration model. The first is Bratman's own account of non-reconsideration policy which he presented in *Intentions, Plans and Practical Reason* (I have discussed this in length previously in this chapter). The second is Richard Holton's account of willful resoluteness which he presented in *Willing, Wanting, Waiting*. Holton claims that we, as human beings, have access to a special kind of intentions (which he calls resolutions) that have certain features, most important of which is non-reconsideration. Additionally, Holton argues that we possess the executive capacity of *willpower* which we mostly use in cases of temptation and weakness of will.<sup>71</sup> Holton's account is, in terminology that I have presented in the first chapter, a Kantian account of diachronic rationality and it has an important influence on Bratman's account of diachronic rationality.

Bratman argues for the rational non-reconsideration model by using the example which I have named *An evening with Ann* (I have addressed this example in detail earlier in this chapter). Bratman presents the example in the following manner (which is, at least on the surface, similar to the way in which authors who discussed the dynamic choice problem viewed diachronic rationality).

Ann decides at t1 to A at t2. She, at t1 has good reasons, the ones that matter to her, to A at t2. There are two possible scenarios that can occur. First: she, at t2, does not reconsider and goes through with her initial plan to A. Second: she, at t2, does reconsider and (possibly in the light of new information) rethinks her initial plan and chooses not to A (Bratman 152: 2018).

The first is the case of rational non-reconsideration and the second is the case of rational reconsideration. Bratman argues for the former or for the rational non-reconsideration model.

The Rational non-reconsideration model consists of four parts, namely:

1. Pragmatic edge
2. Snowball effect
3. Anticipated regret
4. Agent's standpoint.

Pragmatic edge

---

<sup>71</sup> This is by no means an exhaustive description of Holton's account. For more see (Holton 2009).

Pragmatic edge is simply the notion that there is a variety of pragmatic reasons to accept the Rational non-reconsideration model. Firstly, we are beings who are limited in all of our resources (cognitive capacity, processing power). Secondly, we are beings who are susceptible to short-term distortions in our attitudes (temptations, weakness of will, addictions of any kind). Thirdly, there are clear advantages, when it comes to coordination with other agents or ourselves (persistently over time), to having the matter settled in advance.

#### The snowball effect

The snowball effect is the notion that every relevant step that an agent takes from  $t_1$  to  $t_2$  brings her closer to the idea that doing A at  $t_2$  is a better option for her than not doing A at  $t_2$ . Bratman takes the example of an agent traveling to Boston from New York. If an agent decides at  $t_1$  to go to Boston and then at some point between  $t_1$  and  $t_2$  buys airplane tickets, her “scale” is tilted more to *going to Boston* than to *not going to Boston*. Every relevant action that the agent takes will tilt her more to doing the action which she initially planned. Along the way, the agent constantly has more reasons to honor her initial intention.

#### Anticipated regret

Anticipated regret is the notion that I have discussed in detail previously in this chapter, so I will briefly summarize the key features. The notion of anticipated regret is, according to Bratman, an extremely common sense idea that an agent should refrain from doing something, even though she wants to do it now, because she will regret it later. There are two preconditions for anticipated regret: agent’s awareness of impending temptation and agent’s capacity to form policy-like intentions. Let us see it on the example of *An evening with Ann*. Ann is aware that she will be tempted during the evening, specifically after she has had one beer. She is also capable of forming a one-beer-in-the-evening policy. How should Ann’s decision-making process be carried out? There are three relevant points in time for Ann. The time when she forms the intention ( $t_1$ ), the time when she is tempted to drink another beer ( $t_2$ ), and the future time when she regrets drinking another beer that evening ( $t_3$ ). If this is the way that Ann is capable of framing her decision problem, then there are good reasons for her to conform to the rational non-reconsideration model. What are those good reasons? The first reason is that Ann knows that her desire to change her mind at  $t_2$  is a temporary shift in her preferences and does

not reflect what she truly wants. The second reason is that there is evidence in favor of her non-reconsidering. The evidence is her future regret. Ann now knows that she will regret her decision in the future and she can use this knowledge of her future regret as a reason not to reconsider at the time of temptation.

### Agent's standpoint

Agent's standpoint is another concept that I have established and explained previously in this chapter, so I will just briefly touch upon it. The concept of agent's standpoint or "where she stands" is the idea that, while at any point an agent has a multitude of attitudes, desires and intentions at her disposal, she has the ability to wholeheartedly identify with a certain subset of those attitudes which enables her to act. That being said, an agent's standpoint is not simply her identification with a specific set of attitudes but more *holistic properties of the agent's psychological economy* (Bratman 2014). In one sense it is comparable to an agent's perspective on matters relevant to her and in another sense it is similar to the concept of diachronic will (Velleman 2000, Ferrero 2006, 2009, 2010, 2012). Bratman presents the concept of agent's standpoint in the form of a standpoint function. A standpoint function is a function from the agent's overall psychological profile to where she stands (Bratman 162: 2018). There are two main standpoint functions (although the list is not exhaustive): the *time-slice evaluative* standpoint function and the *cross-temporally structured* standpoint function. The former states that the agent makes decisions simply by looking at her current attitudes or options. The latter standpoint function enables the agent to make decisions based on her previous commitments and/or future prospects. Bratman argues that it is of greatest importance for agents to have a clear and coherent standpoint on matters relevant for their actions. It is a norm of practical rationality called *Synchronic Standpoint Rationality* (the norm is explained in detail in a previous paragraph in this chapter). Let us return to our *An evening with Ann* example. What should Ann do? For a start, she is rationally obligated to have a clear and coherent place to stand or a standpoint function. If that standpoint function is *cross-temporally structured*, i.e., she values and takes into account her previously formed commitments and future prospects, then she has good reasons to conform to her policy of one beer plus a good book in the evening. As we have seen, the rational non-reconsideration model does not offer any diachronic norms of rationality. The purpose of the RNR model is not to force agents to act in ways contrary to their current desires and preferences, but to allow them the possibility to choose contrary to their current desires and preferences. There are five reasons for accepting the RNR model or

the model of practical decision making which allows an agent to choose contrary to her current preferences. The first reason is that it is pragmatically advantageous to do so. Because of their cognitive, processing and other limitations, human beings make plans in order to coordinate their own activities and activities with others. It is in our interest to have a model like RNR which allows us to have stability and inertia in our everyday lives. The second reason is the snowball effect. When we are “on our way” to a desire destination or goal, it seems that the closer we are to that goal, the more reasons we have not to reconsider. We have more reasons to see our plans unfold to their end. The third reason is future regret. We can use awareness of our possible future regret as evidence or reason for choosing contrary to our current preferences or desires. This is most evident in the case of temptation (as in the example of *An evening with Ann*), but it can be applied to a broader set of cases. Lastly, we have agent’s standpoint. An agent needs to know where she stands regarding the matter relevant for her action. That standpoint needs to be clear and coherent. There are different standpoints, but in this case the relevant one is the cross-temporally structured standpoint which allows agents to conform to the RNR model and choose contrary to their current preferences. All of these reasons serve as support to the plausibility of the RNR model.

### 2.5.5 The strategy of self-governance

The idea behind the strategy of self-governance is quite simple: we, as planning agents, value governing our own lives and this value gives us good reasons to accept some account of diachronic rationality. Bratman clearly states the three presuppositions for his account of diachronic rationality. Firstly, it applies only to those agents who are already planning agents (as I have demonstrated in the planning theory of intention earlier in this chapter). Secondly, Bratman’s account of diachronic rationality is not the only way to think about diachronic agency and diachronic rationality (for comparable alternatives see (Ferrero 2010, 2012, Holton 2009)). And lastly, Bratman’s account does not consider trivial or tragic cases (Sophie’s Choice of 1979).

Bratman presents his account of diachronic rationality – *Diachronic Plan Rationality* – as a set of norms that a reflective rational agent should accept (those norms are synchronic as well as diachronic). There are several things that we need to accept before addressing *Diachronic Plan Rationality*:

- The planning theory of intention

- The rational non-reconsideration model
- The notion of synchronic and diachronic self-governance
- The notion of a practical standpoint (“where one stands”).

I have discussed these ideas at length throughout this chapter so I will just briefly summarize them. The planning theory of intention states that human beings have the capacity or ability to formulate and execute plans and that this ability is crucial in explaining our practical reasoning, practical rationality and decision making. The rational non-reconsideration model states that there is a way in which an agent can rationally choose contrary to her current preferences (which in turn explains the cases of temptation and weakness of will). The notion of self-governance states that we, as human beings, value governing our lives. Moreover, we can govern ourselves at a particular time (synchronic) and over time (diachronic self-governance). Lastly, the notion of a practical standpoint tells us that we have a set of attitudes (although it is more a holistic property of an agent’s psychological economy) which we as agents identify with at any particular moment and we choose to act according to our practical standpoint. These ideas are fundamental to Bratman’s *Diachronic Plan Rationality*. Now I will address the norms of this account.

*Practical Rationality/Self-governance (PRSG): If S is capable of self-governance it is, defeasibly, pro tanto irrational of S either to fail to have a coherent practical standpoint or to choose in a way that does not cohere with her standpoint.* (Bratman 211: 2018).

The idea of the norm is quite simple: an agent, if she is capable of self-governance (both synchronic or diachronic), needs to

(a) *have a coherent practical standpoint*

(b) *needs to choose in a way that coheres with her practical standpoint.*

Self-governance is, as I have discussed in depth in this chapter, a value of governing our own lives. We, presumably, value governing our own lives, i.e., having a certain amount of stability and coherence in our thoughts and actions. The *Practical standpoint* norm states that if we are capable of that, then we need to have a coherent standpoint. A standpoint or “a place to stand” is the idea (also described thoroughly in this chapter) that an agent has a thought or an attitude which she identifies with regarding the matter relevant for her action. We can have, for instance, a plan-infused practical standpoint (Bratman argues in favor of those) in which we

are concerned with future prospects and past commitments. Alternatively, we can have a time-slice evaluative standpoint in which we are concerned only with our current thoughts and actions. The *Practical standpoint* norm simply affirms that whatever an agent's practical point is, she needs to have one regarding some relevant matter. Additionally, the norm states that an agent should "put her money where her mouth is" and act according to her practical standpoint. There are two caveats to this norm (as to other diachronic norms of rationality). The *Practical standpoint* norm holds only pro tanto and defeasibly. The *pro tanto* clause simply implies that the norm is somewhat immune to the paradox of the preface objection, while the *defeasibly* clause states that the norm holds until there is relevant new information or reasons in play.

*Practical Rationality/Self-Governance-Planning Agency (PRSG-P): If S is a planning agent who is capable of self-governance it is, defeasibly, pro tanto irrational of S either to fail to have a coherent practical plan-infused standpoint or to choose in a way that does not cohere with her plan-infused standpoint. (Bratman 213: 2018).*

The first thing that we notice is that this norm, *Practical plan-infused standpoint*, is built on the *Practical standpoint* norm. Everything is identical except now we are discussing planning agents and their plans or plan-infused standpoints. While the *Practical standpoint* norm addressed any kind of agent (planning agents and non-planning agents alike), the *Practical plan-infused standpoint* norm is predicated on the idea that we are sufficiently convinced that there are good reasons for us to be planning agents (as I have discussed in this chapter). The *Practical plan-infused standpoint* norm states that a planning agent should have a practical plan-infused standpoint. That standpoint is not time-slice evaluative, but instead, in some sense, diachronic in nature. It is a standpoint in which an agent takes intentions to be plans and plans have something to do with our future conduct.

*Plan consistency and coherence (PCC): If S is a planning agent who is capable of self-governance it is, doubly defeasibly, pro tanto irrational of S to have plans that are inconsistent or means-end incoherent, given her beliefs. (Bratman 214: 2018).*

Firstly, we should stress that this norm is predicated on the two previous norms: *Practical standpoint* and *Practical plan-infused standpoint*. Secondly, what this norm "brings to the table" is that an agent, if a planning agent and if capable of self-governance, needs to have her plans consistent and means-end coherent. Bratman has presented the norms named *Intention consistency* and *Means-end coherence* before (I have discussed this intensively in this chapter)

as fundamental norms of practical rationality and decision making. He was criticized, most notably by Joseph Raz and Niko Kolodny, that these norms present no “real” normative value and have merely pragmatic benefits to our everyday lives. The norm, *Plan consistency and coherence*, is an attempt to address this criticism. There are normative reasons to accept *Intention consistency* and *Means-end coherence* and that is, if the agent is a planning agent and capable of self-governance, the value of self-governance.

*Diachronic Plan Rationality (DPR): If S is a planning agent who is capable of diachronic self-governance then the following is, defeasibly, pro tanto irrational of S:*

*(a) S is engaged in a planned temporally extended activity that has so far cohered with both synchronic and diachronic self-governance.*

*(b) Given her present standpoint, a choice to continue with her planned activity would cohere with that standpoint and so cohere with her continued synchronic self-governance and, in part for that reason, with her diachronic self-governance. And yet*

*(c) S makes a choice that blocks her continued diachronic self-governance. (Bratman 217: 2018).*

There is a lot to tackle here. Let us begin with a general overview. Firstly, DPR is predicated on the abovementioned norms. Secondly, this is the first (of the two) genuine diachronic norms in Bratman’s diachronic account of rationality. Thirdly, it is a modest norm of rationality since doubly defeasible and extremely narrow. The norm addresses only those agents who are *planning agents, capable of self-governance* and already *engaged in a planned temporally extended activity*. Now, let us discuss *Diachronic Plan Rationality* in depth. The norm consists of two main conditions, if we exclude the ones that are the same as the conditions of the previously mentioned norms of Bratman’s account. Condition (a) simply states that the agent needs to be involved in a temporally extended activity which is relevantly coherent (synchronically and diachronically). This is a *historical condition*, as Bratman calls it. The second condition, condition (b), is more controversial. It states that an agent’s *choice to continue* needs to align with her *present standpoint*. Now the question arises: what if an agent’s *present standpoint* does not lead to her *choice to continue*? In order to solve this problem Bratman, as a *Kantian* in respect to diachronic rationality or at least heavily inspired by *Kantians* (Ferrero 2010, 2012, Velleman 2000, Holton 2009), invokes the concept of willpower. But how does willpower play a role in explaining condition (b)? While Ferrero and Velleman focus on willpower itself (the specific term they use is *diachronic will*), Bratman

argues that willpower is actually a consequence of our self-governance and that this is the way agents connect the *present standpoint* and the *choice to continue*. This, of course, leaves us with the last and final question: why should not the agent block her *continued diachronic self-governance* as it is described in (c)? In order to answer this properly, we should take a closer look at Bratman's last diachronic norm, *Rational End of Diachronic Self-governance*.

*Rational End of Diachronic Self-governance (REDSG): If S is a planning agent who is capable of diachronic self-governance then it is pro tanto irrational of S to fail to have an end of diachronic self-governance.* (Bratman 220: 2018).

On the surface, this norm might seem too strong or even strange. Bratman defends the norm on several grounds. Firstly, he argues that in *the spirit of inference to the best normative explanation* (Bratman 2018), there is a rational pressure to accept the value of self-governance. Secondly, in virtue of being planning agents, we are led into the strategy of self-governance by a meaningful interpretation of consistency and coherence of our plans. Thirdly, the norm *Rational End of Diachronic Self-governance* is quite weak. It does not involve trivial or tragic cases and it does not apply to agents who are not already planning agents. Also, the norm can be satisfied by different agents in different ways relative to their respective standpoints. The main focus of the norm is to put rational pressure in favor of willpower in temptation and non-comparability cases. In other words, the scope of the norm is quite narrow.

Bratman's account of diachronic rationality, *Diachronic Plan Rationality*, is built on two ideas: the planning theory of intention and the rational non-reconsideration model. The planning theory of intention states that human beings have the capacity to formulate and execute plans and that plans are fundamental for coordination with ourselves (at different times) and with others in order to achieve complex and temporally distant goals. The rational non-reconsideration model is based on the idea that agents can rationally choose contrary to their present/current desires or preferences. This allows agents to successfully navigate through situations of temptation and weakness of will.

The central idea behind *Diachronic Plan Rationality* is the notion of self-governance. Self-governance, inspired by the works of Harry Frankfurt and Gary Watson, is a non-instrumental value of coherence that we have in regard to running our own lives. The idea is that we value stability and coherence in governing our own lives freely and reflectively as opposed to wondering willy-nilly from one unfinished project to another.

On that basis, Bratman presents the five norms of his diachronic account of self-governance. Those are as follows: *Practical standpoint*, *Practical plan-infused standpoint*, *Plan consistency and coherence*, *Diachronic Plan Rationality* and *Rational End of Diachronic Self-governance*. The first three norms are synchronic norms and the last two are diachronic norms of practical rationality. All of these norms are extremely modest and weak. In other words, they are all defeasible and pro tanto norms. Together with the concepts of practical standpoint and synchronic and diachronic self-governance, these norms form most of the account of diachronic rationality, i.e., Diachronic Plan Rationality.

## 2.6 Chapter Summary

In this chapter I presented Bratman's *planning theory of intention* and Bratman's *Diachronic Plan Rationality*. Planning theory of intention, although not diachronic account of rationality, is a foundation for most contemporary accounts of rationality (Ferrero 2009, 2010, 2012, Holton 2009, Broome 2013, Snedegar 2017). Bratman argued against belief-desire model of practical rationality ((argued for by his predecessors Davidson (1980) and Anscombe (2000)) and for the idea that intentions matter when we discuss philosophy of action, philosophy of mind and philosophy of practical rationality. Intentions, although sometimes can be focused on the present, are usually future focused attitudes. Bratman's idea is that we should think of intentions as plans. Plans are intentions that have certain properties to them; stability, inertia, partiality, commitment and forward looking. We, as human beings, make plans. As Bratman would put it, we are planning agents. This is a uniquely human ability that we do not share with other animals. There are two reasons why we make plans. The first reason is because we are cognitively limited beings and continuous deliberation and redeliberation of the same issue would be cognitively demanding and extremely unpractical. The second reason is that we need a way to coordinate with other agents and with ourselves in different times and the best way to do this is by planning ahead.

On the basis of planning theory of intention Bratman is developing his diachronic account of rationality; *Diachronic Plan Rationality*. Bratman claims that we should agree that agents value governing their own lives and he calls that value *self-governance*. Self-governance serves as a reason to accept Bratman's classic practical norms of rationality *Means-End Coherence* and *Stability of Intention*. As a diachronic account of rationality, *Diachronic Plan Rationality* is focused on agent's coherence and stability of her intentions and attitudes over time. The

diachronic norm that Bratman is proposing is Diachronic Plan Rationality norm. It is a modest and defeasible norm that takes into account the unpredictability of future state of affairs of the world.

## **3 A case against diachronic rationality: Tragedy, Options and Time-slice Rationality<sup>72</sup>**

In this chapter I am going to present Brian Hedden's synchronic account of diachronic rationality, *Time-slice Rationality*. This will serve two main purposes. First, it will be an exploration into the account of rationality, specifically, Time-slice Rationality, which stands on purely synchronic grounds. Second, it will pose a formidable challenge to my thesis, *Humans have a capacity for diachronic agency and we can make rational assessments of that agency* (more on the second part than on the first), which I will address later.

### **3.1 Theoretical background and the main idea of Time-slice Rationality**

Brian Hedden's account of diachronic rationality, Time-slice Rationality, is the most extensive and comprehensive account of rationality that stands explicitly on the synchronic norms of rationality. The account of Time-slice Rationality was born in the paper called *Time-slice Rationality* from 2015. This makes the account extremely young and possibly underdeveloped (at least when compared to the accounts of Michael Bratman and John Broome who have had over thirty years for the development and refinement of their respective accounts). Nevertheless, Time-slice Rationality is a rich and powerful account with fresh and new ideas that are fairly influential. It has challenged some authors to develop their accounts which explicitly deny the existence of diachronic norms of rationality, such as *Time-slice epistemology* (Moss 2015), and impacted other authors to use ideas from the Time-slice Rationality account and apply it to other fields (Builes 2019, Doody 2019).

---

<sup>72</sup> I will not be addressing the entirety of the Time-slice Rationality account. The Time-slice Rationality account is a rich and extensive account that encompasses both theoretical and practical rationality. On the theoretical side of rationality, Brian Hedden (founder of Time-slice Rationality), argues for the abandonment of two diachronic principles (norms/requirements): conditionalization and reflection. He replaces those principles with new ones: synchronic conditionalization and uniqueness. Although some authors (Bratman 2014) do occasionally venture into this domain whilst arguing from a domain of practical rationality, most authors, specifically those who discuss practical diachronic rationality, do not mention these principles (Velleman 2000, Ferrero 2006, 2010, 2012, 2014, Holton 2009, Broome 2013). Since Time-slice Rationality is an account of both theoretical and practical rationality, I will focus (as I have explained in detail in the first chapter) on aspects of Time-Slice Rationality that are relevant to practical rationality.

Using terminology presented in the first chapter, from the historical view, Time-slice Rationality is located in the third phase or *diachronic rationality as a problem of the nature and existence of diachronic norms* (although it has several strong connections to the first phase or *diachronic rationality as a problem of dynamic (sequential) choice* and I will explain why this is the case in this chapter). From the contemporary view, Time-slice Rationality is the *antirealist* account of diachronic rationality, which means that the proponents of Time-slice Rationality claim there to be no genuine norms of diachronic rationality.

The language of Time-slice Rationality is extremely synchronic-centered. This means that it uses terms that cater to our synchronic intuitions. These are mostly the terms from decision theory, such as *preferences*, *utilities* and *choices* (although there are others like *options* which are unique to the Time-slice Rationality account). Conversely, the language of diachronic accounts (such as Bratman's account of *Diachronic Plan Rationality*, Broome's *Persistence of intention* or Ferrero's account of diachronic rationality) is mostly diachronic-centered. This means that these authors use terms that cater to our diachronic intuitions, such as *agency*, *intention* and *narrative* in their respective accounts. Although there is a certain terminological divide, all of these authors (Bratman, Broome, Ferrero, Holton, Hedden) agree that they all talk about the same thing – the notion of diachronic rationality.

The general idea of Time-slice Rationality is staggeringly simple: there are no genuine diachronic norms of rationality. There are no good reasons or arguments why someone should accept the existence of diachronic norms and abide by them. When there is a case of an apparent violation of a diachronic norm of rationality, there are two possible explanations. Either that norm is actually reducible to a synchronic norm of rationality or that norm is not actually a norm of rationality. These claims are highly controversial for at least two reasons. Firstly, there are already comprehensive and extensive accounts of diachronic rationality (Bratman 2018, Broome 2013, Holton 2009). Secondly, if there are actually no diachronic norms of rationality, that seems to leave a huge explanatory gap when discussing short- and long-term decision making of real agents. An almost uncontroversial fact is that we make *plans*, *resolutions*, *commitments* and *promises* and are usually able to stick to them despite weakness of will or other temporary distractions. The burden is, at least partially, on Hedden to explain why we seemingly do these things and why we value them. In his account of Time-slice Rationality, he proposes an alternative view on these phenomena and offers explanation for them.

There are two main influences on the formation of Time-slice Rationality. The first influence is explicitly stated at the beginning of Hedden's book *Reasons without Persons* (2015). More specifically, it is Derek Parfit's book *Reasons and Persons* (1984) (even the title of Hedden's

book *Reasons without Persons* is an homage to Parfit's book *Reasons and Persons*). The inspiration for Time-slice Rationality comes, according to Hedden, from Parfit's quote:

*When we are considering both theoretical and practical rationality, the relation between a person now and himself at other times is relevantly similar to the relation between different people* (Parfit 190: 1984).

The entirety of the Time-slice Rationality account is, at least figuratively, an expansion of this Parfit's quote.

The second major influence is Robert Stalnaker's paper *Extensive and strategic forms: Games and models for games* (1999) in which he connects the problem of dynamic choice (which is still dominated at that time and in some sense even today) with the problem of personal identity over time. The problem of personal identity over time and the way it relates to our ability to rationally make decisions is a starting position for Hedden's Time-slice Rationality.

### **3.2 The rationality of Time-slice Rationality**

In order to address the Time-slice Rationality account, first we need to know what Hedden's take on the notion of rationality is. According to Hedden, we frequently evaluate a wide range of different things as either rational or irrational. For example, we can evaluate the rationality of *people, dispositions, habits, emotions, and even laws, city layouts, voting systems, arguments, and conversations* (Hedden 10: 2015b). This is not what the Time-slice Rationality account is about. The Time-slice Rationality account pertains to the evaluation of agent's beliefs, credences, preferences and actions (I will focus my discussion on preferences and actions). The question arises: How do we evaluate agent's preferences and actions? Hedden offers a simple example to clarify his point:

*Suppose that your friend has a headache, and you have some pills that you justifiably believe to be pain relievers. But you're wrong. They are really poison. Given that you want to help your friend, you rationally ought to give him the pills, even though they will in fact do him harm. You would be quite irrational if, despite your confidence that giving him the pills will relieve his headache and your desire to help him, you neglected to offer him the pills. Of course, when your friend winds up writhing around on the floor and foaming at the mouth, you will quite rightly regret offering the pills, but this does not mean that your initial decision*

*was irrational. It just means that being rational is no guarantee of success in your endeavours* (Hedden 1: 2015b).

There are three important aspects that we can take from this in respect to Hedden's view on rationality. Firstly, he is clearly inspired by the famous *gin and tonic* case presented by Bernard Williams in his influential paper *Internal and External Reasons* (1981). Secondly, Hedden is an outspoken proponent of internalism in regard to the nature of rationality and reasons (although he claims that the Time-slice Rationality account is compatible with weak externalism). Lastly, Hedden argues for the subjective notion of rationality or the notion of *subjective ought*. This means that rationality is not reducible only to things that are true nor is rationality merely a question of satisfying one's desires and preferences. Rationality is concerned with what an agent should do in respect to her perspective on the world, i.e., on evidence available to her at that point. This summarizes the main idea of Hedden's approach to rationality. Now we shall explore of what exactly Hedden's general approach to rationality consists.

According to Hedden, there are three roles that rationality<sup>73</sup> should play: an *evaluative role*, a *predictive and explanatory role* and an *action-guiding role*.

The evaluative role of rationality states that when an agent violates a principle of rationality, she is open to a specific form of criticism. What kind of criticism is that? In Hedden's words it is the following:

*If you are irrational, you are doing something that even by your own lights is a mistake. You are making a mistake that you are in some sense specially placed to recognize and correct. Being irrational is a matter of making a mistake that is in some sense internal to you and not just the result of the world being uncooperative* (Hedden 10: 2015b).

In other words, being irrational means that an agent is in conflict with herself and is, in some sense, aware of that fact. The two main aspects of the evaluative role of rationality are *internalism* and *evidentialism*.

---

<sup>73</sup> Although Hedden does not directly affirm this, his roles of rationality are heavily inspired by *dimensions of rationality* presented by José Luis Bermúdez in his influential book *Decision Theory and Rationality*. Bermúdez's dimensions of rationality are: *an action-guiding dimension, a normative assessment dimension and a predictive and explanatory dimension*. As we can clearly see, there is no denial that Hedden's roles of rationality are inspired by Bermúdez's dimensions of rationality. That being said, although the terms are almost identical (evaluation and assessment are synonyms in this sense), there is substantial difference between the two. While Bermúdez's dimensions of rationality are quite neutral in respect to the internalist versus externalist debate, Hedden's roles of rationality are built on internalist grounds (for more on *dimensions of rationality* see (Bermúdez 2009), especially the first chapter).

The second role that rationality should play is prediction and explanation. We can, with some degree of certainty, predict that people will do what they rationally ought to do. Also, we can explain people's behavior by pointing out that they are mostly rational and take evidence into account when making decisions. Of course, this predictability stands *ceteris paribus*. If there is evidence that someone has behaved irrationally in the past, we will not be able to predict her behavior now. This role stands on the presupposition that agents generally do act rationally. The last role of rationality is the action-guiding role. A theory of practical rationality should be able to instruct agents what they should do, i.e., what they *ought* to do. What an agent rationally *ought* to do is sensitive to the available evidence and her perspective on the world and is not based only on the external facts of the world.

### 3.3 Motivations for Time-slice Rationality

There are two main motivations for developing the Time-slice Rationality account. The first is the problem of personal identity and the second is internalism. According to Hedden, the puzzles about the nature of personal identity should have no bearing upon the notion of rationality. The norms or requirements of rationality should be entirely *impersonal* (which is one of the norms of the Time-slice rationality account that we will explore extensively later). Hedden presents his version of a familiar problem of personal identity:

#### *Teletransportation*

*You enter the teletransporter. The machine scans you, identifying and recording the exact molecular structure of your body. Then, the information is sent to Australia, and your body is destroyed just as a molecule-for-molecule copy of your body is created in Sydney. (Hedden 15: 2015b).*

The main question arising from this example is the following: are the person who entered the *teletransporter* and the person created in Sydney the same person? There is no easy answer to this question. There is a wide range of theories aiming to answer this question. Probably the most prominent ones are the *psychological continuity view* (Shoemaker 1970, Parfit 1984) and the *biological view* (Olson 1997a, 1997b). The former would claim that the person who entered the *teletransporter* and the one created in Sydney are in fact the same person because they have

a continuity of mental states which links them together into one person. The latter would claim that they are in fact not the same person but two different persons because they occupy two different biological bodies. Hedden's point is the following. Whatever the correct answer to the puzzle of personal identity is, it should be completely irrelevant when discussing the notion of rationality, i.e., what an agent rationally *ought* to do. Even if someone did commit themselves to a certain view on the nature of personal identity, that would create more problems than it would solve. Firstly, the accounts of rationality which stand on the presupposition that a certain metaphysical theory of personal identity is true, are inherently weaker than the accounts of rationality that do not. Secondly, there are problems of personal identity that are not solvable by "choosing" one of the dominant views regarding the problem. As an example, Hedden presents *Double Teletransportation*.

### *Double Teletransportation*

*One person (call her "Pre") enters the teletransporter. Her body is scanned. Then, at the instant her body is vaporized, the information about her molecular state is beamed to two locations, Los Angeles and San Francisco. In each city, a molecule-for-molecule duplicate of Pre is created. Call the one in Los Angeles "Lefty" and the one in San Francisco "Righty." Lefty and Righty are each qualitatively just like Pre is before her body is vaporized. (Hedden 16: 2015b).*

There are no easy solutions to this problem that we could "choose". The main reason is because "the original person" can be "duplicated" almost infinitely many times. The sensible thing to do, according to Hedden, is to eliminate personal identity from the domain of rationality. Rationality should be *impersonal* and *time-slice* focused, not dependent on a *temporally extended person*.

The second motivation for Time-slice Rationality nicely builds on the first – internalism. What attitudes should an agent rationally hold, should depend on her mental states. The notion of synchronicity and internalism, according to Hedden at least, go hand in hand. The reason for this is because our current mental states are, by their very definition, something that is available to us now. Although internalism is a powerful reason to accept Hedden's account, Time-slice Rationality is not incompatible with the externalist notions of rationality. The only thing that the externalist has to concede in order to accept Time-slice Rationality is that facts about one's

past matter for what her attitudes *ought* to be now. The form of internalism that Hedden argues for most is the account of *mentalist internalism*. In Hedden's words, it is the following:

### *Mentalist Internalism*

*What you rationally ought to believe, desire, or do supervenes on your mental states* (Hedden 23: 2015b).

There are several things that we need to address here. Firstly, we should turn to the concept of supervenience. The concept of supervenience is defined as a metaphysical and/or conceptual determination-relation (Horgan 1993). The relation states that a set of properties A supervenes on a set of properties B if and only if a change in the set of properties A effects a change in the set of properties B (Bennett and McLaughlin 2018). In the context of rationality, this means that any requirements/norms of rationality have to *supervene* on agent's mental states and not be some external facts that the agent cannot be aware of. Secondly, Hedden argues that *mental states* should be considered crucial when discussing rationality. What an agent *ought* to do cannot be reducible to a physical system or perceptive apparatus. Although these things play an important explanatory role in discovering why beliefs are true or reliable, they in themselves cannot tell us whether an agent's beliefs or actions are rational or not. There is a lot of challenges when subscribing to a position as contested as internalism (although externalism is highly opposed as well). The main challenges include *Brain in the vat* cases (BIV), Davidson's *Swampman* and, what Hedden calls, "extreme" cases of internalism. The first is an account of internalism which claims that what an agent *ought* to believe supervenes on her intrinsic physical properties. The second account claims that an agent should have perfect access to her mental states. Hedden's response to these worries is his claim that *mentalist internalism* is a modest form of internalism. He grants to the proponents of BIV cases and the *Swampman* case that these two individuals do not have the same mental content and argues that this is compatible with his account because what an agent *ought* to believe is based on the effect it has on her current mental states. Subsequently, he rejects "extreme" cases of internalism. Hedden's rejection of the internalism of intrinsic physical properties is based on the grounds that agent's mental states are dependent not only on the intrinsic physical properties but on the facts about the environment as well. The rejection of the perfect-access internalism is based on

Williamson's anti-luminosity argument.<sup>74</sup> Hedden retains the modesty of his *mentalist internalism*. According to his account, an agent does not need to have infallible access to her mental states or have her mental states supervene on the intrinsic physical properties. Rational belief or action of an agent depends on the mental states which she currently has access to – no more and no less.

We have seen two principal motivations for the Time-slice Rationality account, those being the problems of personal identity and internalism. Now I will present the Time-slice Rationality account in terms of rational requirements/norms.

The main aim of Hedden's Time-slice Rationality is to develop an alternative approach to theoretical and practical rationality. Hedden's goal is to create an account that can challenge the dominant accounts of diachronic rationality, such as Bratman's *Diachronic Plan Rationality* and Broome's *Persistence of Intention* (Hedden 2015a, 2015b). Time-slice Rationality aims to do "more with less". In other words, Hedden's account promises to have the same explanatory and normative power as his diachronic counterparts but with fewer rational requirements/norms. As mentioned before, the Time-slice Rationality account is, at least in its core idea, extremely simple: there are no genuine diachronic norms of rationality. But if there are no diachronic norms of rationality, what norms does Hedden propose? Time-slice Rationality consists of only two<sup>75</sup> rational requirements/norms: *Synchronicity* and *Impartiality*. Now we will take a closer look at these norms.

### *Time-Slice Rationality*

- *Synchronicity: All requirements of rationality are synchronic.*
- *Impartiality: In determining how you rationally ought to be at a time, your beliefs about what attitudes you have at other times play the same role as your beliefs about what attitudes other people have.* (Hedden 9: 2015a).

---

<sup>74</sup> For explanation of Williamson's anti-luminosity argument see (Williamson 2000).

<sup>75</sup> The Time-slice Rationality account actually consists of more than just synchronicity and impartiality. There are really four norms of the Time-slice Rationality account: *Synchronicity*, *Impartiality*, *Synchronic conditionalization* and *Uniqueness*. The reason why I do not mention or go into the latter two is the following. As I explained in the first chapter and at the beginning of this chapter, I am concerned with *practical diachronic rationality*. This means that my focus is on intentions and actions, not beliefs. Synchronic conditionalization and Uniqueness are replacements for the standard diachronic norms of conditionalization and reflection and belong entirely in the domain of theoretical rationality. This is the reason why I will focus my attention exclusively on Synchronicity and Impartiality.

There are three levels at which we will discuss these norms. The first level aims to inspect the implications of these norms. The second level intends to analyze the motivational reasons for accepting these norms. Finally, the third level aspires to look into the aim of these norms, i.e., what problems these norms are trying to solve.

*Synchronicity* is not actually a rational requirement or a rational norm *per se*, but a meta-rationality norm or a second-order rationality norm. Synchronicity does not tell agents how to act, what to believe or what attitudes one *ought* to have. Also, it does not tell agents how not to act, what not to believe and what attitudes one *ought* not to have. Synchronicity just states that the norms of rationality *ought* to be synchronic norms of rationality. *Impartiality* is a norm of rationality which states that agent's past beliefs are irrelevant in regard to current rational assessment. But agent's past beliefs are not simply irrelevant, they are irrelevant in a special way. Agent's past beliefs *ought* to be relevant in the same way that the current beliefs of other agents *ought* to be relevant for how she *rationally ought to be at the time*. Both of these norms are, on their own, underwhelming. They do not offer a lot of substance in terms of what an agent actually *ought* to believe or do. Hedden offers a grounding for these norms and that is *evidentialism*. Evidentialism is a thesis in epistemology, philosophy of rationality, philosophy of science (and other related areas), which states that agent's beliefs are justified in respect to the amount of evidence that supports those beliefs (Feldman and Conee 1985). Hedden's take on evidentialism is more internalist-based, so his would probably claim that agent's beliefs are justified in respect to the evidence available to her at this particular time (Hedden 2015a, 2015b). Hedden's claim is that the internalist form of evidentialism goes hand in hand with synchronicity because evidence is something available to agents right now. So, both synchronicity and impartiality are deeply rooted in this form of evidentialism.

What are the motivational reasons for accepting Synchronicity and Impartiality? According to Hedden, there are two main reasons for accepting synchronicity: internalism<sup>76</sup> and evidentialism. Hedden's claim is that what an agent rationally *ought* to believe or do supervenes on her mental states. Also, what an agent rationally *ought* to believe or do should in some capacity be available to her at that time. Internalism, developed in this way as mentalist internalism, leads directly to the acceptance of synchronicity as norm of rationality, according to Hedden. The idea behind this line of argumentation is the following. The only thing currently internally available to an agent are the beliefs, attitudes and actions now available to her (this is almost tautologically true because *currently* and *now* are used in the same sense here). If we

---

<sup>76</sup> I should note that here we are using Hedden's *mentalist internalism* which I explained earlier in this chapter.

accept *ought implies can*<sup>77</sup> (which states, roughly, that if there is a norm of rationality, then the agent must be able to conform to it), then the only norms of rationality are the ones that are synchronic.

The reasons for accepting Impartiality are also clear and known as the problems of personal identity. There are multiple problems of personal identity (as I have discussed in detail previously in this chapter) and they should, according to Hedden, bear no weight on the discussion about the philosophy of rationality. This is an indirect challenge aimed at authors who argue that there are norms of diachronic rationality (the challenge is targeted directly towards Bratman's and Broome's accounts of diachronic rationality, but it is even worse for *Kantians* (Ferrero, Holton)). The main problem is that the proponents of diachronic rationality accounts have to commit themselves to a *persistent* notion of personal identity, while a synchronist (like Hedden) is able to remain neutral in regard to the metaphysical discussion of the nature of personal identity. The Time-slice Rationality account stands to have less norms and a greater explanatory power than the rival (diachronic) accounts. Impartiality comes into play quite nicely. Agent's past beliefs should be relevant for her current rational assessment as much as beliefs of other agents at this current moment. Impartiality represents a more favorable alternative to the accounts (mainly diachronic accounts) tied up with, in Hedden's words, "weird metaphysics" (Hedden 2015a, 2015b).

Lastly, we will address what kind of problems Synchronicity and Impartiality (i.e., the Time-slice Rationality account) are trying to solve. There are two scenarios which should inspire us to think more like *time-slice agents*. The underlying theme of both examples is *changing one's mind without a reason*.

The first scenario is:

### *Career Decisions*

*Julie is in college trying to decide what career to pursue. Early in her sophomore year, she wants to study medicine and become a surgeon. But by the next week, this desire fades and she desires nothing more than to become a journalist. Shortly thereafter, she is fully committed to becoming a biologist. And she continues changing her mind throughout college. Moreover, it was not as though these shifts are the result of her learning that she is*

---

<sup>77</sup> There are many different instances of *ought implies can* and the classical form goes back to Kant. For more on *ought implies can* see (Copp 2003).

*repulsed by blood, or that she objects to the superficial media coverage so commonplace nowadays, or that she feels claustrophobic spending all day in a lab. She just ... changes her mind.* (Hedden 3: 2015a).

The second scenario is:

*The Career Counselor's Office*

*The college career counselor's office is filled with students seeking advice about how to choose their studies to match their career goals. The counselor is currently talking to a sophomore who wants to become a doctor and needs to know which courses to take to get into med school. Next in line is a prospective journalist who wants to know if working on the university newspaper will help her chances. After her is a student who wants to be a biologist but wants more details about job prospects.* (Hedden 3: 2015a).

There are several important things that Hedden is trying to highlight here. First, the standard view<sup>78</sup> (which consists of diachronic accounts of rationality like Bratman's *Diachronic Plan Rationality* and Broome's *persistence of intention*) should instruct us, according to Hedden, that *Julie* from the *Career Decisions* scenario is irrational. She drops her beliefs, changing her mind capriciously and without a reason which is deemed as an instance of irrationality. The same view (diachronic) would tell us that there is absolutely nothing irrational in the second scenario, *The Career Counselor's Office*. The scenario itself is almost trivial; it is a description of a usual working day of a *college career counselor*. There is no account of rationality (synchronic or diachronic) which would consider anything to be irrational in *The Career Counselor's Office* example. Second, the Time-slice Rationality account allows us to consider and acknowledge that there is actually nothing wrong, in terms of rationality, in the first scenario. People, according to Hedden, frequently change their minds without any apparent reason and there is nothing rationally wrong with them. *Julie* from *Career Decisions* is simply *changing her mind* throughout her college years. Lastly, Hedden would like us to consider the following. The reason why *Julie* is not irrational is not just because of our (alleged) intuitions about the case. The reason why she is not irrational is because her case, as far as rationality is concerned, is the same as *The Career Counselor's Office*. In both scenarios there are decisions

---

<sup>78</sup> Hedden's actual main point is the attack on *conditionalization* and *reflection*, but they are not the principal focus of my investigation here.

made about careers and there are different options for those decisions. The only difference is that in *Career Decisions* there is only one person in play (*Julie*) and in *The Career Counselor's Office* there are multiple students who have different ideas about the career they wish to pursue. We have seen Hedden's arguments why the concept of personal identity should not play any role in the discussion about rationality. If we accept this, then there is apparently no difference (regarding rationality) between these two scenarios. As I mentioned before, no one would claim that there is something rationally wrong in *The Career Counselor's Office* case and subsequently there is nothing rationally wrong in the *Career Decisions* case.

### 3.4 Arguments for the Time-slice Rationality account

We have discussed motivations, requirements and goals of the Time-slice Rationality account. Now we need to address the elephant in the room – arguments in favor of Time-slice Rationality. As I have mentioned at the beginning of this chapter, there are comprehensive and extensive accounts of diachronic rationality (Hedden explicitly mentions Bratman's and Broome's accounts) and without them (or some equivalent account like Ferrero's or Holton's) there is a certain explanatory gap that needs to be addressed. Hedden addresses this problem on two levels. Firstly, he argues that the problems of personal identity should lead us to abandon the notion of personal identity when discussing rationality and accept Impartiality as norm of rationality. Secondly, he claims that a specific kind of internalism – *mentalist internalism* – should lead us to accept Synchronicity as norm of rationality. But this is clearly not enough to accept Time-slice rationality as viable model of rationality. To accept the Time-slice Rationality account we need arguments. There are two arguments presented by Hedden in favor of Time-slice Rationality: *the Diachronic Tragedy Argument* and *The argument from options*<sup>79</sup>.

#### 3.4.1 The Diachronic Tragedy Argument<sup>80</sup>

---

<sup>79</sup> While the *Diachronic Tragedy Argument* is a term used by Hedden in this way (Hedden 75: 2015b), *The argument from options* is my term for describing Hedden's argument.

<sup>80</sup> There are two main arguments in favor of Time-slice Rationality: The diachronic tragedy argument and The argument from options. The first is aimed at beliefs (theoretical rationality) and the second is aimed at action (practical rationality). Although I said that I would not engage in discussions about theoretical rationality, I also believe that The diachronic tragedy argument can be illuminating for discussions in the domain of practical rationality.

The main idea behind the *Diachronic Tragedy Argument* is the following. There are clear cases in which an agent violates some diachronic norm of rationality without being irrational but simply tragic. What does it mean for an agent to be involved in a diachronic tragedy? According to Hedden, there are two kinds of tragedies. The first is a tragedy in which the agent's misfortune is anticipated by the audience. The second is a tragedy foreseeable by the agent herself. The former is a topic of discussion in the philosophy of literature and the latter is a problem in the philosophy of rationality. When the agent decides to venture on a path which she knows to have a bad outcome (in her own view) for her, then she is, in some sense, irrational. The agent has every opportunity to avoid the bad outcome but she does not do so. From the standard view (which is how Hedden calls the accounts of diachronic rationality), this behavior is classified as diachronically irrational. Hedden will claim, on the contrary, that agents are sometimes involved in a series of actions that are diachronically tragic but not necessarily irrational.

Hedden's argument is that the rational *ought* cannot be applied to a *sequence of actions*, but can instead be applied only to a particular action. That is why agents in those cases are not irrational, but simply the protagonists of *tragic sequences* of actions. In order to explain the argument, we need to establish two things: the meaning of the terms and the structure of the argument.

Firstly, we shall address the notion of *tragic sequence*. A tragic sequence is the kind of sequence of actions A1 that an agent prefers at all times during that sequence rather than performing some other sequence of actions A2 over A1. Tragic sequence gives rise to tragic attitudes. Tragic attitudes are those types of attitudes in which an agent prefers each member of a particular sequence at the time but does not prefer performing the sequence as a whole (Hedden 2015a, 2015b). There is a fundamental underlying question here. What does it actually mean for an agent to *prefer each member of a sequence* but *not prefer the sequence as a whole*? How is that possible and, even if it is, how is an agent not irrational in that case? In order to answer these questions, we need to look at an example. Hedden uses *The Russian Nobleman* case, inspired by the eponymous Derek Parfit's example.<sup>81</sup>

### *The Russian Nobleman*

*You will receive an inheritance of 100,000 rubles at age sixty. Right now, you have the*

---

<sup>81</sup> For more on Derek Parfit's *Russian's Nobleman* see (Parfit 1984).

*option (call it Donate Early) of signing a binding contract which will require 50,000 rubles to be donated to left-wing political causes. No matter whether you take this option, you will at age sixty have the option (call it Donate Late) of donating 50,000 rubles to rightwing political causes. (No greater donation is permitted under Tsarist campaign finance laws.) Right now, you most prefer donating 50,000 rubles to left-wing causes and nothing to right-wing causes. But you also prefer donating nothing to either side over donating 50,000 rubles to each side, as the effects of those donations would cancel each other out. (Hedden 79-80: 2015b).*

There is a lot to unpack in this example so we will examine it step by step. The general narrative of *The Russian Nobleman* states the following. The nobleman in the example is a fervent leftist in his youth (the example is structured from the present point of view so the nobleman's youth is described as *right now*). The nobleman will become a fervent rightist in his later years (*at the age of sixty*). He will have a gradual change of heart regarding his political views over the course of his life and he is aware of this fact. The nobleman will receive an inheritance of 100,000 rubles at sixty years of age. He can sign a contract binding him to donate half the money (50,000 rubles) to left-wing political causes right now (this option is called *Donate Early*). In the latter part of his life, he will receive an option to donate 50,000 rubles to the right-wing political causes, which matches his preferences at that time (this option is called *Donate Late*). The problem is the following. Acting in accordance with his preferences, the nobleman becomes a protagonist in a tragedy of his own making. In his youth, the nobleman prefers the option *Donate Early* and in his later years he prefers the option *Donate Late*, but ***at no point in time does the nobleman prefer Donate Early and Donate Late***. The nobleman is engaged in a *tragic sequence*. A tragic sequence is, as we have described earlier, a sequence of actions in which an agent prefers at all times some other sequence available to her. The sequence that is available to the nobleman (*Donate Early* and *Donate Late*) is a tragic sequence. The reason why this is a tragic sequence is because it is clearly and predictively disadvantageous to him and he is aware of that fact. The nobleman prefers at all times a different sequence than the one he is in. That sequence is not to donate to anyone, i.e., the sequence (not *Donate Early* and not *Donate Late*).

This is the crux of Hedden's *Diachronic Tragedy Argument*. The nobleman has a change of heart (he changes his mind) without any conclusive reason. To make matters worse, the nobleman in the example is perfectly aware that this is happening.

According to Hedden, the proponents of diachronic accounts of rationality should argue that there is a dire need for some kind of diachronic norm of rationality here. They should also

claim that the nobleman in the example is irrational. He is presumably irrational if he has violated some diachronic norm of rationality. Hedden argues that it is actually hard to infer what this apparent diachronic norm should be. The nobleman is acting against his interests because he *also prefer(s) donating nothing to either side over donating 50,000 rubles to each side, as the effects of those donations would cancel each other out* (Hedden 80: 2015b). The fact that the nobleman is acting against his best interests does not make him irrational but simply tragic.

We can acknowledge the *Diachronic Tragedy Argument* on a simpler example. Hedden uses the case of an agent who wants to quit smoking but knows that in the near future she will be tempted to smoke.

*The way not to smoke*<sup>82</sup> (The *Diachronic Tragedy Argument* simpliciter)

*Suppose that at present you want to quit smoking, but you believe that within a couple hours you will want a cigarette. You have the option right now of paying someone \$50 to prevent you from buying any cigarettes. This in effect closes off your future options, so that the only thing you can do later on is not smoke. But whether you wind up wanting a cigarette (as you now believe you will) or not, you wind up with a suboptimal outcome, since no matter what, you always prefer not smoking to not smoking plus being \$50 poorer. In this case, you wind up worse off because you are willing to pay to narrow down the options you will have in the future* (Hedden 81: 2015b).

The point that Hedden is making here is the same as in *The Russian Nobleman* but the example is simpler. *The smoker* quits smoking and simply prefers not to smoke. What she ends up doing is not smoking and not having \$50. The reason why this happens is because of the smoker's *weakness of will*. That being said, the smoker still prefers not smoking and not losing \$50 to not smoking and losing \$50. The smoker, just like the nobleman in previous example, is not irrational but simply engaged in a tragic sequence.

Hedden uses the same argumentative structure (an agent not being consistent with her preferences over time is not irrational but simply tragic) in other types of examples. Those examples are *money pump arguments* (or intransitive preferences), *imprecise preferences* and

---

<sup>82</sup> The term *The way not to smoke* is mine and does not appear in (Hedden 2015b) as such.

*imprecise credences*, and *infinite decisions*.<sup>83</sup> All of these types of problems are in, what I defined in Chapter 1, the first phase of diachronic rationality: ***diachronic rationality as a problem of dynamic (sequential) choice***. Hedden argues that these cases are examples of tragic attitudes and not irrationality.

There are two conclusions that Hedden draws from the *Diachronic Tragedy Argument*. The first is that there are no diachronic norms of rationality. The second conclusion is that the rational *ought* cannot be applied to, what Hedden calls, *sequences of actions*. The reason for this is because an agent can simply be in a tragic sequence and not necessarily irrational. These sequences of actions can be interpreted as *plans* (Bratman 1987, 2014, 2018), *resolutions* (Holton 2009), or *persistent intentions* (Broome 2013), and they are useful in at least a twofold manner: pragmatically and morally. Hedden does grant the notion that there are pragmatic benefits to thinking ahead and having stability of intentions over time as described in (Bratman 1987, 2014). Likewise, Hedden vouchsafes that there are relevant moral considerations regarding commitments and promises, but that both of these aspects (pragmatic and moral) have nothing to do with rationality and rational considerations.

### 3.4.2 The argument from options

While the *Diachronic Tragedy Argument* is, at least partially, focused on the theoretical aspects of rationality, *The argument from options* is focused on the practical aspects of rationality. The idea behind *The argument from options* is the following. Agents make decisions about their lives. Those decisions are best viewed as *options*. What agent's options are, supervenes on her mental states. Agent's mental states are available to her only at the current (present) time. In this regard, options are decisions that are available to an agent only at this (present) time. The subjective rational *ought* can only be applied to those decisions which are available to the agent. The only thing available to an agent, in regard to practical rationality, are her options. Options are, by definition, *time-slice* or synchronic. The conclusion is that the subjective rational *ought* cannot be placed on any diachronic notion of rationality.

There is a lot to address here. Let us start with the notion of the subjective *ought*. As we have discussed at the beginning of this chapter, the subjective *ought* is a form of rational obligation sensitive to the agent's perspective of the world. The example that Hedden uses is the following.

---

<sup>83</sup> For more on these examples see (Hedden 2012, 2015a).

*Your friend has a headache, and you have some pills that you justifiably believe to be pain relievers. But you're wrong. They are really poison. Ought you give the pills to your friend?* (Hedden 92: 2015a).

According to Hedden, there are two ways in which we can answer this question: *yes*, and *no*. If the answer is *yes*, then we are talking about the *objective ought*. In other words, we are discussing what the agent should do taking into consideration what the world actually is like. If the answer is *no*, then we are talking about the *subjective ought*. In other words, we are discussing what the agent should do taking into consideration her knowledge and perspective on the world. Hedden argues that the *subjective ought* has a central role in explaining human rationality. The subjective ought is central to fulfilling the three theoretical roles of rationality: a guidance role, an evaluative role and a predictive and explanatory role.<sup>84</sup>

The concept of agent's options is predicated on the notion of the subjective ought. The reason for this is because agent's options are based on facts available to her and what is available to the agent is predicated on the way in which she sees the world or her perspective on the world. Bearing this in mind, we now need to establish what exactly options are. The concept of options entails the following.

*Let me emphasize that I am using the term "option" as a technical term here.*

*An agent's options, on this usage, are the things which are evaluated by the correct decision theory, whatever that may be, such that the option that gets ranked highest by our decision theory is the one that the agent rationally ought to perform.*

*In this way, we can say that options are the things which in the first instance an agent ought rationally to do.* (Hedden 92: 2015a).

According to Hedden, the technical term "option" is a thing evaluated by the correct decision theory. This puts Hedden's *option* very close to the standard term of *preference* from decision theory. The question arises: what is the difference between Hedden's *option* and the classical term *preference* that we can find in decision theory? The first difference is that what an agent's options are, depends on the uncertainty of the world and ultimately options are only those decisions that the agent is currently able to make. The second difference is that options are a

---

<sup>84</sup> These roles are explained at length at the beginning of this chapter.

narrower set of decisions than classical preferences. The reason for this is because options are exclusively decisions which are available to the agent at this present moment while preferences are not (at least in theory). The similarities are, obviously, that both *options* and *preferences* are grounded in some underlying model of decision theory.

Hedden argues for a theory of options. A theory of options must satisfy two desiderata: first, an option is something that an agent is able to do and second, options supervene on the agent's mental states.

### *Desideratum 1*

*If a proposition P is a member of a set of options for an agent S, then S is able to bring about P.* (Hedden 98: 2015a).

*Desideratum 1* is in essence a simple reframing of the *ought implies can* principle. But why is this principle important to Hedden's theory of options? There are two main reasons. The first reason is that *Desideratum 1* effectively fulfills the roles of rationality. *Desideratum 1* fulfills the evaluative role of rationality by protecting the agent from rational criticism for something she is not able to bring about. The predictive role of rationality is fulfilled by the claim that we should not be able to predict something that the agent is not able to do. Lastly, action-guidance is fulfilled by the claim that the agent cannot be guided by the things she cannot bring about. The second reason is that *Desideratum 1* supports synchronic framing of rationality. If an agent is able to *bring about P*, then she is able to bring about P at this particular moment. Hedden argues that, in essence, *Desideratum 1* or the *ought implies can* principle supports the synchronic notion of rationality, more specifically, his Time-slice rationality account. The second desideratum in Hedden's theory of options is based on the notion of supervenience. In Hedden's words:

### *Desideratum 2*

*If something is a set of options for an agent S, then it is a set of options for any agent with the same mental states as S. What an agent's options are supervenes on her mental states.* (Hedden 99: 2015a).

*Desideratum 2* is a simple supervenience thesis. If something is to be considered an option for an agent, it needs to supervene on her mental states. According to Hedden, there are several reasons to accept *Desideratum 2*. The first reason, as we have discussed at the beginning of

this chapter, is internalism, specifically mental internalism for which Hedden argues. The second reason is the claim of *Desideratum 2* that in order to identify and rank her options, those options need to be available to the agent. The third reason is that agent's options should be sensitive to evidence and the best way to achieve this is to accept that agent's options supervene on her mental states.

In order to successfully argue for his theory of options, Hedden needs to find *the right kind* of options. By *the right kind* of options, he means the notion of options which satisfies *Desideratum 1* and *Desideratum 2*. There are three theories of options that Hedden considers in great detail and ultimately rejects.

The first theory states that options are *actual abilities of agents*. This theory is rejected because it does not satisfy *Desideratum 2*. In other words, agent's actual abilities do not supervene on her mental states, but they additionally depend on agent's physical state and her immediate environment. Hedden's conclusion is that *options as abilities* should be rejected.

The second theory states that options are *believed abilities of agents*. This theory is rejected because it does not satisfy *Desideratum 1*, i.e., the *ought implies can* principle. In other words, an agent can believe that she should do something but not actually be able to do it (physically or in some other sense). Hedden's conclusion is that *options as believed abilities* should be rejected.

The third theory of options states that options are *known abilities of agents*. This theory is rejected because it hinges on the notion of knowledge being a mental state. In other words, Hedden would like to remain neutral on the matter of knowledge either being or not being a mental state. In order to solve this problem, Hedden advances his *Desideratum 2* to *Desideratum 2+*. *Desideratum 2+* states the following.

*Desideratum 2+*

*What your options are supervenes on your present non-factive mental states.* (Hedden 103: 2015a).

Hedden upgrades *Desideratum 2* to *Desideratum 2+* for mainly one reason. The reason is that the norms of rationality should reference concepts like *belief* and *action* and not go any deeper in the reductionist sense. Additionally, *Desideratum 2+* fulfills the roles of rationality much better. By rejecting all three theories of options, Hedden argues for a "correct" theory of options. That theory claims that options should best be viewed as decisions. The following is Hedden's proposal.

### *Options-as-Decisions*

*A set of propositions is a set of options for agent S at time t if it is a maximal set of mutually exclusive propositions of the form S decides at t to  $\phi$ , each of which S is able to bring about.* (Hedden 106: 2015b).

There is a lot of content here so we will break it down into several parts. Firstly, we should note that Hedden is claiming options to be *decisions*. This is, in some sense, analogous to *intention formation* present in Bratman's, Broome's and Holton's accounts of diachronic rationality. Intentions and options are both, in that sense, decisions. Contrary to *intentions*, which have the properties of stability and commitment (Bratman 1987), *options* are decisions that are *open to an agent*. By *open to an agent*, Hedden means that *options* are available to an agent at this current point in time and that she can act on them. In addition, *options* are "mental tries" that an agent performs at a particular point in time. The concept of a "mental try" ensures *Desideratum 1* or the *ought implies can* principle. In other words, an agent is under no rational obligation to actually do the thing she decides to do. She is only under the rational obligation to *try* to do the thing she decides to do.

Secondly, the notion of a maximal set of mutually exclusive propositions simply means that when an agent decides to  $\phi$ , she decides *only to  $\phi$*  and nothing else (relevant to her practical decision making, of course). In that regard, *S decides at t only to  $\phi$*  is, in some sense, compatible with Bratman's *Intention consistency constraint*.

Thirdly, the notion of *Options-as-Decisions* is extremely synchronic. There are no intervals such as "from  $t_1$  to  $t_2$ " mentioned here. On the contrary, Hedden specifically claims that agent's *options* are those propositions "located" at a specific time  $t$ . We can acknowledge this by his phrasing *at the time t* and *at t*. For Hedden, there is no other time but now (at least in regard to the notion of rationality).

Fourthly, the notion of *Options-as-Decisions* is based on the presupposition of internalism, more specifically, *mental internalism*. We can observe this by Hedden's narrowing of a set of propositions that an agent is *able to bring about*.

Lastly, there are several conclusions that Hedden draws from the notion of *Options-as-Decisions*. Firstly, *Options-as-Decisions* satisfies both desiderata. Secondly, *Options-as-Decisions* cannot include temporally extended actions. The reason for this is that the problematic notions of personal identity should not have any weight when discussing the

rationality of beliefs and actions. Thirdly, *Options-as-Decisions* is exclusively a time-slice theory of options which focuses primarily on things available to an agent at her current time.

### 3.4.3 The two arguments recap

Hedden presents two arguments in favor of his Time-slice rationality account: the *Diachronic Tragedy Argument* and *The argument from options*.

The *Diachronic Tragedy Argument* states that there are situations where agents are diachronically inconsistent, i.e., they choose contrary to their best interests (even by their own lights). In his main example, *The Russian Nobleman*, the nobleman chooses in the way clearly bad for him (by “bad” Hedden means disadvantageous in the sense of decision theory). But the nobleman, according to Hedden, is not irrational but simply tragic. The nobleman is a “victim” of a tragic sequence. He is clearly not maximally efficient but he is also not irrational. The point that Hedden makes with his *Diachronic Tragedy Argument* is the following. Agents are sometimes diachronically inconsistent. There are multiple reasons for this: weakness of will, intransitive preferences, change of heart, etc. One cannot be regarded as irrational only in virtue of being diachronically inconsistent. One of the reasons for this is the problem of personal identity. If we were to conclude that someone is irrational simply because she is diachronically inconsistent, then we would need to ask the following question. With what is the agent inconsistent? The answer is with her past preferences and that necessarily entails commitment to a certain metaphysical stance on personal identity. Hedden claims that rational requirements should be independent of questions about personal identity so he concludes that when agents are diachronically inconsistent they are not irrational but simply tragic.

*The argument from options* states that the subjective rational *ought* (the internalist conception of *ought*) cannot be attributed to temporally extended actions. The reason for this is because it invokes the problem of personal identity. The subjective rational *ought* should be applied to options. Options are preferences available to an agent at her present (current) time. By that line of reasoning, the only rational norms or requirements that can be placed on agent’s beliefs and actions should be synchronic.

### 3.5 Chapter summary

In this chapter I have presented the most comprehensive synchronic account of rationality – *Time-slice Rationality*. The Time-slice Rationality account is an account of rationality that denies the existence of any genuine diachronic norms of rationality. The presupposition of Time-slice Rationality is a form of evidentialism (Conee and Feldman 2004). The Time-slice Rationality account consists of two norms or requirements of rationality: *Synchronicity* and *Impartiality*. The synchronicity norm states that all requirements of rationality are synchronic. The impartiality norm states two things. First, agent's past attitudes are irrelevant in determining what the agent *ought* to do right now. Second, personal identity is irrelevant in determining what the agent *ought* to do.

According to Hedden, the founder of the Time-slice Rationality account, there are two main reasons to accept his account: the problem of personal identity and internalism. Discussion about rational norms and requirements (theoretical or practical) should not hinge on the notion of personal identity. One of the main problems with diachronic accounts of rationality is the fact that they rely on a certain account in the ontology of personal identity to be true, specifically, that some form of personhood based on *persistence over time* be true (for example, psychological-continuity views: (Parfit 1971, 1984, 2007, 2012; Perry 1972; Shoemaker 1970, 1984)). According to Hedden, this is wrong. Rationality (or rational norms) should not be held hostage by a certain account in the ontology of personhood which is exactly what the proponents of diachronic rationality are doing (Broome 2013; Bratman 2010, 2014, 2018; Holton 2009; Ferrero 2006, 2009, 2010, 2012). The second reason to accept Time-slice Rationality is internalism (although one can be an externalist and also accept Time-slice Rationality). We cannot expect agents to follow certain norms if they actually cannot follow those norms. In other words, Hedden claims that we should abide by the *ought implies can* principle (or *Desideratum 1* of his account). Norms of rationality that are presented to agents need to be available to them (via supervenience). This, according to Hedden, naturally leads to the rejection of diachronic norms of rationality. The reason for this is because what is available to agents supervenes on their mental states and it is available to them strictly at ***this current point in time***.

Hedden presents two main arguments in favor of the Time-slice Rationality account: the *Diachronic Tragedy Argument* and *The argument from options*. The *Diachronic Tragedy Argument* states that when the agent is diachronically inconsistent, i.e., she changes her

attitudes over time (in cases where it is clearly disadvantageous to her even by her own lights), she is not being irrational but simply tragic. *The argument from options* states that the *subjective rational ought* (the one sensitive to the way the agent perceives the world) can only be imposed on those preferences available to the agent. Preferences available to the agent are those preferences available to her at this particular point in time (otherwise there would arise a problem with personal identity and internalism). Preferences available to the agent at this time are called *options*. Subjective rational *ought* can only be imposed on options and options are, by definition, synchronic (available to the agent only at the current time). Hedden's conclusion is that no diachronic norm of rationality can be imposed on agents.

## 4 The last stand<sup>85</sup>: In defense of the diachronic notion of rationality

In the last chapter, I will be defending my thesis: *Diachronic agency can be rationally assessed in the way in which synchronic agency is assessed*. I will present a defense of my thesis in two stages. In the first stage, I will argue against the *Time-slice Rationality* account (this account has been presented in the previous chapter). One of the core claims of the Time-slice Rationality account is that there are no genuine diachronic norms of rationality. In order to defend my thesis, I will need to show that the Time-slice Rationality account does not work (or is at least substantially inadequate in explaining human agency and practical rationality). I will achieve this in two ways. Firstly, I will argue that motivations for Time-slice Rationality are misguided at best (and simply wrong at worst) and that arguments in favor of Time-slice Rationality are inadequate. Secondly, I will present arguments against the Time-slice Rationality account which are inspired by other authors who have criticized Hedden's theory in a similar way (Döring and Eker 2017, Snedegar 2017, Lenman 2017).

In the second stage I will argue in favor of the notion of diachronic rationality. I rely on Bratman's account of diachronic rationality (which I presented in detail in Chapter 2 and the accounts of Kantians like Luca Ferrero and Richard Holton.

Lastly, I will address some interesting conundrums that arise when we accept the notion of diachronic agency and diachronic rationality. I will offer my thoughts on these conundrums and the way in which I believe the future research of diachronic rationality is heading.

### 4.1 General overview and framing of the problem

The thesis that I am defending is the following: *Diachronic agency can be rationally assessed in the way in which synchronic agency is assessed*. There are two presuppositions incorporated in my thesis. The first is that humans have the ability to engage in temporally extended activity and the second is that we can make rational assessments of human behaviour. Human beings have the ability to make plans, commitments and promises and *stick to them*

---

<sup>85</sup> The title "*The last stand*" refers to Bratman's notion of *the place to stand* or *standpoint function* discussed in Chapter 2.

when they are affected by temptation, weakness of will, reluctant change of heart, or other temporary shifts of preferences. Fortunately, there are several accounts that endorse the presupposition that humans have a capacity for diachronic agency (Bratman 2012, 2014, 2018; Broome 2013; Ferrero 2009, 2010, 2012; Holton 2013). Although some accounts of diachronic agency are more controversial than others (such as Ferrero's diachronic account which invokes the concepts of *diachronic will* and *personal narrative*), the idea that we have some ability to engage in activities that are stretched across time is an uncontroversial one. The second presupposition is that we can make rational assessments of human behavior which is also relatively uncontroversial (by rejecting this we would be rejecting the concept of practical rationality altogether). What is not controversial is my thesis which states that we can make rational assessment of diachronic agency, more precisely, that there is such a thing as diachronic rationality. The account of rationality that I will be relying on is Bratman *Diachronic Plan Rationality* account.

Bratman Diachronic Plan Rationality account is based upon Bratman's planning theory of intention. The idea is that we have the ability for diachronic agency because we are fundamentally *limited agents*. We are limited by our cognitive and temporal resources. A real-life agent does not have cognitive capacities to reevaluate a decision every moment from the formation of her decision to the execution of action. She is also limited in her temporal resources, which means that she cannot deliberate and make decisions instantaneously, in other words, it takes time for her to deliberate. Our ability to form intentions, make plans and honor commitments enables us to connect synchronic dots into a diachronic line which we need in order to live stable and coherent lives. We need some way to coordinate with our past and future selves in order to achieve complex and temporally distant goals and we need a way to coordinate with each other's to achieve the goals of the same kind. The best way for us to achieve this type of coordination is by engaging in temporally extended agency.<sup>86</sup>

Bratman's planning theory of intention is arguably an intuitive one and serves as foundation for all contemporary accounts of diachronic rationality (Broome 2013; Ferrero 2010, 2012; Holton 2009; Snedegar 2017). Arguing for the existence of diachronic norms of rationality is somewhat difficult. The question arises: what kind of norms are diachronic norms of

---

<sup>86</sup> Authors who discuss the notion of diachronic agency have different ideas on how we actually engage in diachronic agency. Instrumentalists claim that the notion of intention is enough to carry the concept of diachronic agency (Broome 2013; Snedegar 2017; and in some sense Bratman 1987, 1999), while Kantians claim that we need concepts like *diachronic will* (Velleman 2000; Ferrero 2009, 2010, 2012), *resolutions* (Holton 2009), *narrative* and *autonomy* (Ferrero 2009, 2010) and *self-governance* (Bratman 2014, 2018) in order to capture the notion of diachronic agency.

rationality? I believe there is an intuitive notion that when an agent makes a plan, or a commitment, or a promise, there is some sort of rational pressure on that agent (if nothing relevant for her decision making intervenes) to keep the promise, honor her commitment and finish her plan. But there is a long and dangerous road between an intuitive notion and actually demonstrating that there are diachronic norms of rationality. As I have demonstrated in the previous chapters (Chapter 1 and Chapter 2), there are various accounts that tackle the problem of diachronic norms of rationality. In this chapter, I will evaluate them and side with Bratman's Diachronic Plan Rationality account.

## 4.2 The tragedy of Time-slice Rationality<sup>87</sup>

The Time-slice Rationality account directly contradicts my thesis. I argue that there are, at least some, diachronic norms of rationality and the Time-slice Rationality account directly contradicts this claim by stating that there are no diachronic norms of rationality. There is a clear conflict here. But before we venture into this conflict, I would like to present the three main points of the Time-slice Rationality account:

1. The notion of rationality should be sensitive to the agent's perspective of the world,
2. Problems of personal identity should be divorced from discussions about rationality,
3. The notion of rationality should be the notion of ideal rationality.

These three points, according to Hedden, lead us to accept a purely synchronic account of rationality which is (as we will see) better than a diachronic account of rationality. By the agent's perspective of the world, Hedden refers to *internalism*. Internalism is, according to Hedden, a powerful reason to accept a synchronic account of rationality. Hedden reasons in the following manner: if we accept internalism, we accept that only norms which are *available* to us are legitimate norms of rationality (because rationality should be sensitive to the agent's perspective of the world). The only things that are *available* to an agent are things that are available to the agent at her current time. Therefore, there are only synchronic norms of rationality. Hedden's second point is that all diachronic accounts of rationality have an implicit

---

<sup>87</sup> There are several ways to argue against the Time-slice Rationality account. My approach will be to focus on the part of Time-slice Rationality which addresses practical or instrumental rationality. I am not alone in this endeavor and other authors (Snedegar 2017, Döring and Eker 2017, Lenman 2017) have made similar responses to Time-slice Rationality. My critique is analogous to theirs. There are other approaches that focus their criticism on theoretical aspects of the Time-slice Rationality account. I do not follow in their footsteps. For more on this kind of criticism of the Time-slice Rationality account see (Podgorski 2016a, 2016b, 2017, Titelbaum 2015; Paul 2015).

metaphysical presupposition about the nature of personal identity. That proposition is that personhood persists over time. Hedden argues that the account which is not “held hostage” by one particular position in metaphysics (such being his account), reigns supreme over the ones that are. The third point is that a notion of rationality should be a notion of ideal rationality. This means that concepts like reasoning and reflection have no weight when we discuss norms of rationality. Hedden argues that a theory of rationality should also include beings that are not constrained by time, space and cognitive limitations. A theory of rationality able to include such beings would have, according to Hedden, greater explanatory power than its competitive theories. All these reasons lead us to accept a purely synchronic account of rationality.

Hedden also makes a point regarding the relationship between intentions and the notion of rationality. Intentions do persist over time and that stability of intentions is important and useful, but those facts are completely irrelevant when discussing the notion of rationality. Hedden’s rationale is the following. Intentions do indeed persist over time in the manner described by Broome and Bratman (Broome 2013; Bratman 1987, 1999), but this is a mere causal fact and a *useful part of agent’s mental toolkit*. According to Hedden, the stability of intention is important but not relevant for rationality. Agents are sometimes suboptimal in their decision making (for example, in the case of forgetfulness), but that does not mean they are irrational. Intentions are extremely useful and important *mental sticky notes* (Hedden 125: 2015a). However, intentions cannot be subjected to our rational considerations. Hedden argues that the reason why intentions are not eligible for rational consideration is because intentions are diachronic in nature and not fully accessible to an agent at her present time. We engage in intentions because we are limited in recourses (Hedden 2015a), but that fact has nothing to do with rationality because rationality should apply to ideal agents who do not need intentions in the first place.

#### **4.2.1 The norms of Time-slice Rationality**

The Time-slice Rationality account consists of two main claims. The first claim is *Synchronicity* and the second claim is *Impartiality*. The two claims state the following.

*Synchronicity: All requirements of rationality are synchronic* (Hedden 8: 2015b).

*Impartiality: In determining how you rationally ought to be at a time, your beliefs about what attitudes you have at other times play the same role as your beliefs about what attitudes other people have* (Hedden 9: 2015b).

According to Hedden, there are two reasons to accept both of these claims or the Time-slice Rationality account.

The first reason to accept the Time-slice Rationality account is *internalism* (one particular kind of internalism) and the second reason is the problem of personal identity. The kind of internalism that Hedden refers to is *mentalist internalism*. In Hedden's words it is the following:

#### *Mentalist Internalism*

*What you rationally ought to believe, desire, or do supervenes on your mental states.* (Hedden 23: 2015a).

Hedden is very clear about the connection between this kind of internalism and the Time-slice Rationality account. One does not need to accept internalism in order to accept the Time-slice Rationality account. Externalism is compatible with Time-slice Rationality as long as externalists do not consider facts about the past relevant for current rational considerations. That being said, internalists have *additional* reason to accept Time-slice Rationality because what is rationally relevant for an agent to *believe, desire* or *do* supervenes on her mental states at the current moment. Mentalist internalism, according to Hedden, directly supports *Synchronicity* as norm of rationality.

The second reason to accept Time-slice Rationality is the problem of personal identity. Hedden argues that the notion of rationality should be divorced from any metaphysical stance about the nature of personhood. Time-slice Rationality is neutral in regard to the nature of personhood and any diachronic account of rationality implicitly comprises some metaphysical position of personhood in its account (Hedden 2015a). The problem of personal identity directly leads us to accept *Impartiality* as norm of rationality.

Hedden offers two arguments in favor of the Time-slice Rationality account (when discussing practical rationality): *Diachronic Tragedy Argument* and *The argument from options*. The *Diachronic Tragedy Argument* claims that in some cases of diachronic inconsistencies agents are not irrational but simply tragic. The reason why they are tragic is because of their involvement in a tragic sequence (a sequence in which an agent at all times prefers to be in

some other sequence). The example that Hedden presents (beside the Russian nobleman example) is of a weak-willed smoker.

The smoker decides not to smoke. She knows that she will certainly have a change of preferences and prefer to smoke in an hour. The smoker pays someone \$50 to prevent her from smoking. According to Hedden, the smoker is engaged in a tragic sequence. The reason why is because she would rather be in a sequence where she decides not to smoke and does not pay someone \$50 to prevent her from smoking. There is nothing stopping her to decide not to smoke and not smoke without losing \$50 (but herself). Hedden argues that the smoker is tragic but not irrational. Agents are sometimes suboptimal in their decision making but that does not make them irrational.

The second argument is *The argument from options*. *The argument from options* claims that the subjective rational *ought* (*ought* sensitive to the agent's perspective of the world) can be applied to those propositions which are available to an agent. This rules out intentions and sets of preferences that are stretched over time. The subjective rational *ought* can only be applied to options. Options are preferences that supervene on the agent's current mental states. There are two reasons for this: the *ought implies can* principle and availability. The subjective *ought* should apply to things that agents can actually achieve and it should apply to things that agents have access to. Both of these reasons lead us back to the notion of *options*.

### **4.3 The critique of the Time-slice Rationality account**

The problem with the Time-slice Rationality account is that it is simultaneously too strong and too weak. By *too strong* I mean too demanding and by *too weak* I mean not demanding enough. The account is too demanding in the field of theoretical rationality and I will leave that part aside because it does not directly relate to my research which is focused on practical rationality and the philosophy of action. My contention with the Time-slice Rationality account is that it is not demanding enough in the field of practical rationality and the philosophy of action. The demand that I am addressing here is some kind of a (at least weak) diachronic norm of rationality. There are several problems with the Time-slice Rationality account which I will address shortly, but the crucial flaw of the Time-slice Rationality account is that it fails to identify certain behaviors as irrational, which any reasonable account of rationality should be able to identify as irrational. Those behaviors include reasonless and arbitrary change of mind and severe cases of forgetfulness. Additionally, the case can be made by stating that Hedden

does not engage directly with any practical account of rationality (Bratman's account, Broome's account, or Holton's account), but he does mention them fairly frequently. He does not offer arguments why these accounts fail as accounts of rationality.<sup>88</sup>

Now, I will present several problems with Hedden's Time-slice Rationality.

1. Internalism is compatible with diachronic and synchronic accounts of rationality,
2. The problems with personal identity are equally problematic for diachronic and synchronic accounts of rationality,
3. The notion of rationality should (at least in some sense) reference real-life agents,
4. The underlying assumption of the *Diachronic Tragedy Argument* and *The argument from options* is a formal decision theory (or *the correct form of the expected utility theory*),
5. Time-slice Rationality fails to account for clear and uncontroversial cases of irrationality (reasonless or arbitrary change of mind and severe cases of forgetfulness).

#### **4.3.1 Internalism is compatible with diachronic and synchronic accounts of rationality**

One of the two main motivations for the Time-slice Rationality account is internalism, specifically, mentalist internalism. There is no prima facie reason or argument presented by Hedden as to why exactly internalism presents a good reason to accept the Time-slice Rationality account. Also, there is no relevant logical connection between internalism and the Time-slice Rationality account. The existence of diachronic norms and diachronic accounts of rationality are perfectly compatible with internalism. This is probably the easiest and most straightforward flaw of the Time-slice Rationality account.<sup>89</sup> To be charitable to Hedden's account, he does argue that, when discussing rationality, we should be sensitive to the agent's perspective of the world. The reason why is because, according to Hedden, this is one of the three roles that rationality should play – the action-guiding role (the other two being the evaluative role and the predictive and explanatory role). In Hedden's words:

---

<sup>88</sup> The reason for this is because Hedden concerns himself more with the concepts of conditionalization and reflection (which belong to theoretical rationality) than with the concepts of self-governance (Bratman 2014, 2018), persistence of intention (Broome 2013) and diachronic will (Ferrero 2009, 2010). Still, these facts do not absolve him of criticism.

<sup>89</sup> This problem was also presented by Sabine A. Döring and Bahadır Eker. For more see (Döring and Eker 2017).

*Rationality is a matter of believing and behaving in ways that are sensible, given your perspective on the world* (Hedden 23: 2015a).

While I am sympathetic towards this claim, the *perspective* in this claim refers to *the present perspective* which in turn connects internalism and the Time-slice Rationality account. There are two great problems here. Firstly, how does the agent's perspective connect internalism and Time-slice Rationality? Hedden never answers this question directly. Secondly, why is the agent's *perspective on the world* viewed by Hedden as her *present perspective on the world*? Is it not the case that the agent's *perspective on the world* involves some amount of diachrony in its concept? The agent's *perspective on the world* usually includes immediate past and immediate future. It is certainly not *prima facie* clear that the agent's perspective is a present perspective and Hedden offers no argument why it should be.

It is a fact that the agent's *perspective on the world* changes as the world around her changes and the diachronic norms target exactly those changes and how the agent should adapt to those changes (by, for instance, reacting to new information or evidence). I will reiterate that there is no good reason or presented arguments that connect internalism to a purely synchronic account of rationality like Time-slice Rationality.

#### **4.3.2 The problems with personal identity are equally problematic for diachronic and synchronic accounts of rationality**

Hedden's second motivation for the Time-slice Rationality account is that rationality should be independent from the problems of personal identity. The reason why is because personal identity is sometimes "messy" and "murky"<sup>90</sup> and it is sometimes difficult to determine to whom the norms of rationality apply. The cases that Hedden presents are *Teletransportation* and *Double Teletransportation* (discussed in Chapter 3) and the case of octopuses in which we are not sure how many intelligent agents are in play (Godfrey-Smith 2016). Hedden states that diachronic accounts of rationality endorse a certain ontological position (the position of persistence of personal identity over time), while his account is neutral regarding the ontological debate about the nature of personhood. This fact, according to Hedden, makes his Time-slice Rationality a superior account of rationality.

---

<sup>90</sup> Hedden's expressions.

The first and most important thing to say as a response to Hedden's claim is the following: no contemporary account of diachronic rationality explicitly endorses any ontological claim about the nature of personal identity. While remaining charitable to Hedden, he probably claims that there is an underlying assumption about the nature of personal identity present in the accounts of diachronic rationality, although he never explicitly argues this to be the case. Whether or not there is such an assumption in the diachronic accounts of rationality is a difficult question and my response is as follows. I believe such an assumption exists in, what I termed in Chapter 1, *Kantian accounts of diachronic rationality*. Those being Holton's account with the concepts of *resolutions* and the *non-reconsideration model* (Holton 2009), Ferrero's account with the terms *diachronic will* and *narrative* (Ferrero 2009, 2010, 2012), and Bratman's later account with the notion of *self-governance* (Bratman 2014, 2018). I also believe there is no such assumption in, what I called in the first chapter, *instrumentalist accounts of diachronic rationality*. Those being Broome's account of *persistence of intention* (Broome 2013), Snedegar's account of *filled-in plans* (Snedegar 2017), and Bratman's early account including *stability of intention*, if we can view it as an account of diachronic rationality (Bratman 1987, 1999).<sup>91</sup>

The second point is that it is not clear what the connection between the "murky" and "messy" problems of personal identity and the synchronicity norm is. How do the problems of personal identity lead us to accept *Synchronicity* as norm of rationality? To make matters worse, by analogous reasoning, we can arrive at a contrary conclusion. As Döring and Eker have presented in their critique of Time-slice Rationality, we can look at another case of the problem of personal identity.

*Imagine ... a human being with two personalities, one of which is 'out' or active on even days, and the other of which is active on odd days. Each day at midnight, like clockwork, one personality goes dormant and the other takes over. Call the being who thinks and acts on even days Even and the one who thinks and acts on odd days Odd. ... Suppose that Even cannot recall, on even days, anything that Odd thinks or does on odd days, and that Odd likewise cannot recall on odd days anything that Even does or thinks on even days. ... Even (on even days) is cheerful and gregarious, while Odd (on odd days) is sullen and withdrawn. Even enjoys smoking on even days; Odd detests it on odd days. We may want to add that the even-day and*

---

<sup>91</sup> Bratman does not consider any part of his planning theory of intention as diachronic rationality in any way. But some authors disagree and use some parts of his theory (Holton 2009) or all of his planning theory of intention (Snedegar 2017) to build diachronic accounts of rationality.

*the odd-day thoughts could no more become integrated, through psychotherapy or the like, than your thoughts could become integrated with mine. (Olson 330: 2003)<sup>92</sup>.*

Following Hedden's logic, we could argue in the following way. In this example there are either two separate agents or one agent who is sufficiently disjointed for himself. This is a "murky" and "messy" problem of personal identity. We should look at rationality independently from these "murky" and "messy" problems of personal identity. Whether there are two agents or one agent in play here, there is a need for some norm of consistency and coherence. Because of the way the example is structured, this consistency or coherence needs to stretch over time, more specifically, it needs to be diachronic. Can we now conclude that all norms of rationality are diachronic? In the words of Döring and Eker – *but this is absurd* (Döring and Eker 576: 2017). I fully and wholeheartedly agree with them – this is absurd. But the problem for Hedden is that this is exactly analogous to the way he argues for the Time-slice Rationality account. We cannot arrive at the synchronicity norm from "murky" and "messy" problems of personal identity any more that we can arrive at the norm that all norms of rationality are diachronic from equally "murky" and "messy" problems of personal identity. There is no higher rational ground unspoiled by the murkiness and messiness of personal identity on which one can stand. We are all, both the proponents of the synchronic accounts of rationality and the proponents of the diachronic accounts of rationality, in the same metaphysical mud.

#### **4.3.3 The notion of rationality should (at least in some sense) reference real-life agents**

Hedden's claim is that the notion of rationality should be divorced from (at least) two things: real-life agents and reasoning. Snedegar, in his response to Hedden, puts it nicely and concisely.

*(1) A theory of rationality should be a theory of ideal rationality – it should not take into account our own contingent limitations.*

---

<sup>92</sup> Taken from (Döring and Eker 2017).

*(2) Norms of reasoning would really only apply to limited agents like us. Ideal agents could form beliefs directly in response to the evidence, updating automatically, without going through reasoning processes.*

*(3) So a theory of rationality need not provide norms of reasoning (Snedegar 596: 2017).*

There are two things that I would like to highlight here: the claim that a theory of rationality should be a theory of ideal rationality and the conclusion that a theory of rationality does not need norms of reasoning. Firstly, I would like to say that, while I do disagree with (1), there is nothing inherently wrong with holding that position. That being said, I would like to point out that there are some unfortunate consequences for Hedden's view. We could imagine agents who *respond directly to evidence*, who do not reason, and satisfy all the requirements of Time-slice Rationality, but we would have difficult time calling them rational. Lenman presents two useful examples in his critique of the Time-slice Rationality account.

#### *The case of Professor Instinct*

*Professor Instinct, let's suppose, doesn't go in for any reflective reasoning of any kind ever. He just gets pushed around by his instincts like a brute beast. But his instincts are good instincts. Whoever designed him designed him extremely well. His instincts guide him so reliably that he reliably decides what the uniquely correct utility function says he should decide and believes what the uniquely rational prior probability function tells him to believe on the basis of his evidence. (Lenman 590: 2017).*

#### *The case of Professor Lucky*

*Professor Lucky is a what we might call a Randomizer, someone who arrives at her beliefs and decisions by some entirely random procedure. Of the many possible Randomizers, most do very badly, getting almost everything wrong almost all the time. But there are a small minority who get lucky and do pretty well. A fantastically lucky, very tiny minority, one in a few squillion perhaps, do just perfectly and get from one end of life to the other believing and deciding exactly as the uniquely correct utility function and the uniquely rational prior probability function would tell them to. Professor Lucky, the lucky so-and-so, happens to belong to that very tiny minority. (Lenman 590: 2017).*

Both Professor Instinct and Professor Lucky satisfy the conditions of rationality presented in Hedden's Time-slice Rationality account. But this is strange at best and unacceptable at worst. In the case of Professor Instinct, the professor is doing all the right things (rationally speaking) without reflection or reasoning (or reason for that matter), she simply acts *on an instinct*. There is a long tradition in philosophy which goes back at least to Aristotle that humans are rational beings precisely because we have the ability to reason. This ability (among others) is what separates us from animals which simply rely on their instincts. According to that tradition, the entire notion of rationality is based on the opposition to instinct. If we do not act on reason in this sense, we are not rational. In the case of Professor Lucky, the professor is doing all the right things (rationally speaking) simply by chance. Without reflection or reason (or any thought whatsoever), Professor Lucky stumbles upon believing and doing all the right things. The claim that Professor Lucky (or Professor Instinct for that matter) is rational, makes it at best a hard bullet to bite and at worst an unacceptable bullet to bite. *The right thing to do rationally speaking*, in Hedden's Time-slice Rationality account, means *conforming to the right expected utility theory*. This fact is a problem in itself, as I will demonstrate in the following passage.

**4.3.4 The underlying assumption behind the *Diachronic Tragedy*  
*Argument and The argument from options is a formal decision theory (or  
the correct form of the expected utility theory)***

There are two main arguments that Hedden uses to argue in favor of the Time-slice Rationality account: *Diachronic Tragedy Argument* and *The argument from options*. The *Diachronic Tragedy Argument* states that there are certain situations when the agent is apparently irrational, but she is actually simply tragic and not irrational at all. In the case of a diachronic tragedy, the agent has a *tragic attitude*. Agents who have tragic attitudes are involved in *tragic sequences*. An agent is in a tragic sequence if and only if the agent at every point in time prefers to be in some other sequence available to her. There are two examples of diachronic tragedy that Hedden presents: *The Russian nobleman* and the *weak-will smoker*.

*The Russian Nobleman*

*You will receive an inheritance of 100,000 rubles at age sixty. Right now, you have the option (call it Donate Early) of signing a binding contract which will require 50,000 rubles*

*to be donated to left-wing political causes. No matter whether you take this option, you will at age sixty have the option (call it Donate Late) of donating 50,000 rubles to rightwing political causes. (No greater donation is permitted under Tsarist campaign finance laws.) Right now, you most prefer donating 50,000 rubles to left-wing causes and nothing to right-wing causes. But you also prefer donating nothing to either side over donating 50,000 rubles to each side, as the effects of those donations would cancel each other out. (Hedden 79-80: 2015b).*

The nobleman is indeed involved in a tragic sequence. He prefers at all times to be in a sequence in which he does not *throw away* all of his money by his *change of heart*. The tragedy is substantial because he is aware at every step of the way that he is in a tragic sequence. The inevitable change of heart is predictable to him and he is unable to stop it although he has every ability to stop it. Another example of a diachronic tragedy is the weak-will smoker.

*The way not to smoke*

*Suppose that at present you want to quit smoking, but you believe that within a couple hours you will want a cigarette. You have the option right now of paying someone \$50 to prevent you from buying any cigarettes. This in effect closes off your future options, so that the only thing you can do later on is not smoke. But whether you wind up wanting a cigarette (as you now believe you will) or not, you wind up with a suboptimal outcome, since no matter what, you always prefer not smoking to not smoking plus being \$50 poorer. In this case, you wind up worse off because you are willing to pay to narrow down the options you will have in the future. (Hedden 81: 2015b).*

The weak-will smoker is as tragic as the nobleman. He binds himself not to smoke by *paying someone \$50 to prevent [him] from buying any cigarettes*. We assume that he would prefer to be in a sequence in which he does not smoke and has \$50 but is in effect caught up in a tragic sequence.

The claim that Hedden makes is that the nobleman and the smoker are tragic but not irrational. My response to this claim is the following. I agree with Hedden that there is no irrationality to be found in these cases. The problem that arises is the following. Hedden assumes that the agents, the nobleman and the smoker, are irrational by default because they violate the axioms of the expected utility theory. They violate the axioms of the expected utility theory in a

specifically diachronic way so the proponents of the diachronic accounts of rationality should consider this violation unacceptable. This is an understandable misconception and the place where the structure of different approaches to diachronic rationality which I have presented in the first chapter comes into play. The *Diachronic Tragedy Argument* and *The argument from options*, as we will see shortly, would be effective against the accounts of diachronic rationality from the first phase of diachronic rationality which I termed, *Diachronic rationality as a problem of dynamic (sequential) choice*. In this phase, authors proposed diachronic norms which aimed to establish consistency between sets of preferences at different times in order to avoid violating the independence axiom. There were two major camps: the one advocating the sophisticated choice strategy and the other favoring the resolute choice strategy. The proponents of the sophisticated choice strategy (Stortz 1956, Hammond 1976, Elster 1979, Schick 1986) argued that agents should anticipate their future change of preference and try to avoid it. The most famous case being the story of Ulysses and the Sirens (Elster 1979). Conversely, the proponents of the resolute choice strategy (Machina 1989; McClennen 1990, 1989) claimed that agents should, more or less, be consistent over time with their preferences. Authors from the first phase operated in the conceptual framework of the expected utility theory and their goal was to bring dynamic or sequential consistency into the agent's decision making. Hedden's *Diachronic Tragedy Argument*, and *The argument from options* for that matter, would probably be convincing in the conceptual framework of the first phase because of the common underlying assumption – the expected utility theory. Both Hedden and the proponents of the sophisticated choice strategy and the resolute choice strategy agree that the expected utility theory is the correct way to address the problems of practical or instrumental rationality. But there is arguably an easy way around this: we can simply reject the claim that the expected utility theory is the correct way to address the problems of practical or instrumental irrationality. And the stance that I am taking here is not groundless or controversial (at least as much as any philosophical position is uncontroversial). The expected utility theory has been troubled from the very beginnings by paradoxes (Arrow 1950, Maurice 1979a), by the inability to explain how real people actually make decisions (Tversky 1975, Kahneman & Tversky 1979), and by the incapacity to explain the normative and descriptive sides of rational decision making (Bermúdez 2009).

Now, let us return to Hedden's claim that the nobleman and the smoker are not irrational. Their actions are suboptimal in the light of the expected utility theory. I wholeheartedly agree with him: the nobleman and the smoker are not in fact irrational. There is one more important point that needs to be made here. By rejecting the expected utility theory there is no explanatory

vacuum left in its place. It is not the case that when we throw the expected utility theory out the window there is no way in which we can determine what counts as rationally permissible. Bratman's and Broome's accounts handle instrumental or practical rationality quite well without involving the axioms of the expected utility theory (Bratman 1987, 1999, 2014, 2018; Broome 2013).

The same problem arises in *The argument from options*. The term *option* means the following in Hedden's words.

*Let me emphasize that I am using the term "option" as a technical term here.*

*An agent's options, on this usage, are the things which are evaluated by the correct decision theory, whatever that may be, such that the option that gets ranked highest by our decision theory is the one that the agent rationally ought to perform.*

*In this way, we can say that options are the things which in the first instance an agent ought rationally to do. (Hedden 92: 2015a).*

There are two important things that we should know about *options*. Firstly, *options* are *things which are evaluated by the correct decision theory*. This can be rejected on the same grounds that we rejected the *Diachronic Tragedy Argument*. Secondly, *options* supposedly have these relevant and useful properties of being available to the agent at the current time and of supervenience over the agent's mental states. But there is no reason to think that intentions cannot serve the same purpose. Intentions are surely available to the agent at her current time and, if we accept mentalist internalism, surely supervene on her mental states. Hedden's decisions as options do not have any more or less relevant properties than intentions.

There is an irony to be found here. Hedden's account of rationality is heavily inspired by Bermúdez's *Decision theory and Rationality*. However, the main claim that Bermúdez argues for in that book is that decision theory cannot serve as a proper theory of rationality.

### 4.3.5 Time-slice Rationality fails to account for clear and uncontroversial cases of irrationality (reasonless or arbitrary change of mind and severe cases of forgetfulness)

#### 4.3.5.1 A simple change of heart or a case of severe capriciousness

This is a straightforward flaw of the Time-slice Rationality account. As I mentioned at the beginning of this chapter, the Time-slice Rationality account is at the same time too strong and too weak. I have pointed out that the account is too strong because the reasons for it have more to do with the notion of theoretical rationality.<sup>93</sup> The idea is that there are cases which are undeniably irrational and the Time-slice Rationality account does not consider them irrational. These cases include reasonless or arbitrary change of mind and severe cases of forgetfulness. *An afternoon with Candice* was an example of reasonless or arbitrary change of mind which we have discussed earlier, in Chapter 2. Let us revisit that example.

*An afternoon with Candice*

*Candice decides to go to the post office this afternoon to send out some mailings, but on the way there, she gives up on this end and decides to go buy groceries instead. But on the way to the market, she yet again trades in this end for another: going to hang out with her friend David. But on the way to David's house, she once more changes her mind and intends to spend a relaxing afternoon at home, but by the time she gets home the afternoon is gone and she's accomplished nothing.* (Bratman 83: 2012).

There are two points that I would like to make with this example.<sup>94</sup> Firstly, this is a clear counterexample to Hedden's Time-slice Rationality account. Candice certainly satisfies both norms of the Time-slice Rationality account: *Synchronicity* and *Impartiality*. She also satisfies the general rationality constraint presented by Hedden and that is evidentialism. In a nutshell, Candice is perfectly rational at any given point in time but she is not rational over a certain period of time. It is a simple case of an agent being synchronically rational but diachronically

---

<sup>93</sup> For more on the argumentation why the Time-slice Rationality account is too strong see (Lenman 2017).

<sup>94</sup> I believe Bratman and Brunero would agree with my argumentation here. For the explanation of their particular points and argumentation see (Brunero 2012; Bratman 2012, 2018).

irrational. Which diachronic norm of rationality is actually violated here obviously depends on the account of rationality that one is willing to endorse. I would argue that Candice violates even the modest norms of diachronic rationality such as Bratman's *Diachronic Plan Rationality* norm.<sup>95</sup>

The second point that I would like to make is the following. Candice's irrationality in *An afternoon with Candice* is not a bullet that can be easily bitten. In the example, Candice is changing her mind arbitrarily and in a reasonless manner, spending her afternoon on accomplishing nothing. But we can imagine examples with more troublesome consequences than simply wasting an afternoon.

#### *Mandy's business trip*

*Mandy lives and works in Las Vegas, Nevada. She has a business meeting in Reno, Nevada. She plans a business trip. She forms an intention to go to Reno tomorrow morning with her car and she acts upon that intention. Halfway to her destination, Mandy changes her mind and decides not to go to Reno after all. On her returning to Las Vegas, where she lives and works, she again changes her mind and decides to go to Reno. Mandy changes her mind so many times that she is now without gas and stranded in Mojave Desert. Her life is now in serious danger and she needs help.*

There are several points that I am making with this example. Firstly, *Mandy's business trip* and *An afternoon with Candice* are analogous examples. There are only two variables that are different in these examples: the number of times that the agent changes her mind (Candice changes her mind four times and Mandy changes her mind numerous times) and the severity of the consequences (Candice wastes her afternoon on doing nothing and Mandy puts herself in a life-threatening situation). Secondly, as Hedden puts it, there is no *diachronic tragedy* here. It is a clear and uncontroversial case of irrationality – diachronic irrationality. Thirdly, I believe it is quite hard for the proponents of the synchronic accounts of rationality to deny that the agent is acting irrationally in the *Mandy's business trip* example. There surely must be a point in which we should be able to condemn capricious and erratic behavior like the one I presented in *Mandy's business trip*. Being that Mandy is, much like Candice, perfectly rational in

---

<sup>95</sup> There is an interpretation of *An afternoon with Candice* in which Candice simply violates the *Practical standpoint* norm. But this is not the point that Bratman is trying to make and certainly not the point that I am trying to make. Actually, Bratman and I are claiming that Candice is diachronically irrational.

synchronic terms, we must conclude that Mandy and Candice are diachronically irrational. Lastly, I believe that *Mandy's business trip* evokes the spirit of argumentation made by Bratman's in *An afternoon with Candice*.

#### 4.3.5.2 Remembering to be rational

The second type of cases in which Time-slice Rationality fails to account for clear and uncontroversial irrationality are cases of severe forgetfulness. Firstly, I would like to address the connection between the concepts of rationality and forgetfulness. There is a wide array of stances that authors take when discussing the rationality or irrationality of forgetting. Williamson, who is one of the influencers of Hedden's Time-slice Rationality account, has a famous stance on the connection between rationality and forgetting.

*Bayesians have forgotten forgetting. I toss a coin, see it land heads, put it back in my pocket and fall asleep; once I wake up I have forgotten how it landed. (...) I did my best to memorize the result of the toss, and even tried to write it down, but I could not find a pen, and the drowsiness was overwhelming. Forgetting is not irrational; it is just unfortunate.* (Williamson 219: 2000).

Contrarily, Broome, who has an influential account of diachronic rationality called *persistence of intention*, has an opposite view on the connection between rationality and forgetting. In Broome's words:

*A failure of persistence is a sort of forgetting, so you might think it is a failure of memory rather than of rationality. Memory seems to be a separate faculty from rationality. However, you will at least agree that, if your intentions do not persist, it is a failing of coherence of a sort. Your mental attitudes at one time do not cohere properly with those at another. This at least puts forgetting in the same general area as a failing of rationality. Given that, it does not matter to me whether or not you would naturally classify it as a failing of rationality or of something else. In counting Enkrasia as a requirement of rationality, I am already accepting an expansive notion of rationality, which covers various aspects of coherence among your attitudes. I am happy to let it cover this part of memory too* (Broome 177: 2013).

Here we have two conflicting views regarding the connection between rationality and forgetting. On the one hand, Williamson claims that there is nothing irrational about forgetting because, as he puts it, *forgetting is not irrational; it is just unfortunate*. On the other hand, Broome states that forgetting should be considered a failure of some kind. As he puts it, *it is a failing of coherence of a sort*. In the case of forgetting, agent's mental attitudes do not cohere with each other. Broome concludes that this failure should be considered *a failing of rationality* or a failure of some other sort equivalent to the failure of rationality.

Secondly, I would like to point out that this is not the kind of forgetting that I had in mind when I made the claim that severe cases of forgetting can be used as a counterexample to Hedden's Time-slice Rationality. The cases presented by Williamson and Broome are cases of common everyday forgetting and I hold no hard position on whether this is rational or irrational.<sup>96</sup>

What I am talking about are severe cases of forgetfulness which make agents unable to perform any prolonged action or activity. The example that I will provide is based on the condition of anterograde amnesia. Anterograde amnesia is a medical condition in which the agent is unable to form new memories after the event that has caused amnesia. It can also be defined as *impaired capacity for new learning* (Christine N. Smith, Jennifer C. Frascino, Ramona O. Hopkins, and Larry R. Squire 1: 2014). An agent with anterograde amnesia simply cannot form new memories and, as a result, she cannot connect her attitudes from one point in time to another. Let us take a look at, what I termed, *The Curious case of Leonard Shelby*.

#### *The Curious Case of Leonard Shelby*

*Leonard Shelby had a brain injury while showering two years ago. As a result of that injury, Leonard is suffering from a severe anterograde memory dysfunction. Basically, his short-term memory does not exist anymore. He cannot make any new memories, but he can remember perfectly everything that happened before the accident. He knows who he is, his occupation, his marital status, where he lives, etc. But, because of his condition, his attention can only last for 15 minutes. After every 15 minutes Leonard "resets" to the point immediately after the accident. Leonard's life is structured into a series of 15-minute-long slices that are unconnected to each other. His day-to-day life is extremely unusual. One day he wakes up in a cheap motel, neither knowing where he is nor what he is doing there. When he talks to someone*

---

<sup>96</sup> Although, if I was pushed hard enough, I would obviously have to side with Broome. Broome's account, in this context, is an account of diachronic rationality.

*for too long, he forgets how the conversation started. He is unable even to remember the people that he met the day before. While trying to manage common everyday activities (such as buying groceries or conversing with people close to him), Leonard Shelby's entire life devolves into a paranoid and obsessive puzzle-solving mystery.*<sup>97</sup>

There are several points that I would like to make here. Firstly, this line of argumentation is, in some sense, an extension of the argumentation presented in (Doring and Eker 2017). Secondly, the example that I present is idealized. There is no such thing as *severe anterograde memory dysfunction*. It is an idolized version of anterograde amnesia which is a real medical condition. In reality, anterograde amnesia is usually correlated with retrograde amnesia (Kopelman 1989, Squire and Alvarez 1995, Wickelgren 1989), but anterograde amnesia can sometimes occur without retrograde amnesia (Corkin, Hurt, Twitchell, Franklin, & Yin 1987, Russell and Nathan 1946). The bottom line is that *The Curious Case of Leonard Shelby* is an example of an agent with a severe case of anterograde amnesia but without any sign of retrograde amnesia. He knows everything about himself but nothing about himself at this point in time. Thirdly, the main point that I am making with this example is the following. Leonard Shelby is clearly and uncontroversially irrational.<sup>98</sup> He cannot successfully manage most (if any) of his day-to-day activities. The question arises: what exactly makes Leonard irrational? In every 15-minute slice Leonard is completely rational. He responds correctly to reasons and evidence. His preferences are consistent with the correct expected utility theory at the time, in other words, he chooses the correct options at any given time. Leonard satisfies every synchronic norm/requirement that we can think of, but he is still irrational. The inevitable conclusion is that Leonard violates some diachronic norm of rationality. He is perfectly rational at any point in time but he lacks the ability to meaningfully connect those dots and in turn falls into irrationality, more specifically, diachronic irrationality. There is a certain lack of consistency in Leonard's intentions and actions and that failure of consistency is diachronic. Leonard can make plans, commitments and resolutions, but he has a really hard time keeping them.

Lastly, Leonard is perfectly rational under the Time-slice Rationality account. However, as we have established, Leonard is clearly and uncontroversially irrational, so the conclusion that we

---

<sup>97</sup> This example is inspired by Christopher Nolan's *Memento* (2000), which is based on the concept of a short story titled *Memento Mori* and written by his brother Jonathan Nolan.

<sup>98</sup> If this is not a clear and uncontroversial example of irrationality, we can always imagine more extreme cases. We can imagine an agent who loses all of her memory until the point of accident every two seconds. Another example is of a cute little fish named Dory who introduces herself every two seconds because of her attention span.

have to draw is that *The Curious Case of Leonard Shelby* counts as a counterexample to the Time-slice Rationality account.

#### 4.4 The clash of diachronic rationality accounts

In the first part of this chapter, I have established two main points. The first point is that Time-slice Rationality, as the most extensive account of synchronic rationality, has some serious flaws. The second point that I made was that there is a need for a diachronic account of rationality. In this part, I will examine the strengths and weaknesses of the most influential contemporary accounts of diachronic rationality and side with the one I believe to be the most persuasive.

As I presented in the first chapter, we can classify the accounts of diachronic rationality in two distinct categories: historical overview and contemporary overview. In the historical overview there are three phases:

1. Diachronic rationality as the problem of dynamic (sequential) choice (Hammond 1976, 1988; Levi 1991; Machina 1989; McClennen 1990; Rabinowicz 1995),
2. Diachronic rationality as the problem of understanding the nature of intentions and future-directed attitudes (Bratman 2010, 2012; Gauthier 1997; Holton 2009; Velleman 2000),
3. Diachronic rationality as the problem of the nature and the existence of diachronic norms (Broom 2015; Carr 2015; Noody 2019; Ferrero 2009, 2012; Hedden 2015a, 2015b; Hlobil 2015; Meacham 2010b; Moss 2015; Podgorski 2016a, 2016b, 2017).

At this point, I will offer my critique of the first phase, namely *Diachronic rationality as the problem of dynamic (sequential) choice*, and stress why it may be relevant for the contemporary discussion on diachronic rationality in general. These accounts have been criticized by others (Bratman 1999, Ferrero 2009) and, while my critique is in line with theirs, I will give it a fresh look.

There are two main problems with the accounts of diachronic rationality from the first phase. Firstly, these accounts have no concrete or discernible notion of diachronic agency or temporally extended agency. Secondly, these authors (Strotz 1956; Elster 1979; Hammond

1976, 1988; Levi 1991; Machina 1989; McClennen 1990; Rabinowicz 1995) operate within the framework of the expected utility theory. While they present norms like *resoluteness* (an agent should choose according to her former preferences) and *sophistication* (agents should anticipate their future irrationality and avoid it) in order to save dynamic consistency or rather save agents from dynamic inconsistency, they offer no conclusive reason why agents should abide by these norms. The reason they offer is that agents who are dynamically inconsistent violate the *axiom of independence (substitution)*<sup>99</sup>, but this begs the question why we should abide by the axioms of the expected utility theory which has lost favor in recent years as the correct theory of instrumental rationality (Bermúdez 2009).

The relevance of the accounts from the first phase of the diachronic rationality investigation is the following. Contemporary accounts that address the notion of diachronic rationality directly (Hedden 2015a, 2015b, 2015c)<sup>100</sup> and indirectly (Bermúdez 2018) share the same presuppositions about the notion of rationality as the accounts from the first phase, which is the expected utility theory. Consequently, they suffer the same problems as the accounts of the first phase.

With this matter settled, I can now address the contemporary accounts of diachronic rationality (I am going to cover the accounts from the first and the second phase of diachronic rationality here). As I have argued in the first chapter, there are three distinct positions (camps) in contemporary debate regarding diachronic rationality: *antirealist*, *instrumentalist* and *Kantian*. Antirealist accounts (of which Hedden's Time-slice Rationality is the most extensive one) claim there to be no diachronic norms of rationality. The reason why this fails has been thoroughly explained in the first part of this chapter.

Instrumentalist accounts of diachronic rationality consist of a modest approach to diachronic agency and minimalistic norms of diachronic rationality. Those include Broome's *persistence of intention* account (Broome 2013), Snedegar's *coarse-grained plans* account (Snedegar 2017), and Bratman's *stability of intention* account (Bratman 1987).

Broome's account of diachronic rationality consists of two parts: descriptive and normative. In the descriptive part, Broome elaborates on his notion of diachronic agency (although he does not call it so). Broome's account of diachronic agency is a minimalist one. He claims that it is simply in the nature of intention that it *persists over time* (Broome 177: 2013). In this sense,

---

<sup>99</sup> *Axiom 5. Independence. If A, B, and C are in S,  $A \geq B$  if and only if  $(ApC) \geq (BpC)$ .* (Hastie Dawes 259-260: 2001).

<sup>100</sup> The *axiom of independence (substitution)* is part of Hedden's *Diachronic Tragedy Argument*. For more see (Hedden 87-88: 2015b).

intentions are always diachronic and this fact explains our diachronic or temporally extended agency. In the normative part, Broome proposes a diachronic norm of rationality:

*Persistence of Intention. If  $t_1$  is earlier than  $t_2$ , rationality requires of  $N$  that, if  $N$  intends at  $t_1$  that  $p$ , and no cancelling event occurs between  $t_1$  and  $t_2$ , then either  $N$  intends at  $t_2$  that  $p$ , or  $N$  considers at  $t_2$  whether  $p$ .* (Broome 178: 2013).

There are two problems that I find with Broome's diachronic account of rationality.<sup>101</sup> Firstly, the claim that intentions are diachronic in nature is inefficient to explain human capacity for diachronic agency. As Bratman points out, we have the ability to form future-directed intentions and present-directed intentions (Bratman 1987). Although Bratman downgrades the importance of present-directed intentions for our everyday decision making and action, he does leave conceptual space for present-directed intentions. Let us look at the example presented by David Velleman.<sup>102</sup>

*When the plate of cookies is held out to us, why do we make our minds to take one? Why does not our hand shoot out and grab, as it does when we spontaneously and automatically react to sudden throw?* (Velleman 197: 2007).

The hard question for Broome's diachronic account of rationality is the following. How does the account explain examples like the one above? There is nothing genuinely diachronic in the above example. It seems that it is a case of present-directed intention. If the cookie example is an example of diachronic or temporally-extended agency, then we must sadly conclude that diachronic agency is a trivial concept to begin with. But this does not seem to be the case because of the argumentation presented in other accounts of diachronic rationality (Bratman 2018; Holton 2009; Ferrero 2009, 2010, 2012).

The second problem with Broome's account of diachronic rationality is the normative side. Broome's *Persistence of intention* norm is simply too weak. It is unable to account for the clear and uncontroversial cases of irrationality. I am referring to *An afternoon with Candice*. The question arises: what exactly is the problem with Broome's *Persistence of intention* norm? The problem with the said norm is at the following point: *then either  $N$  intends at  $t_2$  that  $p$ , or  $N$*

---

<sup>101</sup> To be fair to Broome, he does not spend much time elaborating his diachronic rationality account. His account consists of only one chapter in his book *Rationality Through Reasoning*. But then again, to be equally fair to the proponents of diachronic rationality like myself, Hedden (a proponent of a synchronic account of rationality) uses Broome's account as a starting point for his claim that there are no diachronic norms of rationality. For the above reasons, I believe that my critique of Broome is necessary and fair.

<sup>102</sup> Velleman actually uses this example to attack Bratman's theory of intention (unsuccessfully in my opinion), but the example serves its purpose better against Broome's account. For more see (Velleman 2007).

*considers at t2 whether p*. Specifically, the problem lies in *or N considers at t2 whether p*. The “considering” part of the norm allows agents like Candice to get away with irrationality freely. Under Broome’s norm, Candice does not need any diachronic coherence at all, she only needs to “consider” and then do whatever she wants.

Now let us focus our attention on Bratman’s *stability of intention* account. This account is Bratman’s *planning theory of intention* interpreted as a diachronic rationality account. One of such interpretations is Snedegar’s *coarse-grained plans* account. There are, as I see it, two main problems for this kind of *instrumentalist* diachronic accounts of rationality. Firstly, Bratman does not consider his *planning theory of intention* to be an account of diachronic rationality. It can be interpreted as such but it is severely underdeveloped as an account of diachronic rationality. The same can be said for Snedegar’s *coarse-grained plans* account. It consists of one norm of diachronic rationality, the *fill-in* norm<sup>103</sup>. The second problem with these accounts is that they do not properly explain the nature of human diachronic agency. There is no argumentation as to why we engage in plans and why we need them. And these are important questions that need to be addressed (Bratman 2018; Ferrero 2009, 2010, 2012).

We are left with Kantian accounts of diachronic rationality. Firstly, I would like to point out that I do not have any major problems with Kantian accounts of diachronic rationality (Bratman 2012, 2014, 2018; Ferrero 2009, 2010, 2012; Holton 2009). That being the case, I would like to point out some minor issues that I have with Ferrero’s account of diachronic rationality and Holton’s account of diachronic rationality. Ferrero puts a lot of emphasis on the importance and value of temporally extended agency. The best way to conceptually approach our ability to engage in temporally extended activities is by viewing our temporally extended activities as *narratives*. We, as human beings, tell stories to ourselves and others about the structure of our lives and our decision making. Ferrero points out that certain aspects of our temporally extended agency are more relevant than others in the way that would be constructed in a literary narrative. This is, according to Ferrero, the correct way to interpret temporally extended agency. Although I am sympathetic to this line of reasoning, I would like to raise the following questions. What is the connection between interpreting temporally extended agency as personal narratives and practical (instrumental) norms of rationality? Do we really need to interpret temporally extended agency as personal narratives in order to have a satisfactory diachronic

---

<sup>103</sup> As I have presented in the first chapter, Snedegar’s *fill-in* norm states the following. *Fill In: If at t1 you have a partial plan to A at t2, then by t2 you ought to have sufficiently filled in that plan.* (Snedegar 601: 2017).

norm of rationality? It seems to me that a narrative interpretation of temporally extended agency is one of many and not necessarily the only one.

Holton does not consider his account to be a diachronic account of rationality in these terms, although other authors do (Bratman 2018, Hedden 2015b, Ferrero 2009). His account is a reinterpretation of Bratman's *theory of intention* with a focus on concepts like will, commitment and resolutions. This account works well as an account of diachronic rationality and is, beside Ferrero's, the best alternative to Bratman's diachronic account of rationality. We are left with Bratman's *Diachronic Plan Rationality* account.

#### **4.5 Bratman's *Diachronic Plan Rationality* account emerges victorious**

I will not address Bratman's account in detail because I have done so in Chapter 2. I will simply point out some features that make Bratman's *Diachronic Plan Rationality* account stand out from the rest.

Firstly, in my terminology, Bratman's *Diachronic Plan Rationality* account is located in between instrumentalists and Kantians. On the one hand, Bratman would certainly characterize his account as an instrumentalist one in opposition to accounts like Ferrero's. On the other hand, Bratman's account does fulfill all the requirements for Kantians. Bratman finds it necessary to take the concept of temporally extended agency seriously if we want to offer diachronic norms of rationality. In some sense, Bratman's account is "the best of both worlds". His approach and analysis of decision making and acting is an instrumentalist one (because it is based on his *planning theory of intention*) and his account is, at the same time, strikingly Kantian because of the concept of self-governance.

Secondly, Bratman's *Diachronic Plan Rationality* account addresses the notion of diachronic agency and the notion of diachronic norms of rationality effectively. In a descriptive sense, Bratman's account thoroughly explains how humans engage in temporally extended agency. There are two layers to this explanation. The first layer is based on Bratman *planning theory of intention*. Humans are, Bratman argues, future-directed planning agents. The reason for this are our cognitive limitations. We plan because we need to coordinate ourselves with others and with ourselves at different times. In order for this coordination to work, we need stability and commitment to be essential parts of our future-directed intentions and plans. The second layer

is based on Bratman's notion of self-governance. We, as human beings, when intending and acting, do not simply connect one temporal dot with another to achieve consistency. There is an intrinsic value in governing our own lives consistently. And that value is self-governance. From that value, which Bratman hopes we all share, he derives diachronic norms of rationality. Lastly, Bratman's *Diachronic Plan Rationality* account can effectively explain the irrationality in the examples which I have presented in the first part of this chapter: *An afternoon with Candice*, *Mandy's business trip* and *The Curious Case of Leonard Shelby* (the first two are examples of reasonless change of mind and the last is an example of severe forgetfulness). Let us remind ourselves of Bratman's diachronic norm of rationality.

*Diachronic Plan Rationality (DPR): If S is a planning agent who is capable of diachronic self-governance then the following is, defeasibly, pro tanto irrational of S:*

(a) *S is engaged in a planned temporally extended activity that has so far cohered with both synchronic and diachronic self-governance.*

(b) *Given her present standpoint, a choice to continue with her planned activity would cohere with that standpoint and so cohere with her continued synchronic self-governance and, in part for that reason, with her diachronic self-governance. And yet*

(c) *S makes a choice that blocks her continued diachronic self-governance.* (Bratman 217: 2018).

Firstly, we should note that there is a possible interpretation of *An afternoon with Candice* and *Mandy's business trip* in which they are not engaged in temporally extended activities. This interpretation can be made somewhat easily for *An afternoon with Candice* but much harder for *Mandy's business trip*. This is not the interpretation that Bratman or I support. But in such interpretation, Candice and Mandy simply violate the *Practical standpoint* norm.<sup>104</sup>

Secondly, we can clearly see that Candice and Mandy are locally irrational because they make a choice that blocks their continued self-governance. That being said, there is an important thing to point out. *DPR* states that an agent is rationally allowed to change her mind at any point and is allowed to *blocks her continued diachronic self-governance* if she has a reason for it. The norm states that it is *defeasibly, pro tanto irrational* for an agent to choose contrary to

---

<sup>104</sup> *Practical Rationality/Self-governance (PRSG): If S is capable of self-governance it is, defeasibly, pro tanto irrational of S either to fail to have a coherent practical standpoint or to choose in a way that does not cohere with her standpoint.* (Bratman 211: 2018).

her *continued diachronic self-governance*. If an agent has a reason, then she can rationally change her mind. One of the reasons why Candice and Mandy are diachronically irrational is because they have no reason for their respective change of heart by definition. In the case of Leonard, he is certainly diachronically rational within the 15-minute time slice, but it is hard to see how he does not violate *DPR* in a time interval longer than 15 minutes. He certainly makes choices that block his *continued diachronic self-governance* simply because he does not remember his original intention.

#### **4.6 New frontiers: challenges for diachronic rationality and thoughts on further investigations**

As I have demonstrated in Chapter 2 and in the first part of this chapter, Bratman's *Diachronic Plan Rationality* account is the best or, at least, one of the best ways of explaining human diachronic agency and the necessity for diachronic norms. On the descriptive side, Bratman's *DPR* account allows us to explain the nature of our diachronic agency. We are, as human beings, future-directed planning agents who need plans for personal and intrapersonal coordination. On the normative side, the *DPR* account allows us to explain diachronic (practical) norms of rationality. We, as human beings, share the value of governing our own lives in a consistent and coherent manner. And from the value of self-governance arises the *DPR* norm of rationality. Bratman's account is both explanatory and flexible. On the one hand, the account explains why the cases of Candice, Mandy and Leonard are cases of genuine irrationality. On the other hand, the account is flexible enough to allow agents to change their respective minds as many times as they want. The way that agents can rationally change their mind is by having a reason to change their mind. Therefore, agents can change their mind and be rational.

*So, given that your past prior to  $t_2$  is not changeable by you at  $t_2$ , you have at  $t_2$  a reason of self-governance either to retain your prior intention to  $X$ , or relevantly to change your mind about the strength of your grounds for  $X$ . However, it does not in general follow that you have a reason of self-governance simply to retain your prior intention, since you might have conclusive reason instead to change your assessment of your grounds.* (Bratman 84: 2012).

As we can see, on the one hand, we have a reason of self-governance to conform to our prior intention, and on the other hand, we have a choice to change our minds. The trouble for

Bratman's account, in my opinion, starts when we ask the following question: what exactly does it mean that an agent has conclusive reason for a change? It is certainly easy to see that Candice, Mandy and Leonard do not have *conclusive reason for a change*. Candice and Mandy have no *conclusive reason for a change* by definition and Leonard has no *conclusive reason for a change* because he does not remember his original intention. But there are cases in which it is much harder to see the rationality and irrationality of changing one's mind.

*The runaway bride case*<sup>105</sup>

*Mary is engaged to her fiancé John. She sincerely intends to marry him in three months. Mary buys her wedding dress, books the church, sends invitations to her family and friends. She chooses her maid of honor, plans her honeymoon and arranges the music for her wedding. On the day of her wedding, Mary changes her mind and leaves her fiancé John at the altar. (Grčki: Forthcoming).*

The main point of my example is to put some pressure on Bratman's *Diachronic Plan Rationality* account. How should we rationally assess Mary's change of heart? Is she like Candice and Mandy? Does she change her mind capriciously? Bratman's *Diachronic Plan Rationality* account would state the following. Mary is rationally obligated to her commitment to John if she has no *conclusive reason for a change*. But that is the crux of the problem that I am presenting here. Has Mary *conclusive reason for a change*? To be clear, Mary has received no new information about the world between the time she formed the intention and the time she abandons it. There are two logical possibilities here. Either Mary's actions are rational or her actions are irrational. If her actions are irrational, that means she has no *conclusive reason for a change* which puts her in a category alongside Candice and Mandy. This leads one to bite the bullet (at least in some soft sense) because it seems that there are huge differences between Mary's case and cases like Candice and Mandy. If her actions are rational, that means that Mary has found a reason for her change of heart. It is difficult to see how she found a reason if she received no new information.

In recent years, there has been a development in the field of practical rationality that I consider relevant for this kind of problem (*The runaway bride case*). The problem (if we consider it to be one) is: how can Mary find a reason for her action if she has received no new information between the formation of intention and the abandonment of intention? One of the ways this can

---

<sup>105</sup> I can also present a more extreme example of this type. A case in which Mary intends to get an abortion and then changes her mind in the abortion clinic without receiving any new information.

be achieved is by *framing* the same material outcome (or goal) in two different ways. The idea of agents *framing* materially identical outcomes as different goes all the way back to the investigations of Tversky and Kahneman and their *framing effects* (Tversky and Kahneman 1981). Recently, the concept of *frames* has been reintroduced by Bermúdez (2009, 2018) in a manner that is, in a broad sense, related to diachronic rationality.<sup>106</sup> Bermúdez claims that we can rationally have different attitudes towards materially same outcomes. Borrowing from Frederic Schick (1991, 1997, 2003), Bermúdez offers the following example.<sup>107</sup>

*[The soldier] jumped out of the trench and ran along the parapet in full view. He was half-dressed and was holding up his trousers with both hands as he ran. I refrained from shooting at him . . . I did not shoot partly because of that detail about the trousers. I had come here to shoot at ‘Fascists’; but a man holding up his trousers isn’t a ‘Fascist’, he is visibly a fellow-creature, similar to yourself, and you don’t feel like shooting at him. (Orwell 1957, 199) (Bermúdez 86: 2009).*

The point that Bermúdez makes is the following. The term *Fascist* and the term *a fellow-creature* refer to the same object under different descriptions. But those different descriptions are also different *frames*. Moreover, an agent can be rational by *framing* the outcome differently (Bermúdez 2018). One object can have different properties and at certain times some property is more salient than others, so we can rationally choose with respect to that property or *frame*. The question arises: how does this relate to *The runaway bride* problem? Specifically, Mary could find a new reason for abandoning her intention by *framing* her goal (marriage with John) differently. Let us say that John is extremely lazy to do chores. Mary is aware of this fact when she forms her intention to marry him. But, as the time goes by and the wedding day approaches, she begins to wonder. Does she really want to spend the rest of her life with someone who is poor at doing household chores? She has received no new information about her fiancé John, but she focuses her attention on one salient property of John (the fact that he is lazy to do chores) and consequently frames the situation differently. She has acquired a new reason without new information and can reasonably change her mind. This is the main idea behind the recent work by Bermúdez (Bermúdez 2018, Forthcoming). Whether or not this

---

<sup>106</sup> Bermúdez’s main focus is on arguing that dynamic choice counterexamples the claim that decision theory is the correct way to approach practical rationality (2009) and the relation between self-control and framing effect (2018).

<sup>107</sup> The example is taken by Schick from George Orwell’s essay “*Looking back at the Spanish Civil War*” (Orwell 1957).

line of argumentation can be relevantly implemented in the domain of diachronic rationality is a matter of further investigation.

## 5 Conclusion

As I stressed at the beginning, the main aim of this dissertation is to offer a defense of diachronic rationality. My thesis, *Diachronic agency can be rationally assessed in the way in which synchronic agency is assessed*, as indicated in the introduction, is a modest one. It enables us to recognize that sometimes agents are locally irrational in a specific and unique way – in other words, they are diachronically irrational. The severe cases of capriciousness and the cases of agents not being able to form new memories are cases in which agents are irrational in a uniquely diachronic manner. This alone presents a relevant need for the concept of diachronic rationality.

I believe that my thesis underlines some issues that go beyond the realm of the philosophy of practical rationality and the philosophy of action. Those issues include responsible and conscientious behavior of agents. By agents, I mean individual agents as well as public institutions, private institutions, governments or corporations. Intentions matter. Plans matter. If there has been reasonable and relevant amount of deliberation (individual or collective) and a decision (in the form of a plan) has been made, there is a rational (and in some cases even moral) obligation for agents to honor that decision if nothing relevant intervenes. This is easier said than done. Our future is unknown to us. The world can end next Thursday. We are evolutionarily engineered to be present-time biased (we are extremely bad at delayed gratification). Nevertheless, we should be able to find the courage and the conviction to honor our past commitments. From an individual promise made to a friend to a government plan to reduce poverty and cut carbon dioxide emissions, we cannot, either rationally or morally, abandon our plans at the first sign of trouble. Diachronic rationality matters.

## 6 References

- Aesop/Winter, M. 1919. The Aesop for children. Chicago: Rand, McNally & Co. Accessed from *The Library of Congress* on June 15, 2019
- Allais, M. and Hagen, O. Eds. 1979. *Expected Utility Hypothesis and the Allais Paradox*. Reidel.
- Andreou, C. 2007. Understanding Procrastination. *Journal for the Theory of Social Behaviour* 37(2): 183–93.
- Anscombe, G. E. M. 1963. *Intention*. Second edition. Oxford: Blackwell.
- Anscombe, G. E. M. 1983. The Causation of Action. Reprinted in *Human Life, Action, and Ethics*, M. Geach and L. Gormally. eds. 89–108. Exeter: Imprint Academic, 2005.
- Anscombe, G. E. M. 1989. Practical Inference. Reprinted in *Human Life, Action, and Ethics*, M. Geach and L. Gormally. eds. 109–147. Exeter: Imprint Academic, 2005.
- Armstrong, D. 1984. Consciousness and causality. In D. Armstrong and N. Malcolm, eds. *Consciousness and Causality*. Oxford: Blackwell.
- Arntzenius, F. 2003. Some Problems for Conditionalization and Reflection. *Journal of Philosophy* 100 (7): 356-370.
- Arrow, K.J. 1950. A Difficulty in the Concept of Social Welfare. *Journal of Political Economy*. 58: 328–346.
- Audi, R. 1993. *Action, Intention, and Reason*. Ithaca. NY: Cornell University Press.
- Audi, R. 2001. *The Architecture of Reason: The Structure and Substance of Rationality*. Oxford University Press.
- Audi, R. 2004. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton, NJ: Princeton University Press.
- Austin, J. L. Three ways of spilling ink. *The Philosophical Review*. Vol. 75. No. 4. 427-440.
- Bermúdez, J. L. 2009. *Decision Theory and Rationality*. Oxford University Press.
- Bermúdez, J. L. 2019. *Self-Control, Decision Theory, and Rationality: New Essays*. Cambridge University Press.

- Bermúdez, J. L. *The Power of Frames: New Tools for Rational Thought*. Forthcoming from Cambridge University Press.
- Bliss, R. 2014. Viciousness and Circles of Ground. *Metaphilosophy*. 45: 245–256.
- Bradley, R. 2017. *Decision theory with a human face*. Cambridge University Press.
- Bratman, E. M. 1987. *Intentions, Plans and Practical Reason*. The University of Chicago Press.
- Bratman, E. M. 1999. *Faces of Intention Selected Essays on Intention and Agency*. Cambridge University Press.
- Bratman, E. M. 2000a. Reflection, Planning, and Temporally Extended Agency. *The Philosophical Review* 109: 35-61.
- Bratman, E. M. 2000b. Valuing and the Will. *Philosophical Perspectives: Action and Freedom* 14: 249-265.
- Bratman, E. M. 2001. Two Problems About Human Agency. *Proceedings of the Aristotelian Society*. ed. A. W. Price. 101: 309-326.
- Bratman, E. M. 2005. Planning Agency, Autonomous Agency. In James Stacey Taylor, ed. *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge University Press. 33-57.
- Bratman, E. M. 2012. Time, rationality and self-governance. *Philosophical issues*. 22: 73-88.
- Bratman, E. M. 2018. *Planning, Time, and Self-Governance: Essays in Practical Rationality*. Oxford University Press.
- Bricker, P. 2005. The Relation Between General and Particular: Entailment vs. Supervenience. In D. Zimmerman, ed. *Oxford Studies in Metaphysics* (Volume II). Oxford: Oxford University Press.
- Briggs, R. 2009. Distorted Reflection. *The Philosophical Review* 118 (59-85).
- Broome, J. 1999. *Ethics out of economics*. Cambridge University Press.
- Broome, J. 2001. Are intentions reasons? And how should we cope with incommensurable values?. In C. Morris and A. Ripstein, eds. *Practical Rationality and Preference: Essays for David Gauthier*, Cambridge: Cambridge University Press, 98–120.

- Broome, J. 2003. Practical Reasoning. In J. Bermúdez and A. Millar, eds. *Reason and Nature: Essays in the Theory of Rationality*. Oxford: Oxford University Press, 85–112.
- Broome, J. 2006. *Weighing lives*. Clarendon Press.
- Broome, J. 2013. *Rationality Through Reasoning*. Wiley-Blackwell.
- Brunero J. 2012. Instrumental Rationality, Symmetry, and Scope. *Philosophical Studies*. 157: 125-140.
- Brunero, J. and Kolodny N. 2013. *Instrumental rationality*. Stanford Encyclopedia of Philosophy. Accessed September 23, 2019.
- Buchak, L. 2016. Decision theory. In Christopher Hitchcock & Alan Hajek, eds. *Oxford Handbook of Probability and Philosophy*. Oxford University Press.
- Builes, D. 2020. Time-Slice Rationality and Self-Locating Belief. *Philos Stud* 177, 3033–3049.
- Butler, S. 1900. *The Odyssey by Homer*. (translated) CreateSpace Independent Publishing Platform. Accessed March 13, 2019
- Carr, R.J. 2015. Don't stop believing. *Canadian Journal of Philosophy*. 45:5-6, 744-766.
- Christensen, D. 1991. Clever bookies and coherent beliefs. *Philosophical Review* 100 (2): 229-247.
- Christensen, D. 1996. Dutch-Book Arguments Depragmatized: Epistemic Consistency For Partial Believers. *The Journal of Philosophy*, 93: 450–479.
- Copp, D. 2003. 'Ought' Implies 'Can', Blameworthiness, and the Principle of Alternate Possibilities. In David Widerker and Michael McKenna, eds. *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. 265–299. Ashgate.
- Corkin S, Hurt RD, Twitchell TE, Franklin LC, Yin RK. 1987. Consequences of nonpenetrating and penetrating head injury: retrograde amnesia, post-traumatic amnesia, and lasting effects on cognition. In Levin HS, Grafman J, Eisenberg HM, editors. *Neurobehavioral recovery from head injury*. New York: Oxford University Press. 318–329.
- Dancy, J. 2004. *Ethics Without Principles*. Oxford University Press.

- Davidson, D. 1963. Actions, Reasons, and Causes. *The Journal of Philosophy* Vol. 60, No. 23.
- Davidson, D. 1970. How is Weakness of the Will Possible? Reprinted in *Essays on Actions and Events*. 21–42. Oxford: Oxford University Press. 1980.
- Davidson, D. 1971. Agency. Reprinted in *Essays on Actions and Events*. 43–61. Oxford: Oxford University Press. 1980.
- Davidson, D. 1973. Freedom to Act. Reprinted in *Essays on Actions and Events*. 63–81. Oxford: Oxford University Press. 1980.
- Davidson, D. 1978. Intending. Reprinted in *Essays on Actions and Events*. 83–102. Oxford: Oxford University Press. 1980.
- Deroose K. 1992. Contextualism and Knowledge Attributions. *Philosophy and Phenomenological Research*, 52(4): 913–929.
- Donagan, Alan. 1985. *Human Ends and Human Actions: An Exploration in St. Thomas's Treatment*, Milwaukee: Marquette University Press.
- Doody, R. If There Are No Diachronic Norms of Rationality, Why Does It Seem Like There Are?. *Res Philosophica*. 96:2 141-173.
- Döring S. and Eker B. Rationality, time and normativity: On Hedden's time-slice rationality. *Analysis* 77 (3):571-585.
- Elster, J. 1979. *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Cambridge University Press.
- Etherington, D. W. and R. Reiter. 1983. On Inheritance Hierarchies and Exceptions, In *Proceedings of the National Conference on Artificial Intelligence*, Los Altos, Calif.: Morgan Kaufmann.
- Feldman, R. and Conee. E. 1985. Evidentialism. *Philosophical Studies* 48: 15-34.
- Feldman, Richard. 1995. Authoritarian Epistemology. *Philosophical Topics* 23.1: 147-169.
- Feldman, Richard. 2003. *Epistemology*. Upper Saddle River, NJ: Prentice Hall.
- Ferrero, L. 2006. Three Ways of Spilling Ink Tomorrow. In E. Baccharini and S. Prijič-Samaržija (eds.), *Rationality in Belief and Action*. Rijeka: 95-127.
- Ferrero, L. 2009. What Good is a Diachronic Will?. *Philosophical Studies*. 144.3:403-430.

- Ferrero, L. 2010. Decisions, Diachronic Autonomy, and the Division of Deliberative Labor. *Philosophers' Imprint*. 10.2:1-23.
- Ferrero, L. 2012. Diachronic Constraints of Practical Rationality. *Philosophical Issues*. 22.1:144:164.
- Ferrero, L. 2014. Diachronic Structural Rationality. *Inquiry*. 57.3:311-336.
- Frankfurt, H. 1971. Freedom of the Will and the Concept of a Person. *Journal of Philosophy*. 68: 5–20.
- Frankfurt, H. 1987. Identification and Wholeheartedness. In Ferdinand David Schoeman, ed. *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge University Press.
- Frederick, S. and Loewenstein, G. 1999. Hedonic adaptation. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology*. 302–329. Russell Sage Foundation.
- Gauthier, D. 1996. Commitment and choice: An essay on the rationality of plans. In F. Farina, F. Hahn, and S. Vanucci eds. *Ethics, Rationality, and Economic Behavior*. Oxford: Oxford University Press.
- Gigerenzer, G., & Selten, R. 2001. Rethinking rationality. In G. Gigerenzer & R. Selten, eds. *Bounded rationality: The adaptive toolbox*, 1-12. The MIT Press.
- Gilboa, I. 2009. *Theory of decision under uncertainty*. Cambridge University Press.
- Godfrey-Smith, P. 2016. *Other minds: the octopus, the sea, and the deep origins of consciousness*. New York : Farrar, Straus and Giroux.
- Goldman, A. 1986. *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.
- Goldstein, S. 2016. A Preface Paradox for Intention. *Philosophers' Imprint*, 16.
- Grčki, D. 2021. Rational decision making: Changing one's mind. Forthcoming from Prolegomena: Journal of Philosophy.
- Greco J. 1999. *Agent Reliabilism*. In J. Tomberlin (ed.), *Philosophical Perspectives* (Volume 13: Epistemology), Atascadero, CA: Ridgeview Press. 273-296.
- Greco J. 2010, *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*, Cambridge: Cambridge University Press.

- Grice, H. P. 1971. Intention and Uncertainty. *Proceedings of the British Academy*. 5: 263–279.
- Hammond, P. J. 1976. Changing tastes and coherent dynamic choice. *The Review of Economic Studies* 43(1): 159–73.
- Hampton, J. 1998. *The authority of reason*. Cambridge New York: Cambridge University Press.
- Hardt, R. 2018. Storytelling agents: why narrative rather than mental time travel is fundamental. *Phenom Cogn Sci* 17. 535–554.
- Hare, R.M. 1984. Supervenience. *Aristotelian Society Supplementary Volume*. 58: 1–16.
- Harman, G. 1976/1997. Practical Reasoning. reprinted in *The Philosophy of Action*, A. Mele (ed.). Oxford: Oxford University Press, 149–177.
- Hartman G. 1986 Willing and Intending. In *Philosophical Grounds of Rationality*, R. E. Grandy and R. Warner (eds.), Oxford: Oxford University Press. 363–380.
- Hastie, R. and Dawes, R. M. 2001. *Rational choice in an uncertain world: The psychology of judgment and decision making*. Sage Publications, Inc.
- Hedden, B. 2012. Options and the Subjective Ought. *Philosophical Studies* 158, no. 2: 343–360.
- Hedden, B. 2015a. *Reasons without Persons: Rationality, Identity, and Time*, Oxford University Press.
- Hedden, B. 2015b. Time-Slice Rationality, *Mind* 124, no. 494: 449–491.
- Hedden, B. 2015c. Options and Diachronic Tragedy. *Philosophy and Phenomenological Research*, 90, no. 2: 423–451.
- Hedden, B. 2016a. Mental Processes and Synchronicity. *Mind* 125, no. 499: 873–888.
- Hedden, B. 2019a. *Rationality and Synchronic Identity*, Australasian Journal of Philosophy 97, no. 3: 544–58.
- Hedden, B. 2019b. Reasons, Coherence, and Group Rationality. *Philosophy and Phenomenological Research*, no. 3 :581–604.
- Hedden, B. and Greco, D. 2016b. Uniqueness and Metaepistemology. *The Journal of Philosophy* 113, no 8 (2016): 365–95.

- Heeney, M. 2020. Diachronic agency and practical entitlement. *European Journal of Philosophy* 28. 1:177-198.
- Hinchman, E. S. 2003. Trust and Diachronic agency. *Noûs* 37(1), 25-51.
- Hinchman, E. S. 2015. Narrative and the stability of intention. *European Journal of Philosophy*. 23(1), 111–140.
- Hlobil, U. 2015. There Are Diachronic Norms of Rationality. *Thought: A Journal of Philosophy* 4 (1): 38-45.
- Holton R. 2008. Partial Belief, Partial Intention. *Mind*, 117: 27–58.
- Holton R. 2009. *Willing, Wanting, Waiting*. Oxford: Oxford University Press.
- Horgan, T. 1982. Supervenience and Microphysics. *Pacific Philosophical Quarterly*. 63: 29–43.
- Horgan, T. 1993. From Supervenience to Superdupervenience: Meeting the Demands of a Material World. *Mind*. 102: 555–86.
- Hume, D. 1739-40/1975. *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge, 2<sup>nd</sup> ed. revised by P. H. Nidditch, Oxford: Clarendon Press.
- Jeffrey, C. R. 1965. *The Logic of Decision*. New York: McGraw-Hill.
- Kahneman, D. and Tversky, A. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, Vol. 47, No. 2. 263-291.
- Kant, Immanuel. 1785/1956. *Groundwork of the Metaphysics of Morals*. Translated and analyzed by H. J. Paton. New York. Harper and Row.
- Kavka, G. S. 1983. The Toxin Puzzle. *Analysis*. 43: 33-36.
- Kolodny, N. 2005. Why be rational?. *Mind*. 114 (455): 509-563.
- Koons, R. C. 2017. *Defeasible reasoning*. Stanford Encyclopedia of Philosophy. Accessed September 13, 2019.
- Kopelman MD. 1989. Remote and autobiographical memory, temporal context memory and frontal atrophy in Korsakoff and Alzheimer patients. *Neuropsychologia*. 27:437–460.
- Korsgaard, C. M. 2008. The normativity of instrumental reason. In *The Constitution of Agency*, 26-68. Oxford: Oxford University Press.

- Kreps, M. D. 1979. A Representation Theorem for Preference for Flexibility. *Econometrica*, 47: 565-577.
- Lenman, J. 2017. Reasons without humans. *Analysis* 77-3:586-595.
- Levi, I. 1974. On indeterminate probabilities. *The Journal of Philosophy*. 71: 391–418.
- Levi, I. 1986. The paradoxes of Allais and Ellsberg. *Economics and Philosophy* 2: 23–53.
- Levi, I. 1989. Rationality, prediction, and autonomous choice. *Canadian Journal of Philosophy* 19: 339–63.
- Levi, I. 1997. *The Covenant of Reason*. Cambridge: Cambridge University Press.
- Littlejohn, C. 2012. *Justification and the Truth-Connection*. Cambridge University Press.
- Littlejohn, C. 2018. Stop Making Sense? On a Puzzle about Rationality. *Philosophy and Phenomenological Research*.
- Luce, D. and H. Raiffa. 1957. *Games and decisions: Introduction and critical survey*. New York: Wiley.
- Machina, M. 1989. Dynamic Consistency and Non-expected Utility Models of Choice under Uncertainty. *Journal of Economic literature*. 27: 1622-68.
- Maher, P. 1993. *Betting on Theories*, Cambridge: Cambridge University Press.
- Makinson, D. C. 1965. The Paradox of the Preface. *Analysis*. 25: 205–207.
- Memento. 2000. DVD. Directed by Christopher Nolan. Los Angeles, CA. Newmarket Films, LLC.
- McClennen, F. E. 1990. *Rationality and Dynamic Choice: Foundational Explorations*. Cambridge University Press.
- McClennen, F. E. 1998. Rethinking Rational Cooperation. In Leinfellner W., Köhler E, eds. *Game Theory, Experience, Rationality*. Vienna Circle Institute Yearbook.
- McHugh, C. and Way, J. 2018. What is Good Reasoning?. *Philosophy and Phenomenological Research*. 96: 153–74.
- McLaughlin, B.P. 2001. Supervenience. In Smith, ed. *Encyclopedia of the Social and Behavioral Sciences*, Amsterdam: Elsevier, 1142–1147.
- Millgram, E. 1997. *Practical Induction*. Harvard University Press.

- Millgram, E. 2001. *Varieties of Practical Reasoning*. MIT Press.
- Moore, G.E. 1922. *Philosophical Studies*, London: Routledge.
- Moss, S. 2015. Time-Slice Epistemology and Action Under Indeterminacy. In Tamar Szabó Gendler & John Hawthorne eds. *Oxford Studies in Epistemology*. Oxford University Press. 172-94.
- Moyer, M. 2008. Weak and Global Supervenience are Strong. *Philosophical Studies*. 138: 125–150.
- Neumann, J. V. and Morgenstern, O. 1944. *Theory of Games and Economic Behavior*. Princeton University Press.
- Oaksford, M. and Chater, N. 2007. *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.
- Olson, E. 1997. *The Human Animal: Personal Identity Without Psychology*. New York: Oxford University Press.
- Orwell, G. 1957. *A Collection of Essays*. Garden City: Doubleday.
- Parfit, D. 1983. Rationality and Time. *Proceedings of the Aristotelian Society*. 84: 47-82.
- Paul, K. S. 2015. The courage of conviction. *Canadian Journal of Philosophy*. 45:5-6, 647-669.
- Perry, J. 1972. Can the Self Divide?. *Journal of Philosophy*. 69: 463–488.
- Podgorski, A. 2016a. Dynamic Permissivism. *Philosophical Studies* 173(7): 1923-1939.
- Podgorski, A. 2016b. A Reply to the Synchronist. *Mind*. 125(499): 859-871.
- Podgorski, A. 2016c. Dynamic Conservatism. *Ergo* 3.
- Podgorski, A. 2017. Rational Delay. *Philosophers' Imprint* 17.
- Pollock, J. 1974. *Knowledge and Justification*. Princeton University Press.
- Quinn, W. S. 1990. The Puzzle of the Self-Torturer. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*. Vol. 59, No. 1. 79-90.
- Rabinowicz, W. 1995. To have one's cake and eat it: Sequential choice and expected utility violation. *The Journal of Philosophy* 92: 586–620.

- Ramsey, F. P. 1926. Truth and Probability. In *The Foundations of Mathematics and other Logical Essays*, R. B. Braithwaite. ed. London: Routledge & Kegan Paul, 1931.
- Raz, J. 1986. *The Morality of Freedom*. Oxford: Clarendon Press.
- Raz, J. 2005. The Myth of Instrumental Rationality. *Journal of Ethics and Social Philosophy*. 1: 1–28.
- Rescher, N. 1960. Choice without preference. A study of the history and of the logic of the problem of “Buridan's ass”. *Kant-Studien*. 51 (1-4):142-175.
- Resnik, M. 1987. *Choices: An Introduction to Decision Theory*. University of Minnesota Press.
- Russell WR, Nathan PW. 1946. Traumatic amnesia. *Brain*. 69:280–300.
- Sartre, J.P. 1946/2007. *Existentialism is a Humanism*. Tr. Carol Macomber, New Haven: Yale.
- Schick F. 1986. Dutch bookies and money pumps. *Journal of Philosophy*. 83: 112-119.
- Schick, F 1997. *Making Choices*. Cambridge: Cambridge University Press.
- Schick, F 2003. *Ambiguity and Logic*. Cambridge: Cambridge University Press.
- Schick, F. 1991. *Understanding Action*. Cambridge: Cambridge University Press.
- Searle, J. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Setiya, K. 2007a. *Reasons without Rationalism*, Princeton. NJ: Princeton University Press.
- Shoemaker, S. 1970. Persons and Their Pasts. *American Philosophical Quarterly*, 7: 269–285.
- Shoemaker, S. and Swinburne, R. 1984. *Personal Identity*. Blackwell.
- Simon, H. A. 1972. Theories of Bounded Rationality. In C. B. McGuire and R. Radner, eds., *Decision and Organization*. North-Holland Publishing Company.
- Skyrms, Brian. 1987b. Dynamic Coherence and Probability Kinematics. *Philosophy of Science*, 54: 1–20.
- Smith CN, Frascino JC, Hopkins RO, Squire LR. 2013. The nature of anterograde and retrograde memory impairment after damage to the medial temporal lobe. *Neuropsychologia*. 51(13):2709-2714.

- Smokrović, N. 2018. Informal Reasoning and Formal Logic: Normativity of Natural Language Reasoning. *Croatian journal of philosophy*. XVIII. 54; 455-471.
- Snedegar, J. 2017. Time-slice rationality and filling in plans. *Analysis*. 77-3: 595-607.
- Sosa, E. 1991. *Knowledge in Perspective: Selected Essays in Epistemology*, Cambridge: Cambridge University Press.
- Squire LR, Alvarez P. 1995. Retrograde amnesia and memory consolidation: A neurobiological perspective. *Current Opinion in Neurobiology*. 5:169–177.
- Stalnaker, R. 1999. Extensive and Strategic Forms: Games and Models for Games. *Research in Economics*. 53. 293-319.
- Stalzer, K. 2004. *On the Normativity of the Instrumental Principle* (Ph.D. Thesis, Stanford University).
- Strotz, R. H. 1956. Myopia and Inconsistency in Dynamic Utility Maximization. *Review of Economic Studies*. 23: 165–80.
- Teller, P. 1973. Conditionalization and Observation, *Synthese*, 26: 218–258.
- Tenenbaum, S. 2014. Minimalism about Intention: A Modest Defense. *Inquiry*. 57: 384–411.
- Tenenbaum, S. 2018. Reconsidering Intentions. *Noûs*, 52: 443–472.
- Tenenbaum, S. and Raffman D. 2012. Vague Projects and the Puzzle of the Self-Torturer. *Ethics* 123. 86-112.
- Titelbaum, M. 2015. Continuing on. *Canadian Journal of Philosophy*, 45(5-6), 670-691.
- Tversky, A. 1975. A Critique of Expected Utility Theory: Descriptive and Normative Considerations. *Erkenntnis*, 9: 163–173.
- van Fraassen, Bas. 1995. Belief and the Problem of Ulysses and the Sirens. *Philosophical Studies*, 77: 7–37.
- Velleman, J. D. 1991. Well-being and time. *Pacific Philosophical Quarterly*. 72: 48–77 (Reprinted in Velleman 2000).
- Velleman, J. D. 1997. Deciding how to decide. In G. Cullity & B. Gaut (Eds.) *Ethics and practical reason*. Oxford: Oxford University Press. (Reprinted in Velleman 2000).
- Velleman, J. D. 2000. *The possibility of practical reason*. Oxford: Oxford University Press.
- Velleman, J. D. 2003. Narrative explanation. *The Philosophical Review*. 112: 1–25.

- Velleman, J. D. 2006. *Self to self*. Cambridge: Cambridge University Press.
- Velleman, J. D. 2007. What good is a will? In A. Leist & H. Baumann. eds. *Action in context*. Berlin/ New York: de Gruyter/Mouton.
- Victor M, Adams RD, Collins GH. 1989. The Wernicke-Korsakoff Syndrome and Related Neurological Disorders due to Alcoholism and Malnutrition. 2 ed. *Philadelphia*: F.A. Davis.
- Wallace, R. J. 1999. Three Conceptions of Rational Agency. Reprinted in *Normativity and the Will*. 43–62. Oxford: Oxford University Press, 2006.
- Watson G. 1987. Free action and free will. *Mind*. 96: 154-72.
- Wedgwood, R. 2002. *Internalism explained*. *Philosophy and Phenomenological Research* Vol. 65, No. 2. 349-369.
- Wedgwood, R. 2017. *The Value of Rationality*. Oxford: Oxford University Press.
- Weintraub, R. 2012. What can we Learn from Buridan's Ass?, *Canadian Journal of Philosophy*. 42:3-4. 281-301
- White, R. 2005. Epistemic permissiveness. *Philosophical Perspectives* 19 (1): 445–459.
- Williams, B. 1981. *Moral Luck*, Cambridge: Cambridge University Press.
- Williamson, T. 2000, *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Woodard, E. 2019. *Diachronic Normativity: A Puzzle About Fickleness*. NYU Washington Square Circle (Talk 10/2019)
- Zagzebski, L. 1996. *Virtues of the Mind*, Cambridge: Cambridge University Press.

## List of tables and figures

|                |    |
|----------------|----|
| Figure 1 ..... | 24 |
| Figure 2 ..... | 28 |