UNIL | Université de Lausanne

Faculté des lettres

FACULTÉ DES LETTRES

SECTION DE PHILOSOPHIE

The Arrow of Time and Time Symmetry in non-Relativistic Quantum Mechanics

On why we have sound reasons to believe in a quantum arrow of time

THÈSE DE DOCTORAT

présentée à la

Faculté des lettres
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès lettres

par

Cristian Ariel López

Directeur de thèse

Prof. Michael Esfeld
Prof. Olimpia Lombardi

LAUSANNE
2019

UNIL | Université de Lausanne
Faculté des lettres

IMPRIMATUR

Le Décanat de la Faculté des lettres, sur le rapport d'une commission composée de :

Directeurs de thèse :

Monsieur	Michaël-Andreas Esfeld	Professeur, Faculté des lettres, UNIL
Madame	Olimpia Lombardi	Professeure, Université de Buenos Aires, Argentine

Membres du jury :

Monsieur	Christian Sachse	MER, Faculté des lettres, UNIL
Madame	Nélida Gentile	Professeure, Université de Buenos Aires, Argentine
Monsieur	Carl Hoefer	Professeur, Université de Barcelone, Espagne
Monsieur	Karim Thébault	Professeur, Université de Bristol, Royaume-Uni

autorise l'impression de la thèse de doctorat de

MONSIEUR CRISTIAN LOPEZ

intitulée

The Arrow of Time and Time Symmetry in non-Relativistic Quantum Mechanics
On why we have sound reasons to believe in a quantum arrow of time

sans se prononcer sur les opinions du candidat / de la candidate.

La Faculté des lettres, conformément à son règlement, ne décerne aucune mention.

Lausanne, le 7 novembre 2019

Dave Lüthi
Doyen de la Faculté des lettres

Index

Part 3

α.

Introduction

Or what all this is going to be about…

"Inside time there is another time
Ever still
Lacking hours, weight and shadow
Solely alive
Like that old man on the bench
Self-absorbed identical endless
Always unseen
It's transparency"

(Octavio Paz, "The same time")

"Dentro del tiempo hay otro tiempo,
Quieto
Sin horas ni peso ni sombra
Sólo vivo
Como el viejo del banco
Unimismado idéntico perpetuo
Nunca lo vemos
Es la transparencia"

(Octavio Paz, "El mismo tiempo")

We, raised and educated in the Western civilization, believe that past is somehow behind us while future is somewhere ahead. Because of this, we use expressions like "c'mon face the past!" as if one were commanding the other to spatially turn around and see (face) the past; or "winter (or future) is coming" as if things were moving, approaching to us from the front. Notwithstanding how deeply-ingrained these beliefs are in our Western-minded metaphoric mapping of time, some people have come to adopt a quite different world view: you might get a blank stare in return if you command an Aymara (ancient people still living in the Andes highlands) to face the past. Well, that's what she always does: walking back to the future (which lies behind her), always seeing, facing her past. Winter wouldn't be coming *to* her (from the front) but reaching, hunting her from the back, as if their entire metaphoric mapping of time were turned radically upside down[1]. Or would it be rather the other way around and we, Western-minded thinkers, have being having our world view turned upside down all along? Probably neither option, as a good deal of we call 'time' is culturally made and, consequently, it varies along places and time. Nonetheless, beyond what we, as human beings, might come

[1] Take a look at this must-read paper about our metaphoric mapping of time: Núñez, R. and Sweetser, E. (2006). "With the Future Behind Them: Convergent Evidence from Aymara Language and Gesture in the Crosslinguistic Comparison of Spatial Construals of Time". *Cognitive Science* 30: 401-450

to think about our experienced, culturally-made time, Nature might conceal something that we can still call 'time'; time that probably lacks many of the (colorful) culturally-based properties that human beings have largely thought it has. Probably flowless, coming from and going nowhere. Neither back nor in front. Lacking hours, identical even transparent. *And yet*, maybe capturing one of its more irrevocable and persistent (even essential) properties: its *directionality.* And about this the thesis is going to be.

***

Time is bewildering. Take any property one typically ascribes to time and this will surely enclose a deep, largely-discussed philosophical problem. In this light, time looks like a still alive fertile terrain for philosophical inquiry. Beyond its unquestionable philosophical richness, time has a hard-to-handle dual feature, neatly condensed in Augustine's words.

> "What is time? If nobody asks me, I know; but if I were desirous to explain it to one that should ask me, plainly I know not"

On the one hand, time is familiar, too familiar to us. Allegedly, one *knows* what time is because one has daily experience of it. Time is so constitutive of our language, of our knowledge, of our day-to-day lives, of our very existence as human beings that one thinks one has a fair, well-grounded grasp of its essential features. Unquestionably, time is *there*, flying away in front of us –one witnesses how things are unavoidably devoured by time, slipping into non-existence, over and over. It intuitively seems that one knows what time is, and yet philosophical inquiry just comes into play when details need to be spelled out just a tiny bit further.

On the other hand, as one starts looking into our daily experience of time and its alleged properties, we quickly lose the North. In this sense, time is so elusive, that one is rapidly left clueless about where our inquiry should start off and how it should be properly conducted. Is time object of *philosophical* investigation exclusively? Or does it fall better under *scientific* inquiry? If the latter is the case, which science would be the most adequate to delve into the nature of time? Further, various problems around time have been so intertwined that it is already an uphill task to start to disentangle them.

About the topic

The so-called problem of the arrow of time, the philosophical problem I am here mainly interested in, is not alien to this general landscape. In a quite comprehensive way, the problem concerns a simple and intuitively graspable property, to wit:

> Does time (whatever it comes to be) *objectively*[2] instantiate the property of "having a privileged direction"?

By "privileged direction" it is commonly meant that there would be a sharp distinction between the past-to-future direction and the future-to-past one. However, the problem is too vague and abstract to be researched systematically. Which would be the more appropriate way to address the problem? How can we come to know (or to identify) the property of "having a privileged direction"? Something else should be add to the formulation to narrow the problem down.

In effect, such a problem might be investigated along different branches, ranging over psychology and the neurosciences, to pure metaphysics and physics. A longstanding tradition, tracing back to the late nineteenth century, has nonetheless addressed the problem exclusively within *physics* and the problem becomes a problem to be solved within physics ever since. In accordance with this, it may be broadly reworded as following. If physics is in the business of studying how matter moves in space-*time*, well, it is likely that physics has something to say about space-time's structure (particularly, time's) through this study. In accordance with this, it may be broadly reworded as following:

> Does time (whatever it comes to be) *objectively* instantiate the property of "having a privileged direction" *according to some of our best physical theories*?

We are now getting closer to a sense of the overall problem, but we are not quite there yet. One of the central claims of this thesis is that such a way to pose the problem has been understood quite differently by many authors: Though loads of ink have been spilled since the nineteenth century in discussing about the (seemingly) directed nature of time, it is not often clear what the problem is really about. Indeed, the problem of the arrow of time seems to enclose a bunch of closely-related, but conceptually different problems. Literature on the problem of the arrow

[2] My usage of the word "objectively" here refers to a mind-independent property, that is, a property that is instantiated by the world out there and that is independent of any *subjective* feature of whom knows this property. This is close to the Nagelian sense of objectivity: being a property of nature that remains invariant under a change of perspective –that is, it's *independent of* a perspective (the so-known "view from nowhere"). In this case, "having a direction of time" would be a property that is independent of agents knowing such a property and that, at least in principle, it's independent of the perspective from which they know about the direction of time.

of time in physics has not been always careful in keeping the scenario clear. In 1974, John Earman already complained about how poorly understood and how uninformative the problem of the arrow of time had been so far:

> […] "very little progress has been made on the fundamental issues involved in "the problem of the direction of time." By itself, this would not be especially surprising since the issues are deep and difficult ones. (…) It seems not a very great exaggeration to say that the main problem with "the problem of the direction of time" is to figure out exactly what the problem is or is supposed to be!" (1974: 15)

Curiously, others have also raised their voices in complaining about this uneasy situation: what are we really discussing when dealing with the problem of the arrow of time?

A good part of this thesis deals with this imperative clarification: Sometimes progress, especially in foundational matters, can initially come from clarifying the problem at dispute. So as to start off on the right foot, any philosophical inquiry should be conducted in such a way that the problem at stake be cleared out from the outset. One should be able to put the finger on the problem and claim "*this* is what we are arguing about, and this and that have been left out". By drawing a clear-cut line to identify the problem, one is able to establish if there are actually genuine divergent positions about *the same* problem, or if there are simply different problems on the table. I think that many divergences around the nature of time, and time's direction particularly, have come up, and been intensified, because problems were not made clear from the very beginning. Even worse, controversies could seem taking us nowhere as long as they are actually talking about different topics.

Yet, the problem ought to be fenced in even more. Current physics features many theories, some of them incompatible with one another. Hence, the question whether time is objectively directed in current physics, should be rather circumscribed to a specific physical theory. There are some common loci where the question has been largely investigated (e.g., thermodynamics and statistical classical mechanics), but others have instead been scarcely taken into consideration (at least comparatively). Such is the case of *non-relativistic quantum mechanics*: While issues gravitating around the direction of time have greatly involved melting ices, entropy-increasing systems or expanding gases in a box, few things have been said about quantum systems and the directionality of time. Another central claim of this thesis will be that non-relativistic quantum mechanics encloses many more conundrums as to time's direction than usually taken. The guiding question of the thesis would thus be posed as following:

> Does time (whatever it comes to be) *objectively* instantiate the property of "having a privileged direction" *in non-relativistic quantum mechanics*?

There is yet a further point to pay attention to: As it's widely-known, non-relativistic quantum mechanics is a theory of success as well as of scandals. The former is exemplified in the fact that no many theories have come so close to the accuracy of its predictions and come to be so successful in developing new technologies. The latter implies in practice that many philosophers and physicists have nonetheless found the theory (as an epistemic enterprise) deeply disappointing: Not only does it deliver no picture whatsoever of what the world is like according to the theory, but it is neither clear whether we even have a scientific theory at all. They thus found themselves in need of *interpreting* the theory, which in turn paved the way to empirically equivalent, but metaphysically divergent, quantum theories as well as to potentially rival quantum theories. All this prelude relates to my business here because any appropriate inquiry about the arrow of time in non-relativistic quantum mechanics must capture as much as possible this complexity. So, by "non-relativistic quantum mechanics" I will mean both its bare formal apparatus (the *standard* non-relativistic quantum mechanics) and its varied interpretations.

About the content

Having said what this thesis will be about, let's now detail its content. The thesis articulates in three parts, where the third part supposes the second, and this in turn the first. Each part is organized in chapters and its headed by a brief introduction of its content.

The first part seeks to shed some light what the *problem of the arrow of time* is, first from a broad viewpoint and then within physics. By "shed some light" I mainly mean not only to elucidate many of the involved concepts, but also to disentangle and to sort out some subsidiary problems. That's why I mean this part as "*unpacking"*. The content is organized in two chapters. The first one (Chapter I) deals with the problem of the arrow of time from a metaphysical, broad viewpoint and then introduces the relevant physical concepts to approach it from physics. The second chapter (Chapter II) asserts that there are, at least, two versions of the problem. This distinction has been largely overlooked in the literature, and the two versions of the distinction renders conflated. Furthermore, it will point to the relationship between both versions.

The second part focuses on *standard* non-relativistic quantum mechanics. The guiding question here is *whether*, and *in which* sense, non-relativistic quantum mechanics treats time as having an objectively privileged direction. By "standard" non-relativistic quantum mechanics I mean the partially-interpreted formalism free of any so-called interpretation. In such so-fenced-in terrain, the relevant question will thereby have to do with the Schrödinger equation and the notion of time-reversal invariance, managing without any interpretational matter for the time being. The part contains three chapters. The first (Chapter III) introduces, conceptually and formally, the standard non-relativistic quantum mechanics and its symmetry group –Galilean Group. The second chapter (Chapter IV) involves the notions of time reversal and time-reversal invariance in standard non-relativistic quantum mechanics and directly addresses the question of whether The Schrödinger equation is left unaltered under flipping the direction of time. The next chapter (Chapter V) assesses previous results in the light of the arrow of time discussion.

The third and last part complexes the scenario to the extent that interpretations of standard non-relativistic quantum mechanics come into play. Its primary focus will be those interpretations that add further elements to the standard theory and how they deal with the problem of the arrow of time (in either of its versions). Four chapters organize this last part. The first (Chapter VI) overviews how various interpretations come up and in which way their emergence intervenes on the problem of the arrow of time in quantum mechanics. The next chapter (Chapter VII) focuses on the so-called Orthodox Quantum Mechanics. Here the Collapse Postulate is analyzed as a candidate for a quantum arrow of time. The third chapter (Chapter VIII) analyzes spontaneous collapse theories (like GRW-type theories) and to what extent they offer some grounds for a fundamental (structural) arrow of time. The last chapter (Chapter IX) directs its attention at Bohmian Mechanics, the problem of the arrow of time in a Bohmian framework and the role that time reversal plays within the theory.

About the claims to be defended

Introductions of each part will pose clearly the theses to be defended therein. In general, the thesis pursues three aims (following the chapters), articulated in the following claims

(a) At least two distinguishable, though closely-related, problems have historically and conceptually been put under the label 'the problem of the arrow of time': *the problem of a structural arrow of time* and *the problem of a non-structural arrow of time* (or *the problem of the two realms*).

(b) The notions of time reversal and time-reversal invariance play a paramount role in stablishing whether a direction of time (in either sense) holds or not within a physical theory.

(c) The notion of time reversal, however, entails a series of metaphysical, epistemic and methodological assumptions that ought to be properly unveiled.

(d) The orthodox way to understand and to formally define time reversal in non-relativistic quantum mechanics can be, I think quite soundly, challenged along with the widely-extended claim that it turns out to be undoubtedly time-reversal invariant.

(e) In challenging it, whether non-relativistic quantum mechanics is time-reversal invariant depends upon how one comes to conceptually understand and to formally represent time reversal within the theory.

(f) From (e), whether standard quantum mechanics treats time as having an objective, privileged direction of time depends upon how one comes to conceptually understand and to formally represent time reversal within the theory.

(g) From (e) and (f) follows that even standard quantum mechanics in the most elementary situation may reflect a structural, fundamental time asymmetry on which a structural arrow of time may be grounded.

(h) Independently of the above-listed points, whereas some interpretations of non-relativistic quantum mechanics exhibit a structural time asymmetry (paradigmatically, spontaneous collapse theories) by adding further dynamical elements to the standard version, others just manifest a non-structural asymmetry (for instance, orthodox quantum mechanics and Bohmian mechanics).

About the philosophical methodology

Some centuries ago, this thesis would have, plainly and undoubtedly, been a thesis on *Natural Philosophy*. Nowadays, the expression has lost much of its (nostalgic) appealing and would better be reworded as *philosophy of physics, foundations of physics, scientifically-informed metaphysics* or something like that. I feel that my thesis lies somewhere among them, in their interstices. The overall topic tackles a longstanding metaphysical issue to the extent to which it ultimately focuses on what time is, what its properties are. The research, however, doesn't depart from heavenly philosophical claims or from a purely arm-chair attitude, but it by contrast intends to be conducted in an, so to speak, Earthlier way. And that means that it will always keep an eye on what current physics has to say about the matter. That is, it is to be conducted in a scientifically-informed manner. In doing this, it will also deal with some foundational issues and physics-inner problems, combining conceptual and formal tools coming from quantum physics to address an iconic philosophical issue.

I am convinced that our understanding of the natural world depends upon developing interdisciplinary research programs where philosophy and various scientific disciplines converge. This thesis tries to mirror this overarching spirit. Not only have many philosophical concerns experienced a renewed boost when addressed in an interdisciplinary manner, but some conceptual and even genuinely scientific progress has also been made when philosophy has intervened. If science (and physics in particular) is somehow in the business of knowing what the world is like, then philosophy of science, philosophy of physics and a scientifically-informed metaphysics also contribute, actively, to this aim. I here share Hasok Chang's words: philosophy is the continuation of science by other means; means not exclusively confined to an elucidatory, passive task.

Let me add some words about the way in which my arguments, and my enquiry in general, will be ideally executed. Philosophy's terminology is prone to be somewhat coercive. Philosophical arguments aim to *force* others to adopt a given belief. As Nozick neatly puts it "a philosophical argument is an attempt to get someone to believe something, whether he wants to believe it or not". This explains why philosophers typically put so much efforts in producing *knock-down* arguments or to push adversaries out of rationality's limits. I mean this research as an exercise in a more exploratory, so to speak, way of doing philosophy. I don't mean at all that knock-down arguments shouldn't play any role in philosophizing or that they unavoidably incur in some sort of coercive practice. What I rather exercise here is something that weights equally knock-down arguments and exploratory tasks: I don't intend anyone to believe what I actually believe about the theses I will defend here, but rather to consider things from the perspective I will propose. Some of the claims I will argue for are not conventional and somewhat heterodox and I don't expect the reader to be fully persuaded. If persuaded, good. If not persuaded, it's equally valuable that she rethinks of her previously-held beliefs from a different viewpoint and eventually reinforce them. This, in a way or another, will push our understanding of the matter further.

Acknowledgments

This thesis has been thought, discussed and written in different places and times and has in consequence involved many people. To begin with, I'm deeply grateful to the terrific Group of Philosophy of Science of the University of Buenos Aires (Mariana, Hernán, Manuel, Jesús, Guido, Camilo, Martín, Leo), specially, to Sebastián Fortín and Federico Holik for tireless and, in occasions, heated and even frustrating discussions about many (controversial) topics in this thesis. All their remarks were crucial to improve my initial ideas. I'd also like to thank to Lausanne Group (Frida Trotter, Andrea Oldofredi, Christian Sachse, Tiziano Ferrando, Federico Benítez, Dustin Lazarovici and Guillaume Köstner). I felt not only very welcomed when I was there, but I also received many valuable remarks on my work. Many thanks to Frida for her friendship and for revising and commenting some parts of this thesis. And, of course, special thanks to one of my supervisors, Prof. Michael Esfeld, who kindly accepted to supervise this thesis. His support, guide and all his advices were important for the result achieved.

Many pieces of this work were discussed in different personal meetings, visiting, conferences and workshops I attended to during the last four years, so I'm much in debt with all the people I came across in the period and the several commentaries, criticisms and advices I've received from them. Specially, many thanks to Bryan Roberts, Dennis Dieks, Giovanni Valente, Karim Thébault, Pablo Acuña, Diego Romero, Patricia Palacios, Pieter Thyseen, Sylvia Wenmacker, Thomas Ryckmann, Neil Dewar, Wayne Myrvold, Ruth Kastner, Christian Wüthrich, Tim Maudlin, Michael Cuffaro, Roberto Laura, Manuel Gadella, Sebastián Murgue, among many others. This manuscript was substantially improved thanks to the feedback I received from the jury, specially, from Karim Thébault and Carl Hoefer. I'm deeply grateful to them for the careful reading of the first version of this thesis and for all their helpful remarks.

Out of academy, many thanks to my (official) family (my dad, my mom, my sister, her boyfriend, their daughter, their pets and my dog, Ibsen) for their unconditional support and love. To my (non-official) family (my girlfriend's family) and to my lifelong friends (Patricio, Matías, Pili, Cande).

Finally, my deepest and heartfelt gratitude to two persons that are entirely responsible for the person I am and the student I came to be. Olimpia, not only my supervisor but also my friend and quasi-mother, has been in the last eight years an unconditional support for my academic career. She not only taught me how to do philosophy at the highest level (whether I successfully learned it or not is a different issue), but also how to be a *morally* good researcher, how much valuable is the interdisciplinary work and how to struggle with some academic setbacks. Thank to her, I'm where I'm now. Any positive, valuable point in this thesis came surely from her influence. And María José (my eternal partner) who has been my companion for as long as I can remember (fifteen years?), with whom I not only share my entire life, but also my academic life: all I've written, or thought, has always been firstly discussed with and

read by her. Her presence has been so steady in the last fifteen years that I can rightfully consider my research as a team work (I've come to learn as much philosophy of biology as she philosophy of physics!). To them this thesis is humbly dedicated.

Part 1

Introduction

This first part, which contains two chapters, faces the uneasy situation of not counting with a clear-stated formulation of the problem one is to deal with. As one gets into the literature, one rapidly has the felling that under the label “the problem of the arrow of time” a packed bunch of problems (or a “thicket of problems” as John Earman (1974) puts it) and intertwined notions have been indistinctly discussed in the philosophy of physics and of time. So, the first step towards some clarification of the problem is to *un*pack such a thicket. The main aim of this first part is to pose the problem of the arrow of time precisely, offering a rationale for why the problem should be primarily understood the way I shall claim it should be primarily understood. Along with this, there are some further general theses I would like to defend here.

(a) How one conceives the problem of the arrow of time depends on certain metaphysical and epistemic assumptions with respect to the nature of time.

When discussing the direction of *time*, one broadly concerns whether *time* instantiates the property of “being directed” (whatever this may be). The focal point here is that people may come to different understandings of what *time* is. Philosophical and scientific literature has been massive on this issue, as widely known. However, as far as I am aware, literature *on* the problem of the arrow of time has not drawn enough attention to it thus far. As it will become clear afterwards, divergent philosophical stances with respect to the nature of time may lead us to a different understanding of what the problem of the arrow of time is supposed to be. And this is so for a simple reason: there is no assumption-free way to understand the problem, or a non-trivial and comprehensive formulation thereof. My aim here is hence not to pursue this unattainable goal. What I do want to defend is that such commitments are constitutive of how the problem is to be grasped. And it is this acknowledgment which will take us to a far-reaching understating of it.

There is yet another caveat regarding now the *direction* of time. Not only does the formulation of the problem depend upon our previously-taken assumptions with respect to the nature of time, but also upon the status of its *directionality*, that is, if it is either a contingent/non-structural property, or a necessary/structural one. As I will defend along this part is that there are different ways to conceive the *directionality* of time.

(b) One's assumptions with respect to the world's ontology, what role symmetries play in physics, what a law of nature is or is supposed to do, also play an active, though secondary, role in conceiving of the problem.

In posing the problem of the arrow of time, or in developing the tools for addressing it from a physical viewpoint, the literature has frequently resorted to a panoply of expressions or images referring to, for instance, laws of physics' properties (for instance, "being invariant under time reversal"), or to how the *world's content* would look like going backward in time, or to time-reversal invariance as *symmetry* of fundamental physical theories, and so forth. Of course, there is no univocal, straightforward way to understand all these notions or images. And these divergences may lead us to set the problem up differently.

(c) Putting some details aside, there are two ways to understand the problem of the arrow of time in the literature, namely, in terms of *a structural arrow of time* and in terms of a *non-structural arrow of time*

Points (a) and (b) converge on (c). Actually, both ways to understand the problem embody two quite different projects under the same label 'the problem of the arrow of time'. Projects that, at the same time, involve unalike commitments with respect to time, the sort of directionality they are looking into, the status of symmetries, physical theories and physical laws as previously mentioned. These two unalike understandings of the problem lead, I shall propose, to two different formulations that can be tracked in the relevant literature, to wit, *the problem of a structural arrow of time* and *the problem of the two realms*, as I shall call them. To fully display both formulations and to show their relations are the core of the second chapter. In relation to point (c), I shall make two points

(i) First, though both problems should be kept conceptually apart, they are nevertheless related. Particularly, the problem of the two realms already implies an answer to the problem of a structural arrow of time.
(ii) Second, the problem of a structural arrow of time is thereby conceptually prior to the problem of the two realms. Hence, the problem of the arrow of time should be literally and primarily understood in the structural way.

Finally, the last thesis I would like to defend here is primarily methodological.

(d) Certain *requirements* must be satisfied for the formulation of the problem to be philosophically meaningful.

Time exhibits a convoluted dual character. On the one hand, temporal notions are so deeply-ingrained in our language, knowledge and experience of the world that, unwittingly, our inquiry could be being conducted in a temporally-biased way. In other words, as one regards reality from a temporal perspective, one could be mistakenly ascribing time objectives properties that it actually lacks. Huw Price (1996) warns against this sort of anthropocentric approach by proposing an *atemporal view* –the "view from Nowhen". By adopting such an atemporal perspective, one guards oneself not only from subjective temporal biases, but also from falling into fallacies that come from there. On the other hand, there seems to be a sort of opposite stream within philosophers of physics' community in demanding our theories to be maximally symmetric. Since symmetries are playing an increasing relevant role in physical theories, philosophical problems are often formulated in these terms. I contend one should be extremely careful in this maneuver. Unnoticeably, physicists and philosophers frequently disagree on what it is the epistemic and ontological status of symmetries. In order to carry out a proper investigation of time's direction, some methodological requirements will thus be in order here too.

To reach the aims of this first part, and to develop the above-mentioned theses further, the investigation will be carried out in two stages:

- In the first chapter, I shall offer a metaphysical overview of the problem of the arrow of time as well as an analysis of the most relevant notions in the physics literature – those of time-reversal invariance and (ir)reversibility. Firstly, I shall state the most general formulation of the problem plus two different readings of it: structural and non-structural. Secondly, I shall disentangle the notion of time-reversal invariance from the notion of (ir)reversibility and will show how they can be associated with the arrow of time debate. Finally, I will point out that the problem of the arrow of time may be introduced in two conceptually different ways, namely, *the problem of the two realms* and *the problem of a fundamental arrow of time*.
- In the second chapter, I shall develop further these two formulations of the problem and how they can be traced back in the relevant literature. The chapter will formulate both problems accurately and will link to the metaphysical background offered in the first chapter. Broadly stated, I will argue that both problems address the problem of the arrow of time differently, entailing different conclusions with respect to the physical foundations of the arrow of time. At the end of the chapter, I shall suggest that the problem of a structural arrow of time is prior to, and supposed by, the problem of the

two realms. In addition, I will point to the conceptual and formal tools one has at disposal to investigate each of them.

I.

Unpacking the Problem of the Arrow of Time

From Metaphysics to Physics…

Section 1. Unpacking the problem

What is *metaphysically* the quid of the problem of the arrow of time? And, what is the *philosophical* motivation of the problem as addressed in physics? Frequently, the origin of the problem relates to our intuitive notion of a changing world, of a becoming, of a flowing time, or of the passage of time. Intuitively, we believe time goes by. We *feel* the flow of time and our very experience apparently witnesses such an objective flowing time. A longstanding philosophical tradition inspired in Hume's philosophy have thought that philosophy's task is to offer a well-grounded and thorough explanation of our beliefs or ideas based on our experience. Thus, what are the grounds for our deep-seated beliefs about time?

This is, no doubt about it, a fascinating self-contained enterprise as well as an interesting starting point to dig into the nature of our experience of time. However, the problem of the arrow of time as discussed in physics wouldn't quite span the problem of how to explain our *experience* of time in physical terms, or to answer on which grounds our temporal experience arises. From that motivational starting point, the problem should be narrowed down. And this is due to obvious reasons: if the problem of the arrow of time in physics is, roughly, whether our best and fundamental physical theories are capable of picking a privileged direction of time (as mentioned in the Introduction), one could hardly rebuild our complex temporal intuitions by merely pointing out that fundamental physics somehow does pick such a temporal directionality. One can surely set forth a philosophical investigation from people's temporal intuitions, but one also needs to disentangle the different ingredients they involve, such as that of a passage or a flowing time. By doing so, one will be able to single out which ingredient of our experience relates to the problem of the arrow of time as understood in philosophy of physics.

If one in fact has experience of the passage or flow of time, our very idea of such a passage is extremely complex. Setting aside any psychological or neurologically-based mechanism that carry out anything like our alleged experience of time, the very concept of "the passage of time" is philosophically convoluted. Huw Price (2012), for instance, has pointed out three ingredients to the complex "passage package" idea.

(a) The first one is the idea that *present* is an objectively-distinguished moment that continually varies. This thesis is famously known as the '*moving now*' doctrine, or '*presentism*'.[3] Peter Landsberg (1982) illustrates this idea by saying that "the time variable is rather like a straight line on which a point marked "The Now" moves uniformly and inexorably" (1982: 2).

(b) The second ingredient is that of having a *privileged directionality*. This ingredient would allow us to objectively distinguish between the past and the future in so far as the past-to-future direction would look quite different with respect to the future-to-past one.

(c) The third element compounding the passage of time package is that of a flow, that is, time as intrinsically having a flow-like, dynamic character. Price stresses that the most particular "ingredient" of this element is that time would involve a *rate* of change as any flowing stuff does. Furthermore, the flow-like element is entangled to some extent with the others, as a flow-like movement logically implies a preferred directionality, for instance.

Clearly, according to this conceptual reconstruction, defending the idea of the passage of time implies holding all their ingredients altogether. However, one could either raise various suspicions on them or instead defend them individually. One can, for instance, be worried about whether these ingredients in themselves *logically* make sense. For instance, some have argued that the idea of a flowing time (that is, the third ingredient), in so much as it implies the idea of a rate of change, is roundly meaningless. Price (1996) argues as follows:

1. If time really flows, then it would make sense to say *how fast* time flows

[3] Presentism also implies the ontological thesis that *only* the present moment is real. The 'moving now' doctrine is not actually committed to such an ontological thesis (e.g., the moving-spot light theory by Broad 1927).

2. Somebody could reply by saying that the idea makes perfect sense insofar as time flows at one second per second
3. However, to say that time flows at one second per second is ridiculous, because one second per second is not, physically, a rate whatsoever because it is dimensionless.
4. Therefore, time does not flow.

By debunking the idea of a flow of time, Price is in the same movement undermining the whole idea of the passage of time (for discussion on this point, see Maudlin 2002, Raven 2010, and Price 2012 for a reply to Maudlin 2002). It is however clear that one could nonetheless hold the second ingredient and believe that time has a directionality, though it doesn't flow in any relevant sense.

Alternatively, one could object the first ingredient individually, too. Famously, John Ellis McTaggart (1908) argued that the idea of a 'moving now' is logically incoherent. McTaggart introduced to the philosophical literature the 'A-theory' and 'B-theory' of time as different theoretical frameworks to account for the nature of time. Traditionally, an A-theory is defined in terms of the A-series, where each instant has the monadic property of "being past", "being present", and "being future" (A-properties). Philosophers of time generally claim that presentism implies an A-theoretic framework, since events are ordered and located relative to now (see Craig 2000, Zimmerman 2008. For criticisms and discussion see Rasmusen 2012, Tallant 2012 and Deasy 2017). Contrarily, a B-theory holds that only B-series exists, where the ordering of events is based on binary *relational* properties, as "being earlier than" or "being later than" (B-properties).

In a more liberal understanding of these theories of time, A-theory's supporters typically defend a *dynamic theory of time* according to which time objectively flows, whereas B-theory's supporters deny that such a flowing or passing be an objective feature of reality (see Baron 2015). McTaggart's argument seeks to show that time (or genuine change) is unreal by claiming that one of its essential features (its dynamic character) is self-contradictory. Though there are many different versions of McTaggart's argument, the simplest one can be sketched as follows (see McDaniel 2016):

1. Time is real only if genuine change occurs
2. Genuine change occurs only if A-series exists
3. A-series does not exist
4. Time is unreal

There are many subtleties in the argument, but the point to be stressed is that an A-series exists only if events localized in time (or instants of time) can coherently acquire the properties of being past, being present *and* being future. The dynamics of those *absolute* attributes is what constitutes genuine change. But, in so much as those attributes are incompatible with each other, the very idea of an A-series is self-contradictory: times or events cannot instantiate the property of being future *and* past, for example. Naturally, many objections have been raised against the argument. Yet McTaggart's argument is a well-known objection to the logical coherence of the moving-now doctrine, and consequently, to the whole idea of the passage of time.

Above and beyond all these considerations, the most compelling reason for holding the idea of the passage of time is that one *truly* has experience of it. In general, both parties – supporters and deniers of an objective dynamic time– agree on that everybody has genuine experience of the passage (see Schuster 1986, Davies 1995, Price 1996, 2012; Craig 2000, Le Poidevin 2007, Paul 2010, Prosser 2013). This widely-accepted intuitive advantage has been the workhorse for any supporter of an objective passage of time. And to explain *why* it is the case that we truly have such an experience, whereas the passage of time is not real, has been a hard nut to crack for any denier of the passage.

Notwithstanding this widely-shared assumption, some philosophers have recently tried to undermine such a belief by arguing that one should not too easily concede to passing time's advocators that one *genuinely* has experience of the passage. For instance, Akkiko Frischut (2013) claims that conceding such an intuition as obviously true "is a mistake because if we think carefully about what it would mean to experience temporal passage, it is not obvious at all that we do" (2013: 145). Christoph Hoerl (2013) has argued along the same line, though his arguments differ from Frischut's slightly; Gal Yehezkel (2013) has also supported such conclusion but based on stronger reasons: he argues that conceiving time as perceivable is a categorical mistake. Clearly, they attempt to pave the way for an alternative way to debunk dynamical theories of time, undercutting one of the at-first-glance more compelling reasons to believe in it; in doing so, they argue for static accounts of time (although Yehezkel explicitly does not mean to discredit A-theories, 2013: 79): If, to begin with, one does not even have experience of the passage of time, it seems that there is nothing left to hold on for A-theory's supporters, and any motivation to defend the very idea of the passage of time rapidly vanishes.

In any case, one does not need to get that far away to debunk the plausibility of the passage of time. Many of the arguments raised in the philosophy of physics lie much closer to us than logical or metaphysical criticisms. Maybe time passage's ingredients are logically coherent, and even metaphysically possible. Probably they are genuinely part of our day-to-day experience. And yet, such ingredients might either be purely subjective or merely apparent: the passage of time perhaps is not an objective feature of the natural world (or even of any *physically* possible world). Some have argued, for instance, that presentism or the existence of an objective moving now have no place in current physics (see Rietdijk 1966, Putnam 1967, Penrose 1989). For instance, Special Relativity not only treats time as a dimension of a bigger four-dimensional structure (the so-called space-time), but also treats present (as a location in time) as a frame-dependent notion. The point here is that different moving reference systems could disagree on what instant they call 'now' because of the relativity of simultaneity. According to this argument, presentism is not an option for anyone that believes that Special Relativity is (approximately) true. (For an interesting debate see Valente 2012). Therefore, the passage of time would not be a property of the natural world either.

This is the sort of inquiry I am mainly interested in. Taken logical coherence and metaphysical plausibility for granted, one can be rather worried about whether such features have an *objective* correlate with the natural world as described by current physics. The overall strategy would hence consist in taking what our best and fundamental theories teach us about the natural world and using that evidence to argue for or against any of the features of the passage of time. In any case, it should be bear in mind the complexity of the notion of the passage of time and kept clear what features of it one is contending. Naturally, by undermining any of those features (a dynamic now, having a directionality or a flow-like behavior), one is also debunking the whole idea of the passage of time for it implies those. However, it is quite clear that the opposite is not true. Physics might not offer any well-grounded reason to believe in the passage of time as an objective feature of the natural world (for instance, there might be no sound reasons to hold the objectivity of a moving now), but it might well capture any of the other features (having a directionality, for instance).

True. It is not so easy to figure out how to go about know this. Neither is so clear how to translate these concerns into physics' language. Be that as it may, to get on the right track, I think that one should start off by conceptually disentangling the notion of interest from other related notions, and by clarifying how they all relate. This distinction will be twofold. The first part will concern metaphysically useful distinctions in order to narrow the problem down. The second part will involve physical distinctions in order to show how the problem can be articulated within the field of physics.

Section 2. Non-structural and structural metaphysical distinction between both time's directions

Having analyzed the idea of the passage of time, it is clear none of those arguments against the passage of time or the moving now doctrine go against the very idea of an objective direction of time. Neither do they deny its logical coherence or metaphysical plausibility. As mentioned before, even though it is true that the notion of an objective time passage *implies* the idea of an objective directionality, the converse does not follow: it is not true that the existence of a direction of time implies an objective becoming, or the moving-now doctrine, or any dynamical theory of time (i.e., A-theory). Generally, those doctrines require more structure than that needed to hold the existence of an objective directionality. So, it will be useful to be precise about how much structure is required for time to have a directionality. I will get into this next.

On the basis of the above, it follows that an objective direction of time is compatible with non-dynamic theories of time. To see this, let us come back to McTaggart's work. In his 1908 paper, McTaggart also introduced a third series, the C-series. Whereas the relevant distinction between the A-series and B-series is that the former implies a *dynamic* of time and the latter does not, the difference between the B-series and the C-series is that a C-series lacks a *privileged direction* (see Farr 2012). In McTaggart's words: "The C-series, while it determines the order, does not determine the direction" (McTaggart 1908: 462). The three series then involve a different structure of time because they equip time with either metaphysically highly-structured or lowly-structured features. Let us systematize this.

Suppose the following series of instants or events localized in time: $A - B - C - D$. According to A-theory's supporters such a series is dynamic in the sense that each event has the absolute property of being 'the past', 'the present', or 'the future'; and that *absolute* properties' changing is what constitutes the temporal dynamic. The series has a direction too, since A-properties change only in one direction, from being the past to being the future, through being the present. Obviously, an order is implied by the directionality. Then,

A-theory's structure: order + direction + dynamics ($A^{\rightarrow} \rightarrow B^{\rightarrow} \rightarrow C^{\rightarrow} \rightarrow D^{\rightarrow}$)

Where '$N^{\rightarrow}$' stands for an event instantiating temporal absolute properties' changing, and '$\rightarrow$' stands for a privileged direction of time.

As mentioned before, the B-series is not dynamic in this sense: there are not genuine absolute properties such as being past or being future. Rather, A-properties are actually reducible to B-properties, so that events localized in time may be ordered according to 'earlier than', 'later than' or 'simultaneous with' relations. B-theory's supporters give up equipping time with a dynamics but they do consider that time has a *direction* and an *order*

B-theory's structure: order + direction ($A \rightarrow B \rightarrow C \rightarrow D$)

Even though there is no objective passage of time within a B-theoretic framework, there are objective facts conforming to which time has an order and a privileged direction. In this sense, the B-series running as $A \rightarrow B \rightarrow C \rightarrow D$ would be different from the series $D \rightarrow C \rightarrow B \rightarrow A$. One could well say that they are completely different series to the extent that the relational properties 'being later than' or 'being earlier than' are instantiated by different events or times in each case.

Finally, the minimal structure one can go with is that of the C-theory. Within a C-series only order is preserved since there would be no matter of fact that distinguished the $A - B - C - D$ from $D - C - B - A$. One can determine that some event or time, say B, is between others, A and C, but not whether there is a direction: sentences like "after B follows C" are meaningless in a C-theory framework. Therefore, a C-theoretic framework only confers a "*betweeness* ordering" on time's structure.

C-theory's structure: order ($A - B - C - D$)

The quarrel around the direction of time thus concerns whether time comes metaphysically equipped with at least a B-structure. Physics comes in when such a structure can be reflected in any of our best physical theories, offering formally and empirically-based evidence for either option. So, after this brief detour, we have reached a first approximation of what the problem is (from a general perspective): *Whether* time, according to physics, has at least a B-structure; or, equivalently, whether the past-to-future and the future-to-past direction can be *physically* distinguished.

How should one characterize such a distinction? What does it imply metaphysically? Let us take again the series $A - B - C - D$. In order to know whether that series implies a *direction*, one has to be capable of distinguishing the $A - B - C - D$ series from the $D - C - B - A$. If for any matter of fact there is an asymmetry between the two series, in the sense that $A - B - C - D \neq D - C - B - A$, one is allowed to equip the series with further structure: the series running $A \rightarrow B \rightarrow C \rightarrow D$ is different from the series running $D \rightarrow C \rightarrow$

$B \rightarrow A$. And this difference (this *a*symmetry) is what entitled us to ascribe it a direction: Events or times Bs are followed by events or times Cs, not by As. Conversely, if a symmetry between $A \rightarrow B \rightarrow C \rightarrow D$ and $D \rightarrow C \rightarrow B \rightarrow A$ holds ($A \rightarrow B \rightarrow C \rightarrow D = D \rightarrow C \rightarrow B \rightarrow A$), then both series are *identical* in respect of their directionality, that is, they are said to be equivalent: Events or times Bs may equally be followed by either events or times As or Cs. Nothing in the series' structure allows one to draw a distinction between both directions. The series is therefore *undirected*, obtaining in consequence a C-series' structure, $A - B - C - D$.

But what does it mean that both series are identical? This cannot mean that they are *strictly identical* for there would not be two series any longer in that case. Hence, one should preserve the idea that there are still *two* directions (specifically, two *opposite* directions), though identical in *regarding some relevant features*. This primarily means that both series must share some relevant properties, but not others. This reasoning follows from the well-known *Principle of Identity of Indiscernibles* (PII hereafter):

PII Necessarily, for all particulars *x* and *y*, and every universal *Z*, if *Z* is instantiated by *x* if and only if *Z* is also instantiated by *y*, then *x* is numerically identical with *y*.

There are some quarrels about whether PII is universally valid or whether there are sound convincing counterexamples (see Black 1952, Ayer 1954, Hacking 1975, Rodriguez Pereira 2004, Forrest 2016: Section 3). There is also some controversy about what kind of properties should be taken into account so as to get a meaningful, non-trivial version of PII (see Rodriguez Pereira 2006). Notwithstanding these worries, in what follows I will assume that the principle is well grounded in this general sense: the best criterion one has at hand to count and (minimally) distinguish two particulars is by considering whether such particulars share certain set of *relevant properties.*

Metaphysically speaking, the realm of properties is quite plentiful, and I cannot get into details here about its complex geography. But a useful distinction to be drawn is that of *intrinsic* and *extrinsic* properties. In general, entities have some properties purely in virtue of what they are in themselves, that is, without any relation or interaction with anything exterior. But entities also instantiate other sort of properties in virtue of their relations or interactions with external entities. The former sort of properties is generally called "intrinsic", whereas the latter "extrinsic" (also, though arguable, "relational"). Some physical properties, such as having *x* mass, or positive energy are archetypically considered intrinsic properties. Others, such as "being four kilometers away from Rome" or "being taller than my sister", are typically regarded "extrinsic (or relational) properties".

In order to distinguish the above-mentioned series, one should be able to identify the set of *shared* properties (in virtue of which one can say that they are *identical*), and the set of non-shared properties (in virtue of which one can claim that there are yet *two* series with an opposite direction). It is worth bearing in mind that one is mainly interested in identifying an asymmetry between them –asymmetry on which the directions of both series can be based. What I claim is that such a distinction can be drawn in two different ways.

One promissory strategy to do the trick is relying on an *external fact Z* that breaks the identity between the two series. Even though both series are identical with respect to their intrinsic properties (that is, nothing in the series in themselves allows distinguishing them), they are different *in relation to Z*. Therefore, both series pick opposite directions in virtue of an external fact *Z*. Now, it is matter of empirical research to establish which one is, for instance, the series one lives in.

Despite having grounded a distinction, it is quite clear that such a distinction is not intrinsic (indeed, one assumed that both series shared their intrinsic properties). I will call non-intrinsic distinctions '*non-structural'*, in the sense they are not built-in the series. In some sense, the series in themselves would be symmetric and undirected[4]. And yet, one could come to objectively distinguish them for their relation to an external fact *Z* (naturally, if fact *Z* were absent, the distinction could not be carried out). The asymmetry and the direction are clearly introduced by *Z* in an otherwise completely symmetric and undirected scenario.

Non-structural distinction Two particulars (i.e. temporal series) are *non-structurally different* when they share all their intrinsic properties but instantiate unalike extrinsic properties (for instance, they have an unalike relation to an external fact *Z*)

As one here typically deals with physical evolutions ruled by dynamical laws (solutions of equations of motion), boundary conditions are, for instance, the kind of external facts that may introduce in physics a non-structural difference between two perfectly temporal symmetric evolutions. I will get into details about it later on.

[4] Without any reference to a hypothetical external fact *Z*, the series is symmetric since $A \rightarrow B \rightarrow C \rightarrow D$ and $D \rightarrow C \rightarrow B \rightarrow A$ are (in some relevant sense) equivalent.

Conversely, any *structural* distinction between both series' directions will completely come from their *intrinsic* properties, regardless what sort of relations they have with any external fact. The series running $A \rightarrow B \rightarrow C \rightarrow D$ is structurally different from the series running $D \rightarrow C \rightarrow B \rightarrow A$ in virtue of their having distinct *intrinsic properties.*

Structural distinction Two particulars (i.e. temporal series) are *structurally different* when they instantiate at least one unalike intrinsic property.

The idea is that something in the inner nature of events or times is what allows the distinction to be drawn. As mentioned in passing before, in so far as one is dealing with physical evolutions ruled by dynamical laws, something inherent to dynamical laws would be the perfect candidate to do the job of establishing a structural distinction.

Section 3. The direction of…*what*?

I have so far remained silent about the nature of the constituents of the series, or about *what is* being ordered and directed by the series. After all, "having a direction" or "being asymmetric" is a property of the series, but the series is just a representation of *something*. Since our main concern is time within physics, the constituents of the series should fall under physics' description range.

The terminology immediately suggests that the issue is about *time in itself*. If physical events are temporally directed, it is because time in itself is directed. So, when one speaks of "the direction *of* time" or the "asymmetry *of* time", the of-genitive is understood literally. In this sense, event's behavior would supervene upon time's properties. This might however be taken differently. By "the direction of time" or "asymmetry of time", literature has often understood the direction of things *in time,* or the asymmetry of things *in time.* To start with, one should aware of being involving two different issues under the same label. Secondly, in order to avoid any mistake, both issues need be sharply distinguished. And finally, that any eventual relation or identification calls for further justifications. To my mind, the most straightforward way to conceptually carry out the distinction between an asymmetry *of time* and the asymmetry of *things in time* is through a spatial and material analogy.

Think of the following scenario. You come into the room and find a billiard table inside. Imagine that most the billiard balls concentrate at one of corners, surrounding one of the holes. The rest are asymmetrically spread out all over the table. After looking at the properties of the billiard table (for example, checking whether there an inclination, or an odd geometry on the surface), you get to the conclusion that the billiard table is perfectly symmetric, so such strange asymmetric distribution becomes puzzling. Any explanation of such an odd distribution must, you deduce, rely on some external facts that *caused* it. Maybe an external agent willingly arranged the balls that wat before you enter the room. Or, maybe, there is an attractor at one of the holes of the billiard table that pulls the balls towards it at a specific rate according to their distance from the hole. Had you put this scenario to move in time, you would probably see the billiard balls flowing towards one of the holes.

Whatever the details of the explanation you come up with, the very properties of the table are generally left out in the sense that none of them would explain the distribution: the billiard table, you learned, is perfectly symmetric. So, any question as to the asymmetric distribution of billiard balls on the table is strictly a question about its causes. Clearly, it would be a mistake to claim that, after all, the table is asymmetric *due to* the asymmetric distribution of balls over it. Or, at least, further arguments should be provided to support such an identification. If you split the table in two, you would find an asymmetric distribution of balls in both halves, but no difference at all as to their intrinsic properties. Both halves are actually different (in the sense that you have the means to draw a distinction between both halves), though *non-structurally* different.

Now think of a slightly different scenario. You enter the room and see the same curious ball distribution over the billiard table, except this time you find that the table's surface has many irregularities. After an even more careful examination, those irregularities display a canal-like configuration all over the table. Additionally, you discover that table legs are not perfectly equal long: one of them is shorter than the others, so that the table slightly leans on one of its sides. Imagine that you take all the balls and arrange them differently, for instance, you place them at the opposite corner. With some further assumptions, what you will probably see is the balls finding their way to the opposite extreme. After a while, you end up with a very much similar configuration you start with –most of billiard balls concentrated at one of the billiard table corners, and the rest asymmetrically distributed all over the table. No matter how you set up the initial configuration, you will end up with a pretty similar final distribution. In this case, what rather explains such a distribution is exclusively the intrinsic properties of the table, and not any extrinsic factor. Furthermore, when you split the table in halves, you will note that both halves instantiate different intrinsic properties, regardless of the billiard balls on them. They are thereby *structurally* different. And any asymmetric distribution of the table's content will supervene upon the table's intrinsic properties.

Despite time and space having substantive differences, both scenarios shed some light on the difference between the direction *of* time and the direction of things *in time*. Even though one typically starts with a temporally asymmetric sequence of events, there would be two quite different explanations for it: the sequence is temporally directed either because time in itself is structurally directed (in virtue of its intrinsic properties), or because some external fact provides an explanation and causes a directed-like behavior (that is, because the series is non-structurally directed). In the first case, any temporally asymmetric, or directed, behavior of things in time supervenes upon time's structure, and time is *literally* asymmetric or directed. In the second case, any temporally asymmetric or directed behavior of things in time depends on some external factor, and it occurs *in spite of* time being symmetric. In this sense, one can only speak of "the direction of time" merely in a *metaphorical* manner.

Bearing this distinction in mind, let us consider the temporal series again. The asymmetry or direction *of* time refers to a *structural difference* between $A \rightarrow B \rightarrow C \rightarrow D$ and $D \rightarrow C \rightarrow B \rightarrow A$ series. The asymmetry or direction of things *in* time instead refers to a *non-structural difference* between $A \rightarrow B \rightarrow C \rightarrow D$ and $D \rightarrow C \rightarrow B \rightarrow A$, a distinction relying on an external factor *Z*. It is worth stressing two points here. On the one hand, in both cases there *is* a genuine distinction, and thereby, a genuine B-series. That is, one is *capable of* distinguishing both series for there is at least some non-shared property. A structural identity (non-structural difference) does not entail that there is not any difference whatsoever, but it cannot be carried out in virtue of the series' intrinsic properties. But, on the other hand, it is nonetheless clear that there are two quite different worries here, leading to different issues. Thus, one might introduce under the label of "the problem of the arrow of time" two distinguishable worries.

A structural worry Whether time in itself has a structural difference between the past-to-future and the future-to-past direction. This worry is literally about the asymmetry or direction of time in itself, upon which any temporal asymmetric distribution of things in time supervenes.

Or alternatively,

A non-structural worry Whether a temporally asymmetric distribution or temporally directed behavior of things in time can be explained by a non-structural difference *given that time is structurally symmetric and undirected*. As opposed to the first issue, the of-genitive here is merely metaphorical.

Whilst the former concerns a structural distinction mainly involving intrinsic properties, the latter exclusively relates to a non-structural distinction relying on relational properties (that is, in relation to an external fact that breaks the intrinsic symmetry). Now, putting all these pieces together, one can identify at least two distinguishable kinds of *arrows* of time, and consequently two different formulations of the problem of the arrow of time:

Structural arrow of time A temporal series has a structural direction when the two temporally-opposed series $A \rightarrow B \rightarrow C \rightarrow D$ and $D \rightarrow C \rightarrow B \rightarrow A$ can be distinguished by means of a structural distinction.

And,

Non-structural arrow of time A temporal series has a non-structural direction when the two temporally-opposed series $A \rightarrow B \rightarrow C \rightarrow D$ and $D \rightarrow C \rightarrow B \rightarrow A$ can be distinguished by means of a non-structural distinction.

Therefore, that a temporal series be said to exhibit a privileged direction may receive two metaphysically different, though closely-related, meanings: either it exhibits a structural distinction between both directions, or it exhibits a non-structural one. This will naturally have consequences when physics comes into play: different formulations of the arrow of time depend on these distinctions. But before to come to it, I shall get into some details about other useful metaphysical distinctions.

Section 4. Relationalism, substantivalism and supersubstantivalism with respect to time

This section focuses exclusively on the elements of the series in themselves. Elements of the temporal series A, B, C, D were hitherto referred to as 'times' or 'events'. However, the nature of the series may be ontologically interpreted in different ways: when referring to the direction of time, one might be referring either to the direction of *a substan*ce existing independently of its material content, or to the direction of certain temporal relations among the constituents of *the material content* of the world. These alternatives embody completely different metaphysical views with respect to time; views have been played out in the debate among *relationalism* versus *substantivalism* versus *supersubstantivalism.* Literature about the topic is massive, so I will just present the big picture and how it relates to the arrow of time debate.

Substantivalists with respect to time claim that time is an entity that exists independently of events and things placed within it. Additionally, they claim that world's content's temporal behavior supervenes upon time. *Relationalists*, in turn, support the idea that time is nothing over and above temporal relations among events and things. On the contrary, supersubstantivalists claim that there are no events or things in the world, but (space)-temporal points only. The substantivalist ontology features time and matter (whose behavior supervenes upon time), while the relationalist ontology features only things or events, and the supersubtantivalist ontology only (space)-time points. Let us spell them out in tandem.

There are many different sorts of relationalist-like views in metaphysics and philosophy of physics that, in general, share the idea that time is nothing over and above temporal relations among events and things (Benovsky 2010: 492), though they can greatly vary on which is to be considered as objective and fundamental in the physical world (see Sklar 1974, Earman 1989, Pooley 2013, for comprehensive overviews of the different kinds of relationalisms). Furthermore, they can also diverge on how robust the temporal structure (boiled down to relations among things or events) should be. For instance, Julian Barbour and Bruno Bertotti (1982) have argued for a Machian relationalism in physics where an absolute temporal-ordering structure is assumed for classical mechanics (see Gryb and Thebault 2016 for a defense of a Machian-moderate relationalism in quantum gravity). Carlo Rovelli (2002, 2004) has instead argued for a radical relationalism according to which there is not even a fundamental time-ordering structure in quantum gravity. A more robust relationalism is defended by Michael Esfeld et al. as change not only exhibits a temporal order but also a (primitive) direction (Esfeld et al. 2018: 31).

In any case, I will henceforth take a relationalist-like view on time as supporting two theses:

R1 There are only events or physical bodies in the world (which can have intrinsic properties or not), and their (space) temporal relations. There is no external time.

R2 Time is nothing but change. The sort of relation between the physical world and the concept of 'time' is that of a *Leibnizian representation* or a *Machian abstraction*: time is an ideal, unreal entity parasitic on events-things' changing.

According to these tenets, the variable t occurring in most physical theories (setting aside general relativity) is merely an external and unreal parameter, which should not be taken as representing anything with physical meaning.

As in the case of relationalism, there are also many versions of substantivalist-like views on time (see Sklar 1974, Earman 1989, Pooley 2013 for comprehensive overviews of different kinds of substantivalism). Tim Maudlin has famously supported a substantivalist view on space and time (particularly, a Galilean or Neo-Newtonian space-time, see Maudlin 1993), wherein the direction of time is *intrinsic* to space-time itself (see Maudlin 2002: 259). A *sophisticated 'anti-haccceitist' substantivalism* has been defended by Brighouse (1994), Carl Hoefer (1996), Adam Caulton and Jeremy Butterfield (2012), among others, in the context of general relativity. Furthermore, scientific literature typically accounts for time in classical mechanics, special relativity, relativistic and non-relativistic quantum mechanics, and even string theory as a "parameter presumed by, and hence independent of, dynamics" (Huggett, Vistarini and Wüthrich 2012: 242). These theories are all background dependent in the sense that posit a space-temporal structure that lies outside the scope of dynamics[5].

I will thus take substantivalism as the position supporting the following two theses:

S1 Time is a theoretical entity endowed with a structure that is *intrinsic* to it, and independent of change. Temporal relations among events or things are parasitic on this theoretical entity.

S2 Time is not an ideal or representational notion, but it plays a physically meaningful role in explaining different phenomena or in defining dynamical variables. Time cannot thereby be eliminated or boiled down to a dynamical basis.

Naturally, **S1** and **S2** do not entail any commitment with an observable or measurable entity, but regard that time plays a physical role in physical theories that must be considered to be prior to change, and independent of it. The structure of time is fixed absolutely, irrespective of changes in the world, and thus the structure of change supervenes on the structure of time. In this deflated sense, time can be said to be "substantival". Notably, substantivalists do not claim that time is all is out there, denying the existence of chairs, particles and human beings. Substantivalist's ontology is dual: time *and* matter (in a broad sense) exist in the world. Relationalists, instead, support a monistic ontology where there is only matter (with its spatial-temporal relations). In this sense, some have argued that the real opposite extreme to relationalism is not substantivalism, but *supersubstantivalism*.

[5] Certainly, I would be naïve to stem from these theoretical considerations that one must engage a substantivalist metaphysics of space-time: the structure may be otiose and, from a 'more parsimonious' stance in respect of ontology, eliminable. However, it is also true that standard formulations of those physical theories do countenance a substantivalist-like viewpoint.

In general, supersubstantivalism claims that "everything in the world is space-time" (Lehmkuhl 2016: 1). However, such a core claim can receive two readings: one of them states that space-time is the only substance in the world, in the sense of being the only *fundamental* entity of the world. The other one states that space-time is the only *kind* of substance in the world. In the latter, parts of space-time can also be proper substances. One might alternatively claim, in a non-relativistic universe, that space *and* time are the only kind of substances: space (extension) and time (duration). It has been claimed that Descartes and Spinoza held a supersubstantivalist stance with respect to space, and I do not see any reason why this doctrine cannot be also applied uniquely to time. In any case, supersubstantivalism's main tenet is that matter does not belong to the ontology of the world (for discussion about the metaphysics of supersubstantivalism, see Lewis 1986, Sider 2001, and Schaffer 2009. In a relativist framework, see Lehmkuhl 2016).

The relation between space-time and matter is a bit murky within the supersubstantivalist framework. Some have claimed that the relation is one of *identity* (Schaffner 2009), which leads to the identification of matter and space-time. Others, one of *ontological priority*: space-time (or space *and* time) is (are) *ontologically prior* to matter in the sense that the existence of space-time (or space and time) implies the existence of matter, but not the other way around. So, the core claim should be slightly modified: everything *fundamental* in the world is space-time (space *and* time). Be that as it may, different versions of supersubstantivalism have been discussed in the last decades, but to keep thing simple I will take it as holding the following two core claims:

SS1 There is only space-time in the world as fundamental entity. Alternatively, space *and* time are the only kind of fundamental substances in the world

SS2 Space-time (space *and* time) is (are) ontologically prior to matter, in the sense that the existence of the latter ontologically depends on the existence of the former.

The lesson in the previous sections is that, when asking for *the direction* of time, one could be asking in, at least, two senses*:* structural and non-structural. Now, one additional lesson should be taken as to the direction of *time:* Is one actually referring to the direction of the material content of the world, being time an abstract representation of relations among matter or events? Or is one instead minding whether time, as a fundamental and ontologically prior substance, has a privileged direction regardless of the behavior of matter? Does the behavior of matter's dynamic supervene upon time's structure, or is it rather the other way around? The upshot is that different metaphysical stances with respect to the nature of time will crucially hinge on how the problem of the arrow of time is to be addressed in physics. Now, the landscape greatly expands and gets more complex: relationalist, substantivalist and supersubstantivalist stances might significantly disagree on how the notion of structure should be fleshed out –the structure of *what*? Fundamental particle's behavior? Time's structure in itself? As I will show in following chapters, these positions (largely, and surprisingly, overlooked in the literature on the arrow of time) have strongly influenced the understanding of highly-relevant notions in the literature –for instance, and remarkably, the notion of time reversal.

***

Through Section 1 and Section 4, I have delved into the problem of the arrow of time from a wide metaphysical point of view. It was shown that the problem, metaphysically, is *to distinguish* between the past-to-future and the future-to-past directions. Or, to put it differently, to investigate whether or not both directions turn out to be *equivalent* in some respect. Next, I've also shown that one could, metaphysically, be worried about such a distinction in two different ways: either in a *structural* way, that is, in searching for some intrinsic property that distinguishes both temporal directions, or in a *non-structural* way, that is, in drawing the distinction in term of an extrinsic or relational property. Importantly, both ways are useful to the distinction between both temporal directions, though they draw it differently.

This has naturally led us to a twofold understanding of arrow of time in physics: a structural arrow or a non-structural one. Scenario gains complexity when previously-assumed commitments with respect to what time is come into play. Hence, by the 'arrow of time' one might be referring to a varied pack of closely-related, though differentiable, ideas. Physics enters the scene when all these considerations are reviewed in the light of our physical theories. Particularly, the question now is whether physics offers a non-arbitrary criterion whereby the past-to-future and the future-to-past directions of time were distinguished (or shown to be *physically* equivalent). And this enterprise will be shown to be non-univocal either. I shall get into this in Chapter 2, showing that there are actually two stories to be told about the arrow of time in physics, stories that instantiates the metaphysical distinctions introduced thus far. But, before getting into this, I shall introduce some physics on what has been said up to this point along with some additional distinctions.

Section 5. Physics comes in…

If the problem of the arrow of time is metaphysically about how the two directions of time can be distinguished (or showed to be equivalent), on what should such a distinction (or equivalence) be *physically* grounded? We commonly think that our physical theories are more than mere 'cook recipes': they purport to describe the natural world in an approximately true manner. In this respect, if one comes to think that time's direction should somehow be part of, or be reflected in, the natural world, then it is reasonably to suppose that physics is particularly qualified to comment on its nature and properties. Lawrence Sklar (1974) puts it in this way:

> "The usual line of attack is to try to find *some feature* of the *material happenings* of the world which coordinates in one way or another with temporal priority. That is, to find for any two events that are such that one is earlier than the other some other feature which relates them in an *asymmetric way*" (1974: 355. Italics mine)

Sklar's words shed some light on the matter, but not quite enough. What kind of 'feature' should one take into consideration? How should the expression 'material happenings of the world' be interpreted? One could rephrase the quote in terms of the vocabulary employed in previous sections: in addressing the problem of the arrow of time in physics one is worried about whether a series of material happenings of the world $A - B - C - D$ instantiates some physical property that allows ordering them in such a way that the series displays a direction, that is, the material happenings of the world may be univocally ordered according to an 'earlier than…' or 'later than…' relation. In other words, the material happenings of the world can be shown to be represented univocally by the series $A \to B \to C \to D$ and not by the series $D \to C \to B \to A$, having thereby a B-structure. Besides specifying that one is now within the physics realm, by this rewording one has not gone much farther, and is left more or less where one started off at.

Clearly, to start on the right foot, one should begin by identifying the *physical* properties that could do provide a physical foundation for the temporal asymmetry. Sklar, for instance, points out that there are at least two approaches to be followed: a *lawlike* approach and a *de facto* approach (Sklar 1974: 357-358). According to the former, it is matter of a *lawlike* property of the world that the event *A* in the series is prior to the event *B*. Hence, it is matter of a lawlike property that the world has the temporal structure of a B-series, $A \to B \to C \to D$, instead of the opposite, $D \to C \to B \to A$. In turn, the *de facto* approach concerns the identification of an asymmetry in the actual world as a matter of fact, despite the laws governing it allow for a symmetric underlying scenario. According to the de facto approach, there is no lawlike property that distinguishes between the two series; however, a physical theory might offer the means to distinguish them as a matter of fact, as a matter of a non-intrinsic, extra-nomic property. Although both $A \to B \to C \to D$ and $D \to C \to B \to A$ would be equally possible according to physical laws, they are nonetheless different in virtue of a (de facto) external property.

Paul Horwich (1987) has also offered a systematic investigation of time asymmetry (or time anisotropy, in his vocabulary). He begins by pointing that the difference between the two directions of time should be drawn in terms of properties of elements (or of the series). When dealing with objects, the sort of asymmetry one looks for would be that of “asymmetric geometrical object”, for instance, a cylinder squeezing (1987: 39). How can a similar reasoning be applied to time? For any given series of physical events, he proposes three types of properties as candidates:

(a) Particular properties
(b) General properties
(c) Nomological properties

As to the first type, the asymmetry is introduced by claiming that the series $A \to B \to C \to D$ instantiates the particular property of “the future direction goes from *A* to *D*”. This particular property is obviously not instantiated by the series $D \to C \to B \to A$. Hence, there is a (particular) non-shared property that distinguishes the two series, and an asymmetry is so introduced. Despite its effectivity, it looks very much like an ad hoc resource: it needs to rely on properties referring to particulars (particular events *A* and *D*), and only instantiated by *that* particular series. Also, for foundational purposes, thus-introduced asymmetry seems to be based on a by-pointing procedure.

The second type of properties avoids referring to particulars: “All events of a certain kind (all of As) are followed by events of the other kind (all of Bs)”. This clearly introduces a more general and far reaching asymmetry, but it is not substantially different from the previous one. Now, there would be a general property that a direction of time could instantiate that allows ordering *kinds* of events conforming to an asymmetric relation (‘being earlier than…’ or ‘being later than…’). This however does not bypass the fact that there is a temporally-inverted series of events of a certain kind that is also *possible* and turns out, at least in principle, identical to the original one. Horwich adds that a so-defined general property may establish an arrow of time in terms of, for instance, initial conditions for a given series of events of a certain kind (1987: 41).

Finally, the last type of property is the nomological. In this case, the direction of time is directly related to *time-asymmetric* laws of nature. In Horwich's words: "we have here a property that applies as a matter of physical necessity to one direction and not to the other" (1987: 41). Two noteworthy remarks on this. First, one of the series, say $A \rightarrow B \rightarrow C \rightarrow D$, holds by matter of the laws of physics. It is then no longer the case that both series are physically possible, but just one of the series is: Whereas particular or general properties allow distinguishing the two series (i.e. time's directions), never do they make one of the series (physically) *impossible*. Thus, the distinction between the two series is introduced as a matter of *physical necessity* in this case. Second, Horwich claims that a nomological property introduces an asymmetry in *time itself.* Particularly, since the distinction is matter of a law of nature and of physical necessity, it allows one *to infer* an intrinsic difference between the series $A \rightarrow B \rightarrow C \rightarrow D$ and the series $D \rightarrow C \rightarrow B \rightarrow A$ that can only depend on the series' structure.

Sklar's and Horwich's distinctions have many points in common in spelling out what kinds of features are to be sought out for a well-grounded physical distinction between a series like $A \rightarrow B \rightarrow C \rightarrow D$) and one like $D \rightarrow C \rightarrow B \rightarrow A$. Sklar's lawlike approach is clearly quite similar to Horwich's nomological properties. The de facto approach can be analogously related to particular and general properties in Horwich's sense. Yet, despite having gained some clarity on the matter, it is still unclear how these categories should be systematically applied in physics. Overall, the laws of physics, and some features thereof, seem to play a relevant role in establishing a direction of time. But it is true that, when drawing a distinction between both time's direction, other extra-nomic properties can make a good job as well. Additionally, it is worth noticing nonetheless that the laws of physics and extra-nomic resources, though useful for the pursued aim, might render a different physical foundation of the arrow of time. I will spell this out shortly, but it's already evident that laws of nature seem to play a *prior* theoretical role in grounding an arrow of time in physics, whereas extra-nomic properties also do it but in a *derivative* manner.

I shall in general follow Sklar's and Horwich's spirit of identifying the relevant features for an arrow of time in terms either of laws of nature or of extra-nomic properties. But, which features of physical laws should one look at? And which extra-nomic properties are the relevant ones for an arrow of time? In the next section, I shall answer these questions. In the philosophy of physics and physics literature, time-related issues are typically introduced in terms of *time*-

reversal invariance and/or *reversibility*. This has been so for a longstanding tradition in the field that has regarded (ir)reversibility and (non)time-reversal invariance as conceptually correlated, or and even identical, notions. In one way or another, issues around the problem of the arrow of time has been revolving around these concepts virtually ever since its birth. In the following, I will make the case that a good source of the confusions gravitating around has come from *mixing* (ir)reversibility and (non)time-reversal invariance. A good deal of clarity, I hope, will be provided by disentangling both concepts.

Section 6. Disentangling time reversal and (ir)reversibility

Section 6.1. (Ir)reversibility and its many faces

Quarrels around the problem of the arrow of time in physics have largely involved melting ices, gases spreading out in a box, and hot objects getting colder. Basically, discussions have included *thermodynamic processes*, since thermodynamics is the physical theory that describes most of the usual (and alleged) time-asymmetric behaviors one sees in the world. As macroscopic physical systems spontaneously evolve toward equilibrium states, but they never do so away from equilibrium states, thermodynamics seems to offer a suitable criterion to distinguish both directions of time: future-directed evolutions will be those reaching equilibrium states and fleeing away from the non-equilibrium. This sort of processes has been called "irreversible" and it has to a good extent been an essential mark of time asymmetry.

But what does it mean that a physical system be *reversible*? To begin with, it is worth mentioning that there are several senses of reversibility floating around. Reasons for that likely come from the fact that the notion appears in different theoretical contexts with slightly different meanings and mathematical backgrounds. In general, *processes* or *physical* systems are the sort of things said to be *reversible* or *irreversible*, and they significantly vary across physical theories, so the very characterization of the notion will also probably vary accordingly. And even though the relation between statistical classical mechanics and thermodynamics seems to offer the most relevant features of what reversibility means, there is no univocal meaning in there either. Leaving aside any relation to the arrow of time for now, two main senses of (ir)reversibility can be identified within the thermodynamics literature:

(1) (ir)reversibility in terms of *(ir)recoverability*
(2) (ir)reversibility in terms of *equilibrium states*

Based on the work of Sadi Carnot (1824), the main contributions towards a unified theory of heat and its transmission were those made by William Thomson (alias Lord Kelvin) and by Rudolph Clausius. I shall not get into historical details about the development of thermodynamics, but the core meaning of what reversibility is and its senses (in terms of 1 and 2) are already present in their works, so some references will be helpful here. In general, these different senses of (ir)reversibility relate to different interpretations of the second law of thermodynamics, which in its original formulation states the impossibility of performing cyclic processes as heat does not flow from lower to higher temperatures.

Let us start with (ir)reversibility in terms of (ir)recoverability (1). This meaning of irreversibility intends to characterize a panoply of processes that, starting from an initial state S_i and ending with a final state S_f, cannot be *undone*. More specifically, given a process P and its transition from the state S_i to S_f, there is no process P' that restores the initial state S_i completely. In an 1852 paper, Lord Kelvin proposed this notion of irrecoverability to make sense to his view of the second law of thermodynamics in terms of a "universal tendency in nature to the dissipation of mechanical energy". In expounding this 'tendency', he stated that a reversible process implies that the mechanical energy may be *restored* to its primitive condition, and that "when heat is created by an unreversible process (such as friction) there is a dissipation of mechanical energy, and a full *restoration* of it to its primitive condition [is] impossible" (Kelvin 1852).

Max Plank has also rested upon this sense of irreversibility to explain the second law of thermodynamics:

> "a process which can in no way be completely undone is called 'irreversible', all other processes 'reversible'[6]. In order for a process to be irreversible, it is not sufficient that it does not reverse by itself, –this is also the case for many mechanical processes, which are not irreversible–, rather it is demanded that, once the process has taken place, there is no means (…) of restoring exactly the complete initial state" (Plank 1897: 112)

In essence, irreversibility in terms of irrecoverability stablishes *the impossibility* to get back the initial state S_i once the process has taken place completely. More specifically,

Recoverability Given a process P producing the transition (or the change of states) from its initial state S_i to S_f, P is reversible if and only if there is another

[6] Plank employs the German words "*irreversibel*" and "*reversibel*".

> possible process P' which produces the transition from S_f to S_i. Otherwise, P is irreversible.

It is worth noticing a couple of things in this definition. First, as Planck himself noted (1897: 109), this sense of reversibility does not imply that the process P' must trace back step by step the original process P in reverse order, going exactly through the same original states between S_i and S_f. Second, by the word 'completely' is meant that not only must the system come back to its original state, but also all those bodies that have interacted with the system during the process P.

However, this sense of (ir)reversibility in terms of (ir)recoverability is rarely found in the modern literature on thermodynamics and reversibility. As Jos Uffink points out, the meaning of reversibility in terms of equilibrium states (2) "seems to be the most common meaning of the term" (Uffink 2001:317). Modern literature considers that recoverability is rather a subsidiary notion for it can be fully obtained from reversibility in terms of equilibrium states (see Bridgman 1941: 122). Overall, this meaning of reversibility "denote processes which proceed so delicately and slowly that the system remains close to an equilibrium state during the entire process" (Uffink 2001: 317). Rudolph Clausius (1864) seems to have endorsed this meaning of reversibility when claiming that

> "When a change of arrangement takes place in such a way that force and counterforce are equal, the change can take place in the reverse direction also under the influence of the same forces. But if a change takes place in such a way that the overcoming force is greater than that which is overcome, the transformation cannot take place in the opposite direction under the influence of the same forces. We may say that the transformation has occurred in the first case in a *reversible* manner, and in the second case in an *irreversible* manner" (1864: 251)[7]

Some light is shed on the definition of reversibility by explaining that it is reached at *a limit*:

> "Strictly speaking, the overcoming force must always be more powerful than the force which it overcomes; but as the excess of force is not required to have any assignable value, we may think of it as becoming continually smaller and smaller, so that its value may approach to nought as nearly as we please. Hence it may be seen that the case in which the transformation takes place reversibly is a limit which in reality is never reached but to which we can approach as nearly as we please" (ibidem).

Under this sense of reversibility, a reversible process is no longer one in which one can by any means come back to the initial state through a process P', but one that may be reached by a

[7] Original in German. English translation in Uffink 2001, p. 325, fn. 35.

series of processes in which the divergence from the equilibrium state becomes infinitely smaller. Nowadays, this definition of reversibility typically names quasi-static or adiabatic processes.

Reversibility Given a process P producing the transition (or the change of states) from its initial state S_i to S_f, P is reversible if and only if the transition from one state to the other involves an infinitely small negligible divergence from the equilibrium. If that divergence is non-negligible, then the process is irreversible.

There are several obscurities in this sense of reversibility. In an already classical paper, Ernst Rechel (1947) denounces that the understanding of reversibility in the literature is deeply misguided, involving an overlooked deadly paradox: reversible processes are supposed to be those in which changes produce "infinitely slowly", so that the system still remains in equilibrium, though it somehow changes (in virtue of these infinitely small changes). Remarkably, it is supposed that a "infinitely small" disequilibrium is so small that does not matter, but that matters enough to secure that reversible processes are still *processes*; that is, that they *change*, reversibly, in time. The paradox is evident: reversible processes require being somehow in non-equilibrium states and in equilibrium states at the same time.

Rechel claims that this understanding of reversible processes "unfortunately contains a contradiction within itself" (1947: 301), and they should not be even considered theoretically possible, risking the very fundamental concepts of mechanics. John Norton (2016) has put forward a way out for this paradox, defining reversible processes differently. He claims that the label 'thermodynamically reversible process' denotes a set of *irreversible processes* in a thermal system, delimited by the set of equilibrium states in the limit of net driving forces going to zero. The punch line is that the "properties normally associated with a reversible process are recovered from the set of irreversible processes through limit operations" (Norton 2016: 44), where the notion of 'reversible process" must be understood in terms of an approximation, not in terms of an idealization (see Norton 2012 for the difference). In other words, the term 'reversible process' does not any longer refer to a theoretical process, possible according to the theory as an idealization, but "to a set of irreversible process", where the properties "normally attributed to thermodynamically reversible processes are really borne as limiting properties of this set". In other words, reversible processes are imperfect descriptions of real irreversible processes.

As one turns to statistical classical mechanics, the notion of reversibility needs to be revised. Now, thermodynamical systems consist of interacting particles and their mechanical properties, whose behavior is analyzed statistically. Thermodynamically reversible processes defined in terms of equilibrium states as limit cannot any longer remain eternally unchanging in a fundamental sense: the presence of thermal statistical fluctuations may take equilibrium states to non-equilibrium states spontaneously. There is naturally a difference between macroscopic and microscopic systems in this respect: whereas fluctuations are negligible for the formers, they are significantly relevant for the microscopic level. The presence of thermal fluctuations makes impossible any durable equilibrium state, and thereby, any reversible processes at the micro-level. Typically, reversible processes are now said to have pseudo-equilibrium states *at the limit* (Norton 2016).

John Earman (1986a) distinguishes four senses of reversibility that are most directly relevant to statistical classical mechanics and thermodynamics. The first is “the approach to equilibrium in the sense of the approach to the micro-canonical distribution” (1986a: 226), beings the one that captures to great extent what thermodynamically reversible processes are supposed to be when understood in terms of equilibrium states[8]. An isolated system is said to be micro-canonical when its probability density is uniform over the phase space. Hence, the approach of a system to an equilibrium state is its (infinite) approximation to such a uniform distribution. This sense of reversibility is commonly said to imply that *irreversible* processes are those having a limit state it cannot escape from. Along the same line, Olimpia Lombardi and Mario Castagnigno (2008) defines (ir)reversibility in terms of (having) lacking attractors in finite or infinite points of the phase space; this notion of reversibility is generalized by saying that an “evolution is reversible if it has no final (or initial) equilibrium states, namely, a ‘point of no return’” (2008: 75)

As previously mentioned, the notion of (ir)reversibility has been largely tied to that of the direction of time: a reversible evolution (a sequence of states) might happen in the past-to-future direction of time as well as in the future-to-past one. This mainly means that the sequence of its states might happen either as $A \to B \to C \to D$ or $D \to C \to B \to A$. Conversely, an irreversible evolution has a limit state in one direction of time meaning that once the process has taken place, it has already reached a (limit) equilibrium state from which it cannot return to the original state. Note that irreversible evolutions already pick out a direction of time when

[8] The other senses mentioned by Earman relate to finite times and transport coefficients, the approximate validity of the Boltzmann equation, and the asymmetry between past and future (see Earman 1986a: 226)

taking place: when one says that an evolution is irreversible because it has a (limit) final state of equilibrium, one presupposes that the final state lies in the future direction of time. And when one thus says that the irreversible evolution cannot happen in the opposite direction, this means that it cannot happen backward in time, that is, going in the future-to-past direction once the direction had been already fixed[9]. In this sense, an irreversible evolution displays a direction of time and it is with respect to *that* direction (typically the past-to-future one) that the evolution may not happen in the opposite sense of time. It is worth bearing in mind this remark since it will be crucial afterwards.

Many of the conundrums on (ir)reversibility and its relations to the second law of thermodynamics within the framework of statistical classical mechanics follow from the fact that the underlying dynamics (i.e. equation of motion of classical mechanics) is said to be time-reversal invariant, notion that will be introduced in the following sub-section.

Section 6.2. Time-reversal invariance

The other key notion is that of *time-reversal invariance*. The literature about its meaning and implications is abundant, and a good deal of this thesis will deal with the puzzling nature of time reversal in the quantum theory. The seemingly theory-dependent nature of time reversal in physics makes it hard to formulate a single characterization of it, but at least some general things can be said about. Notwithstanding, some fruits can be picked from the tree in order to get some grasp of time-reversal invariance in general and to differentiate it from reversibility.

To begin with, time-reversal invariance is a property of *physical laws*. Usually, physical laws are mathematically represented by (differential) equations of motion, which renders a set of solutions standing for evolutions of physical systems. The term 'time-reversal invariance' expresses, properly, a *formal* feature of an equation of motion, to wit, whether its structure is left unchanged after applying a time-reversal transformation upon it. So, 'time reversal' can be defined as a formal *transformation* acting upon dynamical equations, and 'time-reversal *invariance*' as a property that may be instantiated by those dynamical equations – particularly, the property of being left unchanged after the application of the time-reversal transformation.

[9] Strictly, and from a 'nowhen viewpoint' one should reword this remark as saying that one presupposes that the final state lies *either* in the future direction of time *or* in the past direction. And when it is said that the irreversible evolution cannot happen backward in time, this means that it cannot happen backward (*forward*) for a complete evolution that has already taken place *either* in the past-to-future direction *or* in the future-to-past one.

Yet this characterization is too vague. What does it mean that the equation is left *unchanged*? Plus, what is time reversal supposed to conceptually represent and to do? An intuitive picture of time reversal and time-reversal invariance is typically introduced as following: Suppose one records some process on film, say a full period of a simple harmonic oscillator. Then, one plays the film backward. Does the played-backward sequence look the same according to physics? If it does, the sequence is said to be time-reversal invariant in the sense in which the direction of time seems to play no role in the physical description of the system's behavior. If the inverted sequence does not look the same, then it is non-time-reversal invariant.

Though intuitive and appealing, this characterization of time-reversal is highly confusing, and somehow misguided. To start with, (non)time-reversal invariance should be predicated of the dynamical equation ruling the system's behavior, and not of the system itself. Additionally, what is properly inverted by time reversal is not the particular sequence of states of a given evolution (say, the simple harmonic oscillator recorded in the forward direction of time), but the dynamical equation in itself, that is, parameters, observables and operators that appear in there. But such an equation has *that* evolution as one of its multiple solutions. Time-reversal invariance in consequence has nothing to do with a particular solution of a dynamical equation but with the dynamical equation's structure and with *the class* of solutions allowed by it. In accordance with this, a time-reversal invariant equation of motion will thus render solutions that can be divided into two complementary subclasses, which are symmetric one another regarding the direction of time. This does not mean that all allowed solutions have the same form or features, but that exists a transformation that turns solutions into solutions and non-solutions into non-solutions for either direction of time (or direction of playing the film).

Let us put it more formally. Classical physical theories commonly employ a phase state Γ containing all possible states of a physical system. An instantaneous state is thus represented by a point in Γ. Any physical evolution $\mathcal{E}$ will be represented by a time-parametrized curve in Γ.

$$\mathcal{E} = \{s_t \in \Gamma : t_i \leq t \leq t_f\} \tag{1.1 }$$

The class of evolutions allowed by a physical law $\mathscr{L}$ represents the class of its possible worlds W, that is, those worlds wherein $\mathscr{L}$ is (approximately) true. The question whether or not a dynamical equation $\mathscr{L}$ is time-reversal invariant is the question of whether W can be partitioned into two *symmetric* subclasses, W^f and W^b. If $\mathscr{L}$ is time-reversal invariant, then the time-reversal transformation T maps solutions in W^f onto solutions in W^b. The existence of the bijective map $\mathcal{E} \to^T \mathcal{E}^T$ then secures that W^b includes also physically equivalent possible worlds[10] of $\mathscr{L}$. It follows from this that for any evolution $\mathcal{E} \in W^f$ going forward in time there exists a time-reversed evolution $\mathcal{E}^T \in W^b$ going backward

$$\mathcal{E}^T = \{Ts_{Tt} \in \Gamma : -t_f \leq t_n \leq -t_i\} \tag{1.2 }$$

Note that whether or not this partition is possible will strongly depend on structural features of $\mathscr{L}$: this dual structure of $W = W^f + W^b$ is *produced* by the formal features of the dynamical equation. Mario Castagnino, L. Lara and Olimpia Lombardi (2003), and Castagnino and Lombardi (2009) have called this phenomenon "time-symmetric twins", in the sense that the dynamical equation produces a par of temporally-mirrored solutions, $\mathcal{E} \in W^f$ and $\mathcal{E}^T \in W^b$.

From all this, time-reversal invariance can be broadly defined as:

Time-reversal invariance A dynamical equation $\mathscr{L}$ is *time-reversal* invariant (i.e. invariant under a transformation *T*) iff the class of its allowed solutions can be partitioned into $W^f; W^b$, where if $\mathcal{E} \in W^f$, then $\mathcal{E}^T \in W^b$, the bijective map $\mathcal{E} \to^T \mathcal{E}^T$ exists.

A similar definition can be also introduced by adopting the model-theory vocabulary:

> A law $\mathscr{L}$ is *T*-invariant iff, if m is a model of L, then the time-reversed model $T(m)$ exists and is also a model of L.

This way to pose the meaning of time-reversal invariance is just one among numerous manners to define a symmetry in a physical theory. The above-mentioned definition focuses on whether $\mathscr{L}$ produces time-symmetric twins under the application of *T*. It was mentioned in passing that this is because *T* leaves the *form* of $\mathscr{L}$ *unchanged*: If the form of the equation of motion is left unaltered under the application of the transformation *T*, then it is expected that it also produces solutions when so transformed. In this sense, 'the form of the law' refers to its functional form, expressed in terms of the independent and the dependent variables. The transformation

[10] The notion of "physical equivalence" is too vague. One could say that two models, related by a symmetry transformation, are *representationally equivalent*, in the sense that both are apt to represent a physical situation. In this sense, the models would be equivalent. One could even give a step further and claim that they are then *indistinguishable*. I'm not giving such a step further, and I will just claim that a symmetry yields at least two models that can be used equally to describe the same physical situation (for discussion see Ismael and van Fraassen 2003, Belot 2013, Wallace 2019)

typically leads one to a formally different, though *structurally* equivalent, mathematical expression that also yields solutions. There is much discussion in the literature on symmetries gravitating around what "equivalence" means in this context, what notion of *physical* (if any) is involved here, if a symmetry transformation is expected to do something else than merely leaving the form of the equation unchanged, or what one is supposed to understand by 'the form of the law'. I shall get into some details of these issues afterwards, but, for the meantime, what has been said so far is helpful enough to characterize time-reversal *invariance* in general.

There is no much debate in the philosophy of physics literature on what time-reversal *invariance* means (at least no further debate besides how symmetries in general should be interpreted). Yet, in so much as the time-reversal invariance is a symmetry under an operator that stands for a change of direction of time, a good deal of what *time-reversal* invariance means is encoded in how the action of time reversal is characterized. In general, this is given by the sort of actions that a time-reversal operator is supposed to formally carry out when applied to an equation of motion. The philosophical literature usually sets forth a general time-reversal operator (T henceforth) that is essentially the mapping $T: t \rightarrow -t$. Most people agree on such general characterization. Yet, differences branch out from here on. Should the time-reversal operator feature additional transformations? Does the mapping $T: t \rightarrow -t$ have *any* physical meaning as it stands? Is there a general, all-embracing definition of the time-reversal transformation? Or is the notion rather intrinsically theory-relative? More on this afterwards, too.

How does the notion of time-reversal invariance relate to the direction of time? I will in the next chapter show (Section 2) that time-reversal invariance allows for certain kind of inferences with respect to time's structure framed in the so-known "symmetry-to-reality inference". But let's leave this aside from the moment. In general, a time-reversal invariant law $\mathscr{L}$ produces sequences of states (evolutions) both in the past-to-future direction and in the future-to-past one. If $\mathscr{L}$ rules the transition of states $A \rightarrow B \rightarrow C \rightarrow D$ with $+t$, the time-reversed law $\mathscr{L}^{\mathrm{T}}$ rules a physically equivalent transition $TD \rightarrow TC \rightarrow TB \rightarrow TA$ with $-t$, where TD, TC, TB and TA stand for the time-reversed states corresponding to *D*, *C*, *B*, and *A*, respectively. If $\mathscr{L}$ is time-reversal invariant, and $[A \rightarrow B \rightarrow C \rightarrow D] \in \mathrm{W}^f$, then $[TD \rightarrow TC \rightarrow TB \rightarrow TA] \in \mathrm{W}^b$. Therefore, the class of solutions of $\mathscr{L}$ is equal to $\mathrm{W} = \mathrm{W}^f + \mathrm{W}^b$. The upshot for the arrow of time debate is that the law $\mathscr{L}$ in itself offers no means whereby the past-to-future direction of time can be distinguished from the future-to-past one.: the law is 'blind', so to speak, to the direction of time.

A failure of time-reversal invariance in this sense naturally leads one to the opposite conclusion. Suppose that $\mathscr{L}$ is non-time-reversal invariant. This means that $\mathscr{L}$ does not produce time-symmetric twins, that is, that the class of solutions allowed by $\mathscr{L}$ is *either* $W = W^f$ *or* $W = W^b$, but not both. This means that $\mathscr{L}$ is asymmetric under the transformation: Given two series of states like $A \to B \to C \to D$ with $+t$ and $TD \to TC \to TB \to TA$ with $-t$, $\mathscr{L}$ has uniquely one of the series as solution. Once one of the solutions is taken, the transformation T takes a solution of $\mathscr{L}$ and turns it into a *non*-solution according to $\mathscr{L}$. What this primarily means is that the chosen direction of time matters in the sense of delimitating what kind of evolutions (physically possible worlds) are allowed by the law. If the problem of the arrow of time in general is how temporally-mirrored series of states can be distinguished, non-time-reversal invariance offers a knock-down criterion: only one of the series is *nomologically* allowed.

Section 6.3. Distinction between both concepts

It becomes evident that there are visible differences between time-reversal and reversibility. I'm going to single them out clearly.

To start with, it must be highlighted that (ir)reversibility and (non)time-reversal invariance are properties predicated of different kinds of *items*. Whereas (ir)reversibility is a property that *evolutions* (or their mathematical representation, solutions) instantiate, (non)time-reversal invariance is a property of *physical laws* (or their mathematical representation, equations of motion). In the case of (non)time-reversal invariance, as pointed out by Castagnino and Lombardi, it is *a fortiori* a property of *the class* of solutions of a dynamical equation (Castagnino and Lombardi 2005: 75), in particular, that it can be divided into two symmetric subclasses ($W = W^f + W^b$) as stated before. (Ir)reversibility is rather a property of a *single* physical evolution, and its origin might naturally come from any of the features characterizing *that single* evolution (for instance, boundary conditions).

In the second place, they significantly differ in their *modal* force. The claim of (non)time-reversal invariance is a claim about the class of *possible* worlds of a physical theory. Non-time-reversal invariance hence means that the past-to-future direction (future-to-past direction) is *necessarily*[11] different from the future-to-past direction (past-to-future direction) in the sense that given a world with a fixed direction of time, a physically equivalent possible world with the inverted direction of time is impossible according to a theory's law describing the former.

[11] To be clear, I'm here referring to *physical necessity* as defined by John Earman (1986), that is, what is allowed by the laws of physics of the actual world. Therefore, I'm not talking about logical or metaphysical necessity.

This explains why a claim of non-time-reversal invariance in a theory's laws is modally stronger: It bans any (physically equivalent) possible world with an inversed direction of time. (Ir)reversibility's modal scope is much weaker in as much as (ir)reversibility circumscribes itself to a single complete evolution, i.e., to a single possible world of the theory. Thus, that a single evolution be irreversible says nothing about the rest of the possible evolutions according to a given law, whether they should or not be irreversible. The property of being (ir)reversible need thus not come from an equation of motion but it might easily come from any of the extra-nomic features that characterize a complete single evolution.

Third, one of the most spread notions of reversibility, in terms of equilibrium states, crucially depends on considering it as *a limit* of a series of processes. As briefly mentioned, this may lead to various puzzling issues around the very definition of the term (see Rechel 1947, Norton 2016). Be that as it may, the definition of reversibility, at least in one of its senses, openly relies on an *approximation* or *idealization* upon irreversible processes. Nothing similar happens with (non)time-reversal invariance: as shown, the term is commonly defined within the vocabulary of symmetries and it has nothing to do with particular evolutions, their behaviors or whether they tend to certain states or not.

Though conceptually different, one could come to think that, anyway, both concepts are correlated: Whenever one gets an irreversible evolution, one also gets a non-time-reversal invariant laws underlying, and whenever one gets a reversible evolution, one gets an underlying time-reversal invariant law as well. However, time-reversal invariance and reversibility are not even accidentally or weakly correlated: combination of the four possibilities can be obtained (though some of them are rather unusual). This point has been technically put forward by Aiello et al. (2008: 261-263) (see also Castagnino and Lombardi 2009, fn. 3). Let us take just two cases: one in which there is an irreversible solution, but a time-reversal invariant law describing it, and one in which there is a reversible solution, but a non-time-reversal invariant law.

Following Aiello et al. (2008), an evolution (i.e. a solution of an equation of motion) is said to be reversible when having no attractors in finite or infinite points of the phase space. Consider a pendulum with the following Hamiltonian

$$H = \frac{1}{2}p_\theta^2 + \frac{K^2}{2}cos\theta \qquad (1.3)$$

The equation ruling pendulum's behavior is clearly time-reversal[12] invariant as it produces classes of symmetric solutions with respect to the Θ-axis. Trajectories within the separatrices are reversible to the extent that they are closed curves. But the trajectory corresponding to the (upper) lower separatrix is irreversible since it tends to $\theta = \frac{\pi}{2}$, $p_\theta = 0$ ($\theta = -\frac{\pi}{2}$, $p_\theta = 0$) as $t \to \infty$ ($t \to -\infty$).

Let us now consider the opposite case: a reversible solution and an underlying non-time-reversal invariant law. For this, suppose a modified oscillator with the following Hamiltonian

$$H = \frac{1}{2}p^2 + \frac{1}{2}K(p)^2 q^2 \tag{1.4 }$$

Where $K(p) = K^+ = const$ if $p \geq 0$, and $K(p) = K^- = const$ if $p < 0$. As Aiello et al. show, if $K^+ \neq K^-$, the dynamical equation is clearly non-time-reversal invariant, but each single evolution is closed, so it is reversible.

In a more intuitive picture, Huw Price (1996) neatly illustrates this point by using an attractive analogy. Suppose a corkscrew factory that produces equal amounts of left-handed and right-handed corkscrews. Corkscrew factory's production is symmetric with respect to the classes of produced corkscrews: for each left-handed corkscrew there is a right-handed corkscrew. However, each corkscrew is clearly either right-handed or left-handed. Then, the analogy runs: time-reversal invariant dynamical equations are like such a symmetric corkscrew factory as long as they *produce* temporally-mirrored classes of solutions, even though each single solution may be irreversible in itself.

Finally, it is worth noticing that (ir)reversible evolutions already suppose a chosen direction of time. Claims about irreversible systems tending to equilibrium states previously assume that such states lie *in the future*. And when asking whether a physical system can "escape" from a certain equilibrium state (and if it can thus trace its steps back to the past), one asks whether it can do it given a fixed direction of time[13]. (Non-)time-reversal invariance does not need to presuppose any direction of time. In fact, the question of time-reversal invariance

[12] The equation is time-reversal invariant under a time-reversal operator T defined as performing the transformations $q \to q$, $p \to -p$ and $t \to -t$.
[13] Cf. Footnote 9 in this chapter.

is whether the dynamical equation remains unchanged, or structurally the same, after formally changing the direction of time, whatever it comes to be.

***

To sum up the chapter. If the problem of the arrow of time is a packed cluster of closely-related, though conceptually distinguishable, problems, the first step towards a sharper understanding of it is to unpack it, outlining some different ways in which it can be philosophically understood. I first went through a general metaphysical analysis of the problem stated in general (Sections 1 to 4), and then through an analysis of the most relevant physical notions typically involved in the discussion (Sections 5 to 6). Some take-home messages from the chapter:

- Metaphysically, the problem of the arrow of time is about how the past-to-future direction can be distinguished from the future-to-past one, which was identified with having a B-series structure.
- Metaphysically, such a distinction can be successfully carried out by two means: structural or non-structural
- Hence, the notion of *direction* can be understood in, at least, two different way: either in terms of a *structural arrow of time* or in terms of a *non-structural* arrow of time
- Metaphysically, the notion of the direction of *time* can also be interpreted variedly as the concept of time has received unalike ontological interpretations. Substantivalists and supersubstantavilists will consider that the temporal series stand for an absolute substance upon which material content's changing supervene. Relationalists will deny it: they just stand for temporal relations held by the material content,
- In the philosophy of physics literature, the notions of time-reversal invariance and reversibility have shown to be crucial. Both notions were carefully distinguished as well as their different philosophical implications with respect to the direction of time were clearly depicted.

In the next chapter, I will show how the problem of the arrow of time can be articulated differently in the light of these metaphysical distinctions and physical notions. In a nutshell: I will suggest that two stories can be told about the arrow of time in physics, stories leading to alternative formulations of the problem. One of the stories mainly concerns the notion of reversibility and its relationship with time-reversal invariance and seeks for a non-structural arrow of time. This sets forth the *problem of the two realms*, as I shall call it. The other primarily concerns time's structure and whether it is symmetric under time reversal. This sets forth what I will call the *problem of a structural arrow of time*.

II.

Unpacking the Problem of the Arrow of Time

From Physics Back to Metaphysics…

So far, I have put forward the *metaphysical* structure of the problem of the arrow of time (Chapter I, Sections 1-4), and have introduced some key notions in order to address it within physics (Chapter I, Sections 5-6). But I have not stepped into a proper formulation of the problem *in physics* yet. All what is known up to this point is that the notions of (ir)reversibility and (non)time-reversal invariance play a relevant role when physics comes in the discussion. In the following, I will show how this philosophical problem has been discussed in physics and how such notions relate to each other to formulate the problem in two alternative ways.

The problem of the arrow of time in physics is relatively recent. It was born in the late nineteenth century, going hand-in-hand with the development of thermodynamics and statistical mechanics. Indeed, many of its features have been greatly associated with specific issues laying on the relation between both theories. For many, the problem of the arrow of time is merely yet another chapter in the lengthy efforts to reduce thermodynamics to statistical classical mechanics. However, from the midst of the twentieth century on, the very formulation of the problem started off branching out, and alternative viewpoints of what the problem *is* began to come up. I will along this chapter suggest that two stories can be told about what the problem of the arrow of time *in physics* is supposed to be. I will call one of them "the problem of the '*two realms*'", and the other one "the problem of a *structural* arrow of time". An in-depth characterization of these views, just as their differences, their commitments and their relations, will be the focus of this chapter[14].

[14] Part of the material for Section 2 was published in López, C. (2018). "Seeking for a *fundamental* quantum arrow of time: time reversal and the symmetry-to-reality inference in standard quantum mechanics". *Frontiers in Physics,* doi: 10.3389/fphy.2018.00104

Section. 1. A shattered picture of reality *–the problem of the two realms*

One of the stories goes as follows. Plenty of phenomena seem to occur only in one direction of time, and never in the opposite. For centuries, people *knew*, for example, that heat flowed from hot bodies to cold bodies, whatever their understanding of 'hot' or 'body' were. This kind of phenomena would indicate a *directed flow* of time, intrinsically correlated, or even identified, with the directed natural behavior of phenomena. Even though those phenomena got several types of explanations throughout the centuries, they were overall partial and isolated: a more comprehensive explanation of those phenomena remained concealed up to well into the nineteenth century.

By the early nineteenth century, a unifying theoretical framework entered the European scientific scene: Thermodynamics. Its development during the first decades offered a well-articulated explanation of a panoply of phenomena involving heat flowing, energy, work and so on. Nicolás L. Sadi Carnot (1796-1832), William Thomson (alias Lord Kelvin, 1824-1907), Rudolf Clausius (1822-1888), among others, were the first to put forward the principles of a unified theory of energy, heat and their transmission. The principles of thermodynamics (particularly, the Second Principle) as well as certain notions to describe thermodynamics systems' behavior will play a crucial role in the discussion of the arrow of time from here on.

According to thermodynamics there seems to exist a *universal tendency* in Nature that explains all sort of *irreversible* processes one has experience of. Rudolf Clausius associated this *universal tendency* with a quantity called *entropy*, giving shape to the famous *second law of thermodynamics*. Roughly, it states that the entropy of any properly isolated system will *always* increase until reaching the state of *equilibrium*; from it on, entropy would remain constant ad aeternum. 'Equilibrium' is also another crucial notion in this story. It is worth noticing that the notion of equilibrium and the Second Law were here introduced in a plain time-asymmetric way: equilibrium states, at first glance, only lie in one of the extremes of a thermodynamic evolution (typically, in the future), inexorably guided by this universal tendency. This feature would allow grounding any temporally asymmetric behavior in physics: the direction in which entropy increases *is* the future direction of time.

The idea that the second law of thermodynamics physically grounds the arrow of time has been largely endorsed. It would give us, in principle, a physical criterion to distinguish the past-to-future direction from the future-to-past one. Certainly, there are various interpretations of what the second laws is supposed to mean, and not all of them relate to the arrow of time equally (see Uffink 2001, for instance). But, overall, the explanation was based on the nomological feature of entropy-increasing processes, that is, *irreversible* processes in terms of equilibrium states. Arthur Eddington (1935) and Hans Reichanbach (1956) are iconic advocators of this *thermodynamical* arrow of time. In 1928, Sir Arthur Eddington wrote:

> "*Time's Arrow*. The great thing about time is that it goes on. But this is an aspect of it which the physicist sometimes seems inclined to neglect. In the four-dimensional world . . . the events past and future lie spread out before us as in a map. (…) We see in the map the path from past to future or from future to past; but there is no signboard to indicate that it is a one-way street. Something must be added to the geometrical conceptions (…) (Eddington 1928: 34)

This is how the expression 'time's arrow" entered the literature. Here there are already two useful elements to start setting (one version of) the problem out. On the one hand, the daily impression that time seems to pass by –that time "goes on", or that time "flows", or that it exists a sort of objective "becoming" (Edington frequently uses these expressions interchangeably, see Price 2010. See Chapter I, Section 1 to see the differences among those notions). On the other hand, that a justification of such an impression is missed in any fundamental physical description of such a world. Eddington also thought that such an arrow could be found in the increase of entropy in isolated systems, claiming that the Second Law of Thermodynamics "holds the supreme position among the Laws of Nature" (Eddington 1928: 74).

In short, Eddington claimed that the "going on" of time is physically explained by this asymmetry at the core of thermodynamics. But he was deeply puzzled by the fact that, within physics, some of these basic elements were missing in different theories –in the above-quoted fragment he explicitly refers to special relativity, but the same rationale can be extended to other theories like statistical classical mechanics. The problem is then that there seems to be a *gap* in reality –A gap between our daily experience of (temporally) irreversible phenomena and thermodynamics, and the rest of our best physical theories (special relativity and/or statistical classical mechanics, for instance). For Eddington, physics' task was hence to try to put them back, taking thermodynamics as a guide.

Hans Reichenbach has also developed an account of the arrow of time in thermodynamic terms. He claims that *by definition* one can identify the future direction of time with the direction in which entropy increases:

> "It follows that a time direction can be defined only for sections of the total entropy curve. Here the inequality (14, 6) applies and allows us to give the definition:

DEFINITION. The direction in which most thermodynamical processes in isolated systems occur is the direction of positive time" (1956: 127)

Eddington's and Reichenbach's approaches to the arrow of time are clearly a type of "lawlike approaches", in the sense that the distinction between both directions of time is grounded on a physical law (see below, Section 2). Nevertheless, one should be a bit cautious here. On the one hand, the notion is not necessarily coextensive with that of time-reversal invariance. The second law is not a dynamical equation of motion, so time reversal can hardly be applied therein. It is true that it seems to display a time asymmetry, but it is not clear at all where it comes from. On the other, the conflict between the two sides of the gap seems to be that of having time-reversal invariance laws on the one side, and irreversible processes on the other. The feature of the world that explains those irreversible processes in the thermodynamic framework is the second law, but it is not strictly a break of time-reversal invariance, but one of *a lawlike* irreversibility as Sklar names it (1974: 357).

Given a series of time instants $A \rightarrow B \rightarrow C \rightarrow D$, this *thermodynamic* approach to time claims that such a series can be *by definition* correlated with the thermodynamic series of states $S(a) \rightarrow S(b) \rightarrow S(c) \rightarrow S(d)$, where the transition between states is always, as a matter of law, an entropy-increasing transition $S(a) \rightarrow_L S(b) \rightarrow_L S(c) \rightarrow_L S(d)$. So, given that $A \rightarrow B \rightarrow C \rightarrow D$ is the past-to-future direction of time and holds if an only if $S(a) \rightarrow_L S(b) \rightarrow_L S(c) \rightarrow_L S(d)$ holds, it follows that if $S(d) \rightarrow_L S(c) \rightarrow_L S(b) \rightarrow_L S(a)$ (that is, an entropy-decreasing process) does not holds, then $D \rightarrow C \rightarrow B \rightarrow A$ (that is the future-to-past direction of time) does not hold either. The antecedent is just expressing the core of the second law of thermodynamics. As stressed before, the application of time reversal to the second law is not matter of routine as the second law is not an equation of motion (see Uffink 2001: 315), so one is dealing with a kind of transitions between thermodynamic states that are forbidden by the theory, but not with a genuine case of non-time-reversal invariance.

Many writers have supported the idea that entropy-increasing processes are the sort of adequate sought-for-foundation features for the temporal directionality (famously, Eddington 1928, Reichenbach 1956, Prigogine 1980). Notwithstanding this, the works of Ludwig Boltzmann and James Clerke Maxwell by late nineteenths radically shook the very basis of physics and of any intended thermodynamic arrow of time. Thermodynamics dealt with a remarkably diverse range of *macroscopic* phenomena, from diffusion processes and phase transitions of diverse objects subject to quite varied external conditions to gases' and magnetic systems' behavior. This wide range of macroscopic phenomena displayed the same universal tendency towards equilibrium states. Based on the original attempt to reduce macroscopic thermodynamic phenomena to their microscopic mechanical constituents, the *"Boltzman-Maxwell Program"* was a milestone in the history of physics, which leads to the development of a fully-blooded mechanical theory of thermodynamic process –statistical classical mechanics.

Statistical classical mechanics was overall a result of a strongly reductionist attempt to account for macroscopically observable phenomena in terms of the laws of Newtonian mechanics plus statistical reasoning. The point to be stressed here is that the constituting microphysics (ruled by Newton's law), to which macroscopic phenomena allegedly reduce, did not display any universal tendency towards anywhere. Any micro-process described by statistical classical mechanics looks rather perfectly reversible. But, when large ensembles of mechanical micro-processes were considered, such universal tendency was partially recovered. In any case, it was no longer true that *always* thermodynamic systems approach to equilibrium states.

More fundamentally, the underlying dynamics (basically, Newton's law) are in fact time-reversal invariant. As explained in Chapter I, Section 6.2, this mainly means that they produce a pair of time-symmetric twins, that is, that the structure of their solutions has the form $\mathrm{W} = \mathrm{W}^f + \mathrm{W}^b$, and for any Newtonian evolution $\mathcal{E}_N$ if $\mathcal{E}_N \in \mathrm{W}^f$, then there exists a $\mathcal{E}_N^T \in \mathrm{W}^b$. Naturally, many paradoxes started cropping out in the field, undermining the alleged successful reduction of thermodynamics to classical statistical mechanics. The two most famous ones were the Loschdmidt's paradox (Loschdmitdt 1876/1877) and the Recurrence Paradox by Ernst Zermelo (1896). There are several details to look into here (for discussion and details, see Sklar 1993, Uffink 2001, 2007), but what should be kept in mind is that these paradoxes bring together many of the elements that will set one of the most spread versions of the problem of the arrow of time in physics. In some sense, the problem of the arrow of time has greatly been taken as giving an answer to these paradoxes.

Even though the main aim of the paradoxes was to put into question the allegedly successful reduction of thermodynamics to statistical mechanics by the Boltzmann-Maxwell program, they have had critical consequences for the thermodynamic arrow of time as well. As stated by Lord Kelvin and Clausius, the second law of thermodynamics was an *exceptionless* principle: entropy *always* increases toward the equilibrium. When correlated with the direction of time, this idea naturally captured a privileged, and exceptionless, temporal direction. Yet, as the statistical reasoning applied to thermodynamics entered physics, the exceptionless feature of the second law of thermodynamics came to be considerably weaker –entropy-decreasing processes would be, at least theoretically, possible.

In regarding an ideal gas as viewed by statistical classical mechanics, its seeking-equilibrium behavior could be explained in terms of the collision of its constituent particles plus their probable distribution. The link between the second law of thermodynamics and the statistical mechanics became established by the well-known H-theorem in 1872 (Boltzmann 1872). The Newtonian particles constituting any sample of gas will naturally have a certain distribution and Maxwell had few years before showed how to derive various macroscopic features of the sample from the distribution of particles' velocities. Particularly, he derived a distribution from which the system cannot depart, and his insight was to relate such distributions to equilibrium states. The so-called 'Maxwell distribution' was thereby related to equilibrium states. What the H-theorem showed was that the distribution of particles' velocities of any sample of a gas will either tend to the Maxwell distribution (that is, an equilibrium state) and remain in it when reached. This result was derived considering only the effects of the mechanical collision of particles plus the statistical reasoning.

In the light of this, what Franz Loschmidt observed was that, given that the underlying Newtonian dynamics governing particles' behavior is time-reversal invariant, for every entropy-increasing mechanical process $\mathcal{E}_{S^+}$, there is also an entropy-decreasing process $\mathcal{E}_{S^-}$ that is also a solution of the (time-reversed) Newton's laws. This is actually a two-faced result: the Loschmidt's paradox (or Reversibility objection as it came to be famously known too) entails that

(1) for any given example of a gas, there is another time-reversed gas experimenting an entropy-decreasing evolution or departing from equilibrium if already there,
(2) entropy is not only expected to increase towards the future, but also towards the past.

Let us spell these points out a bit. Suppose a system S in a micro-state a at t_1. The system will evolve to a micro-state b after some time, say, t_2. According to thermodynamics, $S(b)$ should be a state of higher entropy than $S(a)$. The reversibility objection puts the finger upon the time-reversal invariance feature of the underlying laws governing the evolution from $S(a)$ to $S(b)$: by reversing the direction of time (which Loschmidt interpreted as reversing the velocity of the system) what one gets is a time-reversed evolution $T\mathcal{E}_{S^-}$ evolving from a state $TS(b)$ to $TS(a)$, where the entropies of $TS(b)$ and $TS(a)$ are naturally equal to those in b an a respectively.

Clearly, $T\mathcal{E}_{S^-}$ is an anti-thermodynamic behavior (that is, an entropy-decreasing evolution), but one in complete agreement with the underlying mechanical laws. An alternative, though equivalent, way to see at the results is in terms of entropy-increasing processes in both directions of time: if at t_1 there is an entropy-increasing evolution towards the future (one evolving from $S(a) \rightarrow S(b)$), there is (by time-reversal invariance) then a temporally-mirrored entropy-increasing evolution from $TS(a) \rightarrow TS(b)$ heading back to the past (Fig.2.1).

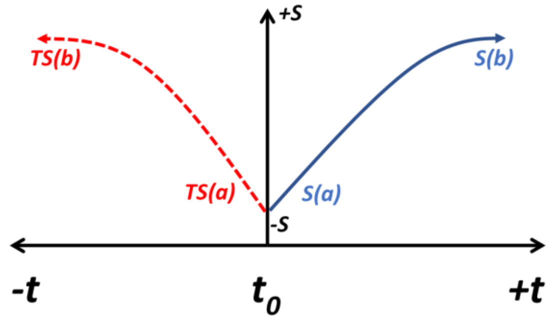

Fig. 2.1

The paradox is indeed shocking. Not only does it directly follow from it that the thermodynamic irreversible behavior cannot be fully captured by classical mechanics plus the statistical reasoning, but also that any intended foundation of the direction of time in thermodynamics is seriously unwarranted. Boltzmann's answer, as well known, was to explain that though different micro-states may realize the same macro-state, some of them can be realized in more many ways than others: equilibrium macro-states can notably be realized by much more different arrangements of the constituent particles than non-equilibrium macro-states. It follows from this that equilibrium macro-states are far *more likely* than non-equilibrium macro-states. In accordance with it, transitions from states of lower entropy to higher entropy are much more likely than transitions from states of higher entropy to lower. And the explanation given becomes just matter of statistics.

Boltzmann's answer doesn't seem to cut the Gordian knot though. On the one hand, it doesn't completely avoid entropy-decreasing processes (these are just much less likely, but not impossible). On the other hand, and probably more important, it is now a mystery why there are (and there were) low entropy states at all, if they are so unlikely. The most likely scenario would be that of a Universe in thermal equilibrium, since it is simply its most likely state. To put it differently, Boltzmann's answer explains why if there is a gradient of entropy, the overwhelming majority of processes will be entropy-increasing processes. But it doesn't explain why there exists such a gradient to begin with![15]

There was another fundamental objection to Boltzmann's explanation of irreversibility: The Recurrence Paradox. This objection was raised by Henri Poincare and Ernst Zermelo in late nineteenth century. In 1889, Poincaré showed that entropy-decreasing behaviors are not only possible, but *inevitable* given an enough amount of time. Poincaré's result pointed to the fact that, excepting a "vanishingly small" number of initial states, a system starting off in a particular mechanical state at t_1, it is bound to return to such an initial state (or to an arbitrarily close one) after enough time. Few years later, Erns Zermelo applied Poincaré's theorem to Boltzman's *H*-theorem: any system starting off in a non-equilibrium state will eventually return to it, despite having reached an equilibrium state during its evolution. Boltzmann's monotonic approach to equilibrium was once again pushed up against the wall.

Both the Reversibility and the Recurrence paradoxes aimed at challenging the very viability of Boltzmann's project. On the one hand, the second law of thermodynamics was no longer an exceptionless principle, but a statistical one: it is true in the overwhelming majority of cases, but not always, challenging the universality of *H*-theorem. On the other, the asymmetry of thermodynamics was completely lost in the middle. And these results came to set one of the more puzzling conundrums in physics, to wit, that features of the macro-level do not seem to be fully compatible with those of the micro-level. The natural world was thus left somehow split into two 'realms': the unobservable reversible micro-physics and the observable irreversible macro-physics. The properties of each world seemed to be at odds with the properties of the other. Even worse, the macroscopic realm was claimed to be reduced to the microscopic one.

Making sense to this shattered picture of reality and closing the "gap" between those two realms have demanded tireless efforts by natural philosophers and physicists up to present. In his task, many foundational and conceptual problems in philosophy of physics have risen. Among them, the problem of the arrow of time; or better, one of the most popular ways to introduce it. The punch line is that if the direction of time is *by definition* the direction in which entropy increases, the paradoxes showed that for any entropy-increasing behavior towards the future, there is also a temporally-inverted increasing-entropy behavior towards the past, wiping off any distinction between the past-to-future direction and the future-to-past one. Furthermore,

[15] A partial and debatable solution to this is to postulate an incredibly low entropy state in the past, tracing back to the origin of the Universe (the so-called Past Hypothesis, see Albert 2000).

the entropy-based direction of time turned out to be matter of probability lacking thereby any forceful foundation.

Hans Reichenbach (1956) offered not only a full-fledged theory of the arrow (or direction) of time in terms of thermodynamics, but also one of the most complete and precise formulations of what the problem is supposed to be. For Reichenbach, the problem of the arrow of time is mainly about a structural distinction between past and future (1956: 27). As he thought that the answer should be nonetheless found in thermodynamics, and that the positive direction of time was the direction in which entropy increases (ibidem. 54), he considered that the major problem to be solved by physicists was the following:

> "The elementary processes of statistical thermodynamics, the motions and collisions of molecules, are supposed to be controlled by the laws of classical mechanics and are therefore reversible. The macro-processes are irreversible, as we know. How can this irreversibility of macro-processes be reconciled with the reversibility of micro-processes? It is this paradox which the physicist has to solve when he wishes to account for the direction of thermodynamical processes and for the *direction of time*" (1956: 109. Italics mine)

In light of this, the problem of the arrow of time has directly to do with the very foundations of statistical classical mechanics and thermodynamics. Reichenbach's proposal consisting in a thermodynamic structure of branching systems aims at solving the problem in these terms (see Reichenbach 1956, Grünbaum 1963. Earman 1974 and Sklar 1993 for criticisms).

The structure of the formulation in terms of a 'tension' between what is given by experience and what is dictated by fundamental physics has also been extended to other theoretical frameworks. And this formulation got canonical in the literature. And philosophers and physicists have largely worked on the problem under this formulation. Much work has been done in the last fifty years in that direction, and naturally, the very formulation of the problem has enormously gotten more accurate. In any case, the problem has crucially depended in every detail on notions like "reversibility", "entropy", "micro-evolutions vs macro-evolutions", "equilibrium", and so forth. And the essence of the problem has been to face the existence of such a *gap*, such a *division* in the natural world. And the essence of any solution to the problem has been to explain the gap. And to make it disappear. Somehow.

Some modern references to the problem typically refer to this gap or apparent contradiction between having time-reversal invariant laws on the one side, and an irreversible or temporally-asymmetric behavior on the other. Huw Price (1996) synthetizes clearly the question of the arrow of time in a general way as following: “why the physical world should be asymmetric in time at all, given the underlying physical laws seem to be very largely time symmetric” (1996: 6).

Craig Callender, in a paper titled “What is the ‘problem of the direction of time’”? says:

> “Concisely put, the problem is that given a non-equilibrium state at t_2, it is overwhelmingly likely that (a) the non-equilibrium state at t_2 will evolve to one closer to equilibrium at t_3, but that due to the reversibility of the dynamics it is also overwhelmingly likely that (b) the non-equilibrium state at t_2 has evolved from closer to equilibrium at t_1” (1997: S225)

This is essentially the Reversibility Paradox as was explained above. According to Callender, then, the problem is that statistical classical mechanics makes a prediction that is radically falsified (1997: S226), so the real problem is how to get rid of option (b). As Price did, Callender assumes that the underlying dynamics is time-reversal invariant, and that one only has powerful evidence for predictions like (a).

David Albert (2000) has also pointed out that there is a *tension* in the discussion on the arrow of time in physics: that between fundamental microscopic physical theories and the everyday macroscopic human experience. Even though Albert thinks that fundamental physics is non-time-reversal invariant (what could give us a hint for a structural arrow of time as I will show shortly), he does think that Newtonian mechanics is fully time-reversal invariant and that our macroscopic experience of time has fundamentally to do with this tense relation between thermodynamics and Newtonian mechanics. Hence, the real puzzle about time is to come up with an explanation of why our daily experience looks so temporally asymmetric, whereas the underlying dynamics (as far as Newtonian mechanics is concerned at least) is temporally symmetric. Indeed, Albert has proposed one of the most spread solutions for this tension up to now: the Past Hypothesis (Albert 2000. See also Callender 2004a-b. For discussion, see Price 1996, Earman 2006, Wallace 2011, Goldstein et al. 2016).

The Past Hypothesis, briefly, claims that the entropy of the universe at the beginning was extraordinarily tiny. It mainly avoids absurd predictions about the past just as it discards option (b) in the Callender’s quote for instance. The motivation for such a (seemingly) ad-hoc resource seems to fit, for instance, in a broader philosophical and scientific framework about how physics explains the actual world: the Mentaculus[16], proposed also by David Albert (see Albert 2015. See also Loewer 2012). And the upshot, as Barry Loewer puts it, is that "if we take the statistical mechanical probability distribution realistically (…) then we *need* to posit the past hypothesis for those probabilities to also ground inferences from the present to the past" (Loewer 2012: 126). I shall spend some words on this solution at the end of this chapter.

To sum up. One of the versions of the problem of the arrow of time in physics has its origin in the puzzling relation between thermodynamics and statistical mechanics, and essentially takes the form of providing a way out to the Reversibility paradox. The real issue here is there is not smooth continuity between the micro-world and the macro-world –the existence of the *gap* between two realms.

The problem of the 'two realms'	(a) Given that the underlying fundamental dynamics of the world is temporally symmetric, why does our world look so temporally asymmetric?
	(b) There is hence a gap between the micro-world and the macro-world.
	(c) the problem at stake is then to close the gap or to bring this shattered picture of reality together.

Note that the first part of (a) is a crucial premise in this formulation of the problem of the arrow of time: though it is not frequently made explicit, the formulation already assumes that there is a sense in which there is no arrow of time in physics given that the fundamental laws are time-reversal invariant. This is the origin of the gap and why one is in need of coming up with an explanation that harmonizes such a shattered picture. To put it informally, this formulation requires an explanation of why it seems that there is an arrow of time, despite knowing that there is a sense (a *fundamental* one) in which there is not.

This is the origin of the most widely-spread versions of the problem of the arrow of time in philosophy of physics: *the problem of the 'two realms'*. It has involved many clever people

---

[16] The 'Mentaculus' provides a kind of probability map of the universe entailing a probability distribution over the set of all possible micro-histories of the universe compatible with a given initial state (Past Hypothesis). Hence, the Mentaculus is a recipe for a fundamental physical theory that explains the reversibility paradox and the second law of Thermodynamics. In a nutshell, the Mentaculus comprehends three ingredients:

(1) The fundamental dynamical laws of physics (in this case, the time-reversal invariant classical laws of physics),
(2) The Past Hypothesis,
(3) A law specifying a uniform probability over the micro-states that realize the very tiny low entropy state at the beginning.

and has undeniably brought about loads of priceless philosophical and scientific work. Notwithstanding this historical fact, others have argued that this formulation of the problem of the arrow of time does not deserve such a name. The real problem has been to a large extent missed: thermodynamics and statistical classical mechanics deal at best with the nature of irreversible processes, but it would be an unjustified jump to draw conclusions about the nature of time from it. In this sense, the project of basing the arrow of time on thermodynamics and/or classical statistical mechanics is too narrow, if not simply off the right track. So, in order to address the problem of the arrow *of* time appropriately, some philosophers have reformulated it in different terms.

Section 2. *Ubi lex non distinguit, nec nos distinguere debemus* –A structural arrow of time

Concisely put, the starting point is that so-introduced formulation of the problem of the arrow of time is subsidiary. The problem of the arrow *of* time *in itself* is conceptually prior and completely independent from any formulation within the framework of thermodynamics and statistical classical mechanics. It would therefore be better introduced it independently of notions such as reversibility, equilibrium states or entropy. Probably, the origin of this alternative proposal is the *Time Direction Heresy*, promoted by John Earman in his 1974 paper. Earman claimed therein that the arrow of time "does not hinge as crucially on irreversibility as the reductionist would have us believe" (1974: 20), but it is rather an "intrinsic feature of space-time which does not need to be and cannot be reduced to non-temporal features" (1974: 20).

This heretic proposal is a powerful reaction to an already deeply-ingrained dogma in the literature: the problem of the arrow of time *is* the problem of the two realms, and consequently, it crucially depends on issues around the foundations of thermodynamics and classical statistical mechanics. John Earman understood that his way to look at the issue minds a *literal* interpretation of the general problem as posed in Chapter I: when asking whether time has a direction or not, one is asking whether time *in itself* has a fundamental direction or an intrinsic asymmetry. Any other sense of temporal asymmetry or time's direction will thus be *metaphorical*, or secondary.

In his 1974 paper, Earman unfolds several problems related to his heretic understanding of the arrow of time. They mostly have to do with determining what the topological structure of space-time is like (whether it is time-orientable, whether it is equipped with a time orientation, whether time itself is anisotropic, and so on). Hence, if one comes to understand the problem in terms of two opposite realms, one is simply missing the point: the problem of the arrow of time *is* about the intrinsic properties of space-time itself, regardless the nature, and the behavior, of its material content. Much of this problem crucially hinges upon how one could come to know what the properties of space-time are, particularly, in its temporal dimension.

Some authors have followed Earman's heretic spirit, focusing on space-time's properties instead of the relation between the micro-realm and the macro-realm. Tim Maudlin (2002), and Mario Castagnino and Olimpia Lombardi (2009) have more recently approached the problem of the arrow of time in those terms. For instance, Castagnino and Lombardi propose a 'global non-entropic approach' to the arrow of time, which mainly consists in looking into space-time's topological properties so as to establish a privileged direction of time capable of accounting for the intuitive asymmetry between past and future. They firstly aim to specify the conditions under which a global arrow of time can be properly settled, mainly based on geometrical properties of space-time as a whole. Then, Castagnino and Lombardi's approach transfers this global feature to the local level, where the arrow of time looks like a non-spacelike local energy flow (2009: 2). In any case, what bears metaphysical priority in the approach is the fact that the arrow of time has primarily to do with intrinsic properties (topological properties) of space-time. So, the problem must be introduced independently of any reference to irreversibility, entropy-increasing processes, and the gap between two realms. Neither does the problem have to do with the concept of time-reversal invariance, even though they argue that the notion of time-reversal invariance and irreversibility should be carefully distinguished from one another.

In turn, Maudlin claims that

> "the *passage* of time is an intrinsic asymmetry in the structure of space-time itself, an asymmetry that has no spatial counterpart and is metaphysically independent of the material contents of space-time. It is independent, for example, of the entropy gradient of the universe" (2002: 259)

In essence, Maudlin considers that the problem of the arrow of time *is* the problem of picking a temporal orientation in a time-orientable space-time. But, unlike Earman's and Castagnino and Lombardi's proposals, Maudlin contends that time-reversal invariance does play a relevant role in picking an arrow of time in so far as the notion allows linking a temporal orientation (a topological feature of space-time) to a property of physical theories' dynamics –whether it is symmetric under time reversal. By testing time-reversal invariance, one gets some information about a temporal orientation. Maudlin is not quite clear as to how time reversal works and what the link between geometry and dynamics is supposed to be. However, a previous work by Paul Horwich (1987) may shed some light on the matter since he also links time-reversal invariance with the direction of time. Particularly, with an intrinsic property of time in itself.

Paul Horwich starts off by pointing out a distinction I made before: that of between an asymmetry *of time* and asymmetries *in time*. The distinction allows raising the following question: there are indeed many asymmetries *in time*, but "do they indicate that time *itself* is asymmetric?" (1987: 1). Perhaps, time itself is perfectly symmetric and the asymmetries in time are caused by other physical elements that do not bear on the nature of time itself. In a broad sense, the problem of the arrow of time mainly concerns the properties of time itself, particularly, if there is an intrinsic similarity between the past and the future directions – intrinsic similarity that must be sharpened in terms of having "the same intrinsic properties expressed as *nomological* properties" (Horwich 1987: 41; see Chapter I, Section 6). Horwich's idea is that any intrinsic or fundamental difference between past and future would manifest itself in some time asymmetry within the laws of physics. And this explains why the notion of time-reversal invariances becomes quite informative.

Doubtlessly, there is plenty of differences among all these heretic-like proposals. For instance, whereas Earman virtually identifies irreversibility and time-reversal invariance as part of the dogma on the arrow of time, the rest of the authors are prone to carefully distinguish them (particularly, Castagnino and Lombardi –who in turn based the distinction on Price 1996). However, whereas Castagnino and Lombardi consider that time-reversal invariance does not play any role in picking an arrow of time, Maudlin and Horwich think that it does. Particularly, Horwich offers a more robust explanation of how the link is supposed to work than Maudlin. Notwithstanding these differences, all of them share the same basic idea: the philosophical inquiry about the arrow of time is primarily about the properties of time itself, regardless (prima facie) its material content. These formulations (either in terms of a time orientation in a relativistic space-time, or its topological properties or its links to time-reversal invariance) are metaphysically and conceptually prior to any formulation in terms of foundations of thermodynamics and classical statistical mechanics, irreversibility or entropy increasing.

However, it is not completely clear so far how this approach to the arrow of time is supposed to work. Neither how the different elements of those proposal are linked to one another. On the one hand, Earman's main lesson is that only properties of space-time matter when the problem of the arrow of time is at stake. On the other, it is not so clear how one can come to know them. With respect to the latter, Horwich mentions in passing that symmetries of a theory's dynamics allow one to *infer* certain properties of the underlying space-time. This idea also underlies the formulation of problem of the arrow of time when introduced in terms of the problem of the 'two realms', giving some hints on how this inferential mechanism is supposed to work. For instance, consider Price's sentence: "to a very large extent, then, the *laws of physics seem to be blind* to the direction of time – they satisfy T-symmetry, as we may say" (1996: 116, italics mine). In virtue of this, Price repeatedly claims that there is no *objective* direction of time: the non-objectivity of the direction of time seems to somehow rest on what he calls "T-symmetry", that is, time-reversal invariance. Though Maudlin (2002) doesn't share the same stance, he does share a similar diagnosis, and resort to the underlying inferential mechanism. He says

> "The usual approach sets the problem as follows: the fundamental physical laws have a feature called 'Time Reversal Invariance'. If the laws are time-reversal invariant, then it is supposed to follow that physics itself recognizes no directionality of time" (Maudlin 2002: 266)

In my introduction of the problem of the arrow of time in terms of the problem of the 'two realms', point (a) evokes the feature of time-reversal invariance at the level of the physical laws already: it says *something* about the nature of time, in particular, that *in some sense* there is no direction of time. Therefore, the implication could be phrased as follows:

> If a physical law *L* is time-reversal invariant, then *that* physical law does not pick, *in some sense,* a direction of time

If one comes to discover that all (or, at least, almost all) fundamental laws of physics are time-reversal invariant, it could be entailed that

> If all our physical laws are time-reversal invariant, then there is no, *in some sense*, a direction of time (according to physics)

And it is for this sort of inferences that the problem of the 'two realms' comes up as described: there is gap *because* one is *in some* sense allowed to draw some conclusions about time's features from the fact that physical laws (in particular, the micro-dynamics underlying the macroscopic behavior) are time-reversal invariant. I think that the underlying premise putting to work this inferential mechanism is the following: symmetries, in particular, space-time symmetries, would shed some light on what the space-time's *structure* is. Hence, by looking into dynamic's symmetries, one can infer space-time's symmetries, and thereby, the space-time's *structure*. Hence, the inference should be reworded as following

> If a physical law *L* is time-reversal invariant, then *that* physical law does not pick a *structural* direction of time
>
> If all our physical laws are time-reversal invariant, then there is no *structural* direction of time (according to physics)

This inferential mechanism is not novel, but it has been frequently used in the philosophy of physics and the metaphysics of science for a long time. As well known, symmetries are playing an increasingly paramount role not only in physical theories, but also in facing many fundamental philosophical problems in philosophy of physics and metaphysics (see for instance, Brading and Castellani 2003, 2007; Baker 2010, Dasgupta 2015, 2016, Caulton 2015). Time symmetry (or time-reversal invariance) is just an instance of this general symmetry-based inference. So, in introducing the problem of the arrow of time as mainly concerning space-time's structure or intrinsic properties of time in itself, the notion of time reversal and time-reversal invariance play the role of testing whether time, according to a given theory, instantiates the property of having a *structural* direction –where the law does not distinguish, neither ought we to distinguish.

There are plenty of references in the relevant literature about the relation between dynamics' symmetries and space-time's structure. H. Mehlberg (1961) says:

> "In mathematical parlance, temporal isotropy in scientific contexts is, therefore, tantamount to the covariance of laws of nature under time reversal. This amounts to asserting that within science time's arrow will have to be rejected if the laws of nature remain unaltered and valid in a universe whose past and future are interchanged with ours". (Mehlberg 1961: 108)

And Paul Horwich (1987) again:

> "In addition, the account should help us to understand why the existence of time-asymmetric laws [i.e. non-time-reversal invariant laws in my vocabulary] is generally taken to guarantee time's anisotropy" (1987: 39).

Lawrence Sklar (1974) is much clear about this link:

> "Some of the important symmetries of physical laws are the invariance of laws under translation in space, invariance under translation in time, invariance under

spatial rotation, invariance under change of inertial frame, invariance under spatial reflection, and so-called time-reversal invariance. All of these invariance principles are intimately related to the structure of the spacetime in which the material happenings of the world occur" (Sklar 1974: 359)

In his famous 1989 book, *World Enough and Space-time*, Earman formulates two symmetry principles working as "adequacy criteria" so as to link the theory's dynamics on one side, and the space-time's structure in which such a dynamics fits, on the other (Earman 1989: 46).

SP1 Any dynamical symmetry of T is a space-time symmetry of T

SP2 Any space-time symmetry of T is a dynamical symmetry of T

Where 'T' is a physical theory. The motivation for introducing such principles is "the realization that laws of motion cannot be written in the air alone but require the support of various space-time structures" (1986: 46). And the point just comes down to the sort of relations among dynamics' space-time symmetries, space-time's symmetries, and its structure.

Jill North (2008, 2009) has devoted considerable efforts to unfold the notion of "physical structure" and its relationship with the symmetries of a given theory. Taking the structure of the space and the space-translation symmetry, she says:

> In applying any transformation to a theory, we hope to learn about the symmetry of the theory, and of the world that theory describes. We do this by comparing the theory with what happens to it after the transformation. If the theory remains the same after the transformation—if it is *invariant* under the transformation—then it is symmetric under that operation. This indicates that the theory 'says the same thing' regardless of how processes are oriented with respect to the structure underlying the transformation. We conclude that a world described by the theory lacks the structure that would be needed to support an asymmetry under the operation. For example, from the space-translation invariance of the laws, we infer that space is homogeneous, that there is no preferred location in space (North 2008: 202)

In a paper exclusively devoted to the notion of structure in physics, she rephrases Earman's adequacy criteria in terms of a methodological principle guiding our understanding of the structure and the ontology of any physical theory:

> "Yet we tend to infer that there is no more structure to the world than what the fundamental laws indicate there is. Physics adheres to the methodological principle that the symmetries in the laws match the symmetries in the structure of the world. This is a principle informed by *Ockham's razor*; though it is not just that, other

things being equal, it is best to go with the ontologically minimal theory. It is not that, other things being equal, we should go with the fewest entities, but that we should go with the *least structure*. *We should not posit structure beyond that which is indicated by the fundamental dynamical law*" (North 2009: 65. Italics mine)

The idea of "going with the least structure" in order to underpin our ontological commitments by means of symmetries has been neatly explained in detail by Shamik Dasgupta (2016). This inferential mechanism going from symmetry to the structure of space-time has been called by him "the symmetry-to-reality inference" [17]. The scheme of the argument as presented by Dasgupta is the following:

(1) Laws *L* are the complete laws of motion governing our world

(2) Feature *X* is variant in *L*

(3) Therefore, feature *X* is [interpretational link]

(4) Ockham's razor applied to *X*

(C) Therefore, *X* is not real

Let us spell this scheme out. The first premises claim that a symmetry, in a formal and general sense, holds for a given set of physical laws (mathematically, dynamical equations of motion). The second premise, particularly indicates, that the physical laws involve a property whose values can vary freely without altering the laws' structure. These are the *formal premises* of the symmetry-to-reality inference. The variant property has then to be interpreted (premise three): variant features can be considered as redundant or superfluous (Earman 1989, North 2009, Belot 2013), non-objective (Weyl 1952, Nozick 2001), or undetectable (Robert 2008, Dasgupta 2016). Premise four is an epistemic premise claiming that it is an epistemic vice to posit a variant feature as part of the theory's structure. From the formal premises, the interpretational link, and the epistemic premise, one is entitled to conclude that the variant feature *X* is not real (or better, fundamental. See fn. 17) in the sense that it is not part of the structure of the reality (or space-time) according to a previously-given set of physical laws – one's metaphysics shouldn't therefore contain variant features as structural or fundamental.

This scheme is general and can be multiply instantiated by specifying the feature X one is interested in. Clearly, one of the instances of this argument is its application to the problem

[17] Dasgupta puts the argument in terms of "reality", but I would rather call it "the symmetry-to-*fundamental*" inference. It seems to me that the argument points out to what one should accept as part of what's fundamental or structural for a physical theory, and not to what is real: there might be real things that aren't fundamental or structural for a physical theory. If one only accepts as real things what falls into the argument as posed it, then one's ontology would be extremely minimal, dooming many non-fundamental things to unreality

of the arrow of time in term of time-reversal invariance. The argument would thus run as follows

(1) Laws L are the complete laws of motion governing our world

(2) The *direction of time* ($+t$ or $-t$) is variant in L

(3) Therefore, the *direction of time* ($+t$ or $-t$) is [interpretational link]

(4) Ockham's razor applied to the direction of time

(C) Therefore, the direction of time is not fundamental

Premises 1 and 2 basically say that a given set of laws of motion are time-reversal invariant, that is, they are invariant under changing the direction of time, $t \rightarrow -t$ (in the sense specified in Chapter I, Section 6.2.) Again, this is a formal result stemmed from applying the time-reversal operator upon such a set of laws. Hence, the direction of time is said to be a variant feature in L. Based on premise one and two, premise three claims that the direction of time is redundant or superfluous (Earman 1989, North 2009, Belot 2013), non-objective (Weyl 1952, Nozick 2001), or undetectable (Robert 2008, Dasgupta 2016). As one should go with the least structure according to premise four, one concludes that the direction of time is not part of the structure of space-time according to the set of laws L.

My claim is that the symmetry-to-fundamental inference for the arrow of time backs up any inference of the type

> If a physical law L is time-reversal invariant, then *that* physical law does not pick a *structural* direction of time

Yet, I think one should be in extremely cautious here, particularly, in evaluating the limits and the scopes of the symmetry-to-fundamental inference. To begin with, such an implication is true only when one is interested in looking for a *structural* arrow of time. Quoting Sklar:

> "Do the proponents of this position really wish to allege that if the laws of nature all do turn out to be time-reversal invariant, our whole impression that the world of events is a world in which an asymmetric temporal priority relates event to event is an illusion?" (Sklar 1974:399)

Even though many have claimed that from time-reversal invariant laws one can conclude that there is no direction of time whatsoever, I think that, when correctly understood, the symmetry-to-fundamental inference just authorizes us to claim that the space-time assumed by a given theory is not equipped with an arrow of time, and not that there is no arrow of time in *any* sense. Or, to put it differently, that the direction of time is not a structural or fundamental property of the world, but a secondary, derivative or an emergent one (Loewer 2012 argues for an emergent arrow of time given that time-reversal symmetry generally holds. See also Callender 2010). In fact, this is what the problem of the two realms assumes: given that there is no structural or fundamental arrow of time (because laws of nature all turn out to be time-reversal invariant), it must come from elsewhere, for instance, from the initial conditions of the early universe.

To get things straight, I shall introduce, and briefly analyze, the implications derived from the symmetry-to-fundamental inference. First, the symmetry-to-reality inference clearly allows us to conclude that

(a) If the set of laws *L* belonging to a theory *T* is non-time-reversal invariant, then there is a *structural arrow* of time according to *T*.

In Maudlin's words: "If laws are non-T-invariant, then the space-time must be equipped with an orientation" (Maudlin 2002: 269). Therefore, time-asymmetric laws are a sufficient condition for a structural arrow of time.

The converse logically follows from (a)

(b) If there is *no structural arrow of time* according to *T*, then the set of laws *L* belonging to a theory *T* is time-reversal invariant.

This implies that time-reversal invariant laws are a *necessary* condition for a non-structural arrow of time, as (b) points out that the structure of space-time is temporally symmetric. As mentioned above, the (non-structural) arrow of time must then be found elsewhere.

Now, one is left with two further options

(c) If there is a *structural arrow of time* according to *T*, then the set of laws *L* belonging to a theory *T* is non-time-reversal invariant.

Though persuasive at first glance, this implication could however be false. Firstly, one can suppose that future physics could show that there is a structural arrow of time with time-reversal invariant laws. After all, time-reversal invariance is just a mean that triggers the inference, but there might be others: the implication would be true only if time-reversal symmetry is the unique way to dig into space-time's property, but this is a highly debatable supposition. Secondly, the implication is false in General Relativity, where the space-time's geometry is strongly related to mass and energy's distribution (I shall add some further words about it afterwards). From (a) and (c) it follows that non-time-reversal invariance is only a sufficient condition for a fundamental arrow of time, but not a necessary one. At least, in general.

Finally, the implication I started with,

(d) If the set of laws L belonging to a theory T is time-reversal invariant, then there is no *structural arrow of time* according to T.

This inference, the converse of (c), is true *so long as* one looks for a structural arrow of time. It shouldn't be taken valid as an inference against an arrow of time in a broader sense (for instance, against a non-structural arrow of time).

To sum up. There is an alternative formulation of the problem of the arrow of time that does not hinge crucially upon notions like reversibility, entropy increasing or upon the foundations of statistical classical mechanics: the problem of a structural arrow of time. Even though there are different ways to introduce this problem, in general all of them agree on that the problem of the arrow of time is mainly about whether space-time in itself is temporally asymmetric. This claims rests on one what I called the "symmetry-to-fundamental" inference. In this respect, the notion of time-reversal invariance becomes crucial to infer space-time's structural properties from dynamics' space-time symmetries.

The problem of a *structural* arrow of time	(a) Does time have in itself the property of being asymmetric/directed? (b) Considering the laws of physics belonging to a theory T, is the structure of the space-time supposed by those laws equipped with a time-orientation? (c) In being backed up by the 'symmetry-to-fundamental' inference, the question is: Is there any non-time-reversal invariant law in physics? Or, is there any physical law that does not produce a par of time-symmetric twins?

It is worth remarking some things before moving on to the comparison between both formulations of the problem of the arrow of time. First, I have mentioned that certain type of implication (in particular (c)) does not work in general relativity. The reason is that in general relativity different solutions to Einstein's field equations entail different space-time's geometrical structures, since they connect with the distribution of mass and energy. And this is so because the matter-energy distribution *is* the space-time geometry. In these cases, any structural temporal asymmetry of space-time along its temporal dimension does not depend only on a formal symmetry but also depends on such a distribution of matter and energy: any variation in the distribution of matter and energy is *also* a variation in the structure (in the geometry) of space-time in itself[18]. Note that this is not the case in the rest of physical theories, where the structure of space-time on which their laws are written down remains fixed in all its models, and any variation on the matter and energy distribution introduces a variation in how the content is distributed without altering the structure of the space-time in itself.

Second, note that this formulation and way to approach the problem of the arrow of time focuses mainly on non-time-reversal invariant laws. The overall idea is that pointed out in (c) above: non-time-reversal invariant laws does not produce a pair of time-symmetric twins. That is, the physical laws belonging to a given theory T do not say the same when running in the backward direction of time or in the forward direction of time (where 'not say the same' might mean that they say nothing whatsoever in one of them). Alberts puts it neatly in saying that

> "(…) if any theory whatsoever offers us *both* predictive *and* retrodictive algorithms, and if those two algorithms happen to be *identical*, and if the theory in question entails that a certain process can happen forward, then it will necessarily also entail that the process can happen backward. *That's* what I'll mean, then, from here on, when I speak of a theory as being invariant under time-reversal. (2000: 14)

Therefore, the question of the *structural* arrow of time is whether the series $A \to B \to C \to D$ and $TD \to TC \to TB \to TA$ (a pair of time-symmetric twins) are produced by the same algorithm. So, there are, at least conceptually, alternative ways to get non-time-reversal invariant laws. For instance, both series could be models of a physical theory but produced by different algorithms: in time reversing a dynamical equation of motion, one gets to a *different* equation that produces the temporally-inverted series. Or, it could be also the case that by time reversing the original dynamical equation one gets to a different equation that does not produce any physical solution according to the theory (for instance, the so-obtained solution contradicts other postulates of the theory). Overall, the idea is the same: non-time-reversal invariant laws naturally and by their own means distinguishes a sequence going backward in time from a

[18] This fact explains why, for instance, Castagnino and Lombardi (2005, 2009) were able to develop a geometrical approach to the arrow of time in general relativity, and to ground a *structural* (non-reductive) arrow of time, despite the time-reversal invariance of the Einstein's field equations.

sequence going forward in time, where the parlance 'running backward' and 'running forward' only makes sense in relation to a time-reversal transformation $T: t \rightarrow -t$ acting upon an equation of motion.

Third. This formulation of the problem requires further conditions to be philosophically meaningful. This will be left fully clear in the following chapters, when addressing a particular case, but it is worth introducing it at this point. There are at least three conditions that should be met in order to formulate the problem in a meaningful manner: (a) the fundamentality condition, (b) the contingency condition, and (c) the co-extensity condition. Let us address them in tandem.

Some philosophers, worried about the problem of the (structural) arrow of time, usually bring up a condition that would make the problem philosophically interesting: only *fundamental* dynamical laws are relevant for the debate. This "*fundamentality* condition" (as I call it) aims at setting aside a huge amount of uninteresting non-time-reversal-invariant laws, and only focusing on force- and interaction-free laws. For instance, Craig Callender (1995) claims that "when asking if the universe is TRI [time-reversal invariant], we desire to know whether it is at bottom TRI" (Callender 1995: 333). Though this could steer the view to an unjustified distinction between "phenomenological laws" and "fundamental" ones (see Hutchinson 1993), the search for a structural arrow of time has in general endorsed the fundamentality condition as good enough. Indeed, the widely-held claim that physics is blind to the direction of time is primarily grounded in that *fundamental* dynamical laws are time-reversal invariant. This condition also links to the symmetry-to-fundamental inference: in discovering the structure of space-time, the relevant laws in premise (1) are only those free of any interaction and force. I think that the reason is clear: a violation of time symmetry in a free-interaction or free-force law can *only* come from the very structure of space-time.

For the question to be conceptually meaningful, a structural arrow of time must also meet another condition: "*contingency".* The formulation of the problem has to assume that there *might* be non-time-reversal-invariant fundamental laws. It must not be the case that every fundamental law is necessarily time-reversal invariant, otherwise, the question pursued by the problem of a structural arrow of time would make little sense. If fundamental dynamical laws *must be* time-reversal invariant, then it is completely pointless to ask *whether* a non-time-reversal fundamental law exists in physics. To be clear: contingency does not mean, obviously, that there *must* exist at least one non-time-reversal invariant fundamental law (a world where all our fundamental laws are actually time-reversal invariant is perfectly conceivable, and the question of the arrow of time would still be philosophically interesting), but that contingency requires that must be at least possible that some fundamental law is non-invariant[19].

Finally, there must be *co-extension* between the concept of time reversal (and time-reversal invariance) and its mathematical representation by means of a time-reversal operator, *T*. By co-extension I mean that there should be only one mathematical representation of the concept of time reversal within a theory. In short: a physical law *L* is time-reversal invariant if and only if it is *T*-invariant. This condition avoids multiple, and divergent, mathematical representations of the same concept of time reversal. In some respect, the very concept of time reversal is completely exhausted by its mathematical representation –time reversal *is* what *T* does. It would be devastating if a physical law *L* is *T*-invariant, but is not, say, *T'*-invariant, where *T* and *T'* intend to mathematically represent the same concept, time reversal. The reason is quite simple: if co-extension fails (that is, if there are at least two formal representations for the concept of time reversal) the sentence "A theory ***S*** is time-reversal invariant" might be true and false *at the same time* and in *the same respect*. If two time-reversal operators claim to formally represent time reversal in the same respect and one states that the theory ***S*** is time-reversal invariant (under its representation of time reversal, *T*), and the other that it is not (under its representation of time reversal, *T'*), then there would not be any way to define whether the sentence is true or false: it would be true if one chooses one of those operators, but false if one chooses the other. The solution is co-extensivity: one concept, one mathematical representation within a theory (at least). Consequently, one should discard one of the time-reversal operators as a fair representation of the concept of time reversal –it might rather be representing a *different* symmetry transformation.

Section 3. Differences between the two approaches to the problem

I have so far argued that there exist two alternative formulations of the problem of the arrow of time in physics: *the problem of the two realms* and the problem of a *structural* arrow of time. With these under our belts, let us now stress some differences. The main point I want to make

[19] I'm here assuming, partially, what John Earman holds about symmetries in general, to wit, that the usual understanding of symmetries principles in physics is that they are contingent (Earman 1989: 121). I don't know if this is actually the "usual understanding" in our days, but I do think that some metaphysical problems demand a contingency condition when posed in terms of symmetries. The problem of a fundamental arrow of time is one instance of this.

is that though both formulations give genuine arrows of time, they are quite different in some relevant features.

To begin with, they are obviously after two quite different kinds of arrows of time. This point links with what was said in Chapter I. The problem of the two realms concerns the physical bases of a non-structural arrow of time, in the sense that it is after some non-intrinsic physical property to distinguish between both time's directions. So-introduced difference typically has relied upon some extra-nomic property that characterizes a single solution (for instance, a highly unlikely initial condition). In general, this extra-nomic property overcomes the gap and explains why one observes what one actually observes. Instead of that, the problem of a structural arrow of time exclusively focus on intrinsic properties of space-time and concerns the whole class of solutions of a set of dynamical equations.

Secondly, there is also a *modal* difference between both approaches to the problem. This modal difference naturally follows from how time-reversal invariance and reversibility have been characterized in Chapter I. In short, whereas the problem of the two realms allows defining a physically *possible* arrow of time, the problem of a structural arrow of time seeks to stablish a physically *necessary* arrow of time. Let's take a closer look at this.

It is typically said that the content of a physical theory is given by the set of *possible* worlds of which the theory is true. Each possible world is a solution of the dynamical equations of the theory. There is some debate about how to zone the space of physically possible worlds according to a theory (see Ruetsche 2013: Chapter 1), but it is standardly assumed that physical laws do the job adequately (see Earman 1986b: 13). Hence, one can define physical possibility as follows

> A possible world *W* is *physically* possible according to a given theory *T* if the laws of *T* are satisfied in *W*

According to the possible-world semantics, a sentence is possibly true if it is true in at least one possible world, while is necessarily true if it is true in all its possible worlds. Mutatis mutandis, this supplies the basis for defining physical possibility and physical necessity as follows:

> *p* is physically possible according to a given theory *T* if and only if it is true in at least one physical possible world of *T*

> *p* is physically necessary according to a given theory *T* if and only if it is true in all physically possible worlds of *T*

Frequently, the truth-makers of physically *possible* descriptions are extra-nomic features of a theory that vary from solution to solution. That is, they are descriptions that are true only under certain extra-nomic conditions (for instance, special initial conditions) and only in some physically possible worlds that instantiate them. Differently, the truth-makers of physically *necessary* descriptions are generally nomic features. When a physical description is physically necessary, one is claiming that its truth stems from the physical laws of the theory at stake.

Let us now bring these pieces together. By a modal difference between both approaches to the problem I mean the following. The problem of a structural arrow of time seeks for a *necessary* arrow of time in the sense that *all* models (*all* possible worlds) of a dynamical equation exhibit a direction of time. The class of possible worlds W (the content of a theory) where a non-time-reversal invariant law *L* is true includes either worlds wherein the direction of time goes from past-to-future (W^f) or worlds wherein it goes from future-to-past (W^b), but not both. This amounts to claiming that a privileged direction of time is a *necessary* feature of a dynamical law *L* as it must hold, once picked it out, in all its possible worlds.

The problem of the two realms rather starts off assuming that the space of possible worlds of a theory has the structure $\mathrm{W} = \mathrm{W}^f + \mathrm{W}^b$, so that it is physically possible the theoretical existence of time-reversed worlds where the same time-reversal invariant laws hold. But, as repeatedly remarked, this does not imply that there is no way of distinguishing both directions of time in a *single* world, say, w_{21}, yielding that an arrow of time is *physically possible* according to a theory *T*. Such a distinction is not matter of the laws, but it might be matter of extra-nomic conditions of the physically possible world w_{21}.

All this clearly leads to consider that, though both formulations of the problem allow distinguishing between the past-to-future direction and the future-to-past direction, they do not do it in the same modal respect. The problem of the two realms seeks for the foundations of a *physically possible non-structural arrow of time* that at least holds in a single solution of the dynamical laws at stake (for instance, the actual world), regardless if there exists an alternative physically possible world, with unalike extra-nomic conditions, where there is no direction of time. The problem of a structural arrow of time is rather after a *physically necessary structural arrow of time* as it holds in all the physically possible worlds of a given law *L*; in virtue of this, the problem of the structural arrow of time is modally stronger that the problem of the two realms.

Another difference between both approaches concerns *the status of physical laws* in the argumentation. In brief, the problem of a structural arrow of time seems to be prone to adopt an ontologically more robust account of physical laws (non-Humean account) than the problem of the two realms. I would like to be quite cautious here: I'm not claiming that there is a closely-knit correlation between a structural arrow of time and non-Humean accounts of laws, and between a non-structural arrow of time and Humean accounts. What I do claim is that there seems to be certain affinity between those pairs of philosophical stances. This claim is not completely new. In a 2012 paper, Barry Loewer argues that there are two views on laws and on the arrow of time that go hand-in-hand: Maudlin's non-Human and non-reductionist account of laws and the search for a structural (or '*metaphysical*' in Loewer's vocabulary) arrow of time; and Albert's Humean-Lewisian account of laws and the entropic explanation of the arrow of time (via The Mentaculus). Loewer's association is a particular case involving only two positions. What I'm arguing for is that it can be generalized involving alternative ways in which the problem of the arrow of time can be introduced and addressed in physics. Let me explain it a little bit further.

It is well known that accounts of laws of nature are quite varied, and literature is massive about it. Broadly speaking, those accounts can be divided into Humean (or regularist) and non-Humean (or realist-neccesitarist). According to the former, there is no fundamental nomological modalities in the world, but just a "vast mosaic" of categorical properties and/or relations instantiated by primitive entities (particles, space-time points or fields) dwelling (or being) the whole space-time. This distribution is typically called the 'Humean mosaic' and what one calls 'laws of nature' is merely an epistemic systematization of its distribution and regularities (see Lewis 1973, 1994; Swartz 1985). Non-Humean philosophers rather claim that there is nomological necessity[20] in the world. All nomological relations displayed in our best physical theories manifest an intrinsic necessity in the natural world (Carroll 1994, Lange 2000, Maudlin 2007 for instance): whereas for Humeans laws of nature supervene upon the Humean Mosaic, for non-Humeans laws of nature govern and constrain matter and events.

[20] Non-humeans can also claim that necessity in the world actually resides in dispositional properties or causal powers (see Shoemaker 1980, Mumford 1998, Bird 1998). But I'm here mainly interested in laws of nature, so I won't consider that account.

One possible way to look at the relation between the formulations of the problem of the arrow of time and the status of laws of physics is the following. A non-structural arrow of time follows from, for instance, a particular distribution or conditions of a given Humean mosaic (as Albert's proposal in general holds). The same laws of nature are however able to systematize varied information compatible with other Humean mosaics, where a non-fundamental arrow of time might not be well-defined. That's why laws of physics do not play any relevant role in the problem of the two realms besides setting the starting point: they are compatible with situations one could not be interested in. Humeans usually want to explain what it is going on in the actual world, regarding *this* Humean mosaic, with this particular distribution, boundary and initial conditions: why do daily-experienced regularities look so temporally asymmetric? Well, Humeans –as Albert does– will come up with an explanation of this *actual* fact based on physics. Regardless what our systematization of such regularities may say to one about other possible world.

Contrarily, a structural arrow of time is not just a property of the actual world, but a property of all possible worlds wherein the laws of physics hold. For non-Humeans, necessary nomic relations between particulars are fundamental, so if those laws fail to be invariant under time reversal, that fact is also giving crucial information about what sort of nomic relations holds in the class of possible worlds of a given theory. Taking Maudlin's wording (Maudlin 2007), if laws of physics *produce* following states, *producing* a physical evolution, a failure of time-reversal invariance says that those laws are unable to produce states (evolutions) in both directions of time. To be clear, given as system s in a state $A(s)$, a law of nature under Maudlin's account takes such a state and transforms it producing a new state $B(s)$, $A(s) \rightarrow_L B(s)$. The $' \rightarrow_L '$ transition says something fundamental about how things in the world behave necessarily according to a theory to which $' \rightarrow_L '$ belongs. Therefore, if the law L is unable to produce the time-reversed transition (that is, to produce the time-reversed next state, $TB(s) \rightarrow_{TL} TA(s)$), then the law is also telling us that the picked direction of time necessarily matters in specifying the sort of states, transitions and evolutions that are possible for the theory. In this sense, the chosen direction of time is structural or intrinsic to the transition. A deflated account of laws would fall short in providing a rationale like this.

There are surely many nuances to be explained in more details. I shall come back to this point when addressing a particular case in quantum mechanics, so I shall now move onto a fourth difference.

One might think that both formulations of the problem run in parallel, and one can equally go with any of them conforming to one's philosophical tastes. Yet, as I have already mentioned in passing, there is a strong relation between both formulations: one of them already assumes an answer to the other, and thereby, it depends on it. In this sense, one of the problems is *prior to*, and conceptually *more fundamental than*, the other. At this point, it should be quite clear that the problem of the two realms already supposes that there is no structural (necessary, objective, fundamental, or intrinsic) arrow of time to the extent that the underlying dynamics is taken to be time-reversal invariant. Notwithstanding this, a distinction between the past-to-future direction and the future-to-past one can be drawn by other means. The problem is thereby to specify those other means, and solutions should come up with a theory of why experience looks so temporally asymmetric despite having a time-symmetric structure underlying. This amounts to a (negative) answer to the problem of a structural arrow of time. That's why I claim that the problem of a structural arrow of time is prior: whatever be the reason one comes to think that the philosophically interesting problem is to provide explanation that closes the gap between the micro-dynamics and the macro-dynamics, the very existence of such a gap (or tension) crucially depends upon having time-reversal invariant laws at the lower level, and therefore, lacking a structural arrow of time.

Final Remarks

The moral of this chapter can be put as following: when asked whether time is physically directed or not, one should be warned that this can be said in at least two ways: Is one asking whether time is *structurally* (intrinsically, fundamentally) directed? Or, is one rather asking whether physics, despite time in itself being symmetric, provides the mean to distinguish the two directions of time? Along this chapter, I've argued why these problems should be carefully distinguished, how they should be distinguished and on which grounds, and under which assumptions, they should be distinguished.

Nevertheless, all this has been so far conducted in a rather abstract, lofty way. To ask whether physics somehow distinguishes between two time's directions is more or less as asking whether Argentina is cold or warm. Well, it depends where. The question in physics is analogously meaningless if a particular theory is not previously singled out. The proper question should rather be whether *this particular theory* exhibits in one sense or the other a temporal asymmetry. The following chapters dive into these queries within *non-relativistic quantum mechanics.*

Part 2

Introduction

The last chapter introduced and developed two different approaches to the problem of the arrow of time, where one of them emerged as more conceptually prior than the other –the problem of a structural arrow of time. As exposed, this problem mainly concerns whether a given dynamical equation is time-reversal invariant. Or, to put it slightly differently, whether it manifests any preference for one direction of time instead of the opposite.

Nonetheless, so formulated the problem cannot take one too far: it needs be zoned and formulated within a concrete theoretical framework: Does *this* dynamical law *L* belonging to *this* theory *T* exhibit a preference for one direction of time since it turns out to be non-invariant under time reversal? This chapter seeks to focus such a question on the quantum realm, in particular, on what I shall call *standard quantum mechanics* (SQM henceforth). As the question of a structural arrow of time aims one's attention at dynamical equations of motion, the question then becomes whether SQM's set of dynamical equations is time-reversal invariant or not.

As widely known, there is no agreement upon what the content of such a set of equations actually is. This is mainly due to long-standing interpretative issues gravitating around the so-called *measurement problem* of quantum mechanics. Quantum mechanics comes equipped with a formalism that, by itself, is incapable of providing a clear-cut picture of what's going on when one carries out measurements involving quantum states. In particular, the wonderfully-designed piece of mathematics entails the existence of quantum systems in a *superposition* of states (that is, a linear combination of the eigenstates of a given observable) that evolve unitarily and deterministically according to the dynamical postulate of the theory. However, when measured, such states seem to undergo a different evolution as one always gets defined states and never states in a superposition. To come up with an explanation of what's going on when quantum systems are measured has proven to be one of the most intractable problems of quantum mechanics. Solutions of the measurement problem are often called "*interpretations* of quantum mechanics". They greatly vary in many aspects but, in general, they have followed two different paths. Many, on the one hand, believe that the mathematical formalism was incomplete, and thereby, it should be supplied with further dynamical conditions so as to get an appropriate physical theory accounting for the relationships between what is obtained after running an experiment and what is predicted by the SQM. Others rather think that SQM was overall right and wasn't in need of any further addition.

One could thus divide the set of dynamical equations of quantum mechanics into two subsets: those belonging to SQM properly, and those belonging to some interpretation that considers SQM to be incomplete. In this part, I will exclusively deal with SQM's dynamics putting aside any interpretation-dependent dynamics for the time being, which will have to wait to Part 3. Hence, I will circumscribe myself to the dynamical postulate of the theory shared common to *all* interpretations – the *Schrödinger equation*. It follows that if SQM exhibits any temporal biase, then it should be mirrored in the formal features of such an equation. Therefore, the question of a structural arrow of time becomes "whether the Schrödinger equation is time-reversal invariant".

This Part 2 is organized as follows. Chapter III begins by laying out the basics of SQM. By this, I will basically understand the formal machinery composed by the Schrödinger equation (the main dynamical postulate of the theory) and Born's rule (through which probabilities enter the theory), plus some basic required assumptions as to how to represent a quantum system, a state and the observables within a formal framework. Next, I will present the symmetry group of the theory –the *Galilean group*. One of the ways to assign fundamental dynamical variables to the mathematical formalism is by considering what quantities remain invariant under Galilei continuous space-time transformations. In this way, physical content goes into SQM and the kinematics is set. This presentation also seeks to get some acquaintance to symmetry-based reasoning and space-time transformations within quantum mechanics

Chapter IV is entirely devoted to the notion of time reversal in SQM. According to a widely-extended position in the field, time reversal must be formally implemented in terms of an *anti-unitary, anti-linear* time-reversal operator. SQM turns out to be invariant under thus-specified time-reversal operator, so it wouldn't exhibit any structural preference either by the past-to-future direction or by the future-to-past one. The general aim of this chapter is to put forward a quite different picture of what's really going on. I will first show that there are actually two alternative ways to formally implement time reversal in quantum mechanics: an *orthodox way* and a *heterodox one*. My overall argument is that, against the widely-extended belief, both approaches can be rightfully defended, though on different grounds. My final aim is to show that the formal implementation of time reversal in standard quantum mechanics presupposes (i) a certain metaphysical background that determines what time reversal is supposed to be in general, and (ii) a certain view of what role symmetries are supposed to play in physical theories. My argumentation will be threefold

(a) The orthodox view endorses a *relationalists metaphysics of time*, and consequently, represents time reversal in terms of motion reversal.

On this metaphysical ground, the orthodox approach emerges more naturally as the right way to formally implement time reversal in SQM. But,

(b) Alternatively, a heterodox approach can be defended as one shifts gears to a *substantivalist metaphysics of time*

On this view, time reversal is formally represented as a geometrical reflection of time in itself, fully captured by the unitary transformation $t \rightarrow -t$.

(c) Whereas orthodox view concludes that SQM is structurally time-reversal invariant, the heterodox view rather concludes that SQM is non-time-reversal invariant

It follows from this that both approaches draw incompatible conclusions as to whether SQM exhibits a *structural* preference for either one direction of time.

Chapter V addresses the relationship between these approaches to time reversal and the problem of a structural arrow of time. I will claim here the orthodox approach's inner mechanism of justification doesn't quite meet some required conditions for the problem of a structural arrow of time to be philosophically meaningful. Thus, the following question arises: could the orthodox approach address the philosophical problem of a structural arrow of time? I will claim that philosophers of time and philosophers of physics face a complex scenario in SQM, where something should be abandoned. To start with, they should endorse either the orthodox approach or the heterodox one. But, (a) if they go with the orthodox, some conditions to formulate the problem of a structural arrow of time are at risk; (b) if they go with the heterodox view, they should conclude that SQM is non-time-reversal invariant, thus standing at odds with what's been mostly held in the field and also upholding that there is indeed a structural arrow of time.

III.

Standard quantum mechanics (*SQM*)

SQM's formalism and symmetry group

Quantum mechanics intends to describe the natural world at a small scale. A *really* small scale. In this attempt, it is one of the most successful physical theories we have ever had. Its success eminently relies on the accurate description and prediction of physical systems' behavior at the atomic level, and thanks to this, on the enormous technological developments it made possible. But, the same time, quantum mechanics is extremely challenging and puzzling –if physical theories not only attempt to describe and to control the natural world, but also to *know* it in any relevant sense, quantum mechanics has hindered any unified view of what the natural world is like at the atomic scale. For this reason, deep philosophical issues gravitate around quantum mechanics since its first formulations, forcing us either to change our previously-conceived beliefs about the natural world, or to think through if something is not going fatally wrong at the very conceptual and formal foundations of the theory.

Though philosophers of physics strongly disagree on how the theory should be properly understood, physicists have been using quantum mechanics for almost a century. They daily learn how to apply it efficiently, how to make predictions and experiments with it. They have developed undreamed technology based on its principles, and also pushed the theory itself forward beyond its limits. So, despite some fierce divergences regarding what sort of world the theory portrays, there are nowadays some broadly-shared tenets that identify the formal machinery of the theory, though in a quite restricted sense. I will refer to this minimal formal machinery as *standard quantum mechanics* (SQM hereafter), but people also frequently refer to it as *ordinary* quantum mechanics (Ruetsche 2013), *bare* quantum mechanics (Wallace 2007) or simply *minimal* quantum mechanics (Schlosshauer 2006). These names try to highlight a pristine sense in which all we can talk about a non-relativistic version of quantum mechanics by referring to its formal apparatus, so that one can more or less get an univocal definition of quantum state, how to formally represent various elements of the theory, what is theory's logical structure, its symmetry group and so forth. This is perfectly fine as long as it is recognized that such a mathematical formalization (plus a minimal interpretation) remains absolutely silent about crucial aspects a full-blooded physical theory should say something about. This minimal sense of scientific theory says nothing, for instance, about what the quantum state refers to, neither about in which way the state of a quantum system relates to measurements' outcomes. To provide an answer to these questions, interpretations of such a silent formalism ought to come in. In any case, one is rapidly stepping beyond SQM's framework.

In this chapter, I shall review some basics elements of SQM without mentioning interpretational issues. It should be clear from the beginning, on the one hand, that such a minimal sense of theory doesn't constitute in any relevant sense a full-blooded physical theory attempting to know what the natural world is like. Notwithstanding, I believe that SQM is already a developed-enough theory that poses some highly relevant, and puzzling, questions about time's direction that have been largely overlooked in the specialized literature. Additionally, it should be also clear that SQM is not a completely interpretation-free formal apparatus. To begin with, quantum mechanics' formalism attempts to be *physical* after all, so it is already *partially* interpreted. This interpretative core may be generally uncontroversial, but already set some basic elements to constitute a minimal notion of what is meant by *physical* within the theory. It is not the case that anything goes within the theory, or that anything is equally allowed. This minimal sense of physical theory already gets rid of some formal elements in order to delimitate what is *physical* and what is physically *possible* according to the theory.

The chapter is articulated in two sections. In Section 1, SQM's formal apparatus is succinctly introduced. In Section 2, SQM's symmetry groups are laid out and used as a methodology to assign fundamental variables to items in the formalism (particularly, to self-adjoint operators). The notion of symmetry and space-time transformation is also presented.

Section 1. SQM's formalism

To understand the conceptual structure of quantum mechanics one needs to firstly see how notions as "state of a system" and "observable quantity" are represented within the theory, and secondly, what laws describe their behavior. That is, to specify its kinematics and its dynamics.

Unfortunately, there are some controversies around the dynamics of the theory across its different interpretations, so it cannot be univocally established within SQM. Despite this, all interpretations agree to at least include the Schrödinger equation in the list.

Though historically quantum mechanics came up in two versions (originally considered as rival theories), there are nowadays several formulations of the formal apparatus of quantum mechanics. Daniel Styer et al. (2002) identify such nine formulations, including the historical matrix mechanics of Heisenberg (Heisenberg 1925) and the wave mechanics of Schrödinger (Schrödinger 1926). There are also different approaches so as to obtain a *quantum* mechanics. For instance, one might start off either by *quantizing* a classical system, which primarily consists in turning classical variables into operators and finding their canonical commutation relations (the so-called *canonical quantization*, see Dirac 1965), or by defining the set of quantum observables in an algebraic fashion (von Neumann 1932/1955, Haag 1992). Be as it may, all these formulations were shown to be unitarily equivalent by John von Neumann (1931) and Marshall Stone (1930)[21], so their physical content is empirically equivalent. According to the so-called Stone-von Neumann theorem, they all are, so to speak, different formal ways to represent the *same theory*.

As just mentioned, quantum mechanics came up in two different ways. The early-twenty-century born theory aimed to provide a unified explanation of varied phenomena that didn't fit in a classical framework, like atomic spectra, radiation phenomena or scattering experiments. In 1925, Werner Heisenberg presented the first attempt towards that direction: the *matrix mechanics*. Seeking to recover Bohr-Sommerfeld's conditions for quantization, Heisenberg's proposal represents position and momentum by means of Hermitian matrices, Q and P respectively, based on the Poisson bracket formulation of classical mechanics. In this picture, observables (represented by matrices) play the leading role as they evolve in time while the state is left static. Few months later, in 1926, Erwin Schrödinger put forward a different theory aiming to reach the same goals than Heisenberg's: the *wave mechanics*. In contrast with the matrix mechanics, Schrödinger placed great emphasis on the evolution of the state, instead of measurable quantities (observables). He rested upon two differential equations, the time-independent and time-dependent Schrödinger equation, the latter describing the evolution of a

[21] To be precise, the Stone-von Neumann theorem showed that Heisenberg's matrix mechanics and Schrödinger's wave mechanics were two equivalent representations of the same theory (of the canonical commutations relations). Many other formulations appeared ever since, all of them equivalent thanks to the same theorem.

complex-valued wave function $\psi\ (x)$. In order to solve the equations, a differential operator *H* (the Hamiltonian) was introduced in the theory.

Setting aside historical details on its birth, present-day textbooks often introduce quantum mechanics in "wave-mechanics fashion", based mainly on the *Hilbert space formalism* and on the "Schrödinger picture", in which the notion of "quantum state" and its dynamical evolution play the leading role in the behavior of quantum systems. The Hilbert space formalism has a long story that goes back to von Neumann's and Dirac's world and has come to be frequently considered *the* formalism of quantum mechanics (see Redei 1998). Therefore, to start specifying a quantum-mechanical description of a physical system one has to provide, at least, three elements: (1) a formal representation of the physical system and its states, (2) some further structure on that representation in order to describe the physical system, and (3) a dynamics determining how the system evolves in time. I will take the formal apparatus of SQM as respecting the following tenets

- Any quantum physical system is formally represented by a complex vector space equipped with an inner product, and which is also complete: any converging sequence of vectors in the space converges to a vector in the space. This space is called the Hilbert space $\mathcal{H}$.
- Any quantum physical system has associated a state bearing information about the quantum system.
- A *quantum state* is represented by means of a density operator ρ, which inhabits the Liouville Space $\mathcal{L} = \mathcal{H} \otimes \mathcal{H}$
- Some quantum states are *pure states*. A pure state is represented by a state vector $|\psi\rangle \in \mathcal{H}$, such that the corresponding density operator is

$$\rho = |\psi\rangle\langle\psi| \qquad (3.1\ )$$

- Any pure state $|\psi\rangle$ (the so-called the wave-function) evolves according to the dynamical postulate of the theory –the Schrödinger equation

$$H|\psi\rangle = i\hbar\frac{\partial|\psi\rangle}{\partial t} \qquad (3.2\ )$$

Therefore, for any initial state $|\psi(0)\rangle$, it will evolve to $|\psi(t)\rangle$

$$|\psi(t)\rangle = e^{-Ht}|\psi(0)\rangle \tag{3.3 }$$

where *H* is the Hamiltonian operator.

- Generic quantum states evolve according to the von Neuman's version of the Schrödinger equation

$$|\psi(t)\rangle = e^{-Ht}|\psi(0)\rangle \tag{3.4 }$$

- Any state vector $|\psi\rangle$ can be represented in different bases of $\mathcal{H}$ (i.e. a set of basis vectors from which is possible to form all possible finite linear combinations of state vectors). Hence, any state vector $|\psi\rangle$ can be written down in terms of the basis $\{|a_i\rangle\}$ of $\mathcal{H}$

$$|\psi\rangle = \sum_{i=1}^{N} \alpha_i |a_i\rangle \tag{3.5 }$$

Where the α_i are complex numbers.

- Any physical magnitude *O*, called *observable*, is represented by means of a (linear) self-adjoint operator $\hat{A}$, which is a map acting on $\mathcal{H}$, $\hat{A}$: $\mathcal{H} \rightarrow \mathcal{H}$ such that for any $|\varphi\rangle, |\psi\rangle \in \mathcal{H}$ and any $\alpha_i, \beta_j \in \mathbb{C}$

$$\hat{A}(\alpha|\psi\rangle + \beta|\varphi\rangle) = \alpha(\mathrm{A}|\psi\rangle) + \beta(\mathrm{A}|\varphi\rangle) \tag{3.6 }$$

Given an operator $\hat{A}$, an eigenket of $\hat{A}$ is a ket $|\alpha\rangle$ in $\mathcal{H}$ upon which $\hat{A}$ acts by

$$\hat{A}|a\rangle = \alpha|a\rangle \tag{3.7 }$$

Where the complex number α is the eigenvalue of $|a\rangle$. Self-adjoint operator's eigenvalues represent the possible values of the magnitude.

- Two observables O_1 and O_2, represented by the self-adjoint operator $\hat{A}$ and $\hat{B}$ respectively) are said to commute if for any $|\psi\rangle \in \mathcal{H}$

$$\hat{A}\hat{B}|\psi\rangle = \hat{B}\hat{A}|\psi\rangle \tag{3.8 }$$

The same can be also expressed by means of the commutator

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} = 0 \quad (3.9)$$

If $\neq 0$, they are said to not commute.

- In SQM, there are some quantities that are canonical conjugated one another (to put it differently, one is the Fourier transformed of the other), and thereby they don't commute. The relations between these quantities have come to be known as the *canonical commutation relations* (CCRs), where for two canonical conjugated quantities $\hat{A}$ and $\hat{B}$

$$[\hat{A}, \hat{B}] = i\hbar \quad (3.10)$$

  This has been frequently interpreted in terms of *mutually incompatible quantities* or as representing incompatible experimental situations.
- In 1926, Max Born proposed a purely statistical interpretation of Schrödinger's wave-function. Beyond interpretation-dependent debates, his work gave rise to the Born's rule, whereby statistics enters quantum theory. Born's rule aims at calculating the probability of getting a certain outcome *if* an experiment were carried out, where such a probability is equal to the modulus square of the wave-function. In the Hilbert space formalism, the probability $P_\psi(a_i)$ of obtaining the eigenvalue a_i of the self-adjoint operator $\hat{A}$ can be computed through

$$P_\psi(a_i) = |\langle a_i|\psi\rangle|^2 \quad (3.11)$$

  Where $|a_i\rangle$ is the eigenvector of $\hat{A}$ corresponding to the eigenvalue a_i. For the case of a generic state ρ, the probability $P_\rho(a_i)$ is given by

$$P_\rho(a_i) = Tr(\rho\Pi_i) = Tr(\rho|a_i\rangle\langle a_i|) \quad (3.12)$$

These tenets constitute SQM's *formal* apparatus. However, as Ballentine stresses (1998: 63), it has by itself little physical content until dynamical variables of the systems are identified with items in the mathematical formalism, particularly, with self-adjoint operators. Much has been written about it, but one way to do it is based on the relations between some fundamental physical variables and space-time symmetries. Any physical theory has its own group of symmetries, that is, a set of transformation that leaves the structure of the theory unaltered (invariant). So, symmetry transformations serve to the purpose of identifying which items in the mathematical formalism (which operators) correspond to the relevant physical properties of systems (dynamical variables).

Section 2. Galilei Group and SQM's physical content

Laws of physics are believed to be symmetric under certain space-time symmetry transformations. These typically include displacing the system in the space or in time, or spatially rotating it. Under such transformations, *fundamental* quantities of the theory are expected to remain invariant, that is, they shouldn't suffer any change depending on which coordinate system one uses to describe them. This belief supposes that a coordinate system is a piece of surplus structure (Baker 2010) and shouldn't alter what is supposed to be *objective* (or *fundamental*), coordinate-independent features of physical systems. Physical quantities that do vary under a symmetry transformation are rather surplus structure, which depends on the coordinate system used to describe the system (for instance, by Belot 2001).

In this light, the strategy to identify fundamental quantities consists in formally representing these geometrical transformations and in testing how a physical theory's structure behaves under their application. Hence, each space-time symmetry operation g will correspond with a transformation of observables ($g: A \to A'$) and of states ($g: |\psi\rangle \to |\psi'\rangle$). Under such a transformation g, an equation of motion is said to be g-invariant (g-symmetric) when the truth of the law is preserved by the transformation (Baker 2010, Dasgupta 2016), that is, when g never takes a model in which the law is true as input and returns a model in which it is false as output (Earman 1989, Roberts 2008, Belot 2013). The concept of group, originally proposed by Galois in early nineteenth century, comes in to supplement the notion of symmetry: a group clusters different transformations into a specific structure[22].

Let's put it more formally. Consider a set $\mathcal{A}$ of objects $a_i \in \mathcal{A}$, a group $\mathcal{G}$ of transformations $g_\alpha \in \mathcal{G}$, where $g_\alpha: \mathcal{A} \to \mathcal{A}$ acts upon the a_i as $a_i \to a_i^{g_\alpha}$.

---

[22] Technically, a group $(\mathcal{G}, \to)$ is a finite or infinite set $\mathcal{G}$ of elements with a binary operation * (the group operation) that satisfies the following four axioms:

- *Closure*: For all g_α and g_β in $\mathcal{G}$, the result of $g_\alpha \to g_\beta$ is also in $\mathcal{G}$.
- *Associativity*: For all g_α, g_β and g_k in $\mathcal{G}$, $(g_\alpha \to g_\beta) \to g_k = g_\alpha \to (g_\beta \to g_k)$
- *Identity element*: There exists an element I in $\mathcal{G}$ such that for all g_α in $\mathcal{G}$, $I \to g_\alpha = g_\alpha \to I = g_\alpha$.
- *Inverse element*: For each g_α in $\mathcal{G}$, there exists an element g_β in $\mathcal{G}$ such that $g_\alpha \to g_\beta = (g_\beta \to g_\alpha = I$

Def. 1. An object $a_i \in \mathcal{A}$ is invariant under the transformation g_α if $a_i = a_i^{g_\alpha}$

In introducing the notion of group, one obtains

Def. 2. An object $a_i \in \mathcal{A}$ is invariant under the group of transformations $\mathcal{G}$ if it is invariant under all the transformations $g_\alpha \in \mathcal{G}$

In physics, the objects to which transformations apply are typically those representing states *s*, observables *O* (represented by self-adjoint operators $\hat{O}$) and differential operators *D*. Each transformation, in turn, acts upon them in a particular way, and each symmetry transformation will be defined depending on how it acts upon those objects.

Dynamical laws of physical theories are often represented in terms of differential equations, which combine the aforementioned objects. So, a dynamical law L is then represented by an equation $E(s, O_i, D_j) = 0$. It follows from what was said above, that a symmetry transformation g_α will act upon the equation by transforming $g_\alpha : \mathrm{s} \to \mathrm{s}^{g_\alpha}$, $g_\alpha: O_i \to O_i^{g_\alpha}$, and $g_\alpha: D_j \to D_j^{g_\alpha}$. Therefore, based on Def.1 and Def. 2, the following definitions can be introduced:

Def. 3. A dynamical law L is g_α-symmetric (i.e. invariant under the transformation g_α) if $E(s^{g_\alpha}, O_i^{g_\alpha}, D_j^{g_\alpha}) = 0$

Def. 4. A dynamical law L is symmetric under the group $\mathcal{G}$ if it is invariant under all the transformations $g_\alpha \in \mathcal{G}$

On this basis, the symmetry group of the theory can be thus defined as:

Def. 5. A group $\mathcal{G}$ of symmetry transformations is said to be the *symmetry group* of a theory if the laws of the theory are invariant under $\mathcal{G}$

In the model-theoretic vocabulary, John Earman (2004) defines a symmetry in terms of the models of a theory. Let $\mathcal{M}$ be the set of the models of a certain mathematical structure, and $\mathcal{M}_L \subset \mathcal{M}$ the subset of the models satisfying the law L. Then, in Earman's model-theoretical terms,

Def. 6. A symmetry of the law L is a map $g_\alpha : \mathcal{M} \to \mathcal{M}$, that preserves $\mathcal{M}_L$, that is, for any $m \in \mathcal{M}_L$, $g_\alpha(m) \in \mathcal{M}_L$ and $g_\alpha(m) = m$.[23]

When represented by a differential equation $E(s, O_i, D_j) = 0$, each model $m \in \mathcal{M}_L$ is represented by a solution $e = F(O_i, s_0)$, corresponding to a possible physical evolution of the system. It follows that

Def. 7. If L is g_α-symmetric (Def. 3), then if $e = F(O_i, s_0)$ is a solution (possible evolution) of L, then $e^{\,g_\alpha} = F^{\,g_\alpha}(O_i{}^{g_\alpha}, s_0)$ is a solution (possible evolution) of L and $e = e^{\,g_\alpha}$.

These definitions capture the more intuitive notion of symmetry in terms of *preserving* the truth of the law, or in terms of transforming solutions into solutions as mentioned above.

Putting aside for the moment any interpretational discussion about the dynamics of quantum mechanics, in *SQM* the evolution of a quantum state ($|\psi\rangle$) is given by the Schrödinger equation as mentioned early, so the above-mentioned definitions have to be now expressed in quantum-mechanical terms. Abstractly, one defines a generic symmetry transformation g_α acting upon the objects of the dynamical equation as $g_\alpha : |\psi\rangle \to |\psi\rangle^{g_\alpha}$, $g_\alpha : O_i \to O_i^{g_\alpha}$, $g_\alpha : \frac{\partial}{\partial t} \to \frac{\partial}{\partial t}^{g_\alpha}$, and $g_\alpha : i \to i^{g_\alpha}$. One thereby obtains that

Def. 8. The Schrödinger equation is invariant under the transformation g_α (it is g_α-symmetric) if $H|\psi\rangle = i\hbar \frac{\partial|\psi\rangle}{\partial t}$ is equal to

$$H^{g_\alpha}|\psi^{g_\alpha}\rangle = i^{g_\alpha}\hbar \frac{\partial|\psi\rangle^{g_\alpha}}{\partial t}$$

Therefore,

Def. 9. The Schrödinger equation is symmetric under the group $\mathcal{G}$ if it is invariant under all the transformations $g_\alpha \in \mathcal{G}$

SQM's symmetry group is the *Galilean group* $\mathcal{G}$, which basically consists of a set of continuous space-time transformations (time-displacement, space-displacement, space-rotation and boost) that leaves the theory unaltered. In interpreting SQM's formalism, one of

[23] It's worth warning the following. Literature is sometimes ambiguous about what one should demand to a symmetry. In some occasions, it's required the original model and the g_α-transformed one to belong to the set of solutions (possible evolutions) of a given equation. But, frequently, something else is also demanded: that the original model and the g_α-transformed one be *equals*. Note that the former is weaker than the later. I will here deal with invariance exclusively. For further details, see López and Lombardi (2019) where we distinguish both notions by calling the former requirement 'covariance' and the latter 'invariance'.

the Galilean group's utilities is that it serves to endow SQM with physical meaning by identifying those $\mathcal{G}$ -invariant objects with fundamental physical quantities. Representing time by the variable $t \in \mathbb{R}$ and position by $r = (x, y, z) \in \mathbb{R}$, the Galilean group $\mathcal{G} = \{T_\alpha\}$ is a group of continuous space-time transformations T_α: $\mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3 \times \mathbb{R}$ such that

Time displacement	$t \rightarrow t' = t + \tau$
Space displacement	$r \rightarrow r' = r + \rho$
Space rotation	$r \rightarrow r' = R_\theta r$
Velocity boost	$r \rightarrow r' = r + \boldsymbol{u}t$

Where $\tau \in \mathbb{R}$ is a real number representing a time interval $\tau = \Delta t$, $\rho = (\rho_x, \rho_y, \rho_z) \in \mathbb{R}^3$ is a triple of real numbers representing a space interval, $R_\theta \in \mathcal{M}^{3\times3}$ is a 3×3 matrix representing space rotation by an angle θ, and , $\boldsymbol{u} = (u_x, u_y, u_z) \in \mathbb{R}^3$ is a triple of real numbers representing a constant velocity.

Since the Galilean group $\mathcal{G}$ is a Lie group, the Galilean transformations T_α can be represented by unitary operators U_α over the Hilbert space $\mathcal{H}$, with the exponential parametrization $U_\alpha = e^{iK_\alpha s_\alpha}$, where s_α is a continuous parameter and K_α is a Hermitian operator independent of s_α and the *generator* of the transformation T_α. Thereby, $\mathcal{G}$ can be also defined by ten generators (K_α) of the group

- One time displacement K_τ
- Three space displacements K_ρ (one for each coordinate)
- Three space rotations K_θ
- Three velocity boosts K_u

The Galilean group can consequently be defined by the commutation relations between its generators

(a) $\left[K_{\rho_i}, K_{\rho_j}\right] = 0$

(b) $\left[K_{u_i}, K_{u_j}\right] = 0$

(c) $\left[K_{\theta_i}, K_{\theta_j}\right] = i\varepsilon_{ijk}K_{\theta_j}$

(d) $\left[K_{\theta_i}, K_{\rho_j}\right] = i\varepsilon_{ijk}K_{\rho_k}$

(e) $\left[K_{\theta_i}, K_{u_j}\right] = i\varepsilon_{ijk}K_{u_k}$

(f) $\left[K_{u_i}, K_{\rho_j}\right] = i\delta_{ij}M$

(g) $\left[K_{\rho_i}, K_\tau\right] = 0$

(h) $[K_{\theta_i}, K_\tau] = 0$

(i) $[K_{u_i}, K_\tau] = iK_{\rho_i}$

Where ε_{ijk} is the Levi-Civita tensor and M is the mass operator $M = mI$ (I is the identity operator and m the mass).

What has been presented so far is just the *geometrical* significance of the operators K_τ, K_ρ, K_θ, and K_u as generators of the symmetry transformations in a Hilbert space. However, no physical meaning has been introduced yet. This is done by ascribing to each operator K_α the representation of a dynamical physical magnitude. This methodology was first introduced by T. F. Jordan (1975) in replacement of the old "canonical quantization" based on the Poison brackets formulation of classical mechanics. For this symmetry-group-based methodology, the dynamics of a closed, constant-energy system free of any external field (i.e. a free particle) is the starting point as it turns out to be fully symmetric under the Galilean Group, being enough to identify operators with fundamental dynamical variables in the following way (for details, see Ballentine 1998: Chapter 2-3):

Energy	$H = hK_\tau$
Three momentum components	$P_i = hK_\rho$
Three angular momentum components	$J_i = hK_\theta$
Three boost components	$G_i = hK_u$

Therefore, the commutation relations are now reworded as:

(a) $[P_i, P_j] = 0$

(b) $[G_i, G_j] = 0$

(c) $[J_i, J_j] = i\varepsilon_{ijk}J_k$

(d) $[J_i, P_j] = i\varepsilon_{ijk}P_k$

(e) $[J_i, G_j] = i\varepsilon_{ijk}G_k$

(f) $[G_i, P_j] = i\delta_{ij}M$

(g) $[P_i, H] = 0$

(h) $[J_i, H] = 0$

(i) $[G_i, H] = iP_i$

It is worth mentioning one further condition to get an adequate interpretation of dynamical variables in *SQM*. The Hamiltonian represents the energy of the system and is the responsible of generating the (infinitesimal) time translation. The spectrum energy is supposed to be always positive, so the Hamiltonian must be *bounded* (from below), that is, $H \geq 0$. It has been largely argued that negative energy states remain undetected and that their existence would necessarily turn matter unstable: physics wouldn't be any longer possible. The Hamiltonian is thus required to be bounded, and to remain bounded after any symmetry transformation, if one wish to get a transformed *quantum* mechanics state.

Having said that, the rest of the physical magnitudes can be straightforwardly defined from those basic ones. For instance,

Three position components	$Q_i = \frac{G_i}{m}$
Three orbital angular momentum components	$L_i = \varepsilon_{ijk} Q_j P_k$
Three spin components	$S_i = J_i - L_i$

All this has to be now applied to the Schrödinger equation. The action of $g_\alpha \in \mathcal{G}$ upon states and observables in the Hilbert formulation can be alternatively expressed as

$$g_\alpha : |\psi\rangle \rightarrow |\psi\rangle^{g_\alpha} = U_{s_\alpha}|\psi\rangle = e^{iK_\alpha s_\alpha}|\psi\rangle \tag{3.13 }$$

$$g_\alpha : O_i \rightarrow O_i^{g_\alpha} = U_{s_\alpha} O U_{s_\alpha}^{-1} = e^{iK_\alpha s_\alpha} O e^{-iK_\alpha s_\alpha} \tag{3.14 }$$

Where the invariance of an observable O under a Galilean transformation amounts to the commutation between that observable O and the corresponding generator K_α.

From all this, one can somehow easily check in which way the Schrödinger equation turns out to be invariant under Galilean group's transformations. One typically begins by (1) pre-multiplying the two members by $U = e^{iKs}$, (2) by adding and subtracting $\left(\frac{dU}{dt}\right)|\psi\rangle$ to the first member, and (3) by using the property $U^{-1}U = 1$. Hence, one writes the Schrödinger equation down as:

$$U\frac{d|\psi\rangle}{dt} + \frac{dU}{dt}|\psi\rangle - \frac{dU}{dt}U^{-1}U|\psi\rangle = UiU^{-1}UHU^{-1}U|\psi\rangle \tag{3.15 }$$

And after applying any of the Galilean transformations as specified above (and recalling the transformations of states and observables), one obtains the equation

$$\frac{d|\psi\rangle^{g_\alpha}}{dt} - \frac{dU}{dt}U^{-1}|\psi\rangle^{g_\alpha} = -i^{g_\alpha}H^{g_\alpha}|\psi\rangle^{g_\alpha} \tag{3.16 }$$

Therefore, invariance can only be obtained if the time-derivative operator transforms as

$$\frac{d}{dt} \rightarrow \frac{d^{g_\alpha}}{dt} = \frac{D}{Dt} = \frac{d}{dt} - \frac{dU}{dt}U^{-1} \;\Rightarrow\; \frac{d|\psi\rangle^{g_\alpha}}{dt} = -i^{g_\alpha}H^{g_\alpha}|\psi\rangle^{g_\alpha} \qquad (3.17)$$

This means that the transformed differential operator $\frac{d^{g_\alpha}}{dt}$ is an invariant time-derivative D/Dt, which make the Schrödinger equation Galilean-invariant as defined above.

In a closed, constant-energy system free from any external fields, H is time-independent and P_i and J_i are constants of motion (see g, h above). Thus, one gets

For time translations $\frac{dU}{dt} = \frac{de^{iH\tau}}{dt} = 0$

For space translations $\frac{dU}{dt} = \frac{de^{iP_i\rho_i}}{dt} = 0$

For space rotations $\frac{dU}{dt} = \frac{de^{iJ_i\theta_i}}{dt} = 0$

The time-derivative turns out to be invariant under all these transformations and, consequently, the Schrödinger equation does too.

For boost transformation the argument needs some adjustments (see Lombardi, Castagnino and Ardenghi 2010 for a careful discussion about this). Briefll, the invariance of the Schrödinger equation implies that the differential operator transforms as $\frac{d}{dt} \rightarrow \frac{D}{Dt}$, which would in turn entail that its invariance under boosts amounts to a sort of "non-homogeneity" of time. So, to get a valid boost-invariant the Schrödinger equation, the transformed time-derivative needs be adjusted so as to compensate the time-dependent transformation of the state.

***

This brief analysis of SQM and its symmetry group showed how some space-time symmetries can be *heuristically* used to *physically* guide theory construction. One started off by considering some geometrical transformations already present in Hamiltonian classical mechanics under which the physics should remain invariant. These assumptions not only set elemental features of SQM, but also some basic features of space-time's underlying structure. In particular, the invariance under some space-time symmetry transformations was required so as to derive basic operators (standing for basic dynamical variables) of the theory. Hence, the Schrödinger equation must remain invariant under those Galilean transformations. Furthermore, space-time should be equipped with a structure conforming to those symmetries. So, one is entitled to consider that space is homogeneous as the Schrödinger equation is invariant under space translations. And so is time as long as the theory was imposed to be time-translation invariant. Relations of commutation or certain conservation principles also follow from taking the theory to be Galilean invariant (see Ballentine 1998: Chapter 3).

Therefore, the assumption of Galilean invariance seems to play a crucial role in the construction of SQM. If you have an isolated physical system, you require the physics describing it to remain invariant under displacing the system a couple of hours (days, years or whatever) forward in time. In fact, you expect your physics to work equally well today and in four weeks in the future, or here at your home as well as there in a lab placed in Hong Kong. Those assumptions seemed also to be defensible to the extent that, for instance, temporally or spatially displaced experiments have generated similar results (putting aside, naturally, different boundary conditions, the presence of external fields, and so on).

Basically, the symmetry group of SQM amounts to the equivalence among inertial reference frames (time-translated, space-translated, space-rotated or uniformly moving with respect to each other). In other words, Galilean transformations do not introduce any modification to the physical situation, but only express *a change in the perspective from which the system is described*. From a broader perspective, as viewed in the Part 1, this relates to the issue of what is the structure of a physical theory, that is, what is objective within the theory, and what elements would represent redundant structure. For the symmetry group of a theory, there seems to be intertwined criteria where certain (space-time) symmetries must be imposed beforehand so as to link basic dynamical variables with the mathematical formalism, while, simultaneously, a certain space-time's structure is being introduced to the theory. All this to get the basic pieces of a viable physical theory.

The symmetry group of a theory seems, certainly, to be necessary *to construct* the theory, that is, to define the fundamental physical magnitudes and to build the dynamical laws[24]. There are however many other symmetries that don't belong to the symmetry group of quantum mechanics. Neither do they share some of their formal properties. For instance, Galilei group's symmetries are continuous, but there are some other symmetries that are rather discrete. Must they be treated equally? Time reversal is one of those cases, so most of what has been just said about symmetries doesn't straightforwardly apply to it. While the above-mentioned

[24] Some have even argued that the symmetry group of the theory is *a priori* (see Dürr and Teufel 2009: 43-44).

explanation seemed to rightly underpin assumptions on Galilei group's symmetries, I think they shouldn't be straightforwardly and uncritically applied to time reversal as well.

Importantly, time reversal encloses even subtler features as it somehow relates to the nature of time and its unavoidable differences with respect to space: whereas one can move freely in all directions of space, it seems that one "moves" in just one direction of time, from past-to-future and never the other way around. Additionally, it is not easy to have a grip on its *geometrical* (and most intuitive) meaning. In the case of the symmetries considered above their respective transformations were given an intuitive geometrical meaning that was formally represented theoretically –one more or less knows what moving in time or space means. But, what does it mean to move *backward* in time? Particularly, what could mean inverting a *quantum* system backward in time? In the next chapter, I will get into these questions.

IV.

The Physics and Metaphysics of Time Reversal

Is the Schrödinger equation time-reversal invariant?

This chapter deals with time reversal in SQM. In particular, it addresses the following question: is the Schrödinger equation time-reversal invariant? According to the current leading approach, the answer is emphatically positive. Furthermore, the leading approach continues, one is advised that the formal representation of time reversal (the *quantum mechanical* time-reversal operator), besides transforming the variable t by $-t$, must meet some further conditions. This is typically condensed in the following dictum: time reversal in SQM can only be implemented by an anti-unitary, anti-linear time-reversal operator. So, according to this *orthodox* approach (OA hence after), the Schrödinger equation is time-reversal invariant because is invariant under a so-specified time-reversal operator.

Along this chapter, I will show that an alternative approach to time reversal in quantum mechanics can be coherently and soundly defended. This *heterodox*, or *heretic*, approach (HA from now on) rather holds that the Schrödinger equation is non-time-reversal invariant *because* it is non-invariant under a unitary, linear time-reversal operator. The theoretical viability of such an approach lies on the metaphysical considerations underlying the construction of the time-reversal operator. What I will, accordingly, argue is that OA heavily relies on a *relationalist metaphysics of time* plus some extra assumptions about time reversal and the role of symmetries in physics. Being all of these assumptions revisable, when one shifts gears and changes the metaphysical background or comes along with an unalike view on symmetries, OA loses much of its persuasive force. Consequently, HA significantly gains relevance, coherence, and feasibility.

The final aim of this chapter is thus to show that there at least exist two opposite approaches to time reversal in SQM, both equally plausible and equally defensible. One's inclinations for either one or another greatly rely upon one's metaphysical sympathies with respect to time, one's general view on time reversal, and one's considerations of the role of symmetries in physical theories. The chapter is thus organized as follows. Section 1 introduces OA and HA. In Section 2, I will present the most common and relevant arguments to support OA. Furthermore, I will argue that a relationalist metaphysics of time underlies them. In Section 3, I will develop HA further and will argue that, when one turns to a substantivalist metaphysics of time, OA's arguments can be neutralized[25]. Finally, in Section 4, I will explore some further assumptions that feed each approach. In particular, I will show that a *particularist, theory-dependent account* of time reversal plus a view taking *symmetries as guides to theory construction* greatly underpin OA, whereas a *universalist, theory-independent account* of time reversal plus a view of symmetries as *contingent properties of the dynamics* strength HA.

Section 1. Orthodoxy and Heresy in SQM: Two approaches to time reversal

The orthodox story typically runs as follow. Most physics textbooks (see for instance Gibson and Pollard 1976, or Ballentine 1998) commonly start off warning against, so to speak, classical expectations that time reversal be implemented by means of a *unitary* time-reversal operator (T_U henceforth), merely mapping $t \rightarrow -t$. In considering the Schrödinger equation (equation 3.1), textbooks typically bring up an *anti-unitary* operator (T_A) that not only does it flip t's sign, $T_A\colon t \rightarrow -t$.

$$TH|\psi\rangle = i\hbar \frac{\partial T|\psi\rangle}{\partial Tt} \rightarrow \; H|\psi\rangle = -i\hbar \frac{\partial |\psi\rangle}{\partial t} \qquad (4.1\,)$$

but also takes the complex conjugate (K).

$$KH|\psi\rangle = K(-i)\hbar \frac{\partial K|\psi\rangle}{\partial t} \qquad (4.2\,)$$

And this is how one gets the time-reversed the Schrödinger equation, wherein $H = H^*$

$$H|\psi^*\rangle = i\hbar \frac{\partial |\psi^*\rangle}{\partial t} \qquad (4.3\,)$$

25 Part of the material for Section 2 y 3 was published in López, C. (2019). "Roads to the past: how to go and *not* to go backward in time in quantum theories". *European Journal for Philosophy of Science*, 9: 27.

Evidently, the Schrödinger equation is *time-reversal* invariant if one takes T_A as the adequate time-reversal operator since 3.1 and 4.3 are symmetric. Note that this quantum-mechanical time-reversal operator is no longer a mapping from t to $-t$ but it features an anti-unitary transformation as well. The explicit factor *i* in the Schrödinger equation plus the requirement that it should leave the modulus of the overlap of two state vectors invariant would, prima facie explain, its anti-unitary form. Hence, T_A can be given with the so-called standard form $T_A = UK$, where U is a unitary operator (more on this below) and K the complex conjugation, $KzK^{-1} = z^*$, with respect to a standard representation. Notably, in the position representation, the time-reversal operator is formally nothing but the complex conjugation operator $T_A = K$.

It is worth explaining a few things further here. First of all, T_A is also demanded *by definition* to change the sign of momentum, $T_A \mathbf{P} {T_A}^{-1} = -\mathbf{P}$. For instance, Leslie Ballentine (1998) while introducing time reversal transformation in quantum mechanics, says

> "The effect of the time-reversal operator T is to reverse the linear and angular momentum while leaving the position unchanged. Thus we require, *by definition*,
> $T\mathbf{Q}T^{-1} = \mathbf{Q}$
> $T\mathbf{P}T^{-1} = -\mathbf{P}$
> $T\mathbf{J}T^{-1} = -\mathbf{J}$
> (Ballentine 1998: 377-378. Italics mine)

One should bear in mind that such a stipulation establishes a smooth continuity between the formal representations of time reversal in classical mechanics and in quantum mechanics: not only does the time-reversal operator transform magnitudes similarly, but it also keeps their fundamental equations of motion invariant. Indeed, the features of the unitary operator *U* are a consequence of the classical conditions for time reversal, namely, $T_A X \, {T_A}^{-1} = X$, $T_A \mathbf{P} \, {T_A}^{-1} = -\mathbf{P}$ and $T_A \sigma \, {T_A}^{-1} = -\sigma$ (see Sachs 1987: 34[26]).

Secondly, a so-defined time-reversal operator is clearly anti-linear and anti-unitary as it takes the complex conjugation and its inverse (X^{-1}) exists. Therefore, it preserves

$$\langle T_A \psi | \, T_A \varphi \rangle = \langle \psi | \varphi \rangle \tag{4.4}$$

[26] T is instead of T_A in the original

I will analyze some relevant arguments in favor of an anti-unitary representation shortly, but in general the literature repeats in different ways the same reasons and requirements an implementation of time reversal should meet.

Finally, under T_A the Schrödinger equation turnouts to be time-reversal invariant. Conforming to my explanation of time-reversal invariance (Chapter I, Section 6.2), this means that the class of solutions of the Schrödinger equation has the structure $\mathrm{W} = \mathrm{W}^f + \mathrm{W}^b$, that is, it produces pairs of time-symmetric evolutions. In particular, if there exists a quantum state evolving forward in time ($|\psi\rangle \in \mathrm{W}^f$), then there must exists too an evolution going backward in time $|\psi\rangle^T \in \mathrm{W}^b$, which is actually an evolution where t has been re-parametrized as $t \to -t$, and $|\psi\rangle^T$ is the complex conjugated state evolving backward, ($|\psi\rangle^*$).

In light of this result, it is commonly claimed that the Schrödinger equation in its simplest form (free fall particle, for instance) does not pick up a structural arrow of time and, thereby, one is allowed to conclude that, in SQM (at least when interpretations are set aside), time's structure is temporally symmetric. This conclusion straightforwardly follows from an instance of the symmetry-to-reality inference as explained in Chapter II, Section 2:

(1) the Schrödinger equation (*SE)* is the (unique) law governing the quantum world

(2) The direction of time ($+t$ or $-t$) is variant in *SE*

(3) Therefore, the direction of time is [interpretational link]

(4) Ockham's razor applied to the direction of time

(C) Therefore, *the direction of time* is not a structural property of the quantum world

Note that premise (2) is backed by the choice of T_A as a genuine formal representation of time reversal in quantum mechanics

So, I will take OA to time reversal in SQM as holding the following theses

OA

(a) Time reversal in SQM must be represented by a time-reversal operator whose form is given by the T_A-operator

(b) The Schrödinger equation is time-reversal invariant because is T_A-invariant

(c) Through the symmetry-to-reality inference, the Schrödinger equation is time-reversal invariant (T_A-invariant), it does not pick up a structural arrow of time.

There is however an alternative view on time reversal that characterizes it in terms of a *unitary* and *linear* time-reversal operator, T_U hereon. According to this approach, a time-reversal operator for quantum mechanics (and likely in any other physical theory) should not involve any other property over and above that of turning t's sign around, $t \rightarrow -t$. This view typically comes up in the literature in either a negative or positive way:

(a) as an unacceptable way to formally represent time reversal in quantum theories, which has to be dismissed in favor of the orthodoxy (see for instance Gasiorowicz 1966: 27), or

(b) as a positive defense of a "more genuine" and "broader" way to formally represent time reversal in physics, and thereby in quantum theories (see Albert 2000, Callender 2000. Costa de Beauregard 1980[27] also defends such a view in quantum field theory).

So, according to HA, when one applies to the Schrödinger equation a unitary time-reversal operator T_U that solely changes the sign of t and of all those quantities expressed as first-time derivative (or non-basic magnitudes, in Albert's vocabulary, see Albert 2000), one obtains the equation (1) again

$$TH|\psi\rangle = i\hbar\frac{\partial T|\psi\rangle}{\partial Tt} \rightarrow \; H|\psi\rangle = -i\hbar\frac{\partial|\psi\rangle}{\partial t} \tag{4.5 }$$

Where $T = T_U$. And that is all what one should expect time reversal to carry out. Period.

A couple of clarifications are in order here. To begin with, as opposed to OA, this heterodox way to define a time-reversal operator does not take into account whether it must change the sign of momentum or not. In fact, it does not, as I will explain shortly.

In the second place, the time-reversal operator is now a unitary and linear operator, T_U, as it does not take the complex conjugation on states. This feature is particularly problematic here for it is what produces a minus sign on the right side of the equation. In other words, as T_U is a unitary-linear operator, then $T_U\mathbf{H}{T_U}^{-1} = -H$, entailing that if $|\psi\rangle$ is an eigenstate of the Hamiltonian with energy E, then the temporally mirrored eigenstate ${T_U}^{-1}|\psi\rangle$ should involve negative energies $-E$. It is quite easy to see that there is an obvious asymmetry in (4.5)

[27] Beauregard also cites Watanabe 1965, and Jauch and Rohrlich 1955 as forefathers of the idea.

(indeed, a very deep and radical one) as one obtains no solution at all when time direction is thus-inverted.

Finally, it follows from above that the Schrödinger equation is non-time-reversal invariant, if T_U is taken as the right time-reversal operator. In contrast with OA, the class of solutions of the Schrödinger equation W now has either the structure $\mathrm{W} = \mathrm{W}^f$ or $\mathrm{W} = \mathrm{W}^b$, but not both. This means that if $|\psi\rangle$ is a solution evolving forward in time ($|\psi\rangle \in \mathrm{W}^f$), then $|\psi\rangle^T$ is not a solution of the Schrödinger equation as $|\psi\rangle^T \in \mathrm{W}^b$. The explanation is quite simple: since T_U inverts the sign of the Hamiltonian ($T_U \mathbf{H}\, T_U{}^{-1} = -H$), if $|\psi\rangle \in \mathrm{W}^f$ features a Hamiltonian bounded from below (interpreted as within the positive spectrum of energy), then its time-reversed wave function $|\psi\rangle^T \in \mathrm{W}^b$ will feature a Hamiltonian unbounded from below (interpreted as within the negative spectrum of energy). But as the Hamiltonian's spectrum must only feature positive energies, and thereby be bounded from below (or rather, have *some* bound), the T_U-transformation necessarily transforms solutions into non-solutions (see Chapter III, Section 1).

In this way, the Schrödinger equation, by its own, manifests a clear preference for one direction of time since it yields solutions either with time increasing or time decreasing, but not with both. As explained in Chapter II, this amounts to a claim for a structural arrow of time: the direction of time is not *a variant* feature in the Schrödinger equation to the extent to which the transformation from t to $-t$ does not leave the equation unchanged. Therefore, the symmetry-to-reality inference inevitably arrives at a positive conclusion: the direction of time *is* a structural property of the SQM in as much it is no longer the case one is positing redundant or non-objective structure.

I will thus take HA as claiming that

HA

(a) Time reversal in quantum theories must be represented by a time-reversal operator whose form is given by the T_U-operator

(b) The Schrödinger equation is non-time-reversal invariant as long as is non-T_U-invariant

(c) By (b) and through the symmetry-to-reality inference, the Schrödinger equation does manifest a structural arrow of time

At first glance, it seems that both approaches hold contradicting theses in so far as they are predicating contradictory properties of the same mathematical object. In fact, the Schrödinger equation cannot be at the same time and in the same respect time-reversal invariant *and* non-time-reversal invariant. If one wants to make sense to the idea of time reversal when applied to SQM, one of these approaches must be necessarily dropped.

I think that this is however too quick a conclusion and misses. To see why, recall that the time-reversal operator plays the role of implementing the concept of time reversal in a formal way. And this concept is intended to be spelled out in the way the time-reversal operator acts upon states and observables. Having said that, it is clear that both theses strictly resort to unalike mathematical representations of the action of inverting the direction of time. To put it in other words, they are not actually leading to contradictory theses; rather, they aim at reversing time in different ways. So, the true point of divergence is thesis (a) in each case: how the concept of time reversal must be formally implemented in SQM.

Formulated in this way, the apparent contradiction fades away, at least, taken in absolute terms. In some sense, both approaches express different things by 'time reversal', which is reflected by their formal implementations (T_U and T_A). Furthermore, the crucial point is now which operator entails a better representation of the concept of time reversal. I think that this point has been largely overlooked in the literature, and it is worth bearing it in mind through the following sections.

Section 2. Underpinning OA

Most reasons to hold OA rely upon giving further arguments to believe in its first thesis (a): T_A is the fairest way to formally realize time reversal in SQM. Despite OA being well-seated in the modern treatment of time reversal in SQM, a thorough analysis of its conceptual grounds is rarely found (exceptions: Sachs 1987 and Roberts 2017). The literature typically addresses the topic already assuming its obviousness. This section thus reconstructs some of the reasons to take OA (particularly, its (a) thesis) as the adequate approach to time reversal in SQM. In supporting (a), one is by the same maneuver grounding (b) and (c) as well.

I think there are at least three closely-related arguments to underpin OA, and to thereby cast HA aside. Indeed, the three share a reductio ad absurdum structure, differing in one of the premises. The first argument has been famously introduced by Eugene Wigner in 1932, and I will call it "two-time-evolution argument"; the second argument relies on the fact the

Hamiltonian's spectrum must remain bounded from below (that is, it must not involve negative energies), which I will call "the Hamiltonian-based argument"; and the third one establishes that momentum (in this case, the momentum operator, $\hat{P}$) must change its sign under time reversal, and I will refer to it as "the momentum-based argument".

2.1 The 'two-time-evolution' argument (or Wigner's general criterion for time-reversal)

Wigner's introduction to time reversal begins by posing a *general criterion* stating that time reversal is a transformation such that, when the following operations are sequentially performed, one obtains the identity. Informally,

time displacement by *t* × time reversal × time displacement by *t* × time reversal = ***I***

And more formally,

$$T\left[U_{\Delta t_2}T\left(U_{\Delta t_1}s_0\right)\right] = s_0 \qquad \text{(General criterion)}$$

Where s_0 is the initial state, and $\Delta t_1 = t_1 - t_2 = -t_2 - t_1 = \Delta t_2$

In general, this statement (as it stands) has been taken as claiming that the time-reversal transformation is an *involution*. That is, when applied twice it is equal to the identity. This mathematical sense of involution is naturally met by any operator that satisfies $X^2 = I$. However, OA is not just supposing that the time-reversal transformation is an involution in this sense, to the extent to which the original state is expected to be obtained after two time displacements (or time evolutions). To put it differently, OA (i.e. Wigner) expects the time-reversal transformation to obtain the original state one started with *after* producing a time evolution going in the future direction of time *and* producing a time-reversed evolution *going backward* in time. What this means is that time reversal ought to generate the second time evolution, that is, a *moving-backward* evolution, through which one should obtain the initial state again. And this is a stronger requirement that a time-reversal operator should satisfy.

So, an appropriate *time-reversal* operator in quantum mechanics must involve the adequate transformation to meet Wigner's general criterion for time reversal. More emphatically, to be a *time-reversal* operator *is* to be an operator that *generates* such a second time evolution going backward as specified above. Additionally, Wigner establishes that any so-defined T has to preserve transition probabilities as well (otherwise, the second time evolution would no longer be possible)

$$|\langle\psi|\varphi\rangle| = |\langle T\psi|T\varphi\rangle| \quad \text{(Preservation of transition probabilities)}$$

There is not much justification in the standard literature on time reversal in SQM of why time reversal ought to preserve the transition probabilities. In general, Wigner postulated that the transition probabilities between two states has an *invariant* physical sense, so any symmetry should preserve them. That is, *if* a symmetry holds, then transition probabilities must be preserved by the symmetry transformation. And in this line, Wigner argues that time-reversal invariance, *if* it holds, must preserve transition probabilities, too.

This condition is one of the ingredients to prove his famous theorem, which in an informal formulation states that a symmetry transformation is either unitary or anti-unitary. However, it has been pointed out that Wigner's overall proof of his theorem was "incomplete" (see Chevalier 2007: 429) and that a correct proof is given by U. Uhlhorn in 1962, who also generalizes the condition of preserving the probabilities. In a nutshell, Uhlhorn's proof replaces the preservation of the transition probabilities by the preservation of orthogonality, that is, that any pair of orthogonal states $\langle\psi,\varphi\rangle = 0$ remains orthogonal under a symmetry transformation S, $\langle S\psi, S\varphi\rangle = 0$. It follows from this that $\langle\psi,\varphi\rangle = \langle S\psi, S\varphi\rangle$ (see Chevalier 2007, Section 5, for a proof of Uhlhorn's theorem). For Chevalier, Uhlhorn generalizes Wigner's proof as he shows that a symmetry transformation preserves the *logical* structure of a quantum theory, as Uhlhorn himself states in the Introduction of his book.[28]

Following the same logic as before, *if* time-reversal invariance holds, then any orthogonal pair of states remains orthogonal under time reversal, that is, if $\langle\psi,\varphi\rangle = 0$, then $\langle T\psi, T\varphi\rangle = 0$. Roberts (2017) notes that Uhlhorn's theorem provides a general answer to why transition probabilities must be preserved under time reversal. His argument is quite simple: orthogonality has nothing to do with time reversal as it just relates to "what is possible in an experimental outcome, independently of their *time* development" (Roberts 2017: 321. Emphasis mine). So, why should one expect that something not related whatsoever to time (as two states being mutually exclusive) be modified by time reversal? As Roberts correctly points

[28] "We shall replace the requirement of the invariance of transition probabilities by the requirement that orthogonal vector rays are transformed into orthogonal vector rays, that is, incoherent states are transformed into incoherent states. By this definition, a symmetry transformation is a mapping preserving the logical structure of quantum mechanics, whereas the definition stated above corresponds to a mapping preserving the probabilistic structure of quantum mechanics" (Uhlhorn 1962: 308)

out, there is no reason for this. Therefore, time reversal ought to preserve orthogonality, and thereby, transition probabilities as well.

Importantly, Roberts stresses that the preservation of orthogonality and transition probabilities should be guaranteed by either an anti-unitary (like T_A) or a unitary time-reversal operator (like T_U). As was mentioned above, the idea is quite simple: if any of these operators bears any relation with *time* reversal, it shouldn't change a condition that is independent of time. *In principle,* I completely agree with this suggestion and I cannot think of any reason why a unitary time-reversal operator would contradict this. However, there are at least two assumptions in Robert's argument: first, and this one comes from Wigner himself, transition probabilities (or orthogonality in Uhlhorn's version) are preserved *if* the law is invariant under the symmetry transformation; (b) that the symmetry transformation still produces a *quantum mechanical* state. With respect to the former point, of course, a unitary symmetry transformation might still preserve orthogonality even if the time symmetry doesn't hold. With respect to the latter, in some very particular cases, it's not just the case the symmetry doesn't hold, but the symmetry transformation doesn't even produce quantum-mechanical states after being applied. And I suspect that something like this is going on here with T_U: the time-reversed state ($T_U|\psi\rangle$) is physically meaningless. Or differently, it's not even a quantum mechanical state. So, why should orthogonality be preserved if the very notion of quantum mechanical state is not even conserved?

To be clear, I think Roberts is quite right in pointing out that (i) if time-reversal symmetry holds, then orthogonality must be preserved, (ii) this is a more general justification of why time reversal must preserve transition probabilities, and (iii) preservation of orthogonality has nothing to do with time, so there would be, in principle, no reason to expect that time reversal changes this. *However*, time reversal does have to specify how states transform upon application and, in this particular case, T_U doesn't transform quantum mechanical states into quantum mechanical states. And one of the main reasons for this is given by the "Hamiltonian-based argument". I will come back to this in Section 3.1.

Let's briefly sum up the "two-evolution-based argument". Beyond this characterization of the *T*-operator in quantum theories, Wigner has so far just specified the general requirements a time-reversal operator must accomplish, but he has remained silent about the specific form of *T*. His famous theorem, as mentioned before in its general form, then proves that *T* must be either unitary (T_U) or anti-unitary (T_A). As T_U fails to satisfy Wigner's general criterion for time reversal, time reversal must be represented by T_A –tertium non datur. To be clear: T_U is discarded as an appropriate representation of time reversal just for it does not satisfy the very definition of time reversal expressed in the above-mentioned conditions. The structure of this reductio ad absurdum argument can be then sketched as follows:

(1) Assume that T_U fairly represents time reversal.

(2a) If T_U fairly represents time reversal, T_U must generate a second time evolution going backward (meeting the general criterion)

(3a) As matter of fact, T_U does not generate the second time evolution in Wigner's general criterion for time reversal.

(C) T_U does not fairly represent time reversal

As long as the form of T can only be either unitary or anti-unitary, it straightforwardly follows that only T_A represents time reversal.

2.2. The Hamiltonian-based argument

Let's move on to the *Hamiltonian-based argument*. This argument relies on an already-mentioned natural assumption (see Chapter III: Section 2): the Hamiltonian's spectrum must remain bounded from below (that is, it must not involve negative energies). Schematically represented,

(1) Assume that T_U fairly represents time reversal

(2b) If T_U fairly represents time reversal, T_U must keep the Hamiltonian invariant

(3b) As matter of fact, T_U does not keep the Hamiltonian invariant as $T_U\mathbf{H}T_U{}^{-1} = -H$

(C) T_U does not fairly represent time reversal

Let me spell premises (2b) and (3b) out. One of the reasons of why the Hamiltonian must remain invariant under time reversal is that, otherwise, an time-reversed evolution would no longer be possible: in order to be able to represent a *quantum* system in a backward-moving evolution, its Hamiltonian must remain within the positive spectrum at any cost (just to provide some references, see Gasiorowicz 1966: 27, Gibson and Pollard 1976: 78, Sachs 1987: 36). As mentioned early, quantum states are said to be "physically meaningless" (in the light of SQM's principles) when they are eigenstates of the Hamiltonian whose spectrum is "unbounded" from below. Putting it drastically, unbounded-from-below Hamiltonians must *not* even be considered as *quantum-mechanics* systems. In a classical textbook, Stephen Gasiorowicz says

> "...we find that [the Schrödinger equation of motion] can be invariant only if
>
> $$THT^{-1} = -H$$
>
> This, however, is an unacceptable condition, because time reversal cannot change the spectrum of *H*, which consists of positive energies only. If *T* is taken to be anti-unitary [….] the trouble does not occur." (1966: 27).

It is worth clarifying that the predicates "positive" or "negative" for the energy spectrum, or "unbounded from below/from above" for Hamiltonians are actually a matter of convention. So, the argument could not hinge on which predicate one adopts to describe the system properly. The real problem is not exactly whether the Hamiltonian is unbounded from below. The problem is if one starts with a Hamiltonian unbounded from above (but bounded from below) *and* ends up with a Hamiltonian unbounded from below (but bounded from above) after a transformation. In some sense, the problem is whether there is a bound *at all*. More precisely, a *specific* Hamiltonian must be bounded (either from above or from below), and the problem would come up if one adopts a transformation that turns a Hamiltonian unbounded from above (bounded from below) into a Hamiltonian unbounded from below (bounded from above), so that Hamiltonians (in general) could adopt either of the bounds.

I will take for granted that a Hamiltonian unbounded from below (but bounded from above and typically interpreted as within the negative spectrum of energy) is physically meaningless, as it is a sort of principle of SQM. As T_U inevitably transforms the system's energy as $T_U\mathbf{H}{T_U}^{-1} = -H$, it follows that it fails to represent time reversal as specified by OA.

Roberts (2017) offers more solid grounds for accepting premises (2b) and (3b). He begins by claiming that "all known Hamiltonians describing realistic quantum systems are bounded from below, which we will express by choosing a lower bound of $0 \leq (\psi, H\psi)$" (2017: 326), and this fact seems to be promoted to a general condition that a time-reversal operator must meet in order for its acceptability, meaning that $\langle\psi, H\psi\rangle$ and $\langle T\psi, TH\psi\rangle$ are both non-negatives. Later on, Roberts demands that there is at least "one realistic dynamical system" that satisfies time-reversal invariance in the sense that satisfies $Te^{itH}\psi = e^{-itH}T\psi$. Roberts makes the point that the time-reversal operator is demanded to be anti-unitary in order to meet these requirements, so for reduction, he assumes that such an operator is unitary. This leads to $itH = -itTHT^{-1}$ and thus to $T\mathbf{H}T^{-1} = -H$. What one finally gets is $0 \leq \langle\psi, H\psi\rangle = -\langle T\psi, TH\psi\rangle \leq$

0 and this result leads to either accept that the Hamiltonian is unbounded from below (what he had previously ruled out) or that the Hamiltonian is the operator zero, which renders triviality. Therefore, by reductio, the time-reversal operator cannot be unitary but anti-unitary.

2.3. The momentum-based argument

Finally, the *momentum-based argument* concerns how time reversal is supposed to act upon certain operators, in particular, the momentum operator. It can be sketched as follows

(1) Assume that T_U fairly represents time reversal.

(2c) If T_U fairly represents time reversal, T_U must transform momentum as $T\mathbf{P}T^{-1} = -\mathbf{P}$

(3c) As matter of fact, T_U leaves momentum invariant, $T_U\mathbf{P}{T_U}^{-1} = \mathbf{P}$

(C) T_U does not fairly represents time reversal

For the most part, the reasons why the sign of momentum should change under time reversal is somewhat unclear in the literature[29]. On the one hand, the reasoning seems to take roots in an analogy with time reversal in classical mechanics. Robert Sachs for instance imposes that "[time reversal must] conform to the requirements of the correspondence principle –namely, operators representing classical kinematic observables must transform under T in a manner corresponding to classical motion reversal." (1987: 34). For Leslie Ballentine, as was shown, time reversal flips the sign of momentum in quantum mechanics *by definition* (1998: 377-378), and, in the same vein, Albert Messiah (1966) simply defines time reversal as transforming r (position) and p (momentum) into r and $-p$ respectively (see Davies 1974: 24-25 for an akin definition).:

> "We are thus led to define a transformation of dynamical variables and dynamical states, which we shall call *time reversal*, in which r and p transform respectively into r and $-p$. ... [T_A] obviously satisfies [the] relations. Therefore, we may take [T_A] as our time reversal operator" (Messiah 1966: 667).

Roberts mentions that "there is a *natural* perspective on the *nature* of time according to which quantities like momentum and spin really do change sign when time-reversed" (2017: 317, italics mine).

[29] Although Bryan Roberts (2017, 2018) has recently offered a thorough justification for this argument. I'll present it and discus it later on (Section 3.3).

Let's take granted for now that it is indeed natural (or intuitive) that the momentum operator changes its sign under time reversal in SQM. In Section 3.3 I will come back to this with a more careful analysis and some philosophical discussion.

2.4. An underlying relationalist metaphysics of time

The above-introduced arguments intend to establish the form of *T* mainly based on formal and physical reasons. Next, I will show that metaphysical commitments with respect to the nature of time underlie and support them as well. Metaphysics comes first in the sense that determines not only what *time* reversal actually *is* but also *upon what* it is supposed to act. And this fact, to my knowledge, has been to great extent overlooked in the current literature on time reversal.

In Chapter I, I introduced three metaphysical views about the nature of time that would play an active role in determining what the problem of the arrow of time is supposed to be about. I will now argue that a relationalist metaphysics of time (as described in Chapter I) underlies OA and its arguments. Recall that I have taken relationalism as holding the following theses:

R1 There are only events or physical bodies in the world (which can have intrinsic properties or not), and their (spatio) temporal relations. There is no external time.

R2 Time is nothing but change. The sort of relation between the physical world and the concept of 'time' is that of *Leibnizian representation* or *Machian-abstraction*: time is an ideal, unreal entity parasitic on events-things' changing.

Therefore, according to these tenets, the variable *t* occurring in the majority of physical theories (setting aside general relativity) is merely an external unreal parameter, which should not be taken as representing something with physical meaning. I am particularly interested in showing how this metaphysical background underlies the defense and characterization of time reversal as T_A in SMQ.

To start with, the $T: t \to -t$ transformation must not be taken too seriously. It would be naïve to take $t \to -t$ as performing a physically relevant action upon dynamical equations. Instead, time reversal should be considered as a "shortcut" standing for a bunch of dynamically relevant transformations. As one is mainly interested in equations of *motion*, the physical meaning of time reversal is entirely exhausted by the dynamically relevant transformations relating to the motion of a system. In a nutshell, time reversal is *nothing but* motion reversal, and thereby the time-reversal transformation should be explicated as a bunch of dynamically relevant transformations that reverses the original direction of the motion. Think of a classical particle moving from point x_1 to x_2 in the time interval $\Delta t = t_2 - t_1$. Saying "I will time reverse the system by applying T" is simply a shortcut for "I will reverse the motion of the system by applying a bunch of dynamically relevant transformations that take the system back to the initial state". *Assuming* that the underlying dynamics in this toy example is classical mechanics, the T-operator should be properly spelt out into the momentum transformation $P: p \to -p$, and the position transformation $X: x \to x$. The transformation $T: t \to -t$ is physically meaningless, and must be merely regarded as a simple re-parametrization of the variable *t*. The *physically* relevant content of time reversal is fully exhausted by the momentum transformation and the position transformation as specified above.

In the light of this, the theory-dependent nature of the time-reversal operator becomes clearer. The dynamically relevant variables related to motion change from theory to theory, so the features of *T* must also change accordingly. That is why the mathematical form of *T* must be in each case figured out by identifying the adequate magnitudes related to spatial and temporal translation so as to reverse them properly (i.e. so as to get the state one began with). The defense of T_A in SQM is just an instance of this broader perspective. Thus, a time-reversal operator from a relationalist metaphysics of time is defined in general as follows

T_{Rel} (a) A physically meaningless re-parametrization of *t* by $T: t \to -t$

(b) A change of all dynamically relevant magnitudes so as to generate a moving-backward system (*a time-reversed evolution going backward to the original state*), which is expressed by extensionally specifying the bunch of dynamically relevant transformations.

This relationalist metaphysical background undoubtedly underlies Wigner's general criterion for time reversal. According to it, a fair characterization of the *T*-operator crucially depends on changing the dynamically relevant variables so as to *guarantee* the existence of the second time translation after applying the first time-reversal transformation. Any intended time-reversal transformation that fails to yield the second time translation will *by definition* be flawed, precisely because it fails to generate a backward-headed movement. The very existence of the second time translation is what precisely ensures that one is really applying a *time (motion)*-reversal transformation correctly. Therefore, the metaphysical reason to discard T_U as a *time*-reversal operator is that it fails to generate the existence of the second time translation, and to thereby reverse motion. In other words, T_U fails to be a *time*-reversal operator because it fails to be a *motion*-reversal operator. Most specialized textbooks take this point for granted in claiming that time reversal is nothing but *motion* reversal. Wigner in passing says that "reversal of the direction of motion" would be a more felicitous expression than time reversal (Wigner 1932: 325). Along the same line, Gibson and Pollard (1976) clarify

> "In this approach we see that no metaphysical notion of reversal of the direction of the flow of time is involved. We are led to consider time reversed processes but not reversal time itself. Although motion reversal and motion reversal invariance would be better names, we shall adhere to the accepted, if imprecise, usage" (1976: 177)

In a modern textbook on quantum mechanics, Ballentine categorically affirms

> "The term "time reversal" is misleading, and the operation that is the subject of this section would be more accurately described as motion reversal. We shall continue to use the traditional but less accurate expression "time reversal", because it is so firmly entrenched" (1998: 377)

Strictly speaking, the notion of time reversal is neither misleading, nor metaphysically misguided: as time is just an abstraction at which one arrives by means of motion (paraphrasing Mach's expression, 1919: 224), time reversal is simply an abstraction of motion reversal. For the relationalist, there is no time reversal symmetry besides motion reversal, but time reversal and motion reversal are one and the same thing.

What about time-reversal *invariance*? The symmetry of time reversal sheds light on the structure of the change in physical theories: there is no further structure of time outside of the structure of change. Time-reversal invariance mainly regards whether physical theories imply that change (as governed by a dynamical law) is necessarily directed. As was shown before, relationalism does not commit a priori to a particular time structure, but this should be unveiled by means of varied time symmetries (time translation, time reversal, time relabeling, and so on). Yet, the central point to be stressed here is that previously-assumed metaphysical commitments with respect to the nature of time are what prescribe *upon what* time reversal is supposed to act so as to generate a moving-backward time translation. As time is considered non-physical, the time-reversal transformation has to act upon dynamically relevant magnitudes in a certain way, and the time-reversal operator *is* essentially the bunch of dynamically relevant transformations that guarantee a reversion of the direction of motion.

Section 3. Shifting gears: a *substantivalist* metaphysics of time for HA and how to overcome OA's arguments.

If someone wants to bring HA back, she firstly ought to neutralize the three arguments presented in the last section. A plausible strategy would be to put forward an alternative metaphysical background: if a relationalist metaphysics underlies, and motivates, OA's arguments, one could instead evaluate them from a different metaphysical stance. In particular, she needs to show that her definition of time reversal is not necessarily committed to accept either (2a), (2b) or (2c), and an alternative metaphysics of time could well do the job. To be clear, what has been proved so far is that *if* a time-reversal operator meets the aforementioned requirements, then it must be an anti-unitary time-reversal like T_A. However, this is a conditional statement: no reasons have been so far provided (beyond the relationalist motivation) to support the requirement that any time-reversal operator in SQM *must* satisfy premises (2a), (2b) and (2c), but if it does, then it is anti-unitary.

Hence, HA's supporters might complain and argue that such assumptions are not legitimate. In this section, I shall introduce some arguments that HA's supporters might put forward in order to neutralize OA's. The maneuver is, I think, primarily metaphysical: If a relationalist metaphysics of time underlies OA's arguments for T_A, when one shifts gears and turns to an alternative metaphysical background, OA's arguments could in fact lose its dispositive force. This section aims to explore such an alternative scenario: Suppose now that one has sound reasons to decline a relationalist basis for time. How should time reversal be now metaphysically and formally characterized? Particularly, as long as the metaphysical background changes, do OA's arguments still stand or rather fall along with it?

In this section, I will argue that the *substantivalist* may endorse a different characterization of time reversal, and thereby, an alternative formalization thereof. As exposed in Chapter I, I will take substantivalism as the position supporting the following two theses:

S1 Time is a theoretical entity endowed with a structure that is *intrinsic* to it, and independent of change. Temporal relations among events or things are parasitic on this theoretical entity.

S2 Time is not an ideal or representational notion, but it plays a physically meaningful role so as to explain different phenomena or to define dynamical variables. Time cannot thereby be boiled down to a dynamical basis.

In the light of this view, how must time reversal, thus, be characterized? On the one hand, as time stands by itself, there are no metaphysical reasons for time reversal to be *explicated* as a bunch of dynamically relevant transformations; quite to the contrary, time reversal is metaphysically and conceptually prior to any other dynamically relevant transformation, and the latter must be specified in function of the former. In which way other magnitudes behave under time reversal follows from what sort of physical and formal relations they hold with time within a specific theory (e.g. whether they are either a first- or second-time derivative, and so on). On the other, as time is now a physically meaningful external parameter, an inversion of the direction of time must mainly mean an inversion of the external parameter itself. This view of time reversal seems to be what Jill North has in mind when she claims:

> "What is a time reversal transformation? Just a flipping of the direction of time! That is all there is to a transformation that changes how things are with respect to time: change the direction of time *itself*" (North 2009: 212. Emphasis added)

The meaning of time reversal is therefore completely exhausted by the transformation $T: t \rightarrow -t$; and, as said above, the rest of dynamical transformations supervene on it. By assumption, this transformation has physical relevance and is not to be considered a non-physical re-parametrization as the relationalist does. A substantivalist *time-reversal* operator can then be defined as follows

T_{Sub} (a) A change of the direction of time $T: t \rightarrow -t$

(b) A change of all magnitudes that are expressed or represented in function of time within the theory

The property (a) is all what one expects a time-reversal transformation must perform from a substantivalist stance. The property (b) is a corollary of (a): whenever you find a magnitude expressed o represented as a first-time derivative within the theory, change its sign accordingly. This property would have two reasons: first, it straightforwardly and formally follows from applying the transformation $t \rightarrow -t$ wherever a t appears in an equation of motion. If my physical theory involves a magnitude where time is a first-derivative, then it does not matter what the magnitude is and how it can be interpreted: the transformation $t \rightarrow -t$ would induce that such a magnitude switches its sign accordingly. Second, if a magnitude is expressed in function of time, it means that the theory supposes that such a magnitude and time stand in a close relation and, thereby, any change of the direction of time should have some effect on that magnitude[30]. This last point may be also expressed in terms of *supervenience*: any change in the variable's sign *supervenes* upon $t \to -t$.

What about time reversal in SQM? To begin with, the T-operator should no longer be required to produce an evolution going backward in time to the original state. Metaphysically, this is clear from a substantivalist metaphysics as defined above: if time is prior and independent of change (or movement), then a specific behavior of change (as a moving-backward physical system) cannot define, or be equivalent to, a reversion of time. The situation is indeed the opposite: time reversal is expressible independently of an inversion of the direction of motion. To define the direction of time in terms of the direction of motion would be as putting the cart before the horse. Under specific circumstances, an inversion of the direction of time *may* lead to a time-reversed evolution that restores the original state; for instance, when the dynamically relevant magnitudes that define the state and/or the direction of motion are first-time derivatives (as velocity in Newtonian classical mechanics). But this is not a desideratum to be held universally and necessarily. It might be the case that time reversal fails to generate such a moving-backward evolution, and then that a change of the direction of time would not lead to an inversion of the direction of motion (as it happens in SQM with T_U). Consequently, most reasons to hold (2a), (2b), and (2c) look less compelling. Let's address each argument in tandem

3.1. Neutralizing the two-time-evolution argument

Under a substantivalist view, the two-time-evolution argument no longer runs so smoothly as before. Particularly, because the premise (2a) does not hold to be true in the light of any instantiation of T_{Sub}. According to Wigner's general criterion for time reversal, the existence of the second time translation must be guaranteed to set the form of the time-reversal operator properly. One of the reasons to discard T_U was exactly it failed to meet such a criterion, but this rationale no longer runs when a substantivalist view on time is assumed instead. In some sense, T_{Sub} is at odds with Wigner's overall characterization of time reversal because it demands further actions and structure to represent time reversal which are actually necessary

---

[30] Suppose that one knows nothing about velocity in Newtonian classical mechanics, just it is defined as $v = \frac{\Delta d}{\Delta t}$. From here, one can infer that, whatever a velocity is, it has some relation with time. So, now one is told to perform the operation $t \to -t$. What would one expect to happen with v after the transformation? By just following the standard transformation rules, one would say that T induces the transformation $v \to -v$, since $Tv = \frac{T\Delta d}{T\Delta t} = \frac{\Delta d}{-\Delta t} = -v$. I think this is the intuition behind the substantivalist definition of time reversal and how it should transform other magnitudes. The point is that magnitudes should only switch their sign if there is an explicit relation with time in its formal expression within a specific theory.

conforming to **S1**, **S2** and T_{Sub}. Under this view, time reversal looks much more like a *reflection* than a mechanism to generate a time evolution going backward towards the initial state. Even though any instantiation of T_{Sub} (as T_U) would be an involution in the mathematical sense, this doesn't imply that *any* application of an involution must also yield an identity when there are two temporal evolutions in the middle. To guarantee this, further structure ought to be imposed on the time-reversal transformation, and that's why any relationalist imposes further conditions on time reversal. But, and this is the point of my argument, the substantivalist might be wary that a reflection needs those requirements. Hence, T_U can be re-established on the proper metaphysical grounds to the extent that fairly represents such a reflection (see Arntzenius 1997 and Savitt 1996: chapter 1 for time reversal as a reflection).

Let me put it a bit differently. Wigner's argumentation invokes two explicit premises: that a suitable time-reversal transformation must be able to restore "the system to its original state" (1932: 326), and that time inversion must flip the *direction of motion* to compensate for the twofold application of *T*. But there is also one *implicit* assumption in requiring that time reversal meets Wigner's general criterion: for a time-reversal operator to be well-defined, the second time translation (from t_2 to t_1) must also evolve according to the Schrödinger equation. Now, suppose that a quantum state $|\psi\rangle$ evolves from t_1 to t_2, according to the Schrödinger equation (first time translation in Wigner's general criterion). At t_2, time reversal is applied upon, according to Wigner, the evolved state. *If* the time-reversal transformation is well-defined, *then* the time-evolved and time-reversed state should evolve from t_2 to t_1 conforming to the Schrödinger equation (second time translation with $|\psi^*\rangle$), which would *intuitively* be equivalent to making the system evolve back with $-t$. According to Wigner, the operation to be applied at t_2 *must* be of such a kind that yields a physically meaningful moving-backward evolution, otherwise, it wouldn't be a solution of the Schrödinger equation.

Yet, is such a structure required to formally represent time reversal? The answer is "Yes" if time reversal *means* motion reversal; but the answer might rather be "No" if one believes that there is no reason to assume that time reversal necessarily means or involves motion reversal. In fact, Wigner's characterization of time reversal is actually a characterization of *motion* reversal as he mentions himself (1932: 325). An HA's substantivalist supporter could argue that demanding time reversal to feature the required structure to represent motion reversal is excessive because time reversal is just supposed to perform the *reflection* $t \rightarrow -t$, and not to grant that the system goes back to its initial state.

So, an available strategy for HA's supporters is that T_A is actually representing a different symmetry transformation than that of time reversal, which is properly represented by T_U. But this strategy introduces a clear-cut distinction between motion reversal and time reversal, which could only rely on a substantivalist background. In this light, a definition of time reversal cannot be boiled down to an inversion of the order of events (or motion).

To sum up. One available strategy to neutralize the two-time-evolution argument is by distinguishing between motion reversal and time reversal at a formal level: T_A simply fails to represent time reversal because it represents a different symmetry transformation –motion reversal. The legitimacy of HA (and thereby of T_U) could start to be considered seriously on this ground. Let's now move on the next argument.

3.2. Neutralizing the Hamiltonian-based argument

As to the demand of preserving the Hamiltonian's energy within the positive spectrum, the HA's counter-argument runs similarly. The essence of such premise is, putting it simply, that the Schrödinger equation generates a physically meaningful solution upon change in the direction of time. So, it is a fundamental feature of T_A in SQM that it precludes negative energy solutions. It would be no possible to generate the second time translation in Wigner's general criterion if the time-reversal operator changes the Hamiltonian's sign.

However, from HA and T_{Sub}, any sign's changing follows from whether physical magnitudes are expressed as time derivative or not (Thesis T_{Sub}-b). The relation between time and a system's energy has been largely puzzling in non-relativistic quantum mechanics. This relates to the viability of a time operator in SQM and how to interpret energy-time uncertainty relation, issues that track back to the work of Wolfgang Pauli (see Pauli 1980 [1958]: 63, fn.2. See also Busch 1990a-b). Undoubtedly, the theory sets up a closely-knit, though unclear, relation between them, and I think the debate around a proper characterization of time reversal is just one of its multiple dimensions. Putting this aside, T_{Sub} demands any observable to change sign under time reversal if it takes time as its first-time derivative. In its simplest form, the Hamiltonian operator can be written down as

$$\widehat{H} = i\hbar \frac{\partial}{\partial t} \tag{4.6 }$$

From T_{Sub}, the Hamiltonian's sign is *directly* inverted under time reversal because it is defined in function of time. Arguing that the time-reversal operator should leave invariant certain quantities in SQM misses the point, since the proper definition of T_{Sub} and the metaphysical background of HA do not reflect such a concern. To be clear: the change of Hamiltonian's sign is just what it is meant to do according to T_{Sub} and HA, and this change is just a particular case of a general rule that is supposed to be applied wherever T_{Sub} holds. T_U is just therefore expressing such characterization.

An HA's defender could also pose the following argument. One of the motivations for not regarding T_U as a well-behaved time-reversal operator is that it renders no solution when the direction of time is inverted. Conversely, the legitimacy of T_A is grounded on the fact that it does turn solutions *into* solutions. But this requirement only makes sense if time reversal aims at generating the second time evolution in Wigner's general criterion for time reversal. As mentioned above, the generation of such a second time evolution is a sine qua non condition for representing time reversal properly as far as T_{Rel} is taken; thereby, the second time translation must also be a solution according to the dynamical equation. But, from T_{Sub}, time reversal is *not* demanded to meet Wigner's general criterion, so the argument is neutralized.

It is worth stressing that SQM is consequently non-invariant under T_U, that is, non-time-reversal invariant from T_{Sub} and HA's perspective. At first glance, this seems to be scandalously problematic, and a compelling reason to throw the whole package away. I think, however, things must be approached from the appropriate angle. The fact that SQM is non-T_U-invariant by no means implies that it is not *motion*-reversal invariant or T_A-invariant. The theory is as motion-reversal invariant as it was from T_{Rel} and OA. The point is that T_A can no longer stand for *time* reversal because, according to T_{Sub} and HA, time reversal and motion reversal are two quite different transformations, and thereby lead to different symmetries. What gets broken is the link between T_A and the *concept* of time reversal, and not the fact that SQM is T_A-invariant. For the substantivalist, T_A should be rather associated with a different concept, and therefore, with a different symmetry.

The rationale is mainly conceptual: HA does not preclude finding a way to represent motion reversal (and T_A is likely the best candidate for it), nor does it argue that T_U is *actually* motion reversal. This would be a rotund non-sense. The philosophical claim is that time and motion have to be distinguished, and thus that an inversion of the direction of time not necessarily collapses with an inversion of the direction of motion. Furthermore, if physical theories are (for any reason) demanded to be motion-reversal invariant (e.g., it could reasonably be a condition one wishes to preserve in any physical theory), *and* time is nothing but motion, then this directly entails that theories must be time-reversal invariant as well. But the implication heavily relies on taking the conjunction as true, and clearly T_{Sub} and HA might remain reluctant to accept it. Under T_{Sub}, it makes full sense that SQM be simultaneously motion-reversal invariant and non-time-reversal invariant[31]. HA seems to lead us to such a scenario.

3.3. Revisiting the momentum-based argument

As it was mentioned in Section 2.3, the transformation rule for momentum as $T\mathbf{P}T^{-1} = -\mathbf{P}$ is typically introduced in non-relativistic quantum mechanics by appealing to its obviousness and to the necessity for correspondence with classical mechanics. And this seems to explain why some definitions of time reversal are sometimes expressly introduced in terms of changing momentum (Messiah 1966, Davies 1974, Sachs 1987) or why it is so "natural" to expect momentum to change its sign under time reversal (Earman 1974, Roberts 2017).

But this argument can be analyzed further by identifying at least two distinct questions therein:

(a) why does momentum change its sign under time reversal in classical mechanics?

And

(b) why must one do the same for the momentum operator in SQM?

The recent philosophical discussion around these questions has been quite subtle, so let's proceed with caution. These questions have been given different answers. Moreover, the question (b) has been impugned by Craig Callender (2000) and David Albert (2000): one must *not* do the same for the momentum operator in SQM. Curiously, Callender backs this claim in Newtonian classical mechanics. As this is one strategy available to HA, let's start with his argument.

Callender claims that there is no need to switch the sign of momentum in SQM because momentum is not a time-derivate magnitude in the theory, but a spatial one. Why does one switch the sign of momentum in a classical context? Callender says that momentum in classical mechanics is defined as a time-derivative magnitude, and the necessity of switching its sign

[31] An akin point seems to have been suggested by David Albert, though his reasons are not completely clear for me. In the case of the Schrödinger equation, Albert says that it is non-time-reversal invariant (2000: 14), though a curious vestige of time-reversal invariance still survives: there is a transformation that "recalls" time reversal though it is not (15). This "partial time-reversal transformation", as he calls it, only concerns particles' position and its invariance is preserved in the vast majority of physical theories. This transformation looks, as far as I can see, pretty much like motion reversal.

follows *logically* (Callender 2000: 263) from it. Let's clarify the argumentation. First of all, Callender refers to *Newtonian* classical mechanics. As any Newtonian state is completely defined in terms of position x and velocity v, the way time reversal should transform the state follows from the fact that the time-reversal operator flips the sign of velocity because velocity is a first-time derivate magnitude. The explanation is that when one transform $t \to -t$, velocity directly transforms as $v \to -v$. This is an *effect* (so to speak) of transforming $t \to -t$. Thus, as long as momentum is defined within the Newtonian classical mechanics as

$$p = mv = m\frac{\Delta x}{\Delta t} \tag{4.7 }$$

it follows that momentum must also change its sign under time reversal. The reason is the same as that for velocity: the transformation $t \to -t$ *induces* the transformation $p \to -p$, because it is what happens when you replace t by $-t$ in eq. 4.7. Momentum in SQM is defined differently: as well known, momentum is now an operator ($\hat{P}$) defined as

$$\hat{P} = -i\hbar\frac{\partial}{\partial x} \tag{4.8 }$$

As matter of fact, momentum is no longer a time-derivative quantity, but is a space-derivative one that plays the role of being the generator of an infinitesimal spatial translation (like in classical mechanics). The Fourier transform of momentum in SQM is, indeed, the position operator, and one can easily transform the momentum-basis into the position basis by the Fourier transform.

Callender's argument relies on two more or less explicit assumptions. The first one is the idea that the definition of time reversal should not involve any reference to how a specific dynamical magnitude must transform under its application. And from this follows the rule that time reversal should only change first-time-derivative magnitude. So, why should one expect a space-derivative magnitude to change sign under *T*? That is the reason why it looks a bit unnatural to reverse the sign of the momentum operator in SQM according to this view: it does not represent a time-derivative magnitude and it is thus out of the range of the time-reversal transformation. The second assumption is Callender's distinction between TRI (time-reversal invariance) and WRI (Wigner reversal invariance): necessitating that time reversal switch the sign of momentum would blurs the difference.

It's easy to see that Callender's account fits pretty well with my distinction between T_{Sub} and T_{Rel}: what he calls TRI is just T_{Sub} and WRI, $T_{Rel.}$ So a defender of HA could challenge the momentum-based argument based on Callender's argument. If Callender is right, the substantivalist argument would run as follows: one shouldn't demand time reversal to either involve a reversion of the direction of motion or indicate how a dynamical magnitude transforms under its application. T_{Sub} only commits to such a minimum time-reversal operator which acts just like a *reflection*, and the rest of transformations simply follows depending on their relationship with time (like velocity and momentum in Newtonian classical mechanics in Callender's argument). A switch of the sign of momentum is in fact required to represent $T_{Rel.}$ But, for the substantivalist, T_{Rel} represents a different symmetry transformation (as WRI with respect to TRI). As T_U is the formal realization of time reversal in SQM and an instance of T_{Sub}, it follows that T_U must not switch the sign of the momentum operator in SQM. And, obviously, to claim that such a reflection is physically meaningless is to accept neither **S1** nor **S2**.

It is worth noticing an additional argument that Callender provides for why the sign of momentum *does* change in SQM (Callender 2000: 264-266), which is based on Ehrenfest's theorem, and wasn't mentioned in Section 2.3. The theorem aims to ground the correspondence between quantum mechanics and classical mechanics by stating that "the centroid of the quantum mechanical probability distribution will follow a classical trajectory" (Callender 2000: 265), which establishes a link between the classical and the quantum probabilities, what makes that the classical equations holds for average quantum values. It follows from this that if one wants to guarantee such a correspondence, momentum should transform under time reversal, because otherwise the time-reversal invariance no longer holds, and the correspondence is lost. In Callender's terms: "Switching the sign of the quantum momentum, therefore, is necessitated by the need for quantum mechanics to correspond to classical mechanics" (Callender 2000: 266).

This explanation, however, requires one to firstly have a good understanding of the rule for momentum in classical mechanics to apply it to quantum mechanics. This is noted by Roberts (2017: 324). If quantum mechanics is supposed to be a more fundamental theory than classical mechanics, it is in fact a bit deceptive this way of justifying the rule for momentum in SQM. But, more importantly, Callender's usage of Ehrenfest's theorem seeks to justify why *WRI* should switch the sign of momentum in SQM and not why *TRI* should do the same. So, mutatis mutandis, this discussion would at best offer some grounds why T_{Rel} must switch the sign of the momentum operator, and not why T_{Sub} should. And, of course, the argument, at least as it stands, not only says nothing about how T_{Sub} should work, but it also assumes Wigner's general criterion for time reversal (as it is valid for WRI, but not for TRI), which the substantivalist might not accept.

But let's come back to Callender's previous argument based on Newtonian classical mechanics against the idea that time reversal (in the sense of TRI) must change the sign of the operator momentum in SQM. This argument evidently relies heavily on that eq. 4.7 is general enough to provide a well-grounded understanding of momentum in Newtonian classical mechanics and why its sign changes under time reversal. But, unfortunately, such a definition is not as general as it could be, since it only holds for a *free* Newtonian particle[32]. This might not be so devastating from Callender's viewpoint: free particles are exactly the kind of entities (and the kind of framework) philosophers are interested in when defining properties at a fundamental level (and, also, physicists, see Ballentine 1998: 77). Also, as I mentioned before, these very simple cases are those of interest when asking whether a theory is structurally time symmetric or not. But there is something even more problematic: this argumentation completely overlooks Hamiltonian classical mechanics. And, if the general argument for switching the sign of the momentum operator is based on the correspondence between quantum mechanics and classical mechanics, it seems plausible to think that it is Hamiltonian classical mechanics where one should look into.

So, one has to rephrase question (a): why does momentum change its sign under time reversal in *Hamiltonian* classical mechanics? To begin with, Hamiltonian classical mechanics is formulated in a different space: an $2n$-dimensional manifold with a symplectic form Ω[33], which is often called 'phase space'. States of systems are points in the phase space ($\varepsilon \in$ Ƥ), where $\varepsilon = (q_i \ldots, q_n;\ p_i, \ldots p_n)$. That is, states of physical systems are specified in term of the components q and p. Then, a dynamical trajectory $s(t)$ is a curve on the phase space that

[32] The reason is that in Hamiltonian classical mechanics momentum and position are now *generalized coordinates* (the generalized position q_i and the generalized momentum p_i). As generalized coordinates, they need not be the ordinary position and the ordinary momentum. In particular, p_i might not match the definition of ordinary momentum of mass times velocity.

[33] A symplectic manifold (Ƥ) is a manifold equipped with a symplectic form (Ω), which is a closed non-degenerate 2-form. Non-degeneracy establishes that the top exterior power of the symplectic form is a volume form, so the symplectic manifolds are necessarily even-dimensional and orientable. Closedness establishes that all symplectic manifolds are locally indistinguishable: they all locally look like an even-dimensional Euclidean space equipped with the $\sum dx_i \wedge dy_i$ symplectic form. (for details, see Ana Canna da Silva 2006: Chapter 3)

satisfies the Hamiltonian equations of motion, for some smooth function $h: \mathrm{P} \to \mathbb{R}$, called the Hamiltonian.

One of the issues is how components q and p must be interpreted. Usually, there are thought of as representing *position* values and *momentum* values respectively. In that case, the time-reversal transformation is supposed to transform, $T: q \to q$ and $T: p \to -p$. This automatically time reverse the trajectories by reversing the states. In the time-reversed trajectory $s^T(-t)$, the same positions are occupied, but with the instantaneous momentum inverted, $\varepsilon^T(q_k, -p_k)$. However, as Roberts points out, this depends on interpreting the components q and p as usual, but this might not be strictly necessary (see Roberts 2018: Section 2.3.2). If this is so, one would firstly need to interpret the components q and p so as to have a clear understanding of how time reversal must act upon them.

One way to avoid this is by viewing the time-reversal transformation as some *antisymplectic* bijection (this has been proposed by Roberts 2018, based on Abraham and Marsden 1978, eq. 4.3.12). Here goes how it would work: by assuming that time reversal is an antisymplectic transformation, one assumes that its application induces a change in the sign of the symplectic form. In better terms, the push-forward of T, T^*, induces the transformation $T^*\Omega = -\Omega$. And here the trick comes in: changing the sign of momentum, $T: p \to -p$, is a case of an antisymplectic transformation, and it leads to $\Omega \to -\Omega$. So, the overall argumentation seems to be the following: time reversal must change the sign of momentum, if time reversal is an antisymplectic transformation (this is assumed). In this case, the form of the time reversal transformation depends on the symplectic form of Hamiltonian classical mechanics.

This is indeed a solid answer to question (a). Now, one has to provide an answer to the question (b): why must one do the same for the momentum operator in SQM? The answer to this question is actually similar to that given previously. Roberts (2018) says:

> "If we interpret Q and P as the 'position observable' and 'momentum observable', respectively, then the time reversal operator is defined by $T\psi(x) = \psi(x)^*$, and the time reversal transformation is given by: $\psi(t) \to \psi(-t)^*$. One can check that transformation has *the effect* of preserving position and reversing momentum: $TQT^{-1} = Q$ and $TPT^{-1} = -P$" (Roberts 2018: 8. Emphasis mine).

So, Roberts says that if one interprets Q and P as position and momentum respectively, then the time reversal operator is given by (T_A). And this transformation has the effect of transforming $TPT^{-1} = -P$. But, at this point, a defender of HA might complain: one first needs some reliable identification of *what* the observables represent. And this could not always be the case. She can thus argue that the formal implementation of time reversal would not only depend on how certain observables transform under its application (which is a relationalist assumption as it was argued previously), but also that the identification of those observables with momentum and position is based on a convention; convention that no longer runs in some contexts. This observation was made by Biedenharn and Sudarshan (1994). They propose a basis-free approach to time reversal that does not depend on the *convention* of interpreting operators Q and P as position and momentum respectively. As Roberts mentions, some cases "do not even admit such operators, in the sense that they do not admit a representation of the canonical commutation relations" (2017: 324). The defender of HA could rhetorically ask: why would the definition of time reversal in general depend on how people agree on identifying certain observables? This sounds pretty much to a theory-relative way to define time reversal, something that the defender of HA could reject.

In his 2018 paper, Roberts seems to take a step further. Though one is not forced to interpret Q and P as position and momentum respectively in SQM, "it is generally assumed that all time reversal operators share a property, which is that of being *antiunitary*" (2018: 8. Emphasis in the original). And, "once we are convinced that time reversal preserves Q and reverses P, then it follows that T must be antiunitary. And, in an irreducible representation of the canonical commutation relations in Weyl form, this T is in fact the unique antiunitary operator (up to a multiplicative constant) that preserves Q and reverses P (Roberts 2017, Proposition 2)" (Ibidem). The footnote 6 in the same page proves that if one assumes that time reversal preserves Q and reverses P, then T cannot be unitary, but anti-unitary. So, the property of being "anti-unitary" seems to play a similar role in SQM, as the property of being "antisymplectic" in Hamiltonian classical mechanics, mutatis mutandis[34]. So, in this case, anti-unitarity would play the role of giving some sort of "persuasive force" to the whole argumentation: one expects that Q and P follow the same transformation rule for time reversal as in Hamiltonian mechanics. The issue is that those formal operators don't need to be interpreted as position and momentum respectively. But one knows that time reversal is *usually assumed* to be anti-unitary (this is the property that all time-reversal operators are supposed to share). So, it turns out that T_A fits perfectly well with what one was expecting: that time reversal (now given by T_A) preserves Q and reverses P.

[34] "However, *as in classical Hamiltonian mechanics*, it is generally assumed that all time reversal operators share a property, which is that of being antiunitary" (Roberts 2018: 8. Emphasis added)

Two remarks are in order here. First, the defender of HA could argue that any appeal to anti-unitarity is not allowed: one of the main reasons to formally represent time reversal in terms of T_A is that the Hamiltonian must be preserved under the transformation (that is, it must still be a bounded-from-below Hamiltonian). But, as I argued previously, HA is not committed to this. So, when it is said that it's generally assumed that all time-reversal operators share the property of being anti-unitary, this claim is supported by the Hamiltonian-based argument. And if the defender of HA is right and not committed to assume this, then she is not committed to assume that anti-unitarity is a shared property of all time-reversal operators. Removing such an assumption, she might conclude, the momentum-based argument loses some persuasive force. To put it differently, the momentum-based argument seems to articulate very well with other claims or beliefs about time reversal in SQM to make its case. However, when one of those beliefs is put into question, the argument demands some adjustments.

The second remark is that the defender of HA would still ask for a conceptual argument to link the transformation $P \to -P$ with time reversal. Why does one expect momentum to switch the sign under time reversal? According to her, observables must transform under time reversal if they have some formal relation with, or dependency on, time. For the substantivalist, any other expectation would be hiding a relationalist-like stance: the expectation seeks to represent motion reversal. Consequently, such an expectation could be rejected. And, at least so far, OA didn't offer an argument in favor of why one should have such an expectation. What it was suggested by OA was that (a) there is some reasonable symmetry transformation that preserves Q and reverses P (and, of course, this is true), (b) that this symmetry transformation must be anti-unitary, (c) that this transformation keeps invariant the Hamiltonian (and this justifies why T must be anti-unitary), (d) finally, that such a quantum symmetry transformation goes along with the classical correspondence of quantum mechanics (and this may be justified by looking at the symplectic form in Hamiltonian classical mechanics and then the case in SQM). However, the defender of HA could still insist: why is one allowed to call *this* transformation 'time reversal'? She can recognize that this is a reasonable symmetry transformation (probably even more interesting than time reversal for practical purposes), but that this is just formally implementing motion reversal, and not time reversal. For her, as it was explained before, the collapse of the notion of time reversal and motion reversal should be very well argued because they are metaphysically different transformations. Defenders of OA have in favor that tradition has called this sort of transformation 'time reversal', but, of course, defenders of HA can well argue that tradition is not source of normativity and doesn't allow solving (or dissolving) a metaphysical difference.

An analogous argument can be put forward for time reversal in Hamiltonian classical mechanics. After all, the form of T relies on the assumption that the transformation is antisymplectic and that the transformation $p \to -p$ leads to an antisymplectic transformation. So, why should one assume that *this* formal transformation deserves the name of 'time reversal'? But I think there is a difference: Hamiltonian classical mechanics is a *formulation* of classical mechanics. And it is generally assumed that Hamiltonian classical mechanics is *equivalent* to any other formulation of classical mechanics, as Newtonian classical mechanics or Lagrangian quantum mechanics (though there is some discussion about this, see North 2009 and Curiel 2014). So, in this case, one could have strong reasons to demand that certain quantities transform in a certain way based on what one knows to happen in other cases, though in a particular formulation the reasons to do that are not so evident (See Curiel 2014: 266-267 for explanation about the physical meaning of momentum in Hamiltonian classical mechanics and its relation with Newtonian classical mechanics). But, one needs some further justifications to jump from here to SQM, which is a completely different theory.

The momentum-based argument is probably the trickiest one for a defender of HA, perhaps because the argument is interweaved with other assumptions and beliefs about time reversal in the traditional literature. Assumptions and beliefs that are in turn based on different arguments. This yields a complex structure hard to dismantle. The defender of HA could neutralize the argument by raising some doubts on some of these assumptions and/or beliefs about time reversal. In fact, when one changes the perspective and turns to a substantivalist view of time reversal, there are some questions that remain unanswered: why should one beforehand expect that a certain magnitude behaves in a specific way, independently of its relationship with time? Or, why should one expect a space-derivative magnitude to change sign under *T*? T_{Sub}'s definition refers *intensionally* to how other magnitudes behave under time reversal: wherever you find a time-derivative magnitude in a theory, invert its sign accordingly. In this particular case, there is no argument against this principle. And probably the most promissory counter-argument is to point out that her metaphysical picture of time is just wrong. In any case, this is enough to make my point: previously-assumed commitment with respect to time *matter* to formally define the time-reversal transformation.

***

This section laid out two accounts for time reversal in SQM –OA and HA. According to the former, time reversal must be formally implemented by an anti-unitary time-reversal operator (T_A). It follows from it that the Schrödinger equation is time-reversal invariant because turns out to be T_A-invariant. I've therein pointed out that a relationalist metaphysics of time underlies OA in so far as it aims to represent time reversal in terms of motion reversal, metaphysically identifying both concepts –to be a time-reversal operator *is* to be a motion-reversal operator, and only T_A fulfils the requirement within SQM. Regardless how widely spread this account has been up to the present, an alternative account, HA, was also presented as feasible. According to it, in taking a substantivalist view of time, time reversal would be better represented by a unitary time-reversal operator in SQM, T_U. As a consequence, the Schrödinger equation turns out to be non-time-reversal invariant because it is non-T_U-invariant. Under this heretic account, motion reversal and time reversal should metaphysically be distinguished: whereas T_U is the appropriate representation of time reversal, T_A stands for motion reversal.

In the following, I'd like to point to two further views that also bear on the way time reversal should be formally implemented in SQM, though they can be easily generalized to the whole of physics. In general, these views chiefly mind the status of symmetries in physics, and the role they play in physical theories. To wit: (a) whether symmetries must be conceived and defined universally and in a theory-independent manner, or are rather forcefully bonded to a particular and theory-dependent view; and (b) whether symmetries are guides to theory construction, constraining theory's dynamics, or contingent properties of theories' dynamics instead. Even if taking side in this issue does not directly and specifically circumscribe the form of the time-reversal operator in SQM, it does become a sort of background that contributes to motivate either OA or HA. In the following section, I will expose and analyze them in tandem. The point I will make along the rest of the chapter is that OA is favored by a view of symmetries that is particularlist, theory-dependent and that considers them as guides to theory construction. Alternatively, a universalist, theory-independent view plus considering symmetries as contingent properties of theory's dynamics work out well for HA.

Section 4. Time reversal: theory (in)dependency and the role of symmetries in physics

In a 2015 paper, Daniel Peterson arises two questions as to the notion of time reversal:

(1) What does the time reversal operator look like?
(2) Which physical theories are time-reversal invariant?

The first question mainly concerns how states should transform under time reversal, and implicitly assumes that it can receive an answer posed in such a general form. The second question concerns which physical theories turn out to be time-reversal invariant under a previously-given characterization of the time-reversal operator. Peterson argues, on the one hand, that the above-mentioned questions go, somehow, hand-in-hand: an answer to one of the questions determines the answer of the other. On the other, he presents two accounts of time reversal that answer those questions, and justify them, differently: an *intuitive account* and a *theory-relative* one. Conforming to the former, any account of time reversal in physics should starts off giving an answer to the first question. The second question will hence be worked out from it. Conforming to the latter, any account of time reversal begins by assuming that a given theory is time-reversal invariant and, from it, the properties of the time-reversal operator for *that* theory are determined.

I think Peterson is quite on the right track in distinguishing two accounts of time reversal in physics: it successfully captures the main features of two broad criteria so as to characterize the time-reversal operator in different physical theories. Nonetheless, I don't fully agree with him on how both accounts are overall characterized. To my mind, both criteria follow not only from taking a decision as to whether the time-reversal operator is theory-relative or not, but also from deciding what heuristic role time-reversal symmetry plays in physics. This aspect, as far as I am aware, is missed in Peterson's analysis. Furthermore, I do not believe either that the name "intuitive" be the more accurate one, as it conceals on how much classical-mechanics-based intuitions the theory-relative account is based. Therefore, I will introduce such criteria differently: a *universalist, theory-independent account* versus a *particularist, theory-dependent account*. Importantly, I will also show how both accounts are articulated along with two general views on symmetries in physics: whether they are *guide to theory construction* or *contingent properties of dynamics*.

Section 4.1 A universalist, theory-independent account *versus* a particularist, theory-dependent account of time reversal.

That the time-reversal operator is theory relative is the current leading view in physics. By 'theory relative' is typically meant that the specific properties of the time-reversal operator have to be worked out within the framework of a physical theory. In some sense, then, the question "how the time-reversal operator is characterized" is pointless, since supporters of this view believe that there exists nothing as a general, all-embracing insight of time reversal in physics, at least in a meaningful sense. That's why any particularist, theory-relative account of time reversal will refuse to give an answer to Peterson's first question –it simply does not have any.

One can use the abstract notion of 'time reversal' in general as a shortcut to refer to a class of particular instances of time-reversal transformations which one must spell out within each theoretical context. The tenet guiding this account would be that any meaningful definition of time reversal cannot be achieved in isolation from a theory's details –the time-reversal operator is built up *from* "geographical" considerations of a physical theory, and never in isolation from them. By "geographical" or "particular" considerations I mean those taking into account the role that variables play in the physical laws, how states and observables are defined within the theory, and the physically possible solutions of a theory's dynamical equations.

The account is thus particularist in the following sense: these geographical details of physical theories matter in the sense that they play an active role in the time-reversal operator construction. When testing whether a given physical theory is time-reversal invariant, one is never in the situation in which a fully-fledged time-reversal operator is available beforehand and can straightforwardly be applied; rather, one has to firstly work it out from the very theory under scrutiny.

But, on which grounds is the time-reversal operator then worked out according to this particularist, theory-dependent view? My claim is that this account is prone to conceive that time reversal is actually motion reversal, and the time-reversal operator ought to specify the conditions to reverse motion. That's why particularist, geographical considerations are so crucial: they determine the very conditions of motion, which significantly vary across theories. And this account goes quite much along with OA.

Particularist, theory-dependent account	(a) The time-reversal operator must be worked out within a theoretical framework (b) There is no general, all-embracing view of what time reversal is in physics, but it depends on a theory's particular features. (c) Those particular features are in general associated with the conditions for motion. Hence, the time-reversal operator aims at representing the required formal features to reverse motion.

Supporters of the universalist, theory-independent account of time reversal would rather be committed to offer a general characterization of time reversal independently of any theoretical framework. That is, one would in principle be able to get some grasp on time reversal before stepping into a physical theory. In this sense, our understanding of time reversal is theory independent –though one needs physics to understand what time reversal is, it applies equally through any particular physical theory. The account seems to be thus guided for lofty concerns about time reversal conducted in abstract, regardless of worldly, particular details about physical theories.

In contrast with the particularist, theory-dependent account, Peterson's first question is utterly meaningful for the universalist, theory-independent view: if one wants to know whether a theory is symmetric under a certain operation, the very operation should then be defined *beforehand*. Upon which grounds such a transformation is based could be controversial. Peterson thinks that the starting point is our intuitions with respect to how states must transform under time reversal in each case. Though I'm partially sympathetic with his point, I believe that intuitions only offer a partial answer: I think that our previously-assumed commitments about time play a guiding role in figuring out how observables and states must transform under time reversal. That is, there is already a metaphysical election when one chooses to go either with T_{Rel} or with T_{Sub}. Whereas, it seems to me, the former will necessarily lead one to take a particularist, theory-dependent account (for, as stated before, the conditions for movement change across theories, and with them, the time(motion)-reversal operator), the latter would pave the way for taking a universalist, theory-independent account.

For the universalist, theory-independent account, one wants to know whether a given physical theory is invariant under a previously-defined operation that all we agree to call 'time reversal'. And when two physical theories are, for instance, time-reversal invariant, they are invariant under *the same* time-reversal operator. Hence, the time-reversal operator is theory independent and unique in the straightforward sense that its formal definition remains unchanged across different theories. This view goes along with HA to the extent to which it represents univocally time reversal as $t \rightarrow -t$ regardless whether or not it guarantees motion reversal.

Universalist, theory-independent account	(a) The time-reversal operator must be worked out independently of any physical theory (b) There is a general, all-embracing view of what time reversal is in physics. (c) The form of the time reversal operator could come out from different sources, but in general they would tend to represent the time-reversal transformation independently of motion reversal.

Contrasting both accounts, at least two features are noteworthy. First, the universalist, theory-independent account preserves the unity of the notion of time reversal because its mathematical representation and definition is always the same regardless of the theory. What the account aims to represent is the *very action* of inverting the direction of time, and as long as both the representation and the action remain the same across theories, unity is preserved. The particularist, theory-dependent account alternatively reaches certain unity because its different formal representations share a Wittgenstenian "family resemblance", given by the re-parametrization of the time coordinate and, more importantly, because they overall aim to reverse the direction of motion. Clearly, what gives unity to all time-reversal operators under this notion is no longer a homogeneous transformation, but certain effects, generally understood as a reversion of motion and its invariance, whatever it means within each specific theoretical framework. Remarkably, the very action (that is, what a time-reversal operator is supposed to do) shifts to the background whereas the physical representation of a past-headed physical solution is brought to the foreground.

Second, there is an interesting sort of trade-off: with the universalist, theory-independent account, while one gains universality in time-reversal operator's definition, one resigns universality in the results: physical theories in fact turn out to be either time-reversal invariant or non-time-reversal invariant under such a way to reverse time. In some cases, a reversion of the direction of motion is produced as an effect of time reversal, but this shouldn't be taken as necessary at all. Conversely, if one believes that one should go with a particularist, theory-dependent account, one renounces to universality in the time-reversal operator's definition while gaining universality in the results –perhaps surprisingly, virtually all physical theories turn out to be time-reversal invariant under such an account of time reversal.

To sum up. My point here is that OA and HA take different sides in the opposition regarding general features of time reversal in physics. They assume different starting points as to how characterize time reversal formally. For this reason, the problem is not so much whether, for instance, a proper characterization of time reversal in SQM should be conducted in a theory-independent fashion or not. The issue is much broader: should the time reversal operator *always* be built up *either* from a lofty, universalist view *or* from a theory-dependent particularist one?

4.2. Symmetries as guides to theory construction *versus* contingent properties of the dynamics

Whereas the first bone of contention is whether time reversal can be meaningfully defined in a full-fledged manner out of particularist considerations of a theory's dynamics, there is a different issue underlying and steering some choices in these accounts. This issue concerns how symmetries are regarded in general. Particularly, what heuristic role they are supposed to play in physical theories: either they play the role of guiding the theory construction in constraining the dynamics or they are rather contingent properties of the dynamics. Time-reversal symmetry may fall into either view. My claim will be that a particularist, theory-dependent account is heavily supported by the first view. The second one, though compatible with both, is required by a universalist, theory-independent account.

The current mainstream takes symmetries to be *principles guiding theory construction.* It demands physical theories to be a priori symmetric under a given symmetry transformation. The trend seems to have been initiated by Albert Einstein in formulating Special and General Relativity and followed by Eugene Wigner (see Wigner 1967. Brading and Castellani 2007: 5.2 offer an in-deep analysis of this view). In analyzing the "theory-relative" accounts, Peterson seems to suggest an akin point in saying

> "These accounts proceed in two seemingly simple but technically complex steps by (1) singling out a particular physical theory which intuitively seems time reversal invariant and (2) determining what features the time reversal operator would need to have in order for the given theory to be time reversal invariant" (2015:47)

I don't think that intuitions play any relevant role in the theory-relative account (except under a very precise sense I will mention shortly). Instead of that, I do think that the account proceeds, greatly, by decree: at the fundamental level, equations of motions *must* be symmetric under various space-time transformation because the space-time must be maximally homogeneous and isotropic. As any deeply-ingrained assumption, it is not easy to find pure instances of it in the literature but here goes some examples.

In inquiring time reversal in classical electromagnetism, Frank Arntzenius and Hillary Greaves describe a "text-book account" of time reversal, according to which time-reversal invariance is previously assumed in order to specify the form of the time-reversal operator within each physical theory. They say

> "Next let us consider the electric and magnetic fields. How do they transform under time reversal? Well, the standard procedure is simply *to assume* that classical electromagnetism is invariant under time reversal. *From this assumption* of time reversal invariance of the theory (…) it is inferred that the electric field E is invariant under time reversal (…)" (Arntzenius and Greaves 2009: 6. Italics mine)

According to them, the specific form of the time-reversal operator will then follow from assuming that the equations of motion (and the theory as a whole) are time-reversal invariant. By the 'specific form of the time-reversal operator' is meant how it transforms states and observables in an equation of motion.

In exposing the foundations of Bohmian Mechanics, Detlef Dürr and Stephan Teufel (2009) address time-reversal symmetry in the following way (also taking classical electromagnetism as an example):

> The primitive variables usually remain unchanged, while secondary variables, those whose role is "merely" to express the physical law for the primitive variables, may change in a "strange" way. A well-known example is Maxwell–Lorentz electrodynamics. The state is $X = (\boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{E}, \boldsymbol{B})$ and
>
> $$(\boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{E}, \boldsymbol{B})^* = (\boldsymbol{q}, -\dot{\boldsymbol{q}}, \boldsymbol{E}, -\boldsymbol{B}),$$
>
> (…). It is clear that $\dot{\boldsymbol{q}} \to -\dot{\boldsymbol{q}}$, but then $\mathbf{B}$ must follow suit *to make the equation time-reversal invariant.* The lesson is that some variables may need to be changed in a strange way for the law to be invariant. But as long as those variables are secondary, there is nothing to worry about. (2009: 47. Italics mine)

This is just a faithful exposition of Arntzenius and Greaves' text-book account of time reversal. For Dürr and Teufel, the '*' denotes an involution, that is, "a representation of time reversal". There are many aspects to look into here. To start with, their mention to primitive variables should be handle with care, for it seems there is no univocal division between "primitive" and "secondary" variables. For instance, for David Albert, the magnetic field is a basic property of classical electromagnetism, and therefore, it should not change sign under time reversal (2000),

$T: B \rightarrow B$. It seems that Dürr and Teufel think differently as they do consider that the magnetic field changes its sign under time reversal. The discussion now changes the angle: it is not anymore on time reversal but whether certain magnitudes should be considered as primitive/basic or secondary/non-basic. Secondly, they notably mention that, in some cases, secondary variables might change "in a strange way". But, "strange" with respect *to what*? More importantly, *why* do they change in a "strange" way? A few lines below one finds the answer: for the law to be (time-reversal) invariant. So, it seems there is an entire physical and formal mechanism aiming to specify the form of the time-reversal operator *in order to* maintain the equation of motion at issue invariant.

More suggestively, Dürr and Teufel add:

> "Let us close with a final remark on time-reversal invariance. One should ask why we are so keen to have this feature in the fundamental laws when we experience the contrary. Indeed, we typically experience thermodynamic changes which are irreversible, i.e., which are not time reversible. The simple answer is that our platonic idea (or mathematical idea) of time and space is that they are without preferred direction, and that the "directed" experience we have is to be explained from the underlying time symmetric law. How can such an explanation be possible? This is at the same time both easy and confusing. Certainly, the difference in scales is of importance. The symmetry of the macroscopic scale can be different from that of the microscopic scale, if the "initial conditions" are chosen appropriately". (2009: 47)

And let me link this paragraph to one on pages 43-44

> "A symmetry can be a priori, i.e., the physical law is built in such a way that it respects that particular symmetry by construction. This is exemplified by spacetime symmetries, because spacetime is the theater in which the physical law acts (as long as spacetime is not subject to a law itself, as in general relativity, which we exclude from our considerations here), and must therefore respect the rules of the theater". (2009: 43-44)

Dürr and Teufel refer to "our platonic idea" of a perfectly symmetric space-time as the reason for assuming, for instance, time-reversal invariance as an a priori symmetry fixing the properties of the "theater". Furthermore, the dynamics ought to respect the rules of the theater, that is, the space-time's symmetries. I'm not saying that most physics and philosophers of physics presuppose a "platonic space-time" as the theater of physics, or that they come to conceive space-time symmetries in the same terms. But I do believe that Dürr and Teufel make very explicit an underlying tendency in physics and philosophy of physics, to wit, that *the* *structure* of the space-time shouldn't manifest any privileged direction of time or intrinsic asymmetry. I will come back to this point afterwards.

Finally, Robert Sachs (1987) imposes as a condition for any time-reversal transformation that

> "In order to express explicitly the independence between the kinematics and the nature of the forces, we require that the transformations leave the equations of motion invariant when all forces or interactions vanish" (Sachs 1987: 7)

For him, time-reversal invariance is a powerful tool in elucidating the nature of the interactions between newly discovered physical systems, and that's why it must be presupposed beforehand: its assumption plays a heuristic role since it leads us to study the nature of forces and interactions and their relations with temporally-asymmetric behaviors of physical system. Anyway, the point to be stressed here is, once again, that time-reversal invariance is established by decree. Interestingly, Peterson's second question ("which physical theories are time-reversal invariant?") becomes uninteresting from this viewpoint, to the extent that its answer is, well, all of them. In so far as time-reversal symmetry is imposed a priori and by decree, all physical theories will trivially be time-reversal invariant.

As I mentioned above, such view, in general and in degrees, pervades philosophy of physics and physics communities. There is however an alternative view which claims that symmetries are, by contrast, a *contingent property of dynamics*. In this view, whether a symmetry holds mainly depends on details of the theory's dynamics, which is prior to the symmetry. To rephrase Dürr and Teufel's wording: the rules of theater are discovered a posteriori, once the characters are on the scene. This is not equivalent to say that, as in General Relativity, matter and energy shape space-time's structure, but that space-time's structure, though fixed and independent of the dynamics, shouldn't be settled from a priori considerations, but its structure is sound by dynamics.

John Earman puts it in the following way. In discussing active and passive transformation within a relationalist framework, he in passing mentions

> "If all these descriptions are equally accurate, it would seem that the symmetry transformation could not fail to be a true symmetry of nature, contradicting the usual understanding that symmetry principles are contingent, that is, are true (or false) without being necessarily true (or false)" (1989: 121)

The risk of the symmetry-as-guide-to-theory-construction view is that it would make time-reversal invariance a *necessary* symmetry, at least so far as fundamental equations of motion are concerned. The symmetry-as-contingent-property view instead leaves the way open to non-time-reversal invariant equations of motion; non-time-reversal invariance might not exclusively depend upon interactions, boundary conditions or forces. Hence, this view takes symmetries as contingently resulting from the dynamics.

Consequently, the space-time structure is no longer an a priori fixed scenario resulting from our idealizations about space-time's structure. The rules of the theater (space-time's symmetries) arise from the characters' properties (the laws of physics' symmetries): the formers are discovered by the latter. This view has been also defended by Harvey Brown (2005) and his dynamical approach to Lorentz covariance in special relativity, for instance. In this light, it makes fully sense to ask which theories are time-reversal invariant, since time-reversal invariance is not any longer demanded as an a priori condition.

There is certainly an acute mystery looming over here. No doubts that space-time's symmetries and dynamics' symmetries are somehow correlated, and that Earman's adequacy principle (see Chapter II, Section 2) should generally hold. But the mystery lays on which one is more fundamental or prior, or which one explains the other. Literature on this topic is abundant and has been drawing philosophers' attention lately (for some discussion see Brown 2005, Lange 2007, Janseen 2009, Acuña 2016, Hicks 2019), and to go deep into it would lead us beyond the topic. Nevertheless, I believe that not only are both views well testified in the literature and that the tension is still quite alive, but also that such a tension has significant impact on other philosophical issues. Time-reversal invariance is, I think, one of them.

***

Let me now bring all the pieces together. I have presented two broad accounts of time reversal in physics: particularist, theory-dependent and universalist, theory-independent. After that, I've suggested that these accounts articulate in different ways to unalike views on symmetries in physics: symmetries as guides to theory construction, on the one side, and symmetries as contingent properties of the dynamics, on the other. What I suggest is that all these elements feed many of the OA's and HA's assumptions on time reversal in SQM. In particular, OA's assumptions are to great extent guided by a particularist, theory-dependent spirit along with considering time-reversal symmetry as an a priori symmetry guiding theory construction.

Differently, HA's assumptions seem to go along with a universalist, theory-independent account plus a view taking symmetries as contingent properties of the dynamics. They thus form sorts of packages of assumptions that, even though imperfectly instantiated, frequently appeared together and provide a background explanation for T_A and T_U particularly.

As argued in Section 2.4, a relationalist metaphysics of time underlies OA, wherein T_A is simply a realization of T_{Rel} in the quantum domain. The specific form of T_A is however influenced as well by a particularist, theory-dependent account (as was showed before) along with considering symmetries as a priori. In other words, T_A transforms observables and states in the Schrödinger equation as it does *because* it aims to maintain the theory time-reversal invariant. That's what it is for. Rephrasing Arntzenius and Greave's words: the form of T_A follows from, firstly, assuming that the Schrödinger equation is time-reversal invariant and, secondly, figuring out how observables must transform to keep it invariant –time-reversal invariance acts like a regulative horizon at which one ought to look when investigating time-reversal transformation's properties.

HA is, contrarily, able to escape from this circle. To start with, HA plus a universalist, theory-independent account of time reversal steer to conceive a single time-reversal transformation that applies everywhere in the same way. Naturally, theories may turn out to be either time-reversal invariant or non-time-reversal invariant under such a transformation. It would be virtually impossible to come to work out a universal time-reversal operator that keeps all fundamental dynamical equations time-reversal invariant. According to its tenets, time-reversal invariance should be considered not as necessary but as contingent. Therefore, the Schrödinger equation may unsurprisingly fail to be time-reversal invariant. And, in fact, it's what actually happens. In this case, the regulative idea is that time reversal aims to represent a theory-independently-devised action (that of reversing time), regardless its potential consequences when applied to each case.

Summing up and what comes next

Along this chapter I've defended the *legitimacy* of two approaches to time reversal in quantum mechanics, each of them standing on quite different metaphysical grounds with respect to time and taking different assumptions with respect to general features of the time-reversal transformation in physics and the role that symmetries should play in physical theories.

Remarkably, I've shown that each approach leads to divergent conclusions as to whether the Schrödinger equation is time-reversal invariant:

(a) According to OA, the Schrödinger equation is time-reversal invariant because it is T_A-invariant, where time reversal *is* motion reversal
(b) According to HA, the Schrödinger equation is non-time-reversal invariant because it is non-T_U-invariant and it is motion-reversal invariant because it is T_A-invariant.

As a result, I've also mentioned that OA would, via the symmetry-to-reality-inference, preclude any structural arrow of time in SQM, whereas HA could solidly defend that the Schrödinger equation does exhibit a privileged direction of time. In a nutshell: for OA, time in quantum mechanics is structurally time-symmetric; for HA, contrarily, time in quantum mechanics is structurally time-asymmetric.

This leaves one in an uneasy situation. Philosophers of physics and philosophers of time could draw any metaphysical conclusion with respect to whether quantum mechanics exhibits a structural arrow of time or not, depending upon one's various commitments with respect to time, time reversal and symmetries in general. Though uneasy, this is not rare whatsoever. Frequently, such assumptions conduct our inquiries towards certain direction, aiming to reach a specific goal or departing from an uncritically-adopted starting point. Certainly, explanations must stop and get somewhere. Nonetheless, such assumptions shouldn't be overlooked. Neither should they block the viability of other paths. Having said that, there is a dimension I didn't take seriously into account so far. After all, time reversal is just a formal device that can be used variedly. That time reversal may be useful to sound time's properties is something one came to assume within the problem of a structural arrow of time and the symmetry-to-reality inference. Plus, I've argued in Chapter II, Section 2, that any approach to the problem must meet some requirements for the problem to be philosophically meaningful. Though I did say that both approaches arrive at divergent conclusions as to whether SQM is structurally time-asymmetric, it has not been showed that both approaches be equally useful and viable for addressing the problem of a structural arrow of time. Slightly differently, it has not been showed that each of them goes along with the conditions for the problem of the structural arrow of time. I will address this issue in the next chapter.

V.

The Arrow of Time and SQM at the Crossroads

The Orthodoxy, Time reversal and Time-reversal symmetry

In Chapter II, I formulated the problem of a structural arrow of time as mainly concerning whether a given dynamical equation is time-reversal invariant. As I've repeatedly stressed, being time-reversal invariant is being invariant under an operator formally representing time reversal. In Chapter IV, I put forward two approaches to work out the right form of the time-reversal operator within SQM. It was in passing suggested therein that OA would arrive at the conclusion that the Schrödinger equation is time-reversal invariant (under T_A), and thereby, that SQM doesn't exhibit any *structural* difference between both directions of time. Or, differently, that SQM is *structurally* time-reversal symmetric. By contrast, HA would arrive at the opposite conclusion to the extent that the Schrödinger equation is non-time-reversal invariant (under T_U). This chapter exclusively deals with the links between the problem of a structural arrow of time, as stated in Chapter II, and the two approaches of time reversal, as laid out in Chapter IV.

This short chapter will be guided by the following recommendation: be cautious when drawing metaphysical conclusions from a formal result. In particular, the aim of the chapter is to argue that the mechanism to justify OA involves some implicit premises that conceptually conflict with some of the conditions that make the problem of a structural arrow of time philosophically meaningful in terms of time reversal. This should discourage the drawing of metaphysical conclusions about time (for instance, whether it is structurally directed or not in SQM) from the claim that the Schrödinger equation is time-reversal invariant (that is, T_A-invariant). Conversely, HA seems to be in a better position to address the problem of a structural arrow of time from as it was formulated (Chapter II). The upshot of this chapter is that this situation sets out a complex scenario where some decisions need to be made.

The chapter is organized as follows. First, I will briefly recall the requirement to formulate the problem of a structural arrow of time. Next, I will reword OA in terms of three conditions (or implicit assumptions) that a time-reversal operator must satisfy in quantum mechanics: physical meaningfulness, kinematic admissibility and reality. In Section 3, I will argue that these three conditions clash the requirements for a structural arrow of time. In particular, I'll argue that OA's conditions and some conditions for the formulation of the problem of a structural arrow of time (particularly, contingency and fundamentality) cannot fit together. If this is true, then philosophers and physicists face a complex scenario where something has to go or be reformulated. I'll lay out this scenario, along with some alternative options, in Section 4. Final remarks at the end.

Section 1. Recalling the three conditions for a structural arrow of time

In Chapter II, the problem of a structural arrow of time was introduced as follows,

The problem of a *structural* arrow of time	(a) Does time have in itself the property of being asymmetric/directed? (b) Considering the laws of physics belonging to a theory *T*, is the structure of space-time supposed by those laws equipped with a time-orientation? (c) In being backed up by the 'symmetry-to-reality inference', the question thereby is: Is there any non-time-reversal invariant law in physics? Or, is there any physical law that does not produce a par of time-symmetric twins?

To put it simpler, the question for a structural arrow of time in physics mainly concerns *whether* a physical theory's dynamical equations are time-reversal invariant or not. As mentioned, non-time-reversal invariant laws would structurally (by their own means) distinguish between past-headed solutions and future-headed ones in so far as the class of their solutions have the form either $W = W^f$ or $W = W^b$, but not $W = W^f + W^b$.

In order to formulate this "whether-problem", three further conditions must be met:

(a) the fundamentality condition,
(b) the contingency condition,
(c) the co-extensity condition.

The first condition demands that only fundamental equations should be taken into consideration. If the space-time's structure presupposed by a theory manifests a temporal orientation, then it should be reflected in free-interaction laws. The second condition requires that it is not the case that a given dynamical equation *must* be necessarily time-reversal invariant[35]. As to the third, it imposes that there should be only one mathematical representation of the concept of time reversal within a theory.

In order to formulate the problem of a structural arrow of time within SQM, one can then say that the problem is about *whether* the Schrödinger equation is time-reversal invariant, and that (a) the Schrödinger equation in its fundamental form should only be taken into account, (b) it must not be the case that the Schrödinger equation is necessarily, or a priori, time-reversal invariant, and (c) the Schrödinger equation should not be time-reversal invariant under a certain time-reversal operator, and non-time-reversal operator under a different time-reversal operator (if this is the case, one of them is not genuinely representing time reversal). If, and only if, the three conditions are satisfied, one can thus give an answer to the whether-question, and consequently, draw metaphysical conclusions as to whether time is structurally directed within SQM.

Section 2. OA's three conditions

For OA, as analyzed in detail previously, a time-reversal operator must be a motion-reversal operator plus a re-parametrization of the variable t ($t \rightarrow -t$). I'm going now to reword the approach as relying on three assumptions demanding three conditions for time reversal.

The first condition strongly relates to Wigner's presentation of time reversal as a transformation that must generate an evolution going backward to the original state (second time evolution in the Wigner's general criterion for time reversal). It follows from this that the second time displacement is also an evolution according the (time-reversed) dynamical equation. Therefore, the second time displacement is *necessarily* also a solution of the Schrödinger equation. For this to be so, the time-reversed Hamiltonian cannot change its sign when time reversed: on the one hand, it would fail to generate a time-reversed *quantum* mechanical evolution, and it wouldn't lead to a time(motion)-reversed solution of the Schrödinger equation, on the other. For this, OA seems to require that the time-reversed dynamical equation must yield a physically meaningful evolution within a quantum-mechanics framework. That's why I call this first condition "physical-meaningfulness condition"

[35] See footnote 19 on page 79.

The second condition follows from Sachs' presentation of time reversal. According to him, the time-reversal transformation, when applied to *fundamental* dynamical equations, must yield them invariant in order to accomplish a "kinematic-admissibility condition". Sachs writes that

> "In order to express explicitly the independence between the kinematics and the nature of the forces, we require that the transformations leave the equations of motion invariant when all forces or interactions vanish" (Sachs 1987: 7).

Moreover, it is also required the time-reversal transformation has to "conform to the requirements of the correspondence principle –namely, operators representing classical kinematic observables must transform under *T* in a manner corresponding to classical motion reversal" (Sachs 1987: 34).

These two requirements should guide the characterization of the time-reversal operator within any theoretical framework. In general, they express a "kinematic-admissibility condition", which is instantiated here in the requirement that the Schrödinger equation must be kept time-invariant in the absence of any force or interaction, and in taking the results obtained in classical mechanics as a starting point for developing concepts in the quantum realm (1987: 9).

In a recent paper, Bryan Roberts offers a careful justification of the orthodox approach to time reversal in quantum mechanics. In his argumentation, he seems to bring up a slightly different condition for time reversal in quantum mechanics. I will call this third condition "the reality condition". In brief, Robert's argument supposes that there is at least one realistic Hamiltonian that is invariant under time reversal. But, T_U fails to guarantee this assumption. Therefore, via Wigner's theorem, the time-reversal operator must be given by T_A.

Section 3. OA and the problem of a structural arrow of time: contingency and fundamentality at risk

In Chapter IV I have set out two cogent ways to formally represent time reversal in SQM –T_A and T_U. My central claim now is that philosophers of physics, philosophers of time and metaphysicians should be cautious when drawing *metaphysical* conclusions about the direction of time (that is, about whether time is *fundamentally, structurally* directed or no) from the fact that SQM is T_A-invariant, therefore, time-reversal invariant. And the problem here is that

Wigner's, Sachs' and Roberts' justification of an anti-unitary time-reversal operator (underpinning OA) rests upon certain assumptions and conditions that don't seem to fit well with the conditions for a philosophically interesting formulation of the problem of a structural arrow of time.

To be clear, my concern is not whether T_A is the correct representation of time reversal or not. My question is whether T_A and, in particular, the claim of T_A-invariance, is a metaphysically substantial claim about the arrow of time in SQM. And *on this* I will cast some doubts. So, according to my stance here, philosophers and physicists that engage with OA would have two options at hand: either they refrain from drawing metaphysical conclusions about the arrow of time in SQM from OA's conclusions about time-reversal invariance, or they revise the formulation of the problem of the arrow of time and its conditions. Either option comes with a cost. In the following, I will back up these claims. But the general moral of the chapter should be read in the following way: the claim that the Schrödinger equation is time-reversal invariant because is T_A-invariant shouldn't be taken metaphysically too seriously. But, if one *wants to* take it seriously, one should either decline OA or revise the formulation of the problem.

I

Recall Wigner's definition of time reversal. He invokes two explicit premises:

(a) that a suitable time-reversal transformation must be able to restore "the system to its original state" (1932: 326),

(b) and that time inversion must flip the direction of momentum to compensate for the twofold application of *T* in Wigner's general criterion.

But there is also one fundamental implicit assumption in requiring time reversal to meet Wigner's general criterion:

(c) for a time-reversal transformation to be well-defined (and to exist at all), the second time translation (from t_2 to t_1) must also be physically possible.

To see how this last assumption works let's suppose that a quantum state $|\psi\rangle$ evolves from t_1 to t_2, according to the Schrödinger equation (first time translation in Wigner's general criterion for time reversal). At t_2, time reversal is applied upon the Schrödinger equation. *If* the time-reversal transformation is well-defined, *then* the time reversion of the time-translated state $|\psi\rangle$ should evolve from t_2 to t_1 also conforming with the Schrödinger equation (second time translation), which would intuitively be equivalent to making the system to evolve again but with $-t$. And here the implicit assumption comes in. According to Wigner, the operation to be applied upon the state at t_2 *must* be of such a kind that yields a physically meaningful moving-backward evolution. To put it differently, the transformation *must* generate a time translation (moving backward) that be also an equivalent model of the Schrödinger equation. Otherwise, the transformation would fail to take the quantum system back to its original state, and it will thereby be ill-defined.

The point is a bit subtle: Wigner's definition of time reversal necessarily implies the *existence* of a (allegedly) *temporally-mirrored model* of the Schrödinger equation; if such a model does not exist, then T would not be a time-reversal transformation. Assume that, at t_2, one applies T_U instead of T_A. As remarked above, the Schrödinger equation will not temporally translate the system back. In fact, it could be argued (and I would agree) what one obtains in that case is not a physical *meaningful* solution of the Schrödinger equation whatsoever.

However, to the extent that this way of reversing time demands *beforehand* a temporally-mirrored model of the Schrödinger equation to exist, it supposes that the class of solutions of the Schrödinger equation under time reversal has a structure like $\mathrm{W} = \mathrm{W}^f + \mathrm{W}^b$, wherein solutions for the first time translation belong to W^f, and solutions for the second time translation to W^b. The time-reversal transformation thus plays the role of *finding* what sort of solutions live in W^b. But this clearly amounts to beforehand demanding that the Schrödinger equation remains invariant under the transformation (it presupposes, at least, that such time-reversal solutions exist, to begin with). In other words, it amounts to demanding that the Schrödinger equation is *necessarily* time-reversal invariant. And this, evidently, clashes the contingency condition for the problem of a structural arrow of time.

I think that two quite different motivations and concerns diverge at this point: Whereas the problem of a structural arrow of time mainly concerns *whether* a fundamental dynamical law is time-reversal invariant or not (which explains why demanding contingency is so important), Wigner's justification of T_A seems to be guided by a quite different concern: *how* a fundamental dynamical equation turns out to be invariant under a transformation that one will call 'time reversal'. What Wigner is really pursuing by this *how*-question is a transformation that *necessarily* guarantees the existence of the temporally-mirrored model, and not whether such model exists or not. Wigner closes his presentation of time reversal in this way

> "Naturally, the considerations of this section do not prove that the quantum mechanical equations are invariant under the operation of time inversion. They do show, however, that if they are, the time inversion operator T_A must be given" (1932: 333).

My point here is that a whether-or-not question (as that motivating the problem of a structural arrow of time) is at odds with Wigner's aims in justifying an anti-unitary time-reversal transformation: Is the aim to figure out the way to take the system back by means of a physically possible evolution according to the dynamical law? Well, if that is the case, then the whether-or-not-question doesn't make sense: the dynamical equation will always be time-reversal invariant. Is the aim to know whether or not there is a model for the fundamental dynamical law when the direction of time is inverted? Well, then the *how*-question is not the right one.

Note that a *how*-problem entitles one to rely on whatever be necessary in order to achieve what one is after. The time-reversal transformation, therefore, will be whatever be necessary in order to get the Schrödinger equation time-reversal invariant. Along this line, Paul Roman (1950) says

> "For the reasons stressed in the introduction to this Section, we should like, even if only by some *artifice*, to achieve full covariance [time-reversal invariance]". (Roman 1961: 267. Italics mine)

The reasons Roman mentions swing from experience-based lessons about invariances to Pythagoras-inspired wishes of perfect mathematical symmetries (see Roman 1961: 216-217). There is nothing per se wrongheaded with this. The issue, I think, is twofold. On one side, if one wills to accept that any *artifice* deserves the name of time reversal in order to fulfill a previously-assumed philosophical view on what symmetries are and what role they should play. In the previous chapter, I've shown that there is no consensus around these questions. On the other, if one wills to accept metaphysical conclusions about time's structure drawn from a goal-oriented artifice. And *this* is what I think one shouldn't.

For these reasons, if T_A is the time-reversal operator that must be given in SQM and it was built up under the assumption that the theory remains invariant under its application, then the problem of a structural arrow of time in terms of (a) whether-or-not question and (b) assuming the contingency condition cannot be addressed –the answer is necessarily determined by construction since the Schrödinger equation couldn't be non-time-reversal invariant in this respect.

II

One could reach the same conclusion by an alternative path. All our experiments run always in one direction of time (positive or negative, but not both) and physical theories receive empirical confirmation with increasing (decreasing) t. The direction of time cannot be, in reality, inverted to see what happens with our experiments and with our physical theories when time runs backward. Intuitively, this does not happen with spatial transformations since one can freely move things in space along its three dimensions and, from this, deduce what would happen if fields, for instance, were removed. With time, one is rather left clueless. So, to figure out what would happen if the direction of time were inverted, philosophers and physicists must of course make some assumptions. The question as to which assumptions one is willing and entitled to accept then becomes relevant. However, I am afraid that OA's conditions for time reversal in SQM make more suppositions that the formulation of the problem and its conditions can tolerate. Roberts' argument for an anti-unitary time-reversal operator in SQM based on a "realistic condition" helps explain the last claim, so let's look at this argument more closely.

Overall, the argument depends on assuming the (possible) existence of at least one *realistic* time-reversed Hamiltonian. Roberts starts off his argument by assuming that one knows almost nothing about the specific form of *T*, only that it is either unitary or anti-unitary. Then, he poses the main premise of his argument

> "There is at least one *realistic* dynamical system that is *T*-invariant" (Roberts 2017: 326. Emphasis added)

Clearly, the relevant word here is 'realistic'. For a system to be realistic, a Hamiltonian with the following features is required:

(a) It is *positive*, that is, $0 \leq (\psi, H\psi)$

(b) It is not the zero operator

This is the definition of 'being realistic'. Why should one accept it? Well, Roberts gives two reasons: first, "all *known Hamiltonians* describing *realistic* quantum systems are bounded from below, which we will express by choosing a lower bound of $0 \leq (\psi, H\psi)$" (*Ibidem*. Emphasis added), which backs feature (a). Second, the zero operator would be trivial since it implies that "no change would occur at all" (*Ibidem*), which in turn backs feature (b). So, the complete assumption is the following

> "At least one of those [realistic] Hamiltonians satisfies the *T*-invariance property that $Te^{itH}\psi = Te^{-itH}T\psi$" (Roberts 2017: 326)

This implies that there exists at least a *realistic* Hamiltonian that is invariant under time reversal. The most promissory candidate for this, though Robert's argument doesn't hinge upon it in any sense, is the free-particle Hamiltonian.

So, from assuming that (1) there exists at least one *realistic* quantum system (positive and non-trivial) *and* (2) it is time-reversal *invariant*, one can determine whether *T* is unitary or anti-unitary. Roberts claims that these assumptions imply that *T* must be anti-unitary, supporting OA, consequently. The proof runs as following:

(i) Let's suppose that *there exists* at least one *realistic* Hamiltonian that is time-reversal invariant (the main assumption)

(ii) The definition of being 'realistic' is that (a) the Hamiltonian is positive and (b) non-trivial.

(iii) Suppose now that 'time reversal' is given by T_U (assumption for reductio ad absurdum)

(iv) This implies (as shown before and as Roberts mentions in his proof) that the Hamiltonians transforms under time reversal as $T\mathbf{H}T^{-1} = -H$

(v) The fourth step in turn implies that there would exist a realistic Hamiltonian that is non-positive, and that $\langle\psi, H\psi\rangle = \langle T\psi, TH\psi\rangle = -\langle T\psi, HT\psi\rangle$; or that the Hamiltonian is the zero operator.

(vi) But this is impossible: premise four is in contradiction with the first and second premises. In particular, a so-transformed Hamiltonian is not realistic since it has to be either negative or the zero operator.

(vii) Also, if the Hamiltonian is negative, then it is non-time-reversal invariant, violating premise (i) again.

(viii) Therefore, *T* is not T_U but can only be T_A

Roberts seems to demand very little. Certainly, if one wants to represent a quantum dynamical system moving backward in time, one has to at least generate the conditions for such system to exist (or to be representable) within a quantum-mechanical framework. And those conditions seem to be well captured by the idea of "being a realistic Hamiltonian". What such a definition does is to guarantee that Hamiltonians (as one *knows* them) are still representable when temporally reversed. And the transformation should capture this.

However, the assumption might look a bit excessive: it might be the case that what it's at stake is *whether* such a representation is possible and *whether* it's legitimate to assume it beforehand. Let me go slower. As it was mentioned before, Roberts starts his argument by supposing we might know *almost* nothing about time reversal, only that it is either unitary or anti-unitary. But, if one follows Roberts' argumentation, one implicitly knew something else about time reversal: whatever the form of T comes to be (either T_A or T_U), one knew that $T: H_{Real} \rightarrow H_{Real}$. In words, that the time-reversal transformation maps realistic Hamiltonians into realistic Hamiltonians, this is, it guarantees that the property of "being realistic" is preserved under its application.

A defender of HA might find *this* implicit assumption a bit excessive and might go on to reject it. Indeed, the implicit assumption that $T: H_{Real} \rightarrow H_{Real}$ can be deemed false to the extent that one technically doesn't know how past-headed Hamiltonians behave (whether time-reversed Hamiltonians are realistic or not). One doesn't even know if even one realistic Hamiltonian is time-reversal invariant. On this basis, the defender of HA can reject Robert's argument by saying that one is not entitled to assume that the time-reversal transformation preserves "reality". Neither would one be entitled to assume that one of those Hamiltonians is time-reversal invariant (the defender of HA might ask: where does such an assumption come from?). From a relationalist metaphysics of time, Robert's assumption looks in fact natural: relationalists have a clear metaphysical picture of what going backward in time means (condensed in Wigner's general criterion). And requiring that the second time translation involves *realistic* Hamiltonians would be crucial for having a meaningful representation of time reversal (otherwise, how would the system go back to the original state?). Also, in order to build the time-reversal transformation up, one can well assume that at least one realistic Hamiltonian be time-reversal invariant.

However, from a substantivalist view, this is not so clear. To begin with, the second-time evolution is not required for time reversal to be well defined. The relevant transformation is the inversion of the direction of time (a reflection), $T: t \rightarrow -t$; how observables transform under such a transformation will depend on the physical theory at stake and how they relate to time therein. Consequently, in this picture, it might be excessive to assume that $T: H_{Real} \rightarrow H_{Real}$. And if one revises Robert's proof from HA's perspective, that is, without being committed to premise (i), one would arrive at the following conclusion: there would exist no realistic Hamiltonian when time reversal is applied; even in the simplest cases. The time-reversal transformation wouldn't preserve the property of "being realistic" (since $T_U: H_{Real} \rightarrow$

H_{Unreal}), and the Schrödinger equation is not thereby time-reversal invariant since it fails to generate bounded-from-below Hamiltonians when T_U-reversed.

Note that the point is not actually that the theory simply fails to yield a *realistic* dynamical system; rather, the theory does not even yield a *quantum* dynamical system. Unbounded-from-below Hamiltonians are, in Roberts' words, "physically implausible". To begin with, "being physically implausible" is generally framed within a physical theory (many events are physically implausible in the light of a certain theory but perfectly plausible within another), so by "physically implausible" I will understand "quantum-mechanical implausible". However, one knows about what's plausible and what's not from "known Hamiltonians" (as one knows that actual Hamiltonians are always bounded from below), and from more general and well-grounded quantum principles that establish that matter would be unstable if Hamiltonians were different (for instance, unbounded from below). To the extent that matter around us seems to be quite stable, then one can fairly conclude that any *quantum* system's Hamiltonian should be bounded from below.

But, what a defender of HA would argue with all this information is that the Schrödinger equation is non-T_U-invariant in such a radical way that it does not even leave room for a realistic Hamiltonian to exist with decreasing t. In order words, the class of solutions of the Schrödinger equation involves either

(a) bounded-from-below/unbounded-from-above (positive spectrum of energy) Hamiltonians in, say, $\mathcal{W}^f$.

or

(b) unbounded-from-below/bounded-from-above (negative spectrum of energy) Hamiltonians in $\mathcal{W}^b$.

As solutions of the Schrödinger equation must only involve Hamiltonians like either (a) or (b), but not both, one of the two subclasses does not truly represent physical solutions within SQM. Hence, one obtains either $\mathcal{W} = \mathcal{W}^b$ or $\mathcal{W} = \mathcal{W}^f$, but not $\mathcal{W} = \mathcal{W}^f + \mathcal{W}^b$. Therefore, the Schrödinger equation is simply non-time-reversal invariant. This, of course, does not mean that T_U *must* be then given; rather. But the point is *whether or not* one has philosophical reasons to reject Robert's main assumption. If one has them, then T_U *may* be given, accepting its consequences.

Now, Roberts recognizes that this conclusion could be drawn in denying his assumption when he says that

> "Denying that time reversal is antiunitary means that a nontrivial realistic quantum system is never time-reversal invariant, under any circumstances whatsoever. Even an empty system with no interactions would be asymmetric in time" (Roberts 2017: 328)

And a few lines below

> "There is little practical use in identifying time reversal with a transformation that does not make any distinctions at all, not even between a free particle and one experiencing an important time-directed process like a weakly interacting meson" (Roberts 2017: 328)

Recall that one of the conditions for the problem of a structural arrow of time (one that has been brought up by Callender) is *fundamentality*. Robert's first quote seems to be reluctant to give the way open for a violation of time-reversal invariance *even* in those simple cases (those cases described by *fundamental* dynamical laws). But, a structural arrow of time would only manifest in those simple cases, and that's why the fundamentality condition is relevant to formulate the problem: the question whether time-reversal invariance holds or not must be aimed at those simple cases. Furthermore, the *contingency* condition is required *there*: it must be the case that even those simple cases might be non-time-reversal invariant. And this shows (again) why OA and the problem of a structural arrow of time seems to not fit together.

The second quote points to a sort of pragmatic stance. At the end of the day, one requires time-reversal transformation to draw distinctions between what seems to be "different sorts of time-reversal violations". This is also aligned to Sachs' kinematic-admissibility condition: in order to identify emergent cases of time-reversal violation, one requires that the simplest cases remain time-reversal invariant. Two remarks to this: first, this position is again at odds with the fundamentality condition. Second, it assumes that it is the time-reversal transformation what makes it possible to draw the differences. I'm not completely sure of this, at least, not in such a general way. A defender of HA could argue that it is actually the theory what should trace the differences between different levels of time-reversal violation. What T_U reveals is just that even in the simplest cases (which is a metaphysical possibility that should be left open) time-reversal invariance fails because time is structurally asymmetric. And this manifests in a "pure way" (so to speak) in the simplest cases. One can retort by saying that OA is unable of capturing any structural or fundamental violation of time reversal. And, again, people interested in the debate on a structural arrow of time could be wary that T_A does the work they require.

Earlier I introduced the distinction between a how-question and a whether-or-not question. Whereas the realistic condition seems to provide an answer to a *how*-question, it is somewhat at odds with the contingency condition for the problem of a structural arrow of time –a *whether*-or-not question. In particular, I believe that the reality condition applies information that can only be obtained from systems actually evolving in one direction of time (say, with t increasing) to the opposite direction of time (say, with t decreasing). To be clear: this is a perfectly legitimate maneuver for addressing a *how*-question like how the Schrödinger equation is left invariant under time reversal. The problem nonetheless arises when a view like OA is endorsed to draw metaphysical conclusions about a structural arrow of time in quantum mechanics, conclusion that answers a *whether*-or-not question. This is also a point that defenders of HA could bring on the table against OA.

Finally, premise (i) and (ii) in Robert's proof seem to assume certain "stability" in SQM under time reversal: by imposing that the time-reversed evolution and the original one share some properties (like in the case where, $T: H_{Real} \rightarrow H_{Real}$, and consequently *both* Hamiltonians are bounded-from-below), one is "leaking" temporally-biased information about future-headed evolutions into past-headed ones; in other words, one is assuming that the theory, to some degree, should work with $-t$ as it does with $+t$. But it is at least imaginable that this be not the case (again, for reasons derived from the contingency condition). The question thus is: How entitled is one to assume about past-headed systems in virtue of one actually knows about future-headed ones? To assume the realistic condition seems to be a sort of violation of Huw Price's nowhere viewpoint (1996), a meta-principle in the arrow of time discussion: the world might be such that matter is only stable when time runs forward but becomes unstable if time runs backward. One should not assume –a priori– that matter behaves equally in both direction of time when one actually only knows how it behaves when time runs forward. One shouldn't even assume that there must be past-headed matter at all. Otherwise, one would be nullifying one metaphysical possibility by imposing that what is actually given is necessarily the case in any other possibility. And allowing for metaphysical possibility is essential to the formulation of the problem of a structural arrow of time.

III

There is yet a third short argument against drawing metaphysical conclusions about time from endorsing OA and its conditions. Now I want to show that the fundamentality condition is also at odds with OA –which brings one back to Sachs' requirement. Indeed, Sachs' requirement entails that any asymmetry of time *must* somewhat *emerge* from interaction and forces. According to Sachs, the purpose of defining time reversal is to understand the intertwined play between dynamics and kinematics, and to explain how a temporally asymmetric world emerges from a temporally symmetric basis. And the explanation in general involves the dynamic's role. This is an assumption that not only already takes for granted that physical laws at the fundamental level are time-reversal invariant, but also that any asymmetry of time can *only* come from forces and interactions. Not only does this establish by decree that there is no structural arrow of time, but also contradicts directly the fundamentality condition, which seeks to set aside equations of motion involving forces and interactions. In brief, if OA has any philosophical implication with respect to the direction of time, it must also decline the fundamentality condition.

IV

My last argument is more general and seeks to show how the role of (a structural) arrow of time is diminished by OA in assuming that time-reversal invariance must overall hold. The point I want to make is that any justification of OA (and thereby of T_A) implicitly regards that a change of the direction of time is, at least, innocuous enough to maintain certain features of a physical theory unchanged. Or, to put it differently, that a change of the direction of time (under some assumptions) would go unnoticed. And this would amount to considering that the word 'structural' is rather superfluous.

One of the ways to see the symmetry-to-reality inference is by considering that if time-reversal invariance overall holds, then the direction of time would be *undetectable* across temporally-switched scenarios. This argument concludes that the direction of time is *epistemically inaccessible*: one would be unable to decide in which scenario one lives. Hence, to any epistemic purpose, a structural arrow of time would be absolutely irrelevant as long as, if real, it turns out to be out of our epistemic accessibility. To make the point clearer, consider the following argument, which was raised by Donald C. Williams (1951) and Huw Price (1996).

Suppose that one lives in a world with a structural direction of time. Also, suppose that the laws governing the world one lives in are, overall, time-reversal invariant. It follows from this that for any physical state S, there is a one-to-one mapping to a time-reversed state S^T (this is what time-reversal invariance assures). Consequently, for a person P, there exists her time-reversed Doppelgänger, P^T. From time-reversal invariance and the existence of the mapping $P \rightarrow P^T$, it follows that the P^T would *surely* have qualitatively identical experiences to *P*, only with the whole process oppositely oriented in time. Therefore, P^T experiences her own life stages in the same order as *P* but observing everything around moving backward from the grave to the cradle. So, she has the same organic and mental processes but oriented differently in time (the infant stage is now the future, instead of being the past). The problem is: how does *P* know that she is not actually P^T, and she was all long mistaken about the direction of time? Given the physical possibility of the existence of such a Doppelgänger, the direction of time seems to be epistemically inaccessible: both have exactly the same evidence about the direction of time. Therefore, it would be better to discard any structural direction of time of our set of beliefs.

Maudlin has, correctly to my mind, replied this argument. In general, Maudlin's reply considers that the argument uncritically presupposes that one can be sure that time-reversed processes will produce the same states (just temporally inverted) as ordinary processes do. And this is not obvious at all: time-reversed processes could turn out to be quite unalike that our processes. The argument aims to show that any structural arrow of time would be inaccessible for us, but one can only achieve this if it is assumed that P^T will still be having organic processes and mental processes *as one knows them* despite being temporally inverted. In some sense, then, a structural direction of time is not so relevant in defining the viability or not of those processes. One does know, for instance, that certain mental processes produce mental state in one direction of time, but why should one be so sure that time-reversed mental processes will equally produce mental states whatsoever?

If the argument aims to show that a structural direction of time is epistemically inaccessible, and therefore dispensable, then it commits a petitio principii –it shows that a structural direction of time is epistemically inaccessible but thanks to guarantee this does not play any relevant role in the very existence of physical and mental states. To put it differently, in imagining time-revered processes and states, the change of the direction of time is innocuous enough in so far as such time-reversed processes and states are expected to behave as they do in the ordinary direction of time. My point is that OA commits the same sin: it uncritically presupposes that the inversion of the direction of time shouldn't transform certain observables for, otherwise, time-reversed states couldn't be produced. This would amount to assuming that a structural arrow of time wouldn't be structural (or fundamental) enough to stablish which states are possible and which ones are not.

Here a more precise case that was analyzed before: in demanding that *realistic* Hamiltonians must exist in the backward direction of time, one is also assuming that the property of being "realistic" keeps invariant under time-reversal transformation. But that property might easily be one of those that one loses if a structural arrow of time genuinely exists. To put it in other way, by assuming that time-reversed Hamiltonians must be realistic (or bounded from below), or that a time-reversed the Schrödinger equation must generates a time-reversed series of quantum-mechanics states, one is not only assuming that certain familiar properties of quantum systems when regarded in one direction of time are equally instantiated in the opposite direction, but also that the direction of time is not structural or fundamental enough to define what is a well-behaved quantum-mechanics Hamiltonian, or which the very conditions for its existence are. The presupposition could be harmless when inquiring the emergence of non-structural arrows of time, or non-fundamental temporally asymmetric quantum processes. But in inquiring whether the Schrödinger equation manifests a *structural* preference for one direction of time instead of the other, such an assumption imposes a bias in favor of a negative answer.

Let me now address the same point from a theoretically broader viewpoint. One has a theory that, let's suppose, works astonishingly well as it stands: SQM. Empirical confirmation for *that* theory comes always from future-directed experiences (by convention). And the theory in itself has been built up in the light of how things behave in, say, the future direction of time (by convention again). One can raise at least two questions regarding SQM and the direction of time:

(a) Is there any acceptable and workable version of SQM with t decreasing?
(b) Is $T(SQM)$ an acceptable and workable version of SQM?

With some caveats and some assumptions, one could with conviction claim that there is in fact an acceptable and workable version of SQM with t decreasing: the one that is obtained by applying a transformation like T_A.

$$T_A(SQM, t) = SQM^*, -t \tag{5.1 }$$

And one gets it by relying upon some artifices (to borrow Roman's expression) that transforms wave-functions and SQM's self-adjoint operators in such a way that the equivalence is satisfied. But, note that one starts with the following situation

$$?^{?}(SQM, t) = SQM^{?}, -t \quad (5.2\ )$$

Where one knows that the equivalence must be satisfied, and that t now goes decreasing, but one knows nothing about which transformation meets the equivalence and how a time-reversed SQM satisfying the equivalence looks. That is what one has to work out so as to answer the first question.

Yet, the second question is entirely different in nature. What is at issue now is

$$T(SQM, t)\ ?\ SQM^{T}, -t \quad (5.3\ )$$

That is, the equivalence is not taken for granted but quite the opposite: it's what is actually at stake. There is naturally a tension here: what is the form of *T*? But the very question demands its form (whatever it come to be) to be theory-independent. And for this, one shouldn't assume that SQM^T will be an acceptable and workable physical theory like SQM: the property of "being a workable theory" could fail to be instantiated by a time-reversed version of SQM. In fact, everything may go fatally wrong when time is reversed: one gets to a theory that doesn't make any physical sense. For instance, one knows that a certain kind of Hamiltonians belongs to SQM (those bounded-from-below/unbounded-from-above). In assuming that that kind of Hamiltonians also belongs to SQM^T is like assuming that the time-reversed theory makes just about much sense as SQM. But this is going too far in one's assumptions: one knows nothing about what SMQ looks like.

Similarly, in assuming that time-reversed ψ represents a quantum-mechanical state, one is going too far: ψ^T belongs to SQM^T but not to SQM. Whether there is a mapping from ψ -solutions to ψ^T-solutions is what one is investigating. And the point is, and from here non-time-reversal invariance comes out, that whereas SQM is an astonishingly successful theory, SQM^T is rather an entirely useless one. Only by asking for SQM^T to be as acceptable and workable as SQM, one gets to SQM^*. But, by this maneuver, one is going too far in one's assumptions about what SQM^T looks like.

Section 4. What the scenario looks like so far?

In the previous sections I showed in which way OA's conditions for time reversal and the three conditions for the problem of a structural arrow of time (particularly, contingency and fundamentality) do not fit together. I believe that, at this point, one is compelled to draw the following skeptic conclusion: by endorsing OA, one is unable to say whether the Schrödinger equation picks a structural arrow of time (if one believes that the claim "the Schrödinger equation is time-reversal invariant" is philosophically interesting). Remarkably, not only does this preclude any answer to the problem of a structural arrow of time, but also it seems to undermine the very formulation of the problem of the two realms since it presupposes that *there is no* structural arrow of time in SQM *given that* the Schrödinger equation is time-reversal invariant. However, as argued, one is not even entitled to claim that the Schrödinger equation is *univocally* time-reversal invariant. One is not entitled to draw such a conclusion with respect to the direction of time either. From OA's position, one can only remain silent. This is so because OA offers no reliable way to argue *for* or a*gainst* a structural arrow of time.

And here is how the scenario looks so far for philosophers and physicists willing to engage in the debate on a structural arrow of time. They either:

(a) accept OA but decline to address the problem of a structural arrow of time and remains reluctant to draw any metaphysical claim about whether time is *fundamentally* asymmetric or symmetric in SQM (e.g., by claiming that the sentence "the Schrödinger equation is time-reversal invariant" is not metaphysically interesting)

Or

(b) abandon OA and turns to an approach like HA in order to provide an answer to the problem, which leads to concluding that the Schrödinger equation is non-time-reversal invariant, and therefore it does manifest a structural arrow of time.

But, is there any other option for OA's defenders willing to address the problem of a structural arrow of time without abandoning their position? One could think that OA's defenders will always have the option of challenging the formulation of the problem, which amounts to challenge some of its conditions. However, some of them are conceptually more essential than others. Particularly, I cannot devise any viable formulation of the problem that circumvents the contingency condition, so this doesn't seem to be the condition one would want to drop. Neglecting the co-extensivity condition would drive one to similar results: if one allows for many representations of time reversal, then there wouldn't be any matter of fact about whether an equation is time-reversal invariant, a property that becomes relative to the chosen representation (this is roughly the current state of affairs in physics).

A more plausible way out for OA's defenders would be to decline the fundamentality condition. For example, arguing that it is too strong, or even worse, that is misguided. So, one could decline that only fundamental physical laws are those relevant for addressing the problem of a structural arrow of time, and this may be given by other means. Keith Hutchison (1993, 1995) has followed this path for classical mechanics. In fact, the distinction between "fundamental laws" and "non-fundamental or phenomenological laws" can be put into question: physicists and engineers routinely apply and deal with "non-fundamental" dynamical equations; physics students have to master calculations featuring "imperfect" springs or different kinds of interactions; most *real* physical systems are described by non-fundamental laws in this sense. Fundamental laws, as Hutchison argues, are laws that merely describe highly-idealized models that share some properties with realistic models but clearly not share many others. To take them as more fundamental or ontologically prior is to condemn physicists' real work to a sort of apparent reality. Intending to find a structural arrow of time in *any relevant sense* in such platonic models is, at best, excessive or, at worse, misguided.

Probably, this is the easiest (if not the only) way out for OA's supporters. Nonetheless it implies deadly consequences for time-reversal *invariance*: as mentioned above, the overwhelming majority of non-fundamental dynamical laws are categorically non-time-reversal invariant. Classical mechanics (paradigmatically, a time-reversal invariant theory) would also fail to be symmetric under time reversal if, for example, one includes dissipative forces in the relevant dynamical equations (this is precisely Hutchison's conclusion). Declining the fundamentality condition would amount to claiming that any dynamical equation is equally good enough to ground a structural arrow of time, which could result not so intuitive. Furthermore, instances of non-time-reversal invariant equations would start to come up and one would be forced to consider all them as relevant for the problem of a structural arrow of time. Thus, the distinction between a structural and a non-structural arrow of time rapidly blurs out.

So, if one workable way around the problem was to decline the fundamentality condition, this maneuver would however imply that there would be as several instances of structural arrows of time as many instances of (non-fundamental) time-reversal invariant physical laws. This doesn't look like a quite promissory scenario for philosophical inquiry around the *foundations* of time. Thereby, when inquiring whether time is structurally directed according to a given physical theory, OA seems to lose, again, its obviousness and persuasive force. So, why don't get rid of OA once and for all?

This would be equivalent to endorse HA. In contrast with OA, HA doesn't presuppose any violation to the conditions for formulating the problem of a structural arrow of time. It doesn't assume that fundamental dynamical equations are a priori and necessarily time-reversal invariant, so that the contingency condition is preserved. Neither does it need assume that all dynamical equations are on equal foot: the distinction between fundamental and non-fundamental laws still runs and can well serve to ground a structural arrow of time. Blatantly, co-extensiveness is also preserved as long as HA claims that *only* T_U *truly* represents time reversal within SQM, whereas T_A rather stands for motion reversal.

In endorsing HA, the straightforward consequence is that the Schrödinger equation turns out to be non-time-reversal invariant, and thereby, to exhibit a structural preference for one direction of time. Moreover, one would, too, be precluding an overarching attitude in physics in respect of the time-reversal operator, to wit, it must be theory relative. Under HA, the time-reversal operator is no longer theory relative, but it instead tries to capture all-sharing, global features of time reversal condensed in the role that the variable t plays in physical theories. Though time reversal can only mean to invert t's sign, the consequences will greatly vary depending on the relation between theory's dynamical variables and time. Difficulties for endorsing HA hence follow from casting away deeply-ingrained suppositions about time reversal in physics and from being mostly inspired in a substantivalist stance with respect to time.

Final Remarks

In this chapter, I have linked what was said about time-reversal and time-reversal invariance in the previous chapter with the problem of a structural arrow of time within SQM. I have argued that OA's inner mechanism of justification (which has some assumptions and imposes some conditions) and the very formulation of the problem conflict one another. Therefore, and this is the moral of the story, a cautious attitude should prevail in drawing any metaphysical conclusion from time-reversal invariance in SQM. Philosophers of physics and time faces thereby the following scenario. They must either

(a) consider the claim "the Schrödinger equation is time reversal invariant (because it is T_A-invariant)" as metaphysically uninteresting, *implicaturing* that OA is not a useful view for shedding light on a structural arrow of time in quantum theories. By doing this, they preserve the orthodox approach to time reversal in SQM and its consequence about time-reversal invariance.

Or,

(b) decline OA and address the problem of a structural arrow of time in terms of HA. By doing this, any claim about (non)time-reversal invariance in SQM would be metaphysically interesting. By doing this, they will be rejecting OA's consequences about time-reversal invariance in SQM as long as SQM turns out to be non-time-reversal invariant *because* it is T_U-invariant.

However, there is a third option for OA's defenders that want to (i) preserve OA as it stands, and (ii) engage in the debate on a structural arrow of time. They may also

(c) decline that the very formulation of the problem of a structural arrow of time be the right one.

I have nonetheless argued that this third option doesn't look, at least at first glance, too promissory.

I have not argued that either OA or the very formulation of the problem of the arrow of time are misguided views taken *by themselves*. On the one hand, OA has steered physicists to a deeper understanding of symmetries in quantum theories and to valuable empirical research. No doubts about it. On the other hand, this formulation of the problem, along with its conditions, has successfully managed to pose the problem of the arrow of time in a systematic, clearer way. However, each view seems to be guided by unalike aims and motivations, and this imposes certain limits when addressing philosophical issues. I think that the possibilities for formulating a philosophically interesting notion of time reversal in physics from which one can learn about time's properties is one of the challenges philosophers of time and of physics should focus on in the future, particularly, in quantum theories.

Part 3

Introduction

Part 2 left us with two take-home messages as to the relation between physics and metaphysics when dealing with the problem of the arrow of time:

(a) An eye should always keep on how physical and formal notions are defined within a theoretical framework since they might trivialize a philosophical problem. In this case, the orthodox definition of time reversal is not conceptually useful to address the philosophical problem of a structural arrow of time. So, one must be cautious with the sort of metaphysical claims about the nature of time being drawn from the often-repeated claim that SQM is time-reversal invariant.

(b) But, going deeper, the very formal definition of time reversal is crucially determined by one's previously-assumed metaphysical and epistemic commitments. This was fairly clear when distinguishing T_{Rel} from T_{Sub} and their instances. In other words, metaphysics plays a central role in determining how time must be reversed and how one comes to formally represent it; role that has been largely overlooked in the literature. Even though the labor of theoretically defining a time-reversal operator is to good extent formal, one *decides* to call *that* piece of mathematics "time reversal", to connect it with a previously-devised concept of what time is and what its reversion amounts to.

It's clear at this point that physics (or mathematics) alone is not only unable to provide an exhaustive answer to the problem of the arrow of time (whichever version one likes to take), but is also unable to set the problem up, even in the very basic sense of getting the relevant notions univocally defined. Neither is it the case that one was given a univocal and clear-cut physical and formal basis one *then* has to interpret so as to strengthen our conceptual understanding of the matter, as if philosophy were the owl of Minerva, always coming too late, entering the scene only when the shades of night are gathering, paraphrasing Hegel's adage. Contrarily, philosophy comes first in the fundamental sense of, for instance, bridging the gap between a piece of mathematics and what its intended referent is supposed to be.

And yet, there is one further sense in which metaphysics plays a crucial role in addressing the problem of the arrow of time in *quantum mechanics*. One is probably more familiarized with this third sense, though it has been scarcely addressed in the literature, to the best of my knowledge, in relation to the arrow of time. This third sense, which I'd like to think through along this third part, has to do with the so-known *interpretations* of SQM. It has been already mentioned in passing that one doesn't get yet any full-fledged physical theory with SQM as it stands; though partially interpreted, SQM is still just a powerful piece of mathematics lacking the relevant elements to get any connection with the physical world out and down there. Before getting ahead ourselves, let's see briefly how this issue pops up.

If one stops for a while and think of SQM as a scientific *theory,* one should be able to at least give an answer to the following question: SQM is a theory…*of what*? Is it merely a theory about outcomes in measurements? Is just a powerful and complex algorithm to work out statistical correlations? If this is so, and I can imagine that many physicists would come along with it, then SQM is merely a "cook book" (to use John Nash's witty expression) that one doesn't really understand and, let me add, one is not supposed to understand. Nonetheless, all this may turn out to be a bit disappointing if one takes SQM to be a *scientific* theory. And if one comes to agree that a *scientific* theory is not merely a repertoire of procedures and rituals providing useful and practical calculations, but it is rather somehow in the business of rendering some understanding of what's going on out there in the natural world, the answer should then be entirely different.

In particular, the answer must involve the necessary elements to bring up a picture of what the world is like according to SQM. But in providing such an answer, one is readily stepping beyond SQM and getting into the realm of its various *interpretations.* If one stops here and think of SQM thoroughly, what one merely knows about it thus far is, first, that SQM is a powerful tool for predicting results in measurements and for accounting for statistical correlations checked by highly-tuned experiments; second, that some no-go theorems have been proved in the last decades constraining, if you'll pardon the repetition, what things don't go within the theory –but remaining absolutely silent about what things *do* go; and, third, that a persistently puzzling issue lies at the core of the foundations of SQM –the so-called *measurement problem*, also popularly known as Schrödinger's cat problem. Interpretations of SQM might be different attempts to overcome the measurement problem and, in this, to tell one a complete *scientific* story about what's going on at the quantum scale.

And all this is crucial for the problem of the arrow of time in the quantum realm because many of the divergent proposals to overcome the measurement problem have *modified*, or *added* stuff to, SQM. In this enterprise, new temporally-asymmetric elements (in any of its versions) might enter the theory. Hence, I shall take for a quantum theory to be a full-fledged quantum theory, it ought to feature, or make explicit, the following points:

(a) SQM's formalism (what was already analyzed in Part 2) and to reproduce usual quantum-mechanical predictions and results.
(b) A claim about whether any modification of, or addition to, SQM is required. Different interpretations can be thus aligned in two sides as to whether SQM is complete or *in*complete.

Some have argued that these two points aren't enough. In order to know what a physical theory is really and *fundamentally* about, and in order to rightly read off the metaphysics from the physics, one needs to postulate a "primitive ontology" (PO) for thus-interpreted quantum theory. Thereby, I shall take for a quantum theory to be a full-fledged quantum theory it ought to also feature

(c) A claim making explicit whether so-interpreted quantum theory has a PO or not.

Along this third part, I shall deal with the problem of the arrow of time (either of its versions) in the framework of some interpretations of SQM. As it is well-known, several foundational problems around how SQM must be interpreted arose a few years after the mathematical formalism were consolidated, and plenty of interpretations have appeared in the field ever since. I don't have enough room to get into much details about them, so I will just consider the most relevant ones in the current literature and those that explicitly modify SQM's formalism.

This third and last part of the thesis is organized as follows. After a brief detour through the measurement problem and how interpretations of SQM appear (Chapter 6), I shall consider, in Chapter 7, a first minimal extension of SQM proposed by Paul Dirac (1930) and John von Neumann (1932), which mainly consists in adding an extra dynamical postulate to SQM, the so-called "Collapse Postulate" (either considered as a physically real process or seen instrumentally as non-physical). This formerly extension of SQM (SQM+CP henceforth) has been so widely-extended in the field that is often confused with SQM itself. This *orthodox* interpretation of SQM (OrQM) has been target of several criticisms and has been ultimately abandoned in the philosophy of physics community, though it is still quite extended in the physicists' community (wherein the collapse postulate is generally seen as instrumental). Furthermore, it has been claimed that SQM+CP provides the means to establish an objective quantum arrow of time. Therefore, it will be useful to introduce the idea of "collapse of the wave function" and whether it delivers a quantum-mechanical base for an arrow of time as a starting point.

In Chapter 8, I shall introduce a more elegant and refined family of collapse theories, the so-called "spontaneous collapse theories". In this chapter, I will exclusively focus on the version originally proposed by Giancarlo Ghirardi, A. Rimini and T. Weber (1985, 1986), and P. Pearle (1989), the so-called "GRW interpretation". Conforming to it, quantum systems not only evolves according to a deterministic and linear equation of motion (the Schrödinger equation), but they also undergo some spontaneous and indeterministic quantum jumps: Given a certain amount of time, a quantum system evolving according to the Schrödinger equation, whose state is represented by a bell-shaped wave function, is bound to undergo a quite different evolution that will localize the system in the space, destroying its bell-shaped probability distribution. In this framework, I shall analyze whether this non-linear and indeterministic dynamical evolution offers the grounds for a structural arrow of time.

Finally, in Chapter 9, I shall address one of the most radical modifications of SQM, *Bohmian mechanics*. Like GRW, Bohmian mechanics considers that SQM's dynamics should be modified. In order to do this, Bohmians introduce the "guidance equation", so that quantum mechanics involves two dynamical laws: the Schrödinger equation, on the one side, and the guidance equation, on the other. But, in contrast with any other interpretation, Bohmian mechanics claims that the quantum state is not fully given by the wave-function alone, but also include particles' positions: quantum mechanics is a theory about particles that always have perfectly defined positions. In some respects, Bohmian mechanics provides a picture of the world that recalls some aspects of classical physics (though, of course, diverges from it in many relevant and deep points): particles with definite positions, following definite trajectories. And both dynamical equations just aim at describing how particles' positions change as time goes by. I shall analyze how Bohmian Mechanics deals with temporal asymmetries within its framework and what role time reversal plays in building up the theory.

VI.

Towards a full-blooded quantum theory

The measurement problem and interpretations

The *measurement problem* can be posed in the following plain way: SQM involves an intrinsic paradox that seems to be unsurmountable from SQM's means alone, namely, that what SQM's dynamics predicts is flagrantly in contradiction with what one observes in experiments. So, it seems that the theory is in need of further resources to escape the paradox; otherwise, it is completely off the right track –its dynamics just yields *wrong* predictions. Let's get deeper into this.

Section 1. The measurement problem and interpretations of SQM

The paradox arises from SQM's dynamics, what it predicts for following times, and what one gets when experiments are carried out. Tim Maudlin (1995) has proposed three ways of understanding the measurement problem: "the problem of the outcomes", "the problem of statistics" and "the problem of effect", the first being the most widely extended. According to Maudlin the measurement problem involves three mutually inconsistent assumptions:

A1. The wave-function of a system is complete.
A2. The wave-function always evolves in accord with a linear dynamical equation (the Schrödinger equation).
A3. Measurements have always determinate outcomes.

How these assumptions are mutually incompatible can be seen in the following familiar example.

Suppose that one conducts an experiment to measure an electron's z-spin, which can adopt any of two values: either z-spin up $|\uparrow\rangle$ or z-spin down $|\downarrow\rangle$. Suppose one also has a (macroscopic) device configured to measure *z*-spin, a "z-device". Measuring the *z*-spin of the electron mainly means that it passes through the *z*-device, and that at the end of the experiment one gets a definite answer about whether the electron is either in a *z*-spin up or in a *z*-spin down state. For instance, this can be straightforwardly realized by a pointer pointing out to the obtained value, "z-spin up" - "z-spin down", on the device's display. So, if the *z*-device measures *z*-spin up, the pointer will point "z-spin up", and likewise when measuring *z*-spin down. This means that measure device's states (either having a pointer pointing to "z-spin up" or to "z-spin down") get correlated with electron's state and its *z*-spin's value. This is, roughly, how a well-behaved *z*-device should ideally work and how a well-configured *z*-experiment should ideally run.

In order to carry out the experiment, the electron must be firstly *prepared* in some state. One doesn't know in which *z*-spin state the electron is, but one knows that it was prepared in a, say, *x*-spin up state. So, the quantum state of the electron before entering the *z*-device is $|\psi(t_0)\rangle_x = |\uparrow\rangle$ in the *x*-spin coordinate. From SQM's formalism, one well knows that a state written down in a *x*-spin basis can be rewritten as a superposition of states in a different spin basis. In particular, the state of an electron in some eigenstate of the observable *x*-spin can be also expressed as being in a superposition of states in the *z*-spin basis,

$$|\psi(t_0)\rangle_z = \sqrt{\frac{1}{2}}(|\uparrow\rangle_z + |\downarrow\rangle_z) \qquad (6.1\ )$$

In this way, so-prepared electron enters the *z*-device at t_0 in a superposition of *z*-spin states. Additionally, the *z*-device has to be also prepared in an initial state, typically called it a "ready state". When the electron enters the *z*-device one gets the composite system consisting of the electron in *x*-spin up state correlated with the *z*-device in a ready state

$$|\psi(t_0)\rangle_z = \sqrt{\frac{1}{2}}(|\uparrow\rangle_z|ready\rangle^D + |\downarrow\rangle_z|ready\rangle^D) \qquad (6.2\ )$$

After running the experiment, as mentioned before, one expects that the final *z*-spin state of the electron gets correlated with the device's pointer pointing to *either* "z-spin up" *or* "z-spin down". That is, one expects that the composite system at t_1 evolves into either $|\uparrow\rangle_z|"z\,up"\rangle^D$ or $|\downarrow\rangle_z|"z\,down"\rangle^D$. This is to be expected simply because that is how experiments have been working so far: pointers have always pointed to definite values (assuming that the measurement device was rightly set up, the experiment carefully carried out, and so forth). And, in fact, one is completely right in assuming it: after running the experiment, that is, after measuring electron's *z*-spin, things passed off as usual: one obtained a definite result for the electron's *z*-spin, say, a pointer pointing to, say, "z-spin up". This is the assumption A3 in Maudlin's first version of the measurement problem.

This is what one always observes. And SQM has been proved to be an incredibly powerful and good piece of mathematics for predicting the probability of obtaining either "z-spin up" or "z-spin down". However, and this is the upshot of all this and from which the measurement problem arises, how things in fact end up being is not what is technically predicted by SQM's dynamics. In other words, things do not turn out the way they *should* according to SQM's dynamics. And here A2 comes in.

The problem is that the dynamics governing the evolution of such a composite system is invariably *unitary* and *linear*. If one stops for a while and think carefully how the composite system should evolve according to SQM's dynamics, one notes that the Schrödinger equation is a purely deterministic, unitary and linear differential equation. This means that if one starts with a state like (6.2), given the adequate interaction between the system and the apparatus, one should after a time interval get a state like

$$|\psi(t_1)\rangle_z = \sqrt{\frac{1}{2}}(|\uparrow\rangle_z|"z\ up"\rangle^D + |\downarrow\rangle_z|"z\ down"\rangle^D) \qquad (6.3\ )$$

And one should get to the same result at t_2 and after three million years. In a nutshell, what SQM's dynamics tells us is that, once the composite system has evolved in a superposition of states, nothing in the SQM's dynamics entitles us to throw one of the terms away. Thinking of it in terms of macroscopic objects, this would amount to a very weird state where, in Albert's words, there would be no matter of fact about where the pointer is pointing to (Albert 1992: 76).

It doesn't make any difference to this result whatsoever if one intends to escape the paradox by introducing further stuff in the chain of superpositions: an aware, competent observer looking at the pointer on a display will also end up in a superposition of mental states according to SQM's dynamics, and a friend in Taiwan that will eventually read an e-mail with the results of the experiment will also end up in a superposition according to SQM's dynamics. No matter what you add to the set up, you only get a larger and larger chain of superpositions, and a weirder and weirder picture of the world consequently. Therefore, something deeply bad underlies SQM as it stands, mainly, regarding A2 and A3.

There are various ways to manifest what is wrong with all this. Some think that the problem lies in a flagrant contradiction between our observations and SQM's dynamics (A2 and A3). At the end of the day, besides their accurate probabilistic predictions, SQM is simply, and deeply, wrong because predicts states that have never been observed (see Albert 1992, Ney 2013). Others express the problem a bit differently: though SQM provides a very effective algorithm to predict macroscopic phenomena, it falls short of being a proper physical *theory* explaining the success of this algorithm (see, for instance, Wallace 2007). Be as it may, the paradox is clear enough for one to become aware of that something must be done if SQM is, in any relevant sense, a scientific theory of nature, conveying positive knowledge about the natural world and about what happens when experiments are carried out. In other words, SQM has to somehow find the way of linking what is observed, on the one hand, with what its dynamics predicts should be observed, on the other. Both things should desirably match.

So, something should be changed or carefully revised. To begin with, it has been supposed that SQM's formalism is overall right. And probably this was a mistake and one should in consequence consider discarding SQM as a whole. This would however be too drastic: after all, it is *this* formalism that has proven to be amazingly successful at experimental predictions. So, it seems not to be a promissory way-out to completely get rid of SQM. As an alternative, it could be considered that experimental results were not-well-confirmed. This would be drastic, too, since they have been performed with higher and higher precision from the very outset of the theory, and superpositions have never been experimentally observed (though Dieks 2019 suggests something slightly different).

So, one should think of modifying either A2 or A3. And this is what various interpretations of SQM basically do. First, one could claim that the quantum dynamics is not fully exhausted by A2: The Schrödinger equation doesn't tell us the whole story. Second, one could rather suggest that A3 should be approached from the right angle, reading what the formalism *really* says. Interpretations of SQM can hence be aligned in two sides:

(a) Those believing that SQM's dynamics is complete (that is, there is no need to either modify or to add further elements to the theory) and that one should just change the angle from which the theory and experiments are typically regarded. They thereby decline A3;

(b) Those believing that SQM's dynamics as stands is *in*complete and must thus be supplemented with further elements, for instance, a new dynamical postulate. They thereby decline assumption A2[36].

David Wallace (2007) says that people aligned with the first option propose a *pure interpretation* of the theory to the extent to which they claim that the solution of the measurement problem doesn't hinge on altering the formalism. Advocators of the second side rather promote, according to him, a *modificatory interpretation* since a full-blooded quantum theory should feature a different dynamic that somehow includes SQM's.

Naturally, particular instances of these general positions come in degree as they, for instance, differ with respect to how much the dynamics must be changed, or how radical the revisions of our view of the world must be for it to harmonize with the quantum theory. However, most interpretations of SQM available in our days can straightforwardly be aligned with either (a) or (b)[37]. For instance,

- Wave-function realists (Albert 1996, 2013, Albert and Loewer 1996, Ney 2012, 2013), Everettians (Everett 1957), many-world's or many-mind's defenders (Deutsch 1986, 2002, Wallace 2012) think that SQM's formalism is complete and indisputably right. The problem is in some way just apparent, since one is not considering experiments through the right perspective. Instead of changing the formalism, one should firstly learn to read the theory correctly and then come up with a story about how observations, as one knows them, come out from SQM. This might imply that some of our philosophical convictions about what the world is like ought to be radically revised (for instance, by accepting that there is no a single world but many, all them comprehended in a Multiverse; or that the real space is not three-or-four-dimensional, but higher dimensional).

[36] Some of whom deny A2 also deny A1 in claiming that the wave-function doesn't specify all of the physical properties of the system, so that the quantum state should be specified by adding something else to the wave-function (e.g. particles' position configuration).

[37] It is worth noting that there is a different general stance with respect to the nature of ψ that has paved the way for purely epistemic interpretations of ψ. In brief, ψ-epistemic view (as they are called) consider that ψ is not an aspect of the quantum reality, but it reflects a subject's incomplete knowledge about an underlying ontic state (see Spekkens 2007, Harrigan and Spekkens 2010). This view paved the way for some interpretations as Quantum Bayesianism or Qbism, where a quantum-mechanical state is just summary of the observers' information about an individual physical system. Information that changes according to how observers get new information by, for instance, measurements (see Fuchs 2010). Even though ψ-epistemic views have been relevant for discussing retrocausality (see Friederich and Evans 2019), any temporally (a)symmetry coming from there would be non-objective, in the ontological sense in which "objectivity" for an arrow of time was defined in the Introduction. That's why I won't consider ψ-epistemic views in this thesis.

- Bohmians (in its different versions, Bohm 1952, Dürr, Goldestein and Zanghi 1992ab, 1997) or collapse theories' supporters (in any of their different kinds, von Neumann 1932, Ghirardi, Rimini, Weber 1986, Pearle 1989) rather believe that the story told by SQM alone is *incomplete* and, when taken literally, plainly wrong. Something else must thus be added to SQM in order for quantum theory to be complete. And this generally means either to modify the dynamics or to add a new dynamical law.

Coming back to my business here, alternative interpretations of SQM may hence come to tell a different story about whether a quantum theory is equipped with a structural arrow of time since the very dynamic varies along with them. That's why interpretations of SQM should be taken into consideration to address the problem properly and to its full extent. I will come back to it in the following chapters. Before that, I'd like to add, and to discuss, a noteworthy ingredient in this picture.

Section 2. Primitive ontology and symmetries

There is yet a further step forward one can take. Some have proposed, and actively promoted, that any interpretation of a physical theory should make also explicit the *fundamental ontology* that such a theory is about. The "primitive ontology approach" (PO), as it comes to be known ultimately, is an attempt to bridge the gap between a physical theory's (partially) interpreted formalism and the ontology of the three-dimensional objects one daily has experience of. Having probably its origin in the notion of "local beables" of John Bell (1987), the approach was firstly introduced by Detlef Dürr, Sheldon Goldestin and Nino Zanghi (1992: Section 2.2), and developed further by Allori et. al (2008, 2014), and Allori 2013.

In a nutshell, the PO posits the "basic kinds of entities that are to be building blocks of everything else" (Dürr, Goldestin, Zanghi 1992: 29). Yet, these "building blocks" cannot be naively read off from a theory's formalism or from textbooks but have to be previously introduced as the referential content of such a formalism. Sometimes a theory's formalism involves mathematical items inhabiting in the phase space or a high-dimensional configuration space. But, according to PO's advocators, it would be misleading to read off the ontology of the actual world from this: if physical theories intend to explain what one macroscopically observes (things like pointers pointing to a certain value on a display), if physical theory intends to provide some positive knowledge about what the three-(or-four)-dimensional world is like, then something else ought to be introduced in order to make completely explicit what the ontology of a physical theory is: the PO is *of what* matter is fundamentally made, and is *about what* the physical theory is. Consequently, any suitable PO will imperiously be:

(a) *microscopic* rather than macroscopic since it needs to specify the building blocks of the macroscopic world,
(b) *three-(or four)-dimensional*, that is, defined in a three(four)-dimensional space rather than in a more complicated high-dimensional space.

This explains why the PO approach has gained so many supporters lately in the field of quantum mechanics: SQM, and many interpretations thereof, would suggest that the theory is actually, e.g., about the *wave-function*, but leaving completely unclear how it connects with the macroscopic world. To put it a bit differently, SQM somehow entails (if you take it seriously) that chairs, dogs, buildings, planets and eggs are fundamentally made of wave-functions. But wave-functions inhabit a high-dimensional space, whereas chairs, dogs, buildings, planets and eggs inhabit a quite different three(four)-dimensional space. As things stands, nothing in the theory allows one to straightforwardly and directly link wave-functions to the macroscopic surrounding stuff. Hence, the theory seems to be in the uneasy situation of having to overcome the gap between its formalism and the real world. Well, the PO approach seeks to puts theory's feet back firmly on the ground by selecting which variables and which elements are to be taken as representing something that exists in the real world. And the theory will simply and ultimately be about such worldly things.

Though the PO approach can well be considered as a useful way to do metaphysics, it has been also suggested that a PO has a strong influence on the relation between a theory's formalism and its symmetries (Allori et al. 2008, Allori 2013), even in relation to the arrow of time (Allori 2019). In this sense, the PO is expected to have some influence on how one comes to understand temporal asymmetries within a physical theory (Allori 2019). It is worth digging into it a bit more deeply because the way in which PO's supporters relate a theory's symmetries to the PO could, to my mind, lead to some misunderstandings about how a symmetry transformation should be settled. Whereas I'd like to make the case that the PO approach could in fact give us some hints about how symmetry transformations must be implemented within a physical theory (under some assumptions), I'm afraid that the current treatment in the literature falls short in strongly tying theory's symmetries and the PO.

In the PO's language, that a physical theory has a given symmetry means that

> The possible histories of the PO, those that are allowed by the theory, when transformed according to the symmetry, will again be possible histories for the theory, and the possible probability distributions on the histories, those that are allowed by the theory, when transformed according to the symmetry, will again be possible probability distribution for the theory

Let me explain the definition a little bit further (I'll be partially following Allori et al. 2008 in this). "The possible histories of the PO" simply refers to the possible evolutions of the PO according to the laws of the theory, representing "physically possible worlds" (in the sense of "*nomologically* possible worlds", Earman 1989) for the PO's behavior or distribution.

Among those possible evolutions, some will be product of a symmetry transformation, for instance, by temporally translating the system forward in time by 5 seconds or by reversing it in time. In this sense, *if* the symmetry holds, the original evolution e will be linked by the symmetry transformation *S* to a *S*-transformed evolution, e^S. Allori et al (2008) says:

> "since the PO is represented by a geometrical entity in physical space (…), space-time symmetries naturally act on it, for example transforming trajectories $Q_i(t)$ to trajectories $\tilde{Q}_i(t)$" (2008: 366).

Even though in some cases it might looks so, the remark may result a bit confusing as it could suggest that the symmetry transformation can be applied *in isolation* to a single trajectory or to the PO directly. The symmetry transformation acts indeed upon variables, observables, states and presumably upon the PO, but always *within* a dynamical equation of motion that, so to speak, *generates* the evolutions. And the symmetry holds *if* the dynamical equation of motion still generates evolutions (that is, series of S-transformed variables, S-transformed observables, S-transformed states and presumably a S-transformed PO) when so S-transformed. To put it in another way: the symmetry transformation doesn't act upon, say, the trajectory $Q_i(t)$ but upon the dynamical equation that produced it. *If* a symmetry holds, then so-transformed dynamical equation would produce something like $\tilde{Q}_i(t)$. A fortiori, $Q_i(t)$ will be $\tilde{Q}_i(t)$ linked by a mapping.

This is so because for some symmetry transformations what it is truly relevant is to transform the dynamical conditions (whatever they can come to be) that make an evolution be possible, and this comes from the dynamical equation of the theory, not from the PO (as Allori et al.'s remark might suggest). Given a PO, say, of a matter-density field, nothing *there* says how the PO ought to evolve in any imaginable situation because the PO is not the sort of stuff that bears such information. The PO left on its own is not the sort of thing that evolves in any relevant sense. The conditions for its possible evolutions must rather come from elsewhere, for instance, from the wave-function and the dynamical laws that, so to speak, "govern" it. And all this altogether must be transformed for one to check whether a symmetry holds or not.

All this matters because it should be left completely clear that the PO cannot be, a priori and without further assumptions, the sort of thing that would guide one to figure out how a symmetry transformation must be defined and applied. In some situations, as I will show later on, the PO could provide us some hints about how a symmetry transformation must be defined and applied within a fully-interpreted quantum theory with a PO, but this, I think, shouldn't be regarded as an overarching rule.

There is one further reason. At the end of the day, as Allori et al. also recognize, one has to figure out how the wave-function (and the Schrödinger equation consequently) must be transformed under a given symmetry transformation. And I've already showed how puzzling can be this in some cases, like time reversal. The authors for instance say

> Notice that $Q_i(t)$ and $\tilde{Q}_i(t)$ may arise in BM (Bohmian mechanics) from different wave functions. In other words, the wave-function must also be transformed when transforming the history of the PO. However, while there is a natural transformation of the history of the PO, there is not necessarily a corresponding natural change of the wave-function (Allori et al. 2008: 366)

Let's stop here for a second and think of *where* such histories (possible evolutions of the PO) come from in Bohmian mechanics (as it is the theory they are considering). As mentioned before, the PO, when left on its own, is unable to generate any evolution, so there is no one single history for the PO if the dynamics is not previously specified. In fact, it's the dynamics who tells the PO how to behave. And this thereby specifies which its possible histories are: otherwise, the PO remains unmoving in a rather Parmenidian changeless world. Yet, those histories can only come from the wave-function since it is the element that plays the nomological role of "governing" (to use Dürr, Goldstein and Zangui's expression) the PO's behavior and of generating its histories. And therefore, there is no one single possible way to transform a history without specifying previously how the dynamics ought to transform, that is, how the wave-function transforms. What the quote seems to rather suggest is that one should first find the way to transform a history of the PO, which is in addition a *natural* way, and then to transform the dynamics accordingly. And this is just to put the chart before the horses.

Let's approach the same point through a different angle. Take a transformation like time reversal and suppose you start off transforming a single history of the PO, which consists of particles plus their trajectories. So, what you want to know is whether for a given trajectory $Q_i(t)$ exists a temporally-inverted trajectory ${Q_i}^T(-t)$ that is also a solution of the (time-reversed) dynamical equations that generated the first evolution. What the authors seem to suggest is that there is firstly a *natural way* of obtaining ${Q_i}^T(-t)$ based on the theory's PO, independently of the dynamical conditions that can ever generate it. But ${Q_i}^T(-t)$ can only come from, or being produced by, either some (transformed) *quantum-mechani*cal dynamical mechanism or some other (naturally-transformed) dynamical mechanism. If the latter, one should explain which it is and comes up with a story that doesn't ruin things completely –a coherent story telling why quantum things going backward in time evolve conforming with a non-quantum mechanism, why this is *natural,* and so forth. If the former, one is taken back to the wave-function, because it is at least one of the ingredients that nomologically produces the trajectories. But, again, nothing in the PO in itself says how the wave-function or the laws governing PO's evolutions should be transformed. This should come from elsewhere.

The point I'd like to make is the following. The PO is completely *inert* without a dynamic specifying with accuracy how it behaves. The PO bears no information about its possible histories, but rather that a theory's dynamics bears it. Hence, when space-time symmetry transformations are at issue, one should be extremely careful about the role that each element is playing within the theory. That is, *what* is after *what*. Suppose that world's fundamental physical theory is "the theory M", which tells one mainly two things: the world is made of particles and their positions (its PO), and particles behave according to a weirdly crazy law that prescribes that particles generally evolve according to the Schrödinger equation, but that when two or more particles are close enough each other (where "close enough" will amount to a new constant of nature), all of them disappears and re-appears in a completely different random place instantaneously. Suppose that one desires to know whether the theory M is temporally symmetric. No matter how much time one spends looking into the inner nature of particles or their positions: nothing there will ever say a thing about what happens with particles and their positions in the following ten minutes, or what happens when two or more particles come to be close one another. In consequence, nothing there will ever either say a thing about whether the theory turns out to be time symmetric or not under a given transformation. What one should instead look into is the weirdly crazy dynamic, which says to the PO how to behave and what its possible histories may ever be.

In a recent paper, Allori (2019) directly links the PO with the arrow of time. In a nutshell, she proposes a PO-based approach to restore time-reversal invariance in quantum mechanics.

The argument runs as follows. Some have argued that quantum mechanics is *not* time-reversal invariant either by considering that the Schrödinger equation is not time-reversal invariant (she only quotes Callender 2000 for this case) or by considering the collapse postulate. For the first case (which I'll briefly consider here), the argument can be sketched in the following way:

1. The complete description of any system is given by its state S (definition of state)
2. What is in the state S represents physical objects (ontological assumption)
3. In SQM the wave-function provides the state S.
4. The wave-function evolves according to the Schrödinger equation when no experiment is performed.
5. The time-reversal transformation T transforms a sequence of states $(S_i), \ldots, (S_f)$ into a temporally-reversed sequence and reverses states (definition of T on sequence of states). $T(S_f), \ldots, T(S_i)$.
6. T changes t in the state into $-t$: $T(S(t)) = S(-t)$ (definition of T on a single state: the ontology does not change depending on whether we look at the forward or backward sequence).
7. A deterministic theory is time-reversal invariant if forward solutions are mapped onto backward solutions (definition of time-reversal invariance)
8. This doesn't happen unless we change the wave-function with its complex conjugate (mathematics)
9. The wave-function is a scalar field (it is a function on configuration space)
10. If premise 9 holds, then there is no reason why wave-function should transform into its complex conjugate under T (physical meaning of scalar field)
11. Therefore, when SQM describes systems that are not observed, it is not time-reversal invariant.

The argument, according to Allori, hinges upon premises 3, 5 and 6, so it can be dismantled by rejecting them. But, before getting into it, why shouldn't one accept the conclusion? Allori seems to find radically problematic "to reject a symmetry". She says:

> "However, I believe that also this view is difficult to defend, in fact, symmetries have always had an important role in theory development of new theories. Also, symmetries are used in theory evaluation as well (…) Thus, all other things being equal, rejecting symmetries seems to be an unnecessarily radical move" (Allori 2019: 4)

I think this is her very starting point: quantum mechanics cannot be non-time-reversal invariant, so time symmetry must be restored by any means. As I showed before (Chapter IV, Section 4), this already implies a certain view of symmetries, namely as guides to theory construction. And as I showed before as well, there is always a formal way to restore a symmetry: abstractly speaking, for any dynamical law one can put some effort to find some transformation that leaves the equation invariant and call it "time reversal". This is not bad per se, though further arguments should be provided for one to be entitled to call *that* transformation 'time reversal'. When metaphysical concerns are at issue, this looks a bit unwarranted (see Chapter V).

But let's move on. Allori claims that tradition (basically, Wigner 1932, Earman 2002, Aharanov and Rohrlich 2005, and Roberts 2017) has generally rejected premises 5 and 6 in the above-sketched argument to restore time-reversal invariance. What does the job is to consider that a wave-function is a ray in Hilbert space, so that it is natural for T to act on the wave-function taking its complex conjugate in order to make the theory time-reversal invariant. This approach however suffers, according to Allori, a problem: "*The mismatch problem*". It says that the backward wave-function is actually a complex conjugated wave-function, whereas the forward wave-function isn't. Therefore, if one accepts that T changes the wave-function into its complex conjugated, then it changes the ontology, since both ontologies (forward and backward) are different (one features a wave-function, the other a complex conjugated wave-function).

I have already analyzed the traditional approach carefully in Part 2, so I won't get into details here. But Allori puts forward a novel ontology-based argument to dismiss it. In particular, the complex conjugation would seem to change the ontology of the theory. Hence, when time reversal implies a reversion of the direction of time *plus* complex conjugation, the ontology changes in consequence depending on the direction of time. However, this is a bit strange. Formally speaking, the wave-function is a complex number. If the wave-function is somehow supposed to give us probabilities (that is, real numbers), one needs to get rid of its complex part. The formal procedure to achieve it is by multiplying the wave-function by its complex conjugate[38]: probabilities are given by the squared norm of the wave-function which is equal to $|\psi(x,t)|^2 = \psi(x,t)\psi^*(x,t)$. But it hardly represents a change in the ontology in any relevant sense: a complex-conjugate wave function is in any respect equivalent to an ordinary wave-function. Furthermore, one can even argue that there are no two different wave-function at all: there are rather two wave-functions and a relation holding between them, wherein one is the complex conjugated of the other and vice versa.

Allori's proposal is to rather reject premise 3: the wave-function is not physically real, so it solely doesn't represent the quantum state. What is rather physically real is represented by

[38] In general, if $z = a + ib$, its complex conjugate is $z^* = a - ib$, where i^*i is equal to -1. The norm of the complex number z is $|z|$ and is positive. Hence, the norm of z must be real and given by $|z| = \sqrt{a^2 + b^2}$. From this, it follows that $|z|^2 = zz^*$.

a spatio-temporal entity given by the PO of the theory. So, to rebut the above-sketched argument one should *only* focus on the PO and somehow figure out how it will transform under time reversal. Only after that, one should transform the wave-function whatever be needed to keep the PO invariant under time reversal. Allori says:

> "the symmetries of the theory should be determined by its primitive ontology, not by the wave-function. It is thus wrong to focus on the fundamental equation for the wave-function when looking at symmetry properties; rather, one should focus on the law of evolution of the primitive ontology" (Allori 2019: 6)

As I mentioned before, the PO on its own says nothing about its possible histories, so it hardly says something about how the time-reversal transformation should act upon them. In any case, one should rather look to the dynamical law governing PO's behavior. However, directly or indirectly, any so-far-proposed law for it depends on the wave-function or whether the Schrödinger equation is time-reversal invariant. I will come back to it in Chapter 9 when analyzing Bohmiam Mechanics, but it's well-known that, on the one hand, if the Schrödinger equation is time-reversal invariant, then so is the guidance equation; on the other, that in order to write down the guidance equation one ought to take T_A, which is justified because it's the only one that leaves the Schrödinger equation invariant (and the guidance equation consequently). Therefore, in one way or another, one should *beforehand* define how the wave-function (and the Schrödinger equation) must transform under time reversal, which takes us back to Part 2. And this is reasonable because the wave-function, directly or indirectly, is what plays a nomological role in the theory.

The moral of the story is the following: What generates the possible histories of the PO is not the PO but the theory's dynamics. So, if one wants to know *whether* there exists a possible history for this particular PO under some circumstances (for instance, whether there exists a symmetric past-headed history), the question must be then aimed at the dynamics that generated such a history, and not at the PO in isolation. Therefore, it cannot be the case that, on the one hand, there exist a natural way of transforming the histories of the PO and, on the other, that the wave-function be allowed to "change in any way, solely determined by its relationship to the PO" (Allori et al 2008: 366). The only way to have a meaningful symmetry transformation is by primarily specifying how the dynamics transforms and then whether so-obtained history of the PO is symmetric with respect to the original one.

This is so far the *par destruens*. This is in which way one shouldn't think that the PO guides or helps the purpose of defining a symmetry transformation. The *par construens*, that is, the sense in which I think that the PO does offer some hints, goes as follows. I mentioned before that Allori et al.'s rationale shouldn't be followed to the letter *as a general case*. Nonetheless, in certain cases, the PO *does* provide some hints for elucidating the relation among symmetries, symmetry transformations and the fundamental ontology.

To put it simply: most space-time symmetry transformations are meant to at first glance be transformation acting upon space-time (or space and time) in itself (themselves) –as Jill North vehemently puts it. Notwithstanding, it would be unpleasantly naïve to take such a first glance for granted: As I've reviewed in Part 1 and Part 2, one can stand on different philosophical sides as to whether time, space or space-time have an independent reality with physical meaning. As it was shown, substantivalists, relationalists and supersubstantivalists dispute about it.

The contend is not merely verbal for whom interested in symmetry transformations since, as I've argued in Part 2, a good deal of how the symmetry transformation is to be formally defined depends upon which of the above-mentioned positions one embraces. And the PO approach, as I will detail afterwards, turns out to be relevant for *relationalists* with respect to time, space or space-time: if time (or space, or space-time) is just an abstraction, and there is no time (or space, or space-time) over and above world's material content and its behavior, then the PO is the most promissory candidate to spell out what such material content actually is. Concretely, if time reversal is a reversion of the material content's behavior in time, then time reversal is primarily a reversion of the PO's behavior in time. Thus, the features of the symmetry transformation will be given mainly by the theory's dynamics *along with* its PO derivatively. In particular, the PO might provide a guide to distinguish between primitive or fundamental variables and non-primitive or derivative variables, which could influence how a symmetry transformation should be applied upon them. And in doing this job, Allori et al.'s remark can have some grip. However, one should be cautious here and not take the approach beyond its capacity, as shown before.

No every interpretation of SQM specifies a PO. On the one side, wave-function realists or Everettians believe that the wave-function is fundamental, that everything out there is fundamentally made of wave-functions. As wave-functions are not the kind of stuff inhabiting a three(four)-dimensional space, it doesn't qualify as a suitable PO. More orthodox-minded quantum theorists (as Copenhagen's followers) don't qualify either as they remain, at best, predominantly agnostic about the fundamental entities of the quantum world. Instead, they are merely confident about measuring devices and their outcomes –that is, about *macroscopic* objects. On the other side, so-called hidden variable theories (as Bohmian Mechanics) or some versions of collapse theories (as any GRW-type theory) do typically bring along a PO, either of particles or of flashes or of a density-matter field.

***

The discussion here has to do with the problem of the arrow of time in a deep, fundamental sense. When discussing whether time is directed *within* a full-blooded physical theory, the enquiry should at least start off from having a clear understanding of *which* is the theory at issue. In the quantum case, a full-blooded theory ought to feature

(a) SQM's formalism (what was already analyzed in Part 2).
(b) Any dynamical modification or complement, if one comes to think that SQM is incomplete.
(c) A PO.

If a good deal of the problem of the arrow of time relates essentially to the theory's *dynamics*, one has to at least take into account the complete set of dynamical laws involved by the (interpreted) theory under scrutiny. After this brief detour for the measurement problem, one finds that many have though of SQM as *incomplete*, so the arguments in Part 2 shouldn't be taken too conclusively by them: Some interpretations of SQM can bring up new elements to evaluate the problem of the arrow of time. In particular, they can provide further dynamical elements to evaluate whether the theory is non-time-reversal invariant. In the event that one comes to conclude that the Schrödinger equation is time-reversal invariant (for instance, if one assumes a relationalist approach to time and thereby formally represents time reversal through some instance of T_{Rel}), then one should carefully pay attention to other potentially time asymmetric elements that the interpretation may be introducing before to conclude that the theory (as a whole) is structurally time-reversal symmetric.

To be clear: what has been showed in Part 2 is, to my mind, an unavoidable result that everyone should take seriously when thinking about the problem of the arrow of time in the quantum realm. In one way or the other, the Schrödinger equation plays a paramount role in the majority interpretations of SQM available in the market at present. As I've argued above, any possible history or evolution comes at least from there. And, insofar as the claim of time-reversal invariance has been shown to crucially depend upon previously-assumed metaphysical and methodological commitments, one lacks any definite and unique answer to the question whether SQM is time-reversal invariant, and thereby, whether it treats both directions of time differently in a structural sense. This is something, I think, anyone worried about the arrow of time in the quantum realm should pay attention to since it sweeps along any interpretation.

What can be said about the arrow of time within interpretations of SQM that *don't* modify the formalism boils down, with a few exceptions, to what has been already said in Part 2, so that I won't address the problem of the arrow of time (in either of its versions) within wave-function realism or any interpretation that takes SQM's dynamics to be complete and that takes the wave-function as the fundamental stuff quantum theory is really about. Many-world interpretation deserves a side remark, though: there has been some discussion about whether the branching-world mechanism that Many-world imposes is temporally asymmetric. At first glance, it does seem to be asymmetric: the future of any branch (in this tree-like structure) will be uncertain, undetermined and multiple, while its past is certain, determined and unique. On a closer inspection, however, this is too quick: the tree-like structure could be just given by our *temporary* way of looking at things. Bell, relying on De Witt, said:

> "But, if I understand correctly, this tree-like structure is only meant to refer to a temporary and rough way of looking at things, during the period that the initially unfilled locations in a memory are progressively filed, labelling the different branches of the tree only by the macroscopic-type variables describing the contents of the locations. When a more fundamental description is adopted there is no reason to believe that the theory is more asymmetric in time than classical statistical mechanics. (…) At the microscopic level there is no such asymmetry in time as would be indicated by the existence of branching and non-existence of debranching. Thus, the structure of the wave function is not fundamentally tree-like." (Bell 1987: 135)

Another relevant concept in the Many-world interpretation is *decoherence*, which can be considered as fundamentally time asymmetric. For instance, Zeh says:

> "This splitting is facilitated by means of decoherence, defined as the dislocalization of superpositions (Sect. 4.3). Because of the locality of interactions, this process describes an effective dynamical decoupling (called a 'branching') of the global wave function into components that are characterized by different quasi-classical (robust and quasi-local) properties – including those of systems that can be regarded as observers. This decoherence has turned out to be the most efficient and most ubiquitous irreversible process in Nature" (Zeh 2001: 128)

However, the extent to which the branching structure and decoherence should be considered as fundamental is not so clear in the literature. For instance, it has been argued that decoherence is *perspectival*, and thereby, non-fundamental (see Bacciagalupi 2007). Within a more ambitious interpretative enterprise, David Wallace has claimed that “decoherence is an emergent process occurring within an already-stated microphysics: unitary quantum mechanics. It is not a mechanism to define a part of that microphysics” (Wallace 2010: 64)

There sure is more to say about Many-world interpretation and the arrow of time, as it seems that the interpretation has introduced some potentially asymmetric mechanisms that are relevant to the debate on a quantum arrow of time. However, it seems to me that all these sorts of time asymmetries are not fundamental, at least not in the sense I have been speaking so far, that is, they don’t point out to an asymmetry *of time in itself*. After all, as Wallace mentions, the branching structure is not fundamental, but emergent. What is fundamental is the unitary quantum mechanics. So, any time asymmetry that comes out from the theory should at first glance be considered as emergent and non-fundamental, consequently. And all this seems to relate to the problem of the two realms, rather than to the problem of a structural arrow of time. Of course, this deserves a careful discussion that, unfortunately, I cannot give here. I’m particularly interested in those interpretations that add some dynamical element to SQM and in whether some of these elements manifest a *structural* asymmetry of time.

Having said that, if one rather regards SQM’s dynamics as incomplete, any complete answer with respect to whether time is structurally directed or not should wait until all the dynamics be fully given. In the following three chapters, I shall solely focus on interpretations following such a path: the so-called Orthodox Interpretation of SQM, the so-called GRW-type theories, and Bohmian Mechanics. Also, these three interpretations have been in one way or another related to the problem of the arrow of time and to the notion of time reversal in the literature.

VII.

Orthodox Quantum Mechanics

Time Asymmetry and the Collapse of The Wave-Function

Let's start with one of the first attempts to overcome the measurement problem. Attempt that has extended widely in the field and firmly become a sort of orthodoxy in the physicist's community even up to the present.

In 1930, Paul Dirac presented one of the first extensions of SQM to bridge the gap between SQM's dynamics and what is observed after experiments: "the Collapse Postulate" (CP henceforth), also called "Projection Postulate". In Dirac's original formulation, CP is introduced as follows:

> "When we measure a real dynamical variable ξ, the disturbance involved in the act of measurement causes a jump in the state of the dynamical system. From physical continuity, if we make a second measurement of the same dynamical variable ξ immediately after the first, the result of the second measurement must be the same as that of the first. Thus, after the first measurement has been made, there is no indeterminacy in the result of the second. Hence, after the first measurement has been made, the system is in an eigenstate of the dynamical variable ξ, the eigenvalue it belongs to being equal to the result of the first measurement. This conclusion must still hold if the second measurement is not actually made. In this way we see that a measurement always causes the system to jump into an eigenstate of the dynamical variable that is being measured, the eigenvalue this eigenstate belongs to being equal to the result of the measurement". (Dirac 1935: 36)

A few years later, John von Neumann (1955[1932]) proposed a model for (ideal) measurements that came to get canonical in the field, giving a more refined version of CP. Von Neumann's main statement of collapse comes in the following way:

> “we have therefore *two fundamentally different types of interventions* which can occur in a system S (…) first, the arbitrary changes by measurements (…). Second, the automatic changes which occur with the passage of time”. (von Neumann 1955: 351. Emphasis added)

Both types of interventions (or processes, as he has also called them) are of a fundamentally different nature: whereas the former is statistical, the latter is causal. Analyzing these two types of processes, von Neumann claims:

> “Why then do we need the special process 1 for the measurement? The reason is this: In the measurement we cannot observe the system **S** by itself, but must instead investigate the system **S+M,** in order to obtain (numerically) its interaction with the measuring apparatus **M**. The theory of the measurement is a statement concerning **S+M**, and should describe how the state of **S** is related to certain properties of the state of **M**”. (Ibidem: 352)

Dirac’s and von Neumann’s introduction of CP gave birth to the quantum mechanics’ *orthodox* formalism: SQM *plus* CP is basically a formal extension of SQM by introducing a postulate that rules a radically different sort of evolution in specific cases, namely, when a measurement process takes place. This reveals that quantum mechanics is somehow a twofaced theory: SQM solely involves a unitary, deterministic, smooth and linear evolution given by the Schrödinger equation; CP rather involves a non-unitary, purely probabilistic, sharp and non-liner evolution given by special physical processes that collapse the quantum state into one of its eigenstates (Paul Dirac’s “jumps”). In this way, SQM gets its first minimal extension capable of getting over the measurement problem by *adding* an entirely different sort of evolution that quantum systems may undergo *under certain circumstances.* So, a full-blooded non-relativistic quantum dynamics is not given by SQM alone, but by SQM+CP. Let’s put all this more formally.

Recall the electron entering a *z*-device, as exposed in the introduction. Once the electron is correlated with the *z*-device, one ended up with a chain of superpositions like (6.3). What Dirac’s and von Neumann’s postulate tells us is that during a measurement the quantum state in such a superposition of *z*-spin abruptly and suddenly collapses into either $|\uparrow\rangle_z|"z\ up"\rangle^D$ or $|\downarrow\rangle_z|"z\ down"\rangle^D$. That is, the smooth and unitary evolutions of the wave-function in such a superposition of *z*-spin state is violently interrupted by the measurement process, and the quantum state undergoes a radically different sort of evolution –a non-unitary, non-deterministic and non-linear one. One hence gets an explanation of why one observes what one actually observes after measurements, despite SQM’s dynamics. And what one observes via CP also accomplishes the quantum probabilities prescribed by SQM.

There are a few things to note here. First, CP is fundamentally a solution to the measurement problem by introducing an extra axiom in the theory, so that it links SQM with what is observed in measurements. In other words, CP aims to explain why one observes states with definite values when SQM's dynamics predicts that, in certain cases, one should rather end up with a superposition. If one dispenses with CP, this gives rise again to the measurement problem, and should thereby put forward an alternative way-out.

Second, these "jumps" or "state-vector collapses", though crucial for the interpretation, have been understood in different ways. As it was described above, when a quantum system interacts with a measurement device a chain of superpositions is produced. But, as it was noted by John von Neumann, the chain of superposition doesn't end at the level of measurement devices, but it would extend to observer's counciousnes as well (known as "von Neumann chain"). So, one of the bones of contention is where and under which circumstances collapses occur. Physicists like John von Neumann and Eugene Wigner thought that collapses were physically real processes that occurred under the presence of an observer's counciousnes (von Neumann 1955, London and Bauer 1939. See Henderson 2010 for analysis and Becker 2004 for discussion on whether von Neumann supports the idea that collapses were physically real processes) and, as in the case of Wigner, at the level of counciousnes (Wigner 1967: 171). On the other hand, Heisenberg believed that humans were not necessary involved in the measurement process. It rather involves, essentially, an interaction between the measured system and the rest of the world (Heisenberg 1958: 54-55). However, there is indeed an epistemic component in any collapse process. According to Heisenberg, the wave function is the sum of the total information one has about a system. Consequently, any measurement-induced collaps relates to a sudden change in the information that an observer has about the system, which should be read in epistemic terms (Heisenberg 1958: 55).

Third, the main CP's motivation boils simply down to the special role that measurements play in the quantum theory. After all, quantum-mechanical results and predictions are about quantum systems *and* measurement devices (as von Neumann stresses). So, even though the Schrödinger equation correctly describes the evolution of any quantum state in isolation, this is unimportant (von Neumann 1955: 357) from the quantum mechanics' complete viewpoint: the theory, as long as it involves *measurements,* must also provide some explanations of why if a quantum state was evolving in a superposition of states, it *collapses* into one of its eigenstates when measured.

Finally, SQM+CP is typically referred to as *Orthodox Quantum Mechanics* (OrQM hereafter) in the literature, one of the most widely-extended, and fiercely objected, interpretations of SQM. Technically, SQM+CP is just a formal extension of SQM by introducing an additional axiom in the formalism as pointed out by Dirac and von Neumann, while OrQM seems rather to involve further interpretational stuff and has CP at its core.

OrQM has also been typically bound to the so-called *Copenhagen Interpretation*. This alleged relation deserves some few words. To begin with, the relation is not so straightforward: as the Copenhagen Interpretation has been notably multifaceted, it is not an easy task to clearly single its main tenets out. An explanation for this is that the interpretation has involved many philosophers and physicists endorsing unalike ideas, which has yielded an amalgam of unclear and loosely related tenets, instead of a systematic, unified corpus. Furthermore, the interpretation has been frequently associated with Niels Bohr's own ideas (its first and main defender), though it is not at all clear either what Bohr's ideas exactly were. For instance, Bohr never talked about "the collapse of the wave function" (if the wave-function is not real for him, how could it collapse?), whereas Heisenberg did. Nonetheless, OrQM's and Copenhagen's seem to have in general followed Heisenberg's ideas on this, rather Bohr's.

In general, OrQM as well as Copenhagians swing from a hiper-operationalist and bare version of Bohr's view (frequently taught in classrooms), where collapse processes are not physically real events, to a more robust physicalist view of the theory where the collapse of the wave-function is a real process and measurements play a crucial physical role in our knowledge of the world. Yet, despite all these divergencies and unclarities, there is at least some central tenets endorsed by virtually any OrQM's supporters (including Copenhagen's). This has been exhaustively discussed in the literature and a careful analysis would take us far beyond the scopes of this chapter (see Cushing 1994, Brock 2003, Aaserud and Heilbron 2013, and Faye 1991, 2014 for details). But it is worth mentioning, and emphasizing, two ideas that have pervaded OrQM and have a "Copenhagen air":

(a) the inner nature of the quantum realm remains veiled to what one is able *to know* and *communicate*: all quantum theory is, and can only be, about measurement outcomes.

Bohr in fact points out that "the procedure of measurement has an essential influence on the conditions on which the very definition of physical quantities in question rests (…), these conditions must be considered as an inherent element of any phenomenon to which the term 'physical reality' can be unambiguously applied" (Bohr 1935: 1025). A self-proclaimed

Bohrian-Copenhagen defender like Anton Zeilinger showily says: “So, what is the of the quantum? (…) I suggest that (…) the distinction between reality and our knowledge of reality, between reality and information, cannot be made” (Zeilinger 2005: 743)

(b) Any meaningful sense of reality (and thus of ontology) can only concern what is knowable by (classical) experiments and communicated by our (classical) language.

This probably just expresses a pragmatic Kantian stance sponsored by Bohr (Brock 2003, Faye 2014), though it has been often interpreted as a manifested *instrumentalist* (or *anti-realist*) posture with respect to quantum theory (mainly to the wave-function). The point has also paved the way to an informational-subjective reading of the quantum theory (being Quantum Bayesianism or Qbism the most serious proponent of this view, Caves, Fuchs and Schack 2002; Fuchs, Mermin and Schack 2014), to an informational reading (Bub 2016, 2018), or to a hiper-operationalist reading of physics portrayed by the well-known lemma: “Shut up and calculate!”

To come back to my business after this brief detour, the question of the arrow of time in the (non-relativistic) quantum realm cannot any more just involve SQM, but it must also include CP in the analysis. And all this is relevant for my purpose here because it has been suggested that CP, in any of its forms, displays a manifest time-asymmetry whereby quantum mechanics would sharply distinguish the past-to-future direction from the future-to-past one. And this would happen *regardless* whether SQM is time-reversal invariant or not. This proposal has come up in various places (it appears, for instance, in Aharanov’s critic 1964 paper as an already widely-extended belief. See also, for instance, Healey 2002, Popper 1982, Penrose 1989, Price 1996, Arntzenius 1997, Lucas 1999, Atkinson 2006, Ellis 2013, Callender 2018: 94 and references therein). For instance, Frank Arntzenius (1997) has defended that collapse theories introduce an arrow of time since

> “such theories say that are invariant forward transition chances for each of the possible initial quantum state to the possible collapsed states after the interaction (…). One cannot add some set of invariant backward transition chances to such theories, while retaining an empirically adequate theory, since the backward transition frequencies in the phenomena are highly non-invariant when one varies the frequencies with which the photons are emitted from the possible sources (1997: S218)

For David Atkinson, the time asymmetry in OrQM (in a more Copenhagen guise) comes from ‘observation’ (Atkinson 2006: 540). In any case, a collapse-induced time asymmetry has been probably popularized within the philosophy of physics field by Roger Penrose (1989) and a quite simple thought experiment. And this is a good starting point to see how time-asymmetric ingredients can be introduced in the OrQM's formalism.

Section 2. OrQM as a time-asymmetric theory: Penrose's thought experiment

Penrose starts by recognizing that OrQM is in fact a time-symmetric theory *but only* regarding the part involving the Schrödinger equation –its *unitary part* (1989: 354); as for the other, its *non-unitary part* given by CP, the theory turns out to be time *a-symmetric.* To show this, Penrose brings up the following thought experiment.

Suppose a lamp *L* at one extreme of the experimental arrangement, and a photo-detector *P* at the opposite extreme. Between them, a half-silvered mirror *M* is placed, which is tilted at a 45° angle to the line from *L* to *P*. Suppose now that *L* randomly emits a photon, which is aimed at *P* to be detected. At *L,* some device registers with high reliability the number of photons emitted given a time interval (see Fig. 7.1).

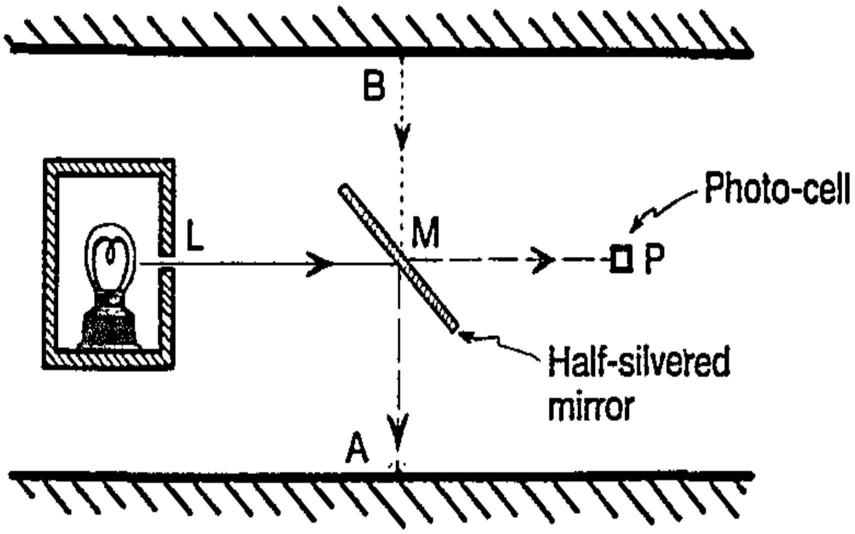

Figure 7.1. Copied from Penrose 1989, p. 357

When a photon is emitted by *L,* the half-silvered mirror *M* can either reflects it or letting it to pass through. Thinking of the experiment in quantum-mechanical terms, when a single photon is emitted, the photon's wave-function "impacts" the half-silvered mirror and splits in two parts: one part is reflected with an amplitude of $\sqrt{1/2}$ and the other passes through with an amplitude of $\sqrt{1/2}$. Until an observation is eventually made, both parts of the photon's wave-function –Penrose stresses– must be considered as "co-existing" in the *forward*-time direction.

Thanks to the statistical postulate of SQM, one knows that the probability that the photon reaches the photo-cell *P* is given by the square of the moduli of the amplitude, $|\sqrt{1/2}|^2 = 1/2$. After some calculations, one can easily answer the following question: "Given that *L* registers, what is the probability that *P* registers?" SQM+CP implies that the probability is exactly 'one-half'. And, after running the experiment many times, one will get *that*: approximately a 50% of the electrons passed through *M* and reached *P*, and the other 50% of them were reflected, reaching the laboratory wall at a point *A*. One can also infer straightforwardly that *if P* didn't register, then the photon hit the mirror and bounced off toward the laboratory wall (point *A*).

One has however assumed that time was running forward: the photon *was* firstly emitted by *L*, *after* a while reached the half-silvered mirror *M* and *then* split in two parts. At the (*future)* end of the experiment, the photon either reached the laboratory wall or was registered by the photo-cell. SQM+CP has provided the formal tools to work out the probabilities of getting a photon registered at *P* and the physical theory to explain what one eventually obtained, assuming that time *runs forward*. In order to know if SQM+CP is time symmetric, one would have to consider if it renders the same results (that is, if it renders the same probability predictions and experimental results) when time *runs backward*. To check this out, Penrose claims one should rather begin with the following (time-reversed) question: "Given that *P* registers, what is the probability that *L* registers?"

Penrose says: "we note that the *correct* experiment answer to this question is not 'one-half' at all, but 'one'" (1989: 358), for if the photo-cell *P* indeed registers, then it is virtually certain that the photon was emitted (and thereby registered) by *L*. So, whenever *P* registers it logically follows that *L* also registered the 100% of times. This is not however what a time-reversed version of SQM+CP retrodicts. It rather retrodicts that if one traces backward in time the photon's wave-function that reached *P*, then it will have one-half of probability of reaching *L*, and one-half of being reflected and of hitting the laboratory wall at a point *B* (the opposite to *A*). In the light of this, Penrose categorically claims that "in the case of our time-reversed question, the quantum-mechanical calculation has given us *completely the wrong answer*" (1989: 358. Italics in the original).

The upshot is that SQM+CP does *not* render the *same* predictions/retrodictions in both directions of time. Therefore, it treats the past-to-future direction and the future-to-past direction differently. Particularly, one can straightforwardly deduce that SQM+CP only works fine in one direction of time, but it does it fatally wrong when time is reversed. Penrose puts it as following. "If we wish to calculate the probability of a past state on the basis of a known future state, we get quite the wrong answers if we try to adopt the standard R [CP] procedure" (1989: 359). The thing is that any measuring device capable of effecting a collapse of the wave-function will be intrinsically time-asymmetric, altering radically any probability calculation when time is reversed.

George Ellis (2013) has put forward a more general argument to show the time-asymmetric nature of CP. Ellis' rationale is mainly based on the fact the quantum states may collapse (when measured) but they never "*un*collapse".

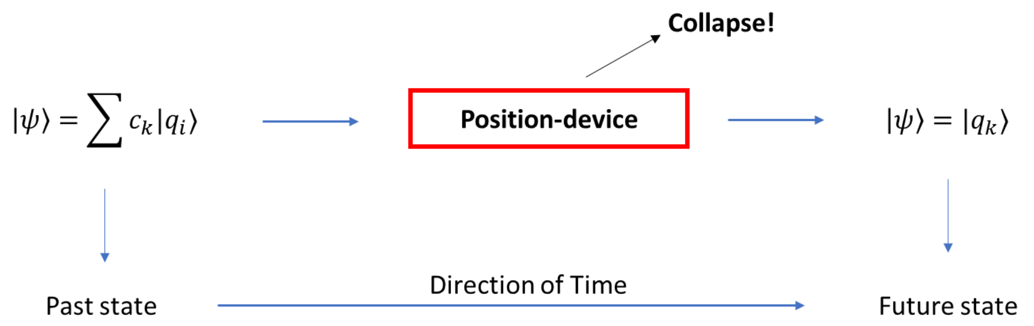

Fig. 7.2.

The process is intrinsically time-asymmetric to the extent to which the eigenstate $|q_k\rangle$ occurs *after* measuring (i.e. collapsing), that is, *after* the superposition. Furthermore, all coefficients in the superposition states one started with have been lost, so the knowledge of the final state says *nothing* about the initial state. To conclude, in Ellis' words, "the process [CP] is where the time irreversibility, and hence the arrow of time, is manifested at the quantum level" (2013: 243).

Huw Price (1996) makes the same case in affirming that any measurement process would, under certain assumptions I'll detail shortly, introduce an "objective asymmetry in the structure of reality" (1996: 207). Assuming OrQM, the state of a quantum system in the period between two measurements reflects the nature of the former instead of the latter. To put it simply, if an electron in a superposition of position states is spatially localized by means of a measurement device, then it will unitarily evolve according to the Schrödinger equation and its state will reflect the fact that the electron *was measured* and *localized* by a position-device. If one decides to a later time measure the electron's momentum with a momentum-device, electron's state won't reflect the nature of the second measurement (lying in its future) but that of former (lying in its past).

When the situation between two measurements is time-reversed, the result is oddly the contrary. What one would then see, according to Price, is the electron's state evolving toward a state associated with a measurement device it is to be involved with in the future (1996: 206). SQM+CP ordinarily takes for granted that a quantum system's state depends upon its past state and its past interactions. The fact that things look very weird when running in the backward direction of time would indicate that a deep time-asymmetry lies at the core of so-interpreted quantum mechanics.

To sum the point up. In introducing an additional dynamical law in the SQM formalism, OrQM would by the same maneuver introduce a time-asymmetric ingredient in the quantum theory. So, one could argue that OrQM defends a *structural quantum* arrow of time, provided that OrQM is the right way to supplement and to understand SQM. In a nutshell, the thesis to defend is that CP turns out to be a non-time-reversal invariant law of the theory to the extent that

(a) OrQM does not render the same probability predictions in both directions of time, and
(b) quantum systems always collapse when measured, but never "uncollapse", according to CP.

To put it in the vocabulary of the Part 2, the structure of OrQM's solutions is *asymmetric* under time reversal since the set of its possible worlds is given by either $\mathcal{W} = \mathrm{W}^f$ or $\mathcal{W} = \mathrm{W}^b$. In W^b one should include evolutions giving us the wrong probability predictions and those quantum systems "uncollapsing" when time-reversed. These are disregarded by the same collapse's mechanism and by the quantum-mechanical statistic one expects to get.

To see this more clearly, take Penrose's thought experiment again: whereas the quantum-mechanical algorithm predicts in, say, the future direction of time that probability of registering a photon in *P* is one-half (agreeing on what one observes), the time-reversed algorithm predicts that probability should also be one-half, when *logically* follows that it would rather be one. Consequently, one is entitled to get rid of one set of solutions (those predicting a result that is *a priori* wrong). And the upshot, and what would explain why OrQM is *structurally* non-time-reversal invariant, is that it turns out to be a predictive but not a *retrodictive* theory. In virtue of this, the theory does treat the past-to-future direction and the future-to-past direction differently: the theory yields the right predictions only in one direction of time, but it flatly fails in the other.

It is worth emphasizing that this result is independent of whether the Schrödinger equation is time-reversal invariant or not. Even though in the event that the Schrödinger equation turns out undisputedly time-reversal invariant (as Penrose indeed assumes), this just says something about the *unitary* part of *SQM*+CP. And the *non-unitary* part, SQM+*CP*, would nonetheless be non-time-reversal invariant. Therefore, the *whole* package *SQM+CP* will turn out non-time-reversal invariant as well.

Section 3. Three strategies to debunk OrQM's time asymmetry

There are at least three assumptions supporting this structural time asymmetry through CP:

(i) CP is *genuinely* a time asymmetric postulate. That is, CP turned out to be time *a*symmetric by means of a reliable way to test time (a)symmetry.
(ii) CP introduces a temporal asymmetry that should be taken as *structural* (fundamental).
(iii) OrQM (and principally the introduction of CP) is the right way to overcome the measurement problem and to get a full-blooded quantum *theory*.

In accordance with these assumptions, there are at least three strategies to debunk OrQM's argument for a structural arrow of time.

(a) To argument that CP is not genuinely (or interestingly) *time asymmetric.*
(b) To claim that, though temporally asymmetric, CP does not introduce a *structural* time asymmetry.
(c) To reject CP.

Through the first strategy, one could argue that the way in which time has been reversed is off the right track, consequently, CP is not genuinely time asymmetric. Through the second, one could rather take for granted that CP is temporally asymmetric, but so-produced time asymmetry is not structural in any relevant sense. Finally, one could alternatively argue that the introduction of CP is an unjustified ad-hoc resource, resulting in more obscurities than clarities. Hence, OrQM is not the right way to overcome the measurement problem, and thereby, to reach a quantum *theory* properly. Any time-asymmetry introduced by OrQM would by the same maneuver fall down.

Let's address them in order.

Section 3.1. OrQM is not genuinely (or interestingly) time asymmetric

Let's focus on Penrose's thought experiment. There are two senses in which Penrose's thought experiment results a bit confusing in showing that CP introduces a genuine and interesting time asymmetry.

First of all, Penrose uses extra-quantum mechanical information when judging whether the theory is time symmetric[39]. When replying the forward-in-time answer "Given that *L* registers, what is the probability that *P* registers?" Penrose appeals to the usual quantum-mechanical expectations. But, when considering the time-reversed answer "Given that *P* registers, what is the probability that *L* registers?" Penrose rather appeals to the non-quantum mechanical answer 'one' judged as "the *correct* experimental answer" (Penrose 1989: 358). And this seems not to be a fair movement.

To begin with, the experiment cannot be in reality carried out *in the backward* direction of time, so one has *to* instead *imagine* what one would expect of so-settled experiment *if* the direction of time were reversed. But, in imagining the time-reversed situation, Penrose leaks non-quantum mechanical information when judging what would be the right answer. Clearly, the (backward) quantum-mechanical answer is that one would obtain one-half of chances of getting the electron registered at *L*, which is equal to that obtained in the original direction of time (that is, through the *forward-in-time-oriented* quantum theory). But such a result would be quite shocking, and somehow anti-intuitive for Penrose, because one *already knows* that *L* always registers, so the *correct* prediction would seem to instead be 'one' rather than 'one-half'. But where does such a knowledge come from? One knows that *because* one has checked it when running the experiment in the original direction of time and *because* one is using extra-quantum mechanical information coming from what is expected in accordance with how the experiment was settled in the future-headed direction of time. Therefore, this asymmetry cannot be genuine or interesting since it is grounded on one's temporally-biased knowledge.

The second sense in which Penrose's argument is confusing relates to how time is being inverted in Penrose's thought experiment. The case is truly tricky because one posseses a sharply-refined receipt of how time reversal should be applied *in abstract* to dynamical equations of motion (as was explained in Chapter II and carefully analyzed in Chapter IV). But in a worldlier situation involving photo-cells or lamps emitting photons, one is left a bit clueless about what a time-reversed experimental setup would look like.

[39] Craig Callender (2000) has also mentioned an akin point, though he doesn't develop it.

Basically, what Penrose does in his thought experiment to reverse time is to imagine the same objects and the same situation but in the reverse order. Let's call Penrose's time-reversal transformation T^P. So, given the (relevant) sequence where the photon is emitted by the lamp *L*, hits the half-silvered mirror *M*, passes through it and reaches the photo-cell *P*

Future-headed sequence $L \rightarrow M \rightarrow P$

T^P produces the (allegedly) time-reversed sequence

Past-headed sequence $T^P(L \rightarrow M \rightarrow P) = P \rightarrow M \rightarrow L$

Furthermore, the question to be responded by the quantum-mechanical algorithm must be temporally reversed accordingly. By simply inverting the sequence as shown above, the terms in the question are interchanged: "Given that *L* registers, what is the probability that *P* registers?" turns into the time-reversed question "Given that *P* registers, what is the probability that *L* registers?"

It is clear in which sense OrQM is time-asymmetric given CP and T^P . What is not so clear is how interesting and legitimate such an asymmetry is, which to a good extent depends on how the operation of time reversal has been applied in the thought experiment. Penrose does not discuss any other alternative, but he seems to uncritically assume that an inversion of the direction of time would amount to simply invert the sequence of states, and then to work out the corresponding probabilities (assuming extra-quantum mechanical information as was shown before). This could be nonetheless a bit sloppy as it has been pointed out by Steven Savitt (1995) and mentioned by Craig Callender (2000). I shall develop this point further.

Savitt displays at least three very broad notions of time reversal in the literature: (i) *time-reversal$_1$*, which amounts to the mapping $T: t \rightarrow -t$; (ii) *time-reversal$_2$*, which not only maps $t \rightarrow -t$ but also temporally reverses the very *states* (and *objects*) of a sequence; and (iii) *time-reversal$_3$*, which captures the essential idea that time reversal is motion reversal, and in consequence, it must retrace the physical system's trajectory. What Savitt concludes is that Penrose's thought experiment comes out non-time-reversal invariant *depending upon* which sort of time-reversal transformation is taken. In particular, it results non-time-reversal$_1$ invariant, but time-reversal$_{2,3}$ invariant.

In this light, Callender asks: "why compare $P(S_i \rightarrow S_f)$ with $P\ (S_f \rightarrow S_i)$ and not with $P\ (S_f^T \rightarrow S_i^T)$?" (2000: 256). In other words, Callender finds suspicious that T^P be the *right* way to reverse time. And the reason is simply that T^P does not transform the states themselves, but leaves them as in the direction of time one started with. So, according to Callender, the genuine time-reversed sequence of Penrose's thought experiment is not given by T^P but by Callender's time-reversal transformation, T^C (which is overall equivalent to *time-reversal$_2$* transformation in Savitt's list).

Past-headed sequence $T^C(L \rightarrow M \rightarrow P) = P^T \rightarrow M^T \rightarrow L^T$

where X^T is a time-reversed state or object in the sequence. Callender gives some hints about how this should be interpreted. He says: "if Penrose is genuinely concerned with TRI [time-reversal invariance], he should treat the *emitter as a receiver* and *vice versa*" (2000: 256. Emphasis added). Therefore, the right time-reversed question to make to the quantum-mechanical algorithm is not "what is the probability that *L* registers, given that *P* registers" but "what is the probability that a time-reversed *L* registers, given that a time-reversed *P* registers". Penrose would thus be addressing his own thought experiment from the wrong angle. And when addressed rightly, time-symmetry (in particular, T^C-invariance) is restored.

Savitt gives further reasons to put Penrose's conclusion into question[40]. Not only does Penrose assumes a debatable view of time reversal (which, according to him, amounts to choosing the video-tape-played-backward version of the thought experiment), but he also takes a *narrow frequentist* way of determining transition probabilities, which only counts the results of one run instead of averaging over the set of all physically possible runs having the same initial (final) conditions. Savitt claims that if a time-reversal transformation like T^C along with a broader view of determining transition probabilities is taken, then the quantum-mechanical algorithm yields the same probabilities both in the forward direction of time and in the backward one (Savitt 1995: 17-18). Therefore, Savitt concludes, Penrose's argument only runs smoothly if a debatable view of time reversal plus an even more questionable way of determining transition probabilities are taken. In changing either of those assumptions, Penrose's argument rapidly breaks down.

Concluding, Penrose's thought experiment holds only *if*, on the one hand, one accepts to leak non-quantum mechanical information into the argument to figure out how the time-reversed probabilities should be worked out and, on the other, if one takes T^P to be a genuine

[40] I shall only mention here the reason related to how transition probabilities must be worked out. For further criticisms and details see Savitt (1995), pp. 15-18.

and reliable time-reversal transformation. Both assumptions are problematic: the first one may lead us to suspect that the source of temporal asymmetry doesn't really come from CP but from the introduction, *by us*, of such extra non-quantum mechanical information in the time-reversed calculations of probabilities. In this way, the temporal asymmetry would be neither genuine or interesting. As for the second assumption, it's true that Penrose introduces a time-reversal transformation (T^P) without further justification. In fact, if the thought experiment is time reversed differently (in particular, T^C-reversed), Penrose's conclusion does not run any longer and one gets no structural arrow of time in any relevant sense.

Section 3.2. Though temporally asymmetric, CP does not introduce a *structural* time asymmetry

Let's now tackle Penrose's argument from a different angle. Remember that the statistical postulate of the theory assumes that if a measurement of the observable A is carried out, it will produce, with a certain probability, one of its eigenvalues a_i as a result. So, if a system is in the state $|\psi\rangle = \sum a_i |a_i\rangle$, then the probability that the eigenvalue a_i of A is found when measured is equal to $P(A = a_i, |\psi\rangle) = |a_i|^2$. In its logical form, the algorithm says

FORWARD $P(measure\ at\ t_2\ value\ a_i\ of\ observable\ A, |\psi\rangle\ at\ t_1) = p$

Borrowing Albert's words, the crucial question here is whether this algorithm (giving us the conditional probabilities of some *later* state given an *earlier* state) works equally well as FORWARD when formulated backward

BACKWARD $P(measure\ at\ t_0\ value\ a_i\ of\ observable\ A, |\psi\rangle\ at\ t_1) = p$

However, as a matter of fact, SQM+CP doesn't give us the probability of earlier states given later states for, as Callender puts it, "the theory is predictive but not retrodictive" (2000: 258). Something like this was probably in Penrose's mind in drawing his thought experiment up. If that's the case, he was right all along in regarding OrQM as a time *a*symmetric theory in so far as the theory is not indeed retrodictive. Satosi Watanabe has also subscribed this conclusion in claiming that "it is precisely irretrodictability what is related to phenomenal one-wayness" (1965: 56)[41]. Hence, the measurement process would inevitably break the time symmetry in

[41] Even though it's true that in general time asymmetry (or non-time-reversal invariance) and "irretrodictability" may come to be thought as two quite different properties, and one could consequently argue that no temporal directionality should be followed from irretrodictability, it has been argued that in some cases, like non-relativistic

the theory as the whole theory (SQM+CP) yields results directly in terms of conditional probabilities for states in the future, but not in the past[42].

The point I want to make here is that though a time asymmetry could be introduced by this mean, it fails to be a *structural* asymmetry. In other words, CP would be at best useful for defining a *non-structural* arrow of time, but it falls short in grounding a structural arrow of time. Following this line, I will put forth two arguments: first, this temporally asymmetry through collapse is necessarily local and cannot be applied to the universe as a whole. Second, it necessarily relies upon extrinsic properties of the dynamics for the time asymmetry to be univocally defined.

Let's start with the first argument. As many times repeated, OrQM assigns a fundamental role to measurements, and thereby, to the *sort of* measurement performed. In fact, quantum theory (as characterized by von Neumann) is, actually, a theory about the measured system *plus* the measuring apparatus. It follows from this that any temporal asymmetry induced by CP makes sense *only if* an external measurement device can be suitably defined. To put it otherwise, *only if* the system is open. However, any so-defined temporal asymmetry will necessarily be local since it is not possible to define it for the universe as a whole. The reason is quite simple: it is not possible to define a measurement device out of the universe, which is, by definition, a closed system. Hence, anyone looking for a *global* and *unrestricted* temporal asymmetry will be deeply disappointed by the CP's temporal asymmetry.

Furthermore, it follows from this that the wave-function of the entire universe never collapses but always evolves unitarily according to the Schrödinger equation. And under the assumption that the Schrödinger equation is time-reversal invariant, it entails that the universe as a whole is blind to any structural direction of time according to this approach. And this scenario looks quite similar to that analyzed before where only a non-structural arrow of time could be defined: one is left with a unharmonized dual reality, where the wave-function of the universe is time symmetric but where one can nonetheless define local temporal asymmetries within measurement contexts.

quantum mechanics, the implication is right. Earman for instance claims that in any statistical theory non-time-reversal invariance directly follows from irretrodictability in the sense of BACKWARD (Earman 1974).

[42] One could come to think that by adding BACKWARD to the theory the problem vanishes. Richard Healey has showed that this cannot be done without trivializing the theory, if it is statistical. For further details and discussion about it, see Healey (1981: 103-108).

One can escape this argument by saying that, though the universe as a whole can never be measured (and thus, it does not collapse onto an eigenstate), this does not entail that the universal wave-function never undergoes any collapses. The idea would be that there are collapses of subsystems everywhere, and *a fortiori* the universal wave-function will change discontinuously. And these changes are specified as only occurring in one temporal direction

[41]. I have two worries with this argument. First, the idea of collapse (in the framework of OrQM), *only* makes sense when a measurement context is properly defined. One could ask: if the universal wave-function eventually collapses, in which basis does it collapse? This question remains unanswered in OrQM. Furthermore, if the universal wave-function changes, according to what type of mechanism does it change? Within von Neumann's framework, there are only two: if the universal wave-function does not collapse, then only changes unitarily according to the Schrödinger equation. Any other type of change would lie out of quantum mechanics. Second, the approach probably falls short in setting the question properly at a cosmological level: when the discussion comes to consider the universal wave-function, the approach does not fit in. But this just solves the dispute by fiat: it is not the case that the universal wave-function never collapses and then is temporally directionless, but it does not even make sense to talk about it within this framework. In any case, the approach is intrinsially local.

The second argument concerns the collapse mechanism itself conforming with CP. The point I want to make is that the transition $uncollapsed\ state \rightarrow\ collapsed\ state$ remains vague and undefined *as long as* the measuring device is not fully specified. Therefore, the CP-induced temporal asymmetry strongly relies upon already having a fully-characterized measurement context. It follows from this that so-introduced arrow of time cannot be *structural* as it doesn't depend exclusively on intrinsic properties of CP's dynamics.

As reviewed before, the collapse is produced when a superposition state, say, $|\psi(t_0)\rangle = \sqrt{\frac{1}{2}}(c_1|a_1\rangle + c_2|a_2\rangle)$, collapses into one of A's eigenstates when measured by an A-device, $|\psi(t_1)\rangle = |a_2\rangle$. This non-unitary transition would define a direction of time in so far as the transition always goes from "uncollapsed/undefined states" to "collapsed/defined states", and never the other way around. However, as it stands, this wouldn't be quite right since the theory *also* describes the transition from "collapsed/defined states" to "uncollapsed/defined states", for instance, when a quantum system is collapsed in *a different basis*. To put it in other words,

[41] Many thanks to Carl Hoefer for raising this remark.

the above-mentioned transition is true *only if* the quantum state is collapsed in the *A* basis, and one knows that only if the measuring device is completely specified *beforehand*. Let's be more precise on this.

Suppose that $|a_1\rangle,\ |a_2\rangle$ are the eigenstates of the observable *A*. Suppose also one designs an *A*-device (a device built to measure the observable *A*). One knows that in measuring the observable *A* through an *A*-device, the quantum state will collapse onto either of its eigenstates. But, by means of SQM, one knows that it is possible to write down the eigenstates of *A* in a different basis, say, *B* with which *A* doesn't commute. So, one can rewrite each eigenstate $|a_1\rangle$ as a superposition of *B*'s eigenstates, $|a_1\rangle = |b_1\rangle + |b_2\rangle$, and the eigenstate $|b_1\rangle$ as $|b_1\rangle = |a_1\rangle + |a_2\rangle$. And the same goes, mutatis mutandis, for $|a_2\rangle$ and $|b_2\rangle$.

It is relatively easy to show that, *without any specification* about what the sort of experiment is to be run, the $uncollapsed/undefined \rightarrow collapsed/defined$ transition

$$|b_1\rangle = |a_1\rangle + |a_2\rangle \rightarrow |a_2 \tag{7.1 }$$

Can be rightfully viewed as a $collapsed/defined \rightarrow uncollapsed/undefined$ transition

$$|b_1\rangle \rightarrow |b_1\rangle + |b_2\rangle \tag{7.2 }$$

where $|a_2\rangle = |b_1\rangle + |b_2\rangle$. The system will thus "uncollapse", so to speak, when measured. This is the case, for instance, when one *prepares* the state before running an experiment in a given basis.

In this light, the transition $uncollapsed/undefined \rightarrow collapsed/defined$ can be fixed univocally only relatively to the sort of experiment is to be run. It follows that the transition is not *intrinsically* asymmetric (that is, it doesn't depend exclusively on the dynamics) but depends on the sort of experiment, which defines the basis in which the state is to be written down. Once this is provided, it's true that the system won't ever uncollapse in *that* measurement context and *that* basis. But, as long as the eigenstates of the system are expressed in a different basis, and the experiment's configuration is modified accordingly, the system may undergo a $collapsed/undefined \rightarrow uncollapsed/undefined$ transition as showed before. Hence, the very collapse mechanism remains undefined without specifying the measurement context adequately. And so does any direction of time depending on it.

This undermines any attempt to ground a *structural*, or even an objective, arrow of time in CP because the collapse is not intrinsic to the system, but it depends upon the chosen measurement device. The approach does give the means to define a *local non-structural* arrow of time relative to a given basis and a measurement device. Let me make the case in epistemic terms. Suppose that the available information is that a quantum state in a superposition of $|a_1\rangle + |a_2\rangle$ was measured by an unspecified measurement device (it could have been either a *A*-device or a *B*-device). Someone told us that the outcome was 'a_2' meaning that the quantum state collapsed onto eigenstate $|a_2\rangle$. The CP-based temporal asymmetry would say that $|a_1\rangle + |a_2\rangle$ came earlier because is an uncollapsed/undefined state, whereas $|a_2\rangle$ came later, after the measurement, because it is a collapsed/defined state.

Nevertheless, one doesn't know whether the measurement device was an *A*-device or a *B*-device. If it was an *A*-device, then it is possible to claim that the state collapsed onto the eigenstate $|a_2\rangle$. And this in fact displays the right sort of temporal asymmetry one was after. But if one doesn't know it, I could lawfully rewrite the states in a *B*-basis, and then obtain an "uncollapsing" scenario. Naturally, if one knows one is dealing with a *B*-device, it would be highly confusing to write the state down in a different basis. But, if one doesn't know it, one could express the state in either of the bases. If one gets to an "uncollapsing" scenario, should one accept that the direction of time was then inverted, that the collapse/defined state was rather a future state and the uncollapse/undefined a past one? This hardly makes sense and seems more reasonable to specify in which basis the quantum state is to be measured. But, in accepting this, any attempt to ground a *structural* arrow of time by means of CP rapidly vanishes, since it would be relative to the (arbitrary) choice of a measurement device. And it is not clear at all how this choice could come to be related to an asymmetry of time in itself.

Section 3.3. Throw CP away!

There is yet a much bigger issue against OrQM and its alleged CP-based quantum arrow of time: that CP, as it stands at least, is an unsatisfactory attempt to overcome the measurement problem. After all, whether CP introduces a quantum arrow of time (in any of its versions) essentially depends on regarding CP as a *fundamental* physical *law* of quantum mechanics. And many authors have in fact rejected this. Literature has analyzed its issues for long, so I won't get into many details here. But, in general, criticisms against CP rely upon two points:

(a) It draws an unnatural distinction between physical systems and measuring devices. Or, to put it differently, it gives an unjustified special role to measurement processes.
(b) CP entails some consequences that render an unsatisfactory picture of what the world is like according to quantum theory.

As to the first point, if CP entails that the collapse of the wave-function is a physically real process, then the following question arises: why are measurement so *physically special* that are capable of inducing abrupt jumps? What about other interventions one may find in nature? As John Bell (1990) remarks, the word ‘measurement’ is not an appropriate concept to appear in any formulation of a (intended) fundamental physical theory. From a broader viewpoint, if a measurement is a physical process *like any other* and quantum theory intends to be a fundamental theory, then it should tell us what a measure is and should describe it in quantum mechanical terms. However, this is not the case: the measurement process appears in the axioms of the SQM+CP, so SQM+CP cannot be used to explain measurements, but measurements rather explain what a quantum theory is! (Dickson 2007: 363).

There are only two way-outs for any defender of CP: either she considers that quantum mechanics is not a fundamental theory in any relevant sense, or that the process of measuring is not like any other physical process. Neither do options look too appealing. Remarkably, the latter implies that measurements are a sort of spooky physical processes, laying beyond any quantum explanation. Furthermore, in relation to the arrow of time, the two way-outs are also puzzling. On the one hand, to the extent that the temporal asymmetry is introduced by CP, if it remains as an odd physical process, so does the temporal asymmetry. On the other hand, if quantum mechanics turns out to be a non-fundamental theory, so is the temporal asymmetry. In any case, the scenario doesn’t look so promissory for any intended structural or fundamental temporal asymmetry. An alternative option would be to discard CP and OrQM altogether. In doing so, the CP-based temporal asymmetry is immediately discarded as well.

As to the second line of criticisms, SQM+CP posits further problems in itself. Not only is its ad hoc nature suspicious, but also the image of what the world is like according to the theory. For instance, David Albert claims that so interpreted, the theory issues a bewildering picture of the quantum dynamics. He says

> “The dynamics and the postulate of collapse are flatly in contradiction with one another (…); and the postulate of collapse seems to be right about what happens when we make measurements, and the dynamics seems to be bizarrely *wrong* about what happens when we make measurements; and yet the dynamics seems to be right about what happens whenever we aren’t making measurements, and so the whole thing is very confusing” (Albert 1992: 79)

Beyond this oddly dual scenario, if the collapse of the wave-function is to be taken as a real physical process, the circumstances under which it occurs should be explained further.

However, as CP stands, it remains silent about the *causes* that produce the collapse: it just says that the quantum state undergoes a collapse when measured, but no mechanism is given (this has been pointed out, for instance, by Lombardi, Fortin, Castagnino and Ardenghi 2011). Further, when applied to entangled compound systems, CP entails that the collapse resulting from measuring upon one of the systems will instantaneously produce sharply-defined values for the rest of the system's components, regardless how far away they are from one another. CP would hence imply some sort of action-at-distance or non-locality.

On empirical grounds, Dennis Dieks stresses that experimental research in the last decades has deadly undermined any motivation of introducing CP. He points out that "Schrödinger cat states, i.e., superpositions of distinguishable quantum states of mesoscopic or even practically macroscopic physical systems are now routinely prepared in the laboratory, and interference between the different terms in the superpositions have abundantly been verified" (Dieks 2019). According to him, this diminishes inductive support to CP to the extent to which superpositions never really collapse. They are just hard to detect in macroscopic systems.

Section 4. Limits and scopes of OrQM's time asymmetry and of the arguments against it

In this last section I'd like to make two remarks about the arguments *against* the CP-induced temporal asymmetry. In particular, I'll first take Savitt's and Callender's arguments against Penrose about how his thought experiment should be properly time reversed. Second, I'll make some brief commentaries about OrQM (particularly on its Copenhagen face), its ontology, and the arguments in favor of a CP-induced arrow of time

As to the first commentary, the point I want to make is that Savitt's and Callender's time reversal proposal not only doesn't do the job better either, but it also begs the question if used to show that temporal symmetry holds. Let's suppose that T^C is the right way to temporally reverse Penrose's thought experiment. As Callender suggests, it implies that one should treat the emitter as receiver and vice versa. To be clear, T^C entails that, for instance,

$$T^C(emitter) = receiver \tag{7.3 }$$

meaning that a time-reversed photo-cell should be treated as an emitter. Thus, the time-reversed sequence

$$T^C(L \rightarrow M \rightarrow P) = P^T \rightarrow M^T \rightarrow L^T \tag{7.4}$$

should be correctly read it as saying that a time-reversed photo-cell emits a photon at t_2, shortly after the time-reversed photon hits the time-reversed mirror, always with t decreasing. By the quantum-mechanical algorithm one well knows that it has one-half chances of passing through the half-silvered mirrored and of reaching the time-reversed lamp, and one-half of being headed to the laboratory wall. Without leaking temporally-biased and extra-quantum mechanical information, one should note that the quantum-mechanical algorithm retrodicts that "one-half of photons reach the time-reversed lamp", so the probability that the time-reversed lamp registers given that the time-reversed photo-cell registers is *exactly the same* to the probability that the photo-cell registers given that the lamp registers. Time symmetry is thereby restored.

I think this argument is flawed in two different senses:

(a) it does not offer a more reliable way to time reverse Penrose's thought experiment. Instead of that, it leads to *destroy* the very objects involved in the thought experiment.
(b) It begs the question if used to claim that time symmetry holds.

These issues follow from two possible readings of Callender's and Savitt's proposal of time reversal, particularly, of how the states in the series must be temporally reversed.

With respect to the first sense, one might ask: how would a time-reversed photo-cell work? One knows how photo-cells works in the ordinary direction of time (conventionally, running forward), but one has no clue about how a time-reversed photo-cell would work if time run backward. Treating the photo-cell as an *emitter* doesn't help so much for a photo-cell is not the sort of things that emits anything. Why is one entitled to suppose that a photo-cell will behave in a completely different way, capable of emitting photons and behaving as an emitter when temporally reversed?

It could certainly be argued that non-time-reversed photo-cells don't emit anything, but time-reversed photo-cells do it. In reply to this, one can say that in fact a time-reversed *something* is gotten, but that something won't be a photo-cell any longer. And this is so because it simply doesn't work like a photo-cell. It seems, hence, that by T^C reversing the direction of time one is, at the same time, *destroying* the very objects and states involved in the sequence.

Photo-cells, if one wish to use the word meaningfully, cannot be the sort of things that emits anything in *either* direction of time as long as one is still dealing with photo-cells in some relevant sense.

Let me put the point slightly differently. Photo-cells may come in various types and instantiate unalike properties, so one can imagine different modal scenarios for photo-cells, altering their properties and their conditions. Notwithstanding this, if one wishes to still refer to photo-cells properly, a certain sub-group of properties must remain fixed and some other properties must be necessarily excluded. Otherwise, one would be unable to identify the objects through different scenarios. My claim is that the property of "functioning like a photo-cell" (that is, the property of behaving as a receiver) must be necessarily fixed and the property of behaving like an emitter must be necessarily excluded. All this in order to identify photo-cells thought different scenarios (or in order to satisfy *trans-worlds identification*, if you like) and to keep talking about photo-cells meaningfully.

I don't see any reason why this should be different in a time-reversed scenario. A time-reversed photo-cell in a past-headed experimental running should emit nothing in as much as a photo-cell emits nothing in a future-headed experimental running. Otherwise, one is rather dealing with a different sort of objects (no a photo-cell) when time is reversed. If object's inner nature can freely vary when time reversed, then time-reversal invariance will practically always follow trivially.

And this is how the second reading of the argument comes into play: one should make explicit how a photo-cell transforms into a different sort of object under time reversal. I think that the only possible choice is to suppose that

$$T^C(P) = L \tag{7.5 }$$

that is, that a time-reversed photo-cell should be treated *as though* if it were a lamp. Analogously,

$$T^C(L) = P \tag{7.6 }$$

Now it seems one is getting somewhere because a lamp is in fact the sort of thing that behaves like an emitter. This could be reworded as following: when time is T^C-reversed, the time-reversed photo-cell is capable of emitting because it *is*, in the opposite direction of time, a lamp.

This reading, however, risks a question-beginning if what is at issue is to test whether time symmetry holds. And this is so because the time-reversal transformation is carefully designed to leave the experimental setup virtually unaltered. So, both situations are bound to be time symmetric because the transformation does nothing if a time-reversed photo-cell behaves like a lamp, and a time-reversed lamp behaves like a photo-cell: one is simply describing the same physical situation in other terms and using other names.

Or, even worse: One is simply marking the states with a “^{T}”. And it is blatant that a graphical mark won’t produce any physical change! Further, the direction of time is assumed to, so to speak, *innocuous* in the description of any physical situation: it just plays the role of adding a mark on the graphical representation of states (indicating that one should read it as specified above) and changes t by $-t$. Nothing else: a transformation of an innocuous direction of time will surely produce a time symmetric scenario. I’m not claiming, at all, that one should thus go with Penrose’s application of time reversal instead. It probably fails in many other respects. However, the alternative proposed by Savitt and Callender doesn’t do it better. Therefore, it cannot be used against Penrose’s.

My second commentary concerns specifically what quantum theory is about according to OrQM. As mentioned repeatedly in previous chapters, for a temporal asymmetry to be fully characterized, it must be considered whether the law at stake is fundamental in some relevant sense, what the fundamental entities of the theory are, and so forth. In this particular case, one can ask: what does it mean ontologically and epistemically that OrQM is temporally asymmetric? Or, what features of OrQM are serious candidates for a full-blooded arrow of time? One could, for instance, show that a physical theory involves certain elements that introduce a temporal asymmetry (for instance, dissipative forces in classical mechanics). Nonetheless, one could also read the theory in such a way that those elements aren’t fully real, or at least, are non-fundamental (for instance, as Callender 1995 argues against Hutchison 1993, 1995 with respect to introducing dissipative forces in classical mechanics’ ontology). In doing so, any temporal asymmetry is deflated. Naturally, one can next dispute about the ontological status of those elements within the theory at stake, but this is a different line of contend.

It was mentioned before that OrQM includes some interpretative stuff linked to what has come to be known as Copenhagen interpretation. One of the essential questions for any interpretation of SQM is: what is the role of the wave-function in the theory? And one of the tenets of OrQM in relation to this question is: any meaningful wave-function talking must be necessarily expressed in terms of measurement's outcomes. In this sense, OrQM tells us that quantum mechanics is not a theory about an *objective* real world composed of particles, wave-functions or whatever, but a theory about what you see in, and know from, the experiments you make. So, any relevant temporal asymmetry for OrQM should be only read off from the measurement's outcomes.

Defenders of OrQM frequently presents themselves as exclusively guided by "facts", by what is known through *direct* experience. Hence, any claim beyond these limits is as obscure as any uncritical metaphysical claim. Along this line, it has been adjudicated to Niels Bohr the phrase "there is no quantum world (…). It is wrong to think that the task of physics is to find out how nature is. Physics concerns what we can *say* about nature" (Petersen 1963: 12), meaning that any question beyond what one can (classically) say just exposes our ideological prejudices or a metaphysical digression. The overall reasoning seems to be driven for a *logical-positivist* philosophical framework, according to which the idea that meaningful assertions must be verifiable by direct observation is a central tenet. And, again, within the quantum theory, all this is only given by measurements' outcomes.

If OrQM is a set of ideas about how the quantum theory should be understood (regardless how controversial its coherence and cohesion are) and a proposal about what the world is like (regardless how minimal its claims are with respect to this), it is quite clear that all there is knowable out there, in one way or the other, boils down to measurements' outcomes. The *fundamental* stuff OrQM talks about is entirely exhausted by measurements' outcomes, the sort of relations and correlations one can discover among them, and the sort of things that bear some referenciality in daily communication employing natural language. Any *quantum* temporal asymmetry can by the same token be only given by an asymmetry displayed in what the quantum algorithm predicts and in what one obtains after measurements. There is nothing else where a temporal asymmetry may come out.

And this takes one back to gravitate around the sort of thought experiments as those analyzed before. Any candidate for a *structural* arrow of time within OrQM framework has to be able to show that the quantum algorithm systematically fails in rendering the same statistics and experiments' outcomes in both directions of time. And this should be considered independently of any alleged behavior of the wave-function: as long as it lies far beyond of what it is directly observed by measurement, any speculation about its properties or nature is just that, non-scientific speculation stepping beyond what knowable. So will be any temporal (a)symmetry coming from there as well. Pieces of mathematics, while remaining just like that, are atemporal items appearing in a formal apparatus, not the sort of things that can manifest a temporal bias. Only those parts of the formalism linked to (knowable) stuff out there in the world are physically meaningful. For OrQM the only knowable stuff in the world is what one can directly observe: once again, measurements' outcomes. But, as showed along this chapter, the kind of temporal asymmetry one may come to find in this context is, at best, a *local non-structural* one.

Final remarks

Along this chapter, I've presented a proposal for a quantum arrow of time based on OrQM: a CP-induced temporal asymmetry. As reviewed, the asymmetry of time would enter into quantum theory through the collapse of the state vector: any measurement performed upon a superposition state forces it to collapse onto one of the components of the superposition. This process is not described by the Schrödinger equation but by a non-unitary and non-deterministic evolution given by CP. In a motto: quantum states collapse but do not uncollapse.

I've also put forward three lines of criticisms to this approach:

(a) CP should not be considered genuinely (or interestingly) *time asymmetric*
(b) Though temporally asymmetric, CP does not introduce a *structural* time asymmetry
(c) OrQM and CP should be rejected

The first line of argumentation focused on that OrQM turns out to be non-time symmetric only under a debatable time-reversal transformation. Time symmetry is easily restored when the time-reversal transformation is taken to also transformed the states of the series. As I argued in Section 4, I've however found this proposal flatly unsatisfactory: either it destroys the objects involved in the series or begs the question when used to show that time symmetry holds. In conclusion, I do think that OrQM offers some means to distinguish the past-to-future direction from the future-to-past one. However, one can thus wonder: in which sense?

The second line of argumentation contends that, if a temporal asymmetry is provided by OrQM, this asymmetry is non-structural. The strong dependence of the time asymmetry on measurement devices' features to determine whether the transition was from an "uncollapsed/undefined" state to a "collapsed/defined" one undermines any attempt to consider any alleged time asymmetry as structural. Furthermore, the asymmetry would be local to the extent to which the wave-function of the whole universe never collapses. All this portrays a scenario quite similar to that of the problem of the two realms, in which any temporal asymmetry is non-structural. For these reasons, in the event OrQM to be taken as an acceptable interpretation, the CP-induced asymmetry will be local and non-structural.

Finally, the last line of argumentation debunks any CP-induced temporal asymmetry by rejecting CP itself. Though CP, and OrQM more generally, achieved a kind of exclusive hegemony in the physics community from the 1930s up to the present, from a more philosophical and conceptual sight the interpretation has been fiercely fought. Besides its inner obscurities and contradictions, any viable formulation of OrQM must in one way or the other include CP. And CP was shown to be not only flatly ad hoc, but also to render an odd picture of the world, if any. No need to show that *if* OrQM turns out to be inadequate, so will be anything following from it.

To sum up. Despite what has been suggested by Penrose, Ellis, and Price, among others, OrQM falls quite short in clearly defining a structural temporal direction. OrQM would in fact provide the means to draw the distinction between both directions of time *locally* and *relatively* to a previously-specified measurement device. This would yield a non-structural arrow of time at best. But, at worse, the questionable principles of OrQM to flesh quantum theory out would steer to not considering the interpretation seriously, disregarding by the same token any temporal asymmetry coming from it.

VIII.

GRW-type theories

Time Asymmetry and Spontaneous Quantum Jumps

OrQM is a (failed) attempt to overcome the measurement problem: SQM+CP, in fact, accounts for what it is observed after running experiments (no superpositions but definite values) despite what the SQM's dynamics prescribes (a superposition of states; for instance, a measuring device in a superposition of states). The issue is that the proposal not only looks unacceptably ad hoc, but it also incurs in the previously-mentioned problems. Nonetheless, this is not the end for *any* collapse theory. Neither does it mean that the notion of collapse be utterly unacceptable.

David Albert (1990) establishes that a *workable* theory of collapse ought to be able to do the following:

(i) Guaranteeing that measurements always have outcomes. That is, that such a so-complemented theory gets rid of any superposition of states when a measurement takes place.

(ii) Preserving the familiar statistical connections between the outcomes of those measurements and the wave-function of the measured system just prior to those measurements.

(iii) Being empirically consistent with the so-far-obtained evidence about the dynamics of physical systems.

Giancarlo Ghirardi (2016) also prescribes a recipe of what a palatable theory of collapse must overcome in order to have a precisely-formulated quantum theory. In particular, it ought to make explicit *how* the process of collapse works, *where* and *when* it is to occur. SQM+CP was unable to provide plausible answers to any of these elementary questions, so any proposal intending to go beyond CP ought to, at least, provide a detailed explanation of the collapse mechanism.

Section 1. The Dynamical Reduction Program: GRW-type theories.

Section 1.1. A *bare* GRW-type theory

The *Dynamical Reduction Program* (*DRP*), wherein CP should be just considered as a first unsuccessful proposal, is an attempt to cook up a workable quantum theory with collapse. Strictly, *DRP* includes any theory that brings up the idea that the wave function undergoes, once in a while or under specific circumstances, some sort of collapse or reduction. And the best candidates for this task up to the present are the so-called *spontaneous collapse theories*, widely-known by the 'GRW' acronym due to one of its first and most successful formulations by Giancarlo Ghirardi, Alberto Rimini and Tulio Weber in 1986.

The GRW-type theories are essentially a modification of the standard evolution law of SQM. Such a modification aims at unifying both microprocesses and macroprocesses in a single dynamic. In a nutshell, the guiding tenet aims at accounting for the reduction or collapse process dispensing notions like 'measurement', 'observer' or 'consciousness'. It rather intends to provide a single dynamic where collapses take place stochastically and spontaneously by modifying the standard evolution given by the Schrödinger equation. Any GRW-type theory endorses the idea that the collapse of the wave-function is not only a real physical process, but that this ought to be understood in an entirely measurement-independent manner and only regarding individual single quantum systems.

Though P. Pearle (1979) and N. Gisin (1984) were the first to entertain and to suggest this idea, the first formulation of a coherent GRW-type theory is the *Quantum Mechanics with Spontaneous Localization* (QMSL), crafted mainly by Ghirardi, Rimini and Weber in 1986 (see Ghirardi, Rimini and Weber 1986, Ghirardi 1995). Spontaneous localization is a working proposal for a stochastic and non-linear *modification* of the Schrödinger equation in order to provide not only an explanation of collapse processes, but also to overcome the measurement problem. The main tenet of the theory is that an individual quantum system (represented by its wave function) *almost* always evolves unitarily, deterministically and linearly according to the ordinary Schrödinger equation but, once in a while, it undergoes a sudden and stochastic "hit" (or "jump", or "collapse") after which the system gets a definite localization (that is, its state collapses onto one of its position eigenstates). In particular, what happens to a quantum system undergoing a collapse is that its wave function gets multiplied by a narrow three-dimensional Gaussian function of the position operator. Consequently, the quantum system gets a precise localization in the three-dimensional space, getting rid of any other term in the superposition.

Despite its at-first-glance simplicity, such a mechanism must be carefully detailed in order to get a workable theory. Let's spell it out a bit more carefully.

From a broad viewpoint, QMSL consists in three tenets: (i) each particle of a system of *n* (distinguishable) particles experiences, with mean frequency λ, a sudden, spontaneous process; (ii) in the time interval between two successive spontaneous processes the system evolves deterministically and linearly according to the Schrödinger equation; and (iii) the sudden spontaneous process is a *localization* described by

$$|\psi\rangle \to |\psi_x^i\rangle = |\varphi_x^i\rangle / \|\varphi_x^i\| \tag{8.1 }$$

$$|\varphi_x^i\rangle = L_x^i |\psi\rangle \tag{8.2 }$$

where L_x^i is the *collapse operator*, a norm-reducing, positive, self-adjoint, linear operator in the *n*-particle Hilbert space, representing the localization of particle *i* around the point *x* (Ghirardi 1995: 174). In more details, the collapse operator is defined as

$$L_x^i = \frac{1}{(2\pi\sigma^2)^{3/2}} e^{-\frac{(\widehat{Q_i}-x)^2}{2\sigma^2}} \tag{8.3 }$$

where $\hat{Q}_i$ is the position operator of system *I* and, importantly, σ is a new constant of nature standing for the width of the multiplying Gaussian. I will be come back to it shortly.

From this, one obtains that the probability density is equal to

$$\rho_i(x) = \|\varphi_x^i\|^2 \tag{8.4 }$$

And all GRW-evolution is thus described by a unified dynamic given by the Schrödinger equation *plus* this stochastic mechanism introduced in it (see eq. 8.5 for seeing this clearly).

Let $|\psi_{t_0}^i\rangle$ be now the initial wave function of a "particle" i at t_0. From t_0 to t_1 (where $t_1 = t_0 + \Delta t_1$ and is chosen randomly) the wave function will evolve according to the usual Schrödinger equation. It is worth noticing that Δt_1 is a random time distributed conforming with the exponential distribution with rate $N\lambda$, where λ is a new constant of nature indicating the frequency that a "jump" or "hit" occurs –of the order of $10^{-16}s^{-1}$– and *N* is the number of quantum systems involved.

At time t_1, the single system *i* suffers an instantaneous collapse with random center *X*. If $|\psi^i_{t_0}\rangle$ was in a superposition of states, one of the terms gets multiplied by a finite number, and all the rest by 0. In this way, all the terms (except that being multiplied by a finite number) vanish and the system gets a precise localization centered at, say, x_k. The probability that the *i*th term gets multiplied by a finite number is thus equal to $|x_i|^2$. Then, the algorithm is iterated: the quantum system evolves again according to the Schrödinger equation up to a time $t_2 = t_1 + \Delta t_2$ where a new collapse occurs, and so forth.

This theory of collapse needs to introduce two new constants of nature, as noted above: (i) the frequency, per unit time, of a collapse (λ), and (ii) the width σ of the multiplying Gaussian curve (the σ founded in the localization operator, see eq. 8.3), whose order is of $10^{-7}m$. The introduction of such constants is required to sharpen the collapse mechanism. Certainly, these constants remain unexplained and should be indeed taken as brute facts about what the world is like at the microscopic level according to a GRW-type theory. Though it might look like unacceptably ad hoc for some, the fact is that any theory needs to leave some hypotheses or assumptions unexplained. The relevant thing, however, is *what* the theory is *able* to explain and whether it accomplishes for what it was formulated. And looking at Albert's and Ghirardi's above-mentioned desiderata, QMSL certainly comes up with a theory of collapse free of the failures of its antecessor (CP). Let's see this closer.

To begin with, the theory predicts that a single quantum system in isolation will suffer a collapse with an extremely low probability in the following seconds. The frequency λ of a collapse for a single quantum system is so extremely low that it can be said. with high certainty, that the system is bound to evolve unitarily, linearly and deterministically for a long time (even though the collapse will necessarily occur in some time). In this sense, an isolated quantum system will virtually always remain unlocalized.

Nevertheless, when it comes to *macroscopic* systems the *same* dynamical mechanism tells us a different, though closely-related, story. And this story links the micro-dynamics with what is eventually observed. Given that a macroscopic system (like a measuring device) consists of an astonishingly large amount of quantum systems (of the order of millions and millions), the probability of undergoing a localization will be extremely high (in the order of having a localization each 10^{-7}seconds). And this is simply explained by the *number N* of quantum systems involved. This entails that, at the macroscopic level, a superposition of states is incredibly short-lived. In John Bell's words, a Schrödinger's cat is not both dead and alive for more than a split of second within a GRW framework. The upshot of all this is that the same mechanism implies, in a nutshell, that whereas a single isolated quantum system will almost never suffer a collapse remaining thus unlocalized, a large enough number of quantum systems (like any macroscopic object) will almost always remained localized in so far as superpositions of states collapse almost immediately.

In this way, QMSL brings forth an elegant and simple explanation of why definite values are always observed, despite SQM's predictions. The theory notably brings forth such an explanation doing without any reference to 'measurements' or 'observers': a measuring process, for instance, is as physically ordinary as any other process, and gets an explanation within the QMSL framework. One can also straightforwardly see how this theory, if taken true, explains *how*, *when* and *where* collapses take place. Quantum mechanics (in its GRW-type interpretation) thus recovers its (alleged) *fundamental* nature.

It is worth adding two further remarks. First, QMSL is not free of criticisms (see Albert and Vaidman 1989, Albert 1990, Shimony 1990) and drawbacks (see Ghirardi 1990). A new GRW-type theory was in consequence introduced few years later by Pearle (1989) and Ghirardi, Pearle and Rimini (1990). This new theory, usually referred to as *Continuous Spontaneous Localization* (CSL), replaces the discontinuous jumps of QMSL by a continuous stochastic evolution in the Hilbert space, also called 'a diffusion process'. One of the benefits of CSL is that it provides a compact mathematical equation for the stochastic mechanism (already presented in QMSL, though less clearly), which is now based on continuous Markov processes in the Hilbert space (Ghirardi, Pearle and Rimini 1990: 179). Furthermore, CSL is a generalization of QSML where the symmetry properties of the wave function, in the case of identical constituents, are respected. The fundamental dynamical equation in CSL is thus a linear stochastic evolution equation for the state vector (see Ghirardi, Grassi and Benatti 1995). In the Stranovich form, the dynamical equation is

$$\frac{d|\psi_w\rangle}{dt} = \left[-\frac{i}{h}H + \sum_i A_i w_i(t) - \gamma \sum_i A_i^2\right]|\psi_w\rangle \tag{8.5 }$$

Note that eq. 8.5 looks quite similar to the ordinary Schrödinger equation with the exception of two terms on the right side, where the quantities A_i are commuting self-adjoint operators and the quantities $w_i(t)$ are c-number stochastic processes with a probability given, roughly, by the previous equations. In the following, I will exclusively refer to eq. 8.5 as *the* dynamical equation of a GRW-type theory.

Secondly, and more interestingly, although a GRW-type theory successfully reproduces all the quantum mechanics results that were obtained up to the present, it is intended to be a *rival* theory to the extent to which some of its predictions diverges from SQM's (see Bassi and Ghirardi 2003, Adler 2007 and Bassi et al. 2013 for reviews). The proposed (crucial) tests are not yet feasible, but as technology is rapidly improving, much of the hopes for its supporters rely on future experiments.

Section 1.2. A *bare* GRW-type theory plus its *Primitive Ontology*

Most GRW theorists agree on this basic formulation. However, if a GRW-type theory is not only meant to be a theory about what is obtained after experiments, this cannot be the whole story. Thus-introduced theory remains utterly silent about what the theory is fundamentally about. This *bare* GRW-type theory would only provide the algorithm to describe the two-faced evolution of the wave-function in a high dimensional space. So, it could outwardly look like the theory is mainly about the wave-function, inhabiting in such a high dimensional space. However, as the theory stands, it is unable to connect with the three-dimensional matter out there (of which cats, chairs and tables are made). So, in order to bridge the gap between the high-dimensional wave-function on the one hand, and the three-dimensional matter around us, on the other, Giancarlo Ghirardi proposed a primitive ontology (PO) for the theory, as was described before (Chapter 6). When equipped with a PO, two versions of a GRW-type theory can be distinguished:

- Bare GRW + PO of *matter-density field* (GRWm henceforth)
- Bare GRW + PO of *flashes* (GRWf henceforth)

An ontology of flashes or of matter-density field are two (physically equivalent) ways to spell out what the fundamental stuff of a GRW-type theory is: either the fundamental constituents of the theory are flashes (events), or a continuous matter-density field. Let's say something more about them.

GRWm was originally proposed by Ghirardi, Grassi and Benatti in a 1995 paper. The introduction of a "mass density in ordinary space" (1995: 7) aimed at overcoming the so-called "problem of the tails" (a drawback raised by Shimony 1990, and by David Albert and Loewer 1990). Briefly, the problem is that even though the localization process leads to wave-functions strongly peaked around the spatial position of the jump, it nonetheless doesn't avoid that the final wave-function be different from zero elsewhere (these are the "tails" of the peaked wave-function). According to Ghirardi (2016), this problem only follows if the "standard probabilistic interpretation of the wave-function" remains unchanged, particularly, in assuming that its squared modulus gives the probability density of the position variable. GRWm's aim is to abandon such an interpretation and to rather turn to a "mass-density ontology".

GRWm suggests that the right attitude towards a GRW-type theory is to regard it as being about a mass-density field $m(x,t)$ for every point $x \in \mathbb{R}^3$ in the physical space and every time t. Those variables are functionals of the wave function and are thereby determined by it, which plays an *ontologically* secondary role just as it explains how such a field behaves in a three-dimensional space. Whereas the wave-function indeed belongs to the GRWm's ontology, it shouldn't be regarded as a fundamental constituent of it. Hence, the microscopic fundamental ontology of GRWm is not of particles or of wave functions but of a *continuous* field whose behavior is given by the wave-function, developing further an idea already suggested by Erwin Schrödinger in the 1930s.

In order to relate the reduction process to the mass-density field, GRWm's proponents introduce a mass-density operator acting upon each point $x \in \mathbb{R}^3$. The stochastic procedure strives to make the mass-density distribution "objectively definite" (to use Ghirardi's word), relating the localization frequency to the mass of the constituents. In Ghirardi's words:

> "what the theory is about, what is real 'out there' at a given space point x, is just a field, i.e., a variable $m(x,t)$ given by the expectation value of the mass density operator $\widehat{M}(x)$ at x obtained by multiplying the mass of any kind of particle times the number density operator for the considered type of particle and summing over all possible types of particles which can be present" (Ghirardi 2016)

There is yet an ontologically quite different version for GRW-type theories. According to it, the fundamental ontology of the theory is one of *flashes*, i.e., *events* occurring in the physical space. This "flashes ontology" was originally put forward by John Bell (1987) as a mean to identify the "local beables" of a GRW-type theory. GRWf, as it has come to be known ultimately, has been developed further by Adrian Kent (1989), Roderich Tumulka (2006) and Allori et al. (2008).

The flashes occur in the physical space (and time) and are mathematically represented by single points in space-time. They are the fundamental constituents of everything out there. A single macroscopic object is neither a bunch of particles, nor a particular configuration of a matter field, but a *pattern of flashes*, where each flash corresponds to one of the spontaneous localizations or collapses of the wave-function. Under this view, three-dimensional objects can be account for in terms of an extremely large amount of collapse processes occurring spontaneously: the larger the number of wave-function's degrees of freedom, the bigger the number of flashes (collapses) (see Allori et al. 2008).

As in GRWm, in GRWf the wave function doesn't belong to the fundamental ontology, but it just plays the role of governing the PO's behavior. In this case, the wave function is the law of evolution for the flashes. As Michael Esfeld and Nicolas Gisin put it:

> "(…) given an initial configuration of flashes, the wave-function enables conceiving histories of the distribution of flashes in space-time and makes it possible to assign probabilities to these histories. Note, however, that on the GRWf theory, only one such history actually occurs, and that the initial configuration of flashes does not determine a unique wave-function; any initial configuration of flashes is compatible with many different wave-functions" (Esfeld and Gisin 2014: 251)

Any fully-interpreted GRW-type theory will therefore have the following ingredients:

(a) A formalism for quantum states and observables (akin SQM)

(b) A dynamic given by the Schrödinger equation plus a spontaneous and stochastic mechanism of localization (or collapse) in it for *any* physical system (eq. 8.5).

(c) Either of the followings POs: (i) a PO of a matter-density field (GRWm), (ii) PO of flashes (GRWf)

Section 2. Time (a)symmetry and the arrow of time in a GRW-type theory

The question of the arrow of time within any GRW-type theory boils naturally down to the question about whether any of the above-listed elements manifests a temporal asymmetry (either structural or non-structural).

As to (a) and (b) *without* the amendment, there is no much to say about it within a GRW-type theory: the question therein mostly reduces to whether the Schrödinger equation *without any addition or modification* turns out to be time-reversal invariant. And this has been already addressed and analyzed in Part 2. If the Schrödinger equation turns out to be non-time-reversal invariant, for any of the reasons I presented before, so will any GRW-type theory be. No further argument or reason is given on GRW's side: GRW adds nothing to the discussion. But, if one comes to agree on any of the reasons that may render the Schrödinger equation time-reversal invariant (i.e. a relationalist stance with respect to time), then GRW-type theories could indeed have something else to say about the arrow of time, which will come from those elements that are specific to it: either (b) *with the amendment* (that is, something like eq. 8.5), or (c).

As I argued in a general fashion in Chapter 6, I'm very suspicious that a PO may introduce on its own, *without* any further assumption, a time-asymmetric feature in the theory. The same logically goes for both a PO of flashes and a PO of a matter-density field. In these cases, the PO alone doesn't bear any information about how the patterns of flashes or the distribution of the matter-density field will behave in time. The PO says only what the building blocks of the universe are, not how they dynamically evolve in time. Both a PO of flashes and of matter-density field on their own give us a static and inert picture of fundamental reality. All this is widely-recognized by GRWf's and GRWm's supporters as they often remark that it is the wave-function *plus* the spontaneous localization process what tell how the patterns of flashes or the mass-density field evolve in time[42]. Nonetheless, it's not the case that the PO says *nothing* whatsoever about time or the property of time-reversal invariance. But I shall address this point afterwards.

This leads us to focus on (b): the Schrödinger equation plus *the spontaneous, stochastic and non-linear localization mechanism* in it (eq. 8.5). It has been often argued that any GRW-type theory offers the more serious and promissory candidate for a *structural* (fundamental) arrow of time in the quantum realm: the time asymmetric ingredient enters the theory entirely through such a stochastic mechanism. And there are broadly two senses in which a GRW-type theory turns out to be interesting for the arrow of time debate:

(i) The spontaneous localization process is structurally (intrinsically, fundamentally) non-time-reversal invariant (see Arntzenius 1997, Callender 2000, Esfeld and Sachs 2011, North 2011 among others)

(ii) A GRW-type theory would offer the means to bridge finally the gap between thermodynamics and statistical mechanics in grounding a temporally asymmetric

[42] There seems to be however a subtle difference between both ontologies. Whereas it's true that a matter-density field without a dynamic is just something that remains inert ad infinitum, an ontology of flashes would be slightly different because flashes are something that happen, something that come about somehow thanks to the dynamics. To put it in other words: whereas a completely inert ontology of a matter-density field is (if you put effort on it) imaginable, an ontology of flashes is impossible because the very events can just come about thanks to the dynamics: an ontology of flashes without any dynamics producing them is, so to speak, no an ontology whatsoever for there is nothing there, just emptiness. In any case, a GRW-dynamics "produces", so to speak, the flashes but they cannot introduce any temporal asymmetry on their own: just the dynamics specifying how patterns of flashes behave in time.

probabilistic tendency for systems to evolve toward equilibrium states (Albert 1994, 2000)

GRW-type theory would thus be killing two birds with one stone: it would, on the one side, exhibit a *structural* time asymmetry at the level of its unified dynamics, entailing that time in itself is temporally asymmetric conforming with the theory (the problem of a structural arrow of time); and it would, on the other, bridge the gap between thermodynamics and statistical mechanics, explaining why one has experience of a temporally asymmetric macro-world (one of the ways to pose the problem of the two realms).

Section 2.1. A GRW-based *structural* arrow of time

Though the Schrödinger equation might turn out to be time-reversal invariant, if one amends it with a new sort of evolution, the new unified dynamics could nonetheless turn out to be *non*-time-reversal invariant. This has been claimed by, for instance, Jill North (2011). She says:

> "Not only is wavefunction collapse governed by a fundamental, indeterministic law on this theory (GRW), but by a fundamental, non-time-reversal invariant law. GRW assigns probabilities to the different possible future wavefunctions that a system's current wave-function could collapse into." (North 2011: 333)

And a few lines below,

> "The theory doesn't assign probabilities to different possible past wave-functions, given a system's current wave-function. The collapse law doesn't say anything about the chances of different past wave-functions." (Ibidem)

Finally,

> "GRW then says that different things can happen in either direction of time: wavefunctions can collapse in accord with lawful probabilities to the future, not the past." (Ibidem)

In relation to the arrow of time, Frank Arntzenius's claim (1997) goes more or less along the same line

> "Thus, if one believes in some "orthodox" collapse version of quantum mechanics, or one of the more recent Pearle-GRW type collapse has reason to believe that time has an objective direction." (S218)

And in the conclusions, he adds

> “The conclusion, therefore, is that whether one should believe that time has an arrow depends on one’s interpretation of quantum mechanics: If one adheres to a theory of probabilistic collapse, one has reason to believe that time has an arrow.” (S222)

That a GRW-type theory is in the position to explain the arrow of time has been explicitly held by Michael Esfeld and Christian Sachse They say

> “This interpretation is in the position to explain the origin of the direction of time. Amending the Schrödinger equation with a stochastic term as GRW does in order to account for state reductions has not only the consequence that the dynamics is indeterministic, but also that it is not time-reversal invariant. In other words, the GRW equation is a candidate for a fundamental *law of nature that is not time-reversal invariant*.” (Esfeld and Sachse 2011: 60)

Craig Callender (2000) holds the same view as well. He puts it as following:

> “On collapse theories, a certain feature of the system (e.g. particle number, mass, being observed) will trigger a non-unitary, indeterministic transition from ψ to one of its components ψ_i. This collapse is not governed by the Schrödinger evolution. And, in general, there is no way of evolving from the collapsed system back to the uncollapsed system with the same chance. According to collapse theories, there is a preferred orientation to time.” (2000: 261)

Callender’s claim extends to *any* collapse theory (as being SQM+CP), including naturally any GRW-type theory.

Let’s think through what means that a GRW-type theory be non-time-reversal invariant in the light of what was said in Chapter 2. The main manifestation of time-reversal invariance is the generation of a pair of time-symmetric twins for any relevant evolution. This readily manifests in the class of possible solutions (or possible worlds) of a theory’s dynamics. In the case of a time-reversal invariant law, the structure of its possible solutions (or possible worlds) is given by $\mathrm{W} = \mathrm{W}^f + \mathrm{W}^b$, where a mapping going from $w_i \in \mathrm{W}^f$ to $w_k \in \mathrm{W}^b$ necessarily exists and W^f and W^b are physically equivalent. It is worth recalling too that this result doesn’t mean that a single evolution, say w_i, be reversible (symmetric) or irreversible (asymmetric).

For a GRW-type theory to be genuinely non-time-reversal invariant, the class of solutions of the eq. 8.5 (that is, the Schrödinger equation plus the spontaneous localization mechanism in it) thus has to have one of the following structure: (a) *either* $\mathrm{W} = \mathrm{W}^f$ *or* $\mathrm{W} = \mathrm{W}^b$ are models of the theory, but *not both*; (b) alternatively, W^f and W^b are models of the theory but $\mathrm{W}^f \neq \mathrm{W}^b$. (a) can be put in words as saying that a GRW-type theory only works (that is, it yields predictions in the form of solutions) in one direction of time, but not in the opposite.

And this is exactly what North means: the collapse law doesn't say *anything* about wave-functions' past. Or, alternatively (b), the theory says *different things* about wave-functions' past and future.

Let me portray this in the following toy model. Imagine a Universe containing only one quantum system in a position superposition state at t_0, $|\psi_0\rangle = \sum_i a_i |a_i\rangle$. The GRW algorithm predicts the following: the quantum system is bound to evolve, *towards the future*, unitarily, deterministically and linearly according to the Schrödinger equation for a good amount of time but there is a *real* probability that in that Universe such a quantum system undergoes a stochastic, indeterministic and non-linear localization process at t_1 (where $\Delta T = t_1 - t_0$ can be equal to 10^9 years!). But, and this is the upshot of non-time-reversal invariance, the GRW algorithm either remains silent or says something different when time runs backward. That is, either it is unable to specify how a time-reversed Universe containing one quantum system in a position superposition state is going to evolve in the backward direction of time, or it says that the quantum system will evolve in a different manner in the backward direction. Let's see how this happens.

Section 2.2. A GRW-based *structural* arrow of time in two senses

Stochastic and spontaneous collapses do seem to introduce a structural time asymmetry in the theory: A GRW-type dynamic intrinsically distinguishes the past-to-future direction from the future-to-past one. However, this is done in two slightly different ways: from an account closer to a *substantivalist view* of time reversal the asymmetry is introduced in the theory through a difference in the patterns of hits in the backward and in the forward direction of time; from an account closer to a Wigner's sense of time reversal and to a relationalist view, the asymmetry is introduced through the idea that the original state one starts with cannot be recovered after a hit. To jump to the conclusions, theory is indeed non-time-reversal in both senses. Notwithstanding, they display a slightly different sort of time asymmetry.

To begin, it is worth mentioning in which sense the collapse mechanism would with precision be time asymmetric. It might be suggested that the spontaneous localization process *in itself*, that is, independently of the Schrödinger equation, is taken to be non-time-reversal invariant. However, GRW's supporters claim that one of GRW's benefits is that any physical system obeys the same single dynamics: the GRW's dynamics is the Schrödinger equation *with* an amendment (that is, eq. 8.5), so it cannot, and shouldn't, be taken independently of the Schrödinger equation. Therefore, the spontaneous localization mechanism is not the right sort of thing that one can temporally reverse *in itself* and independently of the Schrödinger equation. It rather just happens, abruptly, suddenly and instantaneously on a system evolving according to eq. 8.5. What does it thus mean that the localization mechanism occurs backward? By *definition*, the spontaneous localization is something that collapses a quantum state in a superposition onto one of its components (defining a precise position or region in space, for instance). So, its happening backward cannot mean a "spontaneous des-localization" process, but a "spontaneous localization" process onto a backward-in-time-evolving system according to a time-reversed eq. 8.5.

Hence, I think that the right answer to the question "what does it mean that the localization mechanism occurs backward" is that it occurs over a time-reversed wave-function like $|\psi_0\rangle = \sum_i a_i |a_i\rangle$, *evolving backward in time* according to a time-reversed amended Schrödinger equation, that is, it occurs over something like $|\psi_0^*\rangle = \sum_i a_i^* |a_i^*\rangle$ instead over $|\psi_0\rangle = \sum_i a_i |a_i\rangle$.

Having said that, let's come back to the Universe containing only one quantum system in a position superposition state. It was already shown what a GRW-type's dynamic predicts in the forward direction of time. Let's see what happens when time goes backward in the *Wigner's relationalist sense* of time reversal.

Esfeld and Sachs, and also Callender in different parts, claim that *once* a quantum system has undergone a spontaneous localization (that is, it has collapsed in accordance with the stochastic mechanism), it cannot be evolved back to the *un*-collapsed state that *existed* before the first spontaneous localization occurred. And in this sense a GRW-type theory is *non*-time-reversal invariant: it doesn't provide the means to come back to the original state; or, to put it in other words, the quantum state one started with cannot be recovered if it has gone through a spontaneous localization process.

Just-described scenario would look like this. Imagine a Universe much like the previous one. At t_0, the state of the quantum system is in a position superposition state, $|\psi_0\rangle = \sum_i a_i |a_i\rangle$. Let's make *that* system evolve in time conforming to GRW's dynamics. Suppose that such an evolution takes 10^9 years, more or less. So, GRW's predictions are that the quantum system is bound to evolve, *towards the future*, unitarily, deterministically and linearly according to the Schrödinger equation for about 10^9 years and that, in such a span of time, it will highly likely undergo a stochastic, indeterministic and non-linear localization process at, say, t_1.

$$|\psi_0\rangle = \sum_i a_i |a_i\rangle \underset{\sim 10^9}{\longrightarrow} |\psi_1\rangle = |a_k\rangle \tag{8.6 }$$

And one finds, after such a span of time, that the quantum system in fact underwent a collapse process.

Good. Let's ask now the following: does GRW's dynamics allow us to take *that* system back to the original state $|\psi_0\rangle$? Note that the question is not whether the theory's dynamics says something or not about how the quantum system would evolve backward in time, but if it provides any mean for tracing a quantum system back through a GRW's evolution to its original state at t_0. And the answer to this question is that there is no way to generate a GRW's evolution that restores the original state. And this is naturally true for any GRW's evolution can ever be obtained. And this is exactly a violation of time symmetry in the Wigner's *relationalist* sense where what is meant by time reversal is *motion* reversal.

Let's now turn things around. Imagine again the same above-mentioned Universe but now imagine it temporally reversed. Let's thus make *that* temporally-inverted system $|\psi_o^*\rangle$ evolve in the backward direction of time conforming to the GRW's dynamic. Suppose that such an evolution takes 10^9 years in the past, more or less. So, GRW's predictions in such a temporally-reversed scenario are that the quantum system is bound to evolve, *towards the past*, unitarily, deterministically and linearly according to the Schrödinger equation for about 10^9 years and that, in such a span of time, it will highly like undergo a stochastic, indeterministic and non-linear localization process at, say, t_{-1}.

$$|\psi_{-1}^*\rangle = |a_k^*\rangle \underset{\sim 10^9}{\longleftarrow} |\psi_0^*\rangle = \sum_i a_i^* |a_i^*\rangle \tag{8.7 }$$

Good. Let's ask again the same question: does GRW's dynamics allow us to take *that* system forward to the original state $|\psi_o^*\rangle$? the answer to this question is, once again, that there is no way to recover the original state through a GRW-generated evolution, regardless if it lies either in the past or in the future.

Recall Wigner's general criterion for time-reversal invariance. For Wigner, time reversal is meant to be a transformation such that

$$time\ displacement\ by\ t \times time\ reversal \times time\ displacement\ by\ t \times time\ reversal\ = \boldsymbol{I}$$

Within a GRW-type framework, the relevant time displacements are those long enough wherein spontaneous collapses occur. So, let's suppose that during the first time-displacement the system evolves according to the Schrödinger equation for a good amount of time but, at some point, it undergoes a spontaneous collapse. Then time-(motion)-reversal is applied. This means that the time-reversal operator should invert the direction of time and take the complex conjugate over the collapsed state. The system will evolve backward in time (second time-displacement) and it surely undergoes a spontaneous collapse after a while. However, it's blatant that one will end up with a quantum state completely different from the original. The identity could only miraculously be achieved. As the second time-displacement will hardly ever give us the state one started with, the theory is not time-(motion)-reversal invariant in a structural sense.

There is nonetheless a slightly different way in which a GRW-type theory turns out to be temporally asymmetric. And here my second sense of non-time-reversal invariance comes into play, which goes along with a more *substantivalist-like view of time* as long as one doesn't necessarily expect to generate an evolution going backward in time to recover the original state, but to simply reflect the time axis at a given instant and to check how the two directions of time looks according to the theory (in this case, a GRW-type theory, in particular, how things would evolve conforming to eq. 8.5 towards the past and towards the future). So, in the following, let's think that for any reason a substantivalist-minded person thinks that time reversal is implemented by something like T_A (regardless what has been said in Part 2, just for the sake of the current argument[43]), and let's ask now how a GRW-type theory deals with this, and what it adds to.

One of the senses of non-time-reversal invariance (for instance, what North claims) seems to be that a GRW-type's dynamic says *nothing* in the backward direction of time. Thus, this might drive one to believe that a GRW-type theory doesn't work when time reversed just as the ordinary Schrödinger equation didn't work when time reversed by T_U. Nonetheless, this is not quite so: *if* one time reverses the eq. 8.5 by applying T_A (as it is commonly done with the ordinary Schrödinger equation), then what GRW's dynamics predicts in the backward direction of time is that a time-reversed quantum system in the superposition state $|\psi_0^*\rangle = \sum_i a_i^* |a_i^*\rangle$ is

[43] Obviously, if one formally represents time reversal in terms of T_U, eq. 8.5 is as non-time-reversal invariant as the ordinary Schrödinger equation is. But, in this case, what GRW-type theories introduce would add nothing to the argument, since eq. 8.5 would be non-time-reversal invariant for all the reasons mentioned in Part 2. So, let's make the exercise of thinking whether a GRW-type theory introduces new reasons for non-time-reversal invariance *independently of* what has been said before.

bound to evolve, *towards the past*, unitarily, deterministically and linearly according to the Schrödinger equation (strictly, according to eq. 8.5 which behaves almost always much like the (time-reversed) Schrödinger equation) for a good amount of time and that there is a *real* probability that in that time-reversed Universe such a quantum system undergoes a stochastic, indeterministic and non-linear localization process at t_{-1} (where$\Delta T = t_0 - t_{-1}$ can be equal to 10^9 years!). This is basically what eq. 8.5 prescribes.

Note that nothing in a GRW-type theory says that localization processes only occur in the forward direction of time. What a GRW-type theory just says is that those stochastic processes occur, how often they occur, and what one should expect when they occur. But this, and only this, happens in both temporal directions. And the quantum state one started with is completely irrelevant for this as it is independent of any previous (or posterior) state. So, it doesn't matter the temporal instant one stands at, a GRW-type's algorithm will say that your quantum state in a superposition state is bound to evolve, *towards the past and toward the future*, unitarily, deterministically and linearly like a Schrödinger evolution for a good amount of time and that there is a *real* probability that in both Universes (the original and the time-reversed) such a quantum system undergoes a stochastic, indeterministic and non-linear localization process in 10^9 years. Again, this is what eq. 8.5 prescribes.

This doesn't mean, however, that a GRW-type theory is time-reversal invariant. In accordance with what has been said before, a GRW-type theory would in fact have the following structure of solutions under time reversal, $\mathrm{W} = \mathrm{W}^f + \mathrm{W}^b$, but the kicker is that W^f and W^b are bound to bear *completely different models*. To be clear: the theory *does say* something about the past (it yields past-headed models), but the past is bound to be something completely different with respect to what the future will get to be, that is, $\mathrm{W}^f \neq \mathrm{W}^b$. And this is so due to the *stochasticity* of the hits: though collapses are also expected to happen towards the backward direction of time, it would be an enormous miracle that they are produced in the same way, reproducing the same patterns, that in the forward direction of time. If time-reversal invariant means that W^f and W^b are a pair of time-*symmetric twins*, one could say that a GRW-type theory produces pairs of solutions, but they are bound to be asymmetric and not twins at all.

All this is of course possible under the assumption that the right way to time reverse eq. 8.5 is given by something like T_A, which permits the quantum state to evolve backward in time to begin with. This is what firstly generates the pair W^f and W^b: an individual system evolving according to eq. 8.5 will behave almost always as a system evolving according to the ordinary Schrödinger equation. Now, let's put all this in the following way: the stochastic mechanism just randomly "triggers" a collapse both onto the *original* wave function evolving conforming with eq. 8.5 and onto the *time-reversed* wave function evolving conforming with (time-reversed) eq. 8.5. The point is that there seems to be no reason why such a randomly-triggered process couldn't happen in the two directions of time. When time reversal is thought of as a *reflection*, a quantum system could either evolve *towards the future* or *towards the past* following eq. 8.5 and spontaneous localizations are expected to happen in *both* cases, on both sides of the reflection, so to speak. Non-time-reversal invariance comes out from the fact that it would incredibly unlikely be that both sides of the reflection reproduce the same patterns or evolutions (the same possible worlds if you like). Therefore, the pair W^{f} and W^{b} is inherently asymmetric. And the upshot of all this is that a GRW-type theory does provide a *structural* arrow of time even in this sense too.

Section 3. Spontaneous localizations and the foundations of thermodynamics

There is another sense in which GRW-type theories are highly valuable for the arrow of time debate. This sense is that of offering deeper and more fundamental grounds for the thermodynamical temporal asymmetry. In other words, it has been proposed that GRW-type theories would bridge the gap, and relieve tension, between thermodynamics and statistical classical mechanics, that is, it would offer a solution to the *problems of the two realms* as it was presented in Part 1.

The impact GRW-type theories would have on that issue has been extensively investigated by David Albert (1992, 1994, 2000). Not only does Albert believe that a GRW-type theory is a fundamental non-time-reversal invariant theory, but it would also ground the temporally asymmetric probabilistic tendency for systems evolving toward equilibrium. The first step towards a "new universal statistical mechanics" (as he pompously calls his project) is to take GRW-based statistical mechanics seriously.

Now, let's set the case up. Consider a two-body system, where the temperature-difference between both bodies is, say, ΔT. Such a situation is, naturally, compatible with a set of possible microstates $\{C\}$. After some time, the temperature difference between two bodies will have varied. In accordance with the (time-reversal invariant) equations of motions of classical mechanics, those micro-states may evolve towards two quite different kinds of states: those compatible with microstates in which the temperature difference is smaller (that is, the two bodies' temperature approaches one another), and those compatible with microstates in which the temperature difference is bigger (that is, the two bodies' temperature diverges one another as time goes by). Call the first subset of $\{C\}$ "normal" microstates $\{C_n\}$, and the second subset "abnormal" microstates $\{C_a\}$

Recall that the bewildering aspect of this situation is that both microevolutions are equally entailed by the micro-dynamics governing those systems, though there seems to be an enormous numerical imbalance in nature between them that remains unexplained: $\{C_n\}$-type evolutions are so much common than $\{C_a\}$. In fact, as Albert puts it, normal microstates vastly outnumber abnormal microstates not only in general but also in every individual microscopic neighborhood of $\{C\}$ and even within the microscopic neighborhoods of $\{C_a\}$. And there seems to be no answer, nor way-out, within classical dynamics.

Such an answer, or way-out, lies beyond classical dynamics: GRW-type theories, according to Albert, offer the right sort of dynamics capable of doing the job satisfactorily. That is, the right sort of theories capable of answering the above-mentioned questions from an *internal* (asymmetric) *dynamical* cause (Price 2002: 32). The argument runs roughly as follows. The above-sketched situation wherein $\{C_n\}$-type cases vastly outnumber $\{C_a\}$ ones just shows that normal states are much more stable under small perturbations than abnormal ones. And if the two bodies in the above example "were in fact somehow being frequently and microscopically and randomly perturbed, then the temperature of those two bodies would be overwhelmingly likely to approach one another no matter which one of the microstates in $\{C\}$ initially obtained" (Albert 1994: 203). Albert's suggestion is that GRW's jumps (spontaneous localization processes) are exactly the right sort of chancy perturbations one needs to get the job satisfactorily done.

In more details, Albert's suggestion is "that every single one of microstates in $\{C\}$ (and not merely a large majority of them) will be overwhelmingly likely, on the GRW theory, to evolve, over the subsequent ten minutes, into states in which the temperature difference between the two bodies is smaller" (1994: 203). GRW-type theories' dynamics is precisely very good at, according to Albert's thought, getting us from abnormal microconditions to normal ones. And this is possible only because the scales over which "the tiny individual clots of abnormal micro-conditions typically extend are vastly smaller than the scales over which the values of the $P_A(B)$ appreciably vary" (Albert 2000: 155).

In conclusion, the perturbations required to get and to single out the right sort of thermodynamic evolutions seem to be provided, in a genuinely chancy and fundamental manner, by any GRW-type theory. Those GRW-like perturbations (produced by inherently stochastic spontaneous localization processes) don't prohibit transitions to abnormal microconditions but are extremely unlikely in virtue of the GRW's internal asymmetric dynamics. As Callender puts it,

> "The innovation of the theory lies in the fact that although entropy is overwhelmingly likely to increase toward the future, it is not also overwhelmingly likely to increase toward the past (because there are no dynamic backwards transition probabilities provided by the theory)" (Callender 2016)

And so, the theory does not suffer from the problem of predicting entropy-increasing processes, for instance, towards the past.

To sum up. The structure of this "new universal GRW-based statistical mechanics" comes out like the following:

(1) The GRW law of motion for quantum-mechanical wave functions given by the Schrödinger equation plus a stochastic amendment over it (eq. 8.5).
(2) The Past Hypothesis (like the Mentaculus' standard contraption), which postulates a particular low-entropy macroconditions at the beginning of the Universe.
(3) The current state of the Universe as a *contingent fact.*

And this structure would be able to offer a convincing solution of the problem of the two realms, as I've formulated it in Part 1, Chapter II. Let's see it closer.

Remember that the problem of the two realms comes down to three points:

(a) Given that the underlying fundamental dynamics of the world is temporally symmetric, why do we have experience of a temporally-asymmetric world?
(b) There is hence a gap between the micro-world and the macro-world.
(c) To close the gap or to bring this shattered picture of reality together has largely been the problem at stake.

Albert's proposal changes radically the underlying fundamental dynamics: this is no longer the Newtonian (or Hamiltonian) classical mechanics, but a GRW-type one. Remarkably, not only is the dynamics non-time-reversal invariant, but also chances enter the world only once through the GRW-type's dynamics: those tiny and random perturbations account for all the statistics required to explain our world and its temporally asymmetric manifestation. Furthermore, a unified dynamic readily closes the gap between the macroscopic world and the microscopic one: the two above-specified laws govern equally both. Hence, one gets a deep explanation of why one sees what one actually sees: behind any entropy-increasing macroscopic behavior, a stochastic, indeterministic and non-linearly GRW-type dynamics rules. And the same dynamics governs both a single isolated quantum system and a soft drink getting warmer. Each GRW-type evolution incorporates such a built-in arrow across levels.

One further remark. All this task was mostly accomplished by the GRW-type dynamics. But, what about the Past Hypothesis? As to this, its status is somehow fuzzy here. Remember that the idea of postulating a Past Hypothesis aimed at blocking any entropy-increasing evolution towards the past, as falsely predicted by the Newtonian laws of motion. As the underlying micro-dynamics is now a GRW-like dynamics, it seems to be no longer necessary. Callender, however, contends that the Past Hypothesis is required so as "to explain how the universe ever got into a nonequilibrium state in the first place" (Callender 2016). Alberts rather thinks that the Past Hypothesis is here filling an empty space left by the GRW's dynamics: nothing in a GRW-type theory's dynamics says anything about the past of a GRW-type *particular* evolution (because, as it was shown, any GRW-type dynamics is not time-(motion)-reversal invariant in a relationalist sense). So, the Past Hypothesis is not here playing the role of correcting a false prediction towards the past, but filling a space left by the theory's silence about any past instant.

Many objections have been raised against Albert's proposal. Albert himself presents a survey thereof in Albert 2000 (156-160) and he replies them in order therein. There is yet a more conceptual objection raised by Huw Price (2002) against Albert's logic and assumptions. Price's counter-argument goes as follows.

In Albert's proposal, the observed monotonic increase of entropy is grounded on GRW's asymmetric dynamics. This claim seems to assume the following counterfactual

> *If there were no such asymmetric mechanism, the observed phenomena would be different.*

Albert would be presupposing such a counterfactual, at least according to Price's thought, in saying, for instance, that an "extraordinarily tiny and extraordinarily compressed and absolutely isolated gas will have no lawlike tendency whatever to spread out" (Albert 1994: 667). And this is so because, conforming with the collapse-driven statistical mechanics, such gas, in Albert's words, wouldn't have the enough number of constituents to suffer even a single GRW-type collapse[44]. However, the case doesn't run counter to our intuitions and daily experience, because macroscopic systems are big enough to include the sufficient amount of quantum systems undergoing collapses and, hence, explaining the normal behaviors.

If one regards this GRW-like asymmetric mechanism as something can be turned off or turned on, what Albert says is that if one had the means for turning GRW-like collapses off at all, then the general tendency of thermodynamics system to increase would simultaneously cease as well. And this counterfactual (similar to the above-mentioned one) doesn't run counter to our experience. But Price puts the finger on the triviality that Albert is bringing up: every strictly counterfactual conditional doesn't run counter to empirical experience, because it has an antecedent which is contrary to facts (Price 2002: 35).

Furthermore, Albert (uncritically) supposes that the tendency of gases to spread out is "lawlike", and it is this tendency what calls for an explanatory basis: something more basic must be producing such regularity. But, as very well-known, this is highly debatable. One strictly has experience of gases spreading out, not of any "lawlike" tendency. And what calls for an explanation is the gases' spreading out, not any posit lawlike tendency.

Price argues that other approaches, which don't assume such a lawlike tendency, can very well account for the gases' spreading out; for instance, what he calls "Boltzmann's one-symmetry view", according to which, only asymmetric boundary conditions –as an initial state of low entropy– plus a symmetric default conditions –as entropy likely to be high– are required to produce an explanation of the observed entropy-increasing thermodynamic behavior. Thus, the burden of the proof is on Albert's side: he has to show that the causal mechanism "makes a difference". He has to justify the introduction of such a mechanism in the explanation. And, Price concludes, he provides none. So, the best one can say about this is an agnostic claim: one doesn't know whether Albert's proposal works, because one doesn't know whether the GRW-like mechanism makes a difference.

[44] This is actually an objection against Albert's view put forward by Philip Pearle. The idea is, roughly, that tiny gases of around 10^5 constituents will surely spread out after some time, though such a span of time is not enough for such constituent systems undergo any sort of GRW collapse. Albert's reply is that if tiny gases manifest a tendency to spread out, this is due to spontaneous collapses of the wave-functions of the microscopic constituents of the containers of such a sort of gases, which should be taken into account to explain the whole phenomenon (see Albert 1994: 677, Albert 2000: 156-157).

Final Remarks

Along this chapter I've analyzed one of the most promissory candidates for a structural arrow of time in non-relativistic quantum mechanics: a GRW-based structural arrow of time. Whereas it is not true that *every* collapse theory does the job satisfactorily (as shown in the previous chapter), a workable version of it (as GRW-type theories actually are) seems to have some good grips. The reason is that a GRW-type theory poses a dynamic that is intrinsically non-time-reversal invariant in the sense of treating both directions of time differently solely in virtue of its intrinsic properties. I have also shown that relationalist and substantivalist accounts reach the same conclusion, though the very inversion of the direction of time is carried out a bit differently. In any case, a GRW-type theory does seem to be a knock-out case for anyone believing that non-relativistic quantum mechanics is structurally directionless.

In addition, Albert's proposal takes a step forward: it intends to also provide a GRW-based explanation of entropy-increasing macroscopic behaviors. To be clear, a GRW-type theory would not only be able to ground a structural arrow of time but would also provide the means to link the temporal asymmetric dynamics with entropy-increasing behaviors at the macro-level. That's why GRW-type theories kills two birds with one stone: the problem of a structural arrow of time and the problem of the two realms. While I think the first aim is successfully achieved, I'm not completely sure about the latter. The main counter-argument is that put forward by Price: if Albert's proposal works eventually like a "switch", and it holds that a GRW dynamics makes a difference for entropy-increasing macroscopic behaviors, then it is Albert who should show that his proposal not only makes a difference but also is the simplest one that makes a difference. This doesn't seem to be the case (as Price points). However, one benefits is that Albert's proposal is certainly *more unifying and all-embracing* than any other competitor that merely explains thermodynamics behaviors. Boltzmann's view (the one that Price puts as an alternative) is much more modest at its aims.

IX.

Bohmian Mechanics

The Arrow of Time and the Special Role of Time Reversal

Collapse theories, in any of its fully-worked out forms, put forward a thorough explanation about what's going on at the quantum level and how the measurement problem could be overcome, successfully accounting for all quantum-mechanical outcomes one obtains in experiments. And yet, a completely different explanation can be provided about what's going on at the quantum level and how the measurement problem could be overcome. In this alternative explanation, the wave function is not a complete physical description of all the physical properties of a system: *something else* needs to be added to the wave function description so as for the quantum state to account for all physical properties of a system. In doing this, the explanation declines assumption A1 in Maudlin's formulation of the measurement problem. But that's not all: the explanation also demands a *new extra* dynamical equation to be introduced in the formalism, besides the original Schrödinger equation, declining Maudlin's A2 assumption as well. Interestingly, the explanation manages all the issues without any collapse mechanism whatsoever. This alternative explanationh has received varied names: "the pilot-wave theory" (proposed originally by de Broglie), "the Broglie-Bohm theory", "the 'no-hidden variable' interpretation of quantum mechanics", or simply "Bohmian Mechanics".

The idea was firstly introduced by Louis de Broglie in 1927 at the worldwide famous Fifth Solvay Conference (the paper can be found in Bacciagalupi and Valentini 2009, Chapter 2, along with an analysis). De Broglie's seminal theory, known as "pilot-wave theory", proposed that the wave-function could play the role of a "guiding field", adding an equation for the particle motion equivalent to a guiding equation. However, de Broglie couldn't nail down the physical nature of such a pilot wave, and the theory not only was widely rejected by the physicists' community, but even abandoned by him shortly after presented.

Few decades later, in 1952, the idea got a new revival thanks to David Bohm. Not only did he fully understand the relevance of the pilot-wave theory along with its implications, but he also wrote down the theory in a mathematically careful manner. The "de Broglie-Bohm theory", as it came to be known ever since, was developed further in the following decades. The seminal idea remained: the quantum state is not fully described by the wave function alone, but it should feature something else. Furthermore, a full specification of the full quantum state's behavior requires an additional equation, as de Broglie had suggested. John Bell (1987) was a promoter of the de Broglie-Bohm theory during the seventies and eighties. In the nineties, the theory was reformulated differently by Sheldon Goldstein, Dettlef Dürr, and Nino Zanghi (1992, 1996, 1997, 2013), coming to be known simply as "Bohmian Mechanics". I shall exclusively focus on this last modern version of the theory and the discussion it has raised.

To put it simply, Bohmian Mechanics (BM henceforth) is a non-relativistic quantum theory about the *motion* of *particles*. Period. All there is out there are particles. And all one has to say about those particles is how they move in space and time. BM is just the required machinery to account for this simple fact. In virtue of this, as mentioned in passing above, BM presumes fundamentally two things:

(a) Particles have *always* definite positions
(b) The behavior of those particles is fully described by, on the one hand, the Schrödinger equation (without any amendment) and, on the other, a *guidance equation.*

Any other stuff introduced in this story (like wave-functions, a potential field –like in the de Broglie-Bohm version–, or even any other property of a system) requires a watchful interpretational task: any other stuff apart from particles and their motion in space and time one comes up with will probably be just part of the required tools to account for the motion of particles. And all this brings out a fully deterministic version of quantum mechanics, without spontaneous collapses or objective chances.

Let's put some flesh on all it.

I've said that BM affirms that the quantum state is not fully specified by the wave-function alone, but it features extra data –the *positions of particles*. In this way, the complete state of a quantum system with N variables (number of particles) at time t is completely specified by the wave function *and* all the positions of the particles composing the system:

$$(\psi(t), Q(t)) \tag{9.1 }$$

where $\psi(t)$ is the usual quantum state, $\psi(t) = \psi(x_1, \dots, x_N, t)$, and $Q(t)$ represents the actual positions of the particles composing the system, $Q(t) = (Q_1(t), \dots, Q_N(t))$. Since particles always have determined positions and particles' positions vary along time, particles naturally will have *trajectories* as well. And, of course, different initial states will pick up different trajectories for the involved particles. Particles' positions, it is worth stressing it, are an objective feature of the quantum world, independent of any measurement or observation. Why Bohmian Mechanics is (mistakenly) called it "a hidden variable theory" is due to this fact: particles' positions are not included in the purely quantum-mechanical description of a system (that is, its wave function) but are further information that should be add according to Bohmians. For them, naturally, positions are not "hidden" at all but quite the opposite: it's the only sort of property you ever observe (Bricmont 2016: 130). Or, more radically, it's the only property that a quantum system really instantiates (see Esfeld, Lazarovici and Oldofredi 2019).

The Bohmian evolution is given then by two equations:

(a) the Schrödinger equation
$$H|\psi\rangle = i\hbar \frac{\partial |\psi\rangle}{\partial t} \tag{9.2 }$$

(b) the guiding equation
$$\frac{dQ_k(t)}{dt} = \frac{\hbar}{m_k} Im \frac{\psi^* \partial k \psi}{\psi^* \psi} \ (Q_1, \dots, Q_N) \tag{9.3 }$$

Whereas the Schrödinger equation governs ψ's behavior, the guiding equation rules the evolution of the particles' positions. Furthermore, as the second equation shows, it provides the particles' velocities at any t as a function of the wave-function at any t, its spatial derivative and the particles' position at t.

Some remarks are in order. To begin with, it's easy to see that the guiding equation represents a first-order "velocity" formula[45] involving the wave function whose role is to bring

[45] Bohmian mechanics typically introduces its guiding equation as a first-order formula at present, in which velocity is taken to be *fundamental*. But this is not the way in which Bohm wrote down the dynamics. He chose to write it down as a second-order equation, featuring a *quantum potential*. In that version of the theory, particles move around thanks to forces stemming from that quantum potential. In any case, there is an underdetermination problem lurking over: Bohmian explanations don't require the guiding equation to be exactly as it is (either in the Bohm's fashion or Dürr, Goldstein, and Zanghi's). Actually, there are many other candidates for guidance equations working just as well as the "official" one, somehow implying that there are many other candidates for

in a "velocity field" along which particles move. In this sense, the wave function "guides" (or "choreographs", see Dürr, Goldstein and Zanghi 2013: 83) the motion of the particles. Secondly, both equations are perfectly deterministic, meaning that initial conditions determine a unique state at later times. Thirdly, as particles always have well-defined trajectories, they have determinate velocities at all times as well. It follows from this that particles have positions *and* velocities at all times, being seemingly at odds with some SQM's principles (why this doesn't violate Heisenberg's principle, see Dürr, Goldstein, and Zanghi 1992: Section 11, and Bricmont 2016: 159-161).

Fourthly, since the theory is strictly and fully deterministic, any randomness enters it through the initial conditions and arises from averaging over ignorance. For Bohmians, configurations are random in this sense, whose distribution is given by the *quantum equilibrium distribution*, $|\psi(q)|^2$. Fifthly, Bohmian mechanics is a non-collapse quantum theory: collapses occur neither spontaneously nor induced by measurement or observers. The theory nonetheless accepts a sort of "collapse" mechanism, but it should rather be regarded merely as a "pragmatic affair" (Goldstein 2017) or as a "collapse in practice" (for details, see Dürr, Goldstein, and Zanghi 2004). In any case, this pragmatic mechanism doesn't play any fundamental role in the theory, neither must it be considered as a real physical process.

Finally, the status of ψ is matter of some controversies among Bohmians. On the one hand, many have considered it like a *field*. But it's a field without any source and not defined in the physical space $\mathbb{R}^3$ but in the configuration space $\mathbb{R}^{3N}$ (for instance, for a two-particle system, its wave function will be defined in a $\mathbb{R}^3 \times \mathbb{R}^3$ configuration space). Since ψ is a field, it behaves wavily according to the Schrödinger equation. And as ψ's behavior relates to particles' behavior, any alteration of its wavy behavior will too affect the particles' trajectories. Notably, this "affectation" is not produced by any force and occurs, so to speak, instantaneously regardless distances. Stronger positions along this line have considered ψ as real as the classical electromagnetic field. John Bell (1987) famously holds this viewpoint with respect to the wave function. In a widely-known passage, he claims:

> "Note that in this compound dynamical system the wave is supposed to be just as 'real' and 'objective' as say the fields of classical Maxwell theory [...] *No one can understand this theory until he is willing to think of as a real objective field rather than just a 'probability amplitude.*" (Bell 1987: 128; italics in the original)

a Bohm-type theory working just as well as the official one. See Stone 1994, Dürr, Goldstein, and Zangui 1996, Deotto and Ghirardi 1998, Peruzzi and Rimini 2000, and Skow 2010 for details.

David Bohm also seems to have held the same view in claiming that the wave function of an individual electron is "a mathematical representation of an objectively real field", which exerts a force on the particles similar to that exerted by an electromagnetic field on a charge (Bohm 1952: 170). Tim Maudlin (2007) also endorses this ontic Bohm-Bell view of the wave function in BM at present.

On the other hand, some Bohmians have rather promoted the idea that the wave-function is nomological (or quasi-nomological). Dürr, Goldstein, and Zanghi (1997, 2013) endorse this view claiming that the wave function (a mathematical representation) plays a nomic role in the particles' motion, so it should in fact be considered as a highly-relevant piece of information, but not as representing a wavy stuff out there in the world. In Goldstein's words:

> "The fact that Bohmian mechanics requires that one take such an unfamiliar sort of entity seriously bothers a lot of people. It does not bother us all that much, but it does seem like a significant piece of information. What it suggests to us is that you should think of the wave-function as describing a law, not as some sort of concrete physical reality." (Goldstein 2013: 97)

This stance should yet be considered through two angles: what is truly nomological is the timeless wave function of the Universe (**ψ**), whereas temporal worldly wave-functions should be rather regarded as *quasi*-nomological (see Goldstein 2013: 107).

A more minimalist proposal has also been put forward by the so-called Quantum Humeanism (Esfeld 2014). Under a PO approach applicable both to GRWm, GRWf and BM, Esfeld claims that the wave function doesn't belong to the quantum ontology at all. To be clear: it's not the case that the wave function belongs to the ontology but not to the PO, as other have argued. The point is that it doesn't belong to the ontology whatsoever. Curiously, one of the sources for Quantum Humeanism is Bell, who seemed to support the reality of wave function as quoted above. Esfeld counters such a passage with the following one from Bell (1987: 53)

> "One of the apparent non-localities of quantum mechanics is the instantaneous, over all space, 'collapse of the wave function' on 'measurement'. But this does not bother us if we do not grant beable status to the wave function. We can regard it simply as a convenient but inessential mathematical device for formulating correlations between experimental procedures and experimental results, i.e., between one set of beables and another." (Bell 1987: 53)

In contrast, Quantum Humeanists proposes a minimalist ontology where there are fundamentally only point-like particles without natural properties (the so-called primitive stuff), individualized by their spatio-temporal relations. All the rest, properties like mass, energy or being a lepton, is grounded on, and "emerges from", the dynamics.

In the light of what was said so far, it is easy to see that the best candidate for a PO of BM is one of *particles* and their respective *positions* (for other alternatives, see Allori et al. 2008. For discussion on ontological issues around BM, see Esfeld et al. 2013). The wave-function can be also regarded as part of the BM's ontology if one endorses an ontic Bohm-Bell view, though it would hardly be taken as part of the PO for it doesn't inhabit a three-dimensional physical space. Taking a nomological view should also endorse this view inasmuch as laws are not, typically, part of a theory's ontology (for a different view, see Maudlin 2007. He considers that laws are *sui generis* entities and, thereby, might be part of a theory's ontology). In any case, it's clear that particles and their respective positions are the right sort of stuff that can play the role of "local beables" in Bell's sense (1987).

Therefore, the BM's structure (at least as typically understood at present) should include:

(a) $(\psi(t), Q(t)$, as the state of a quantum system *S.*
(b) The SQM's formalism (including naturally the Schrödinger equation).
(c) The guiding equation.
(d) A PO of *particles* and their respective *positions.*

Let's now take a look at BM in action. Consider a two-slit experiment consisting of a source that fires electrons toward a barrier with two slits in it. The electrons get eventually detected on a screen behind the barrier. One well knows how the orthodox story about two-slit-like experiments overall goes. The story that Bohmians tell us is nonetheless quite different. To begin with, remember that BM assumes that electrons *always* have definite positions, so it's a matter of fact that each electron fired by the source and reaching the screen goes through *either* the upper slit *or* the lower one. Furthermore, the theory is invariably deterministic, so each electron hitting the screen has followed a well-defined trajectory completely determined by its initial conditions (that is, by its initial position *and* wave function). Remarkably, slight variations in the initial position will lead to slight variations in the hits on the screen, even if the initial wave function does not vary at all.

The two Bohmian equations govern the electron's behavior through this experiment as follows. On the one hand, the initial wave-function will evolve according to the Schrödinger equation, traveling through both slits and "causing" the known interference pattern. On the other, particles' final positions (the hits on the screen) will exclusively depend on particles' initial positions. Hence, a single electron being fired out by the source will go only through one of the slits, following a deterministic trajectory. However, the particles' behavior is ruled by the guiding equation: assuming that the frequency with which a particle starts out in some region R_1 near the source is equal to $\int_{R1}|\psi(x,t_0)|^2$, the guidance equation shows that the frequency with which an electron can be found in a region R_2 near the screen is equal to $\int_{R2}|\psi(x,t)|^2$. As particles are somewhat guided by the wave-function as well, the final hits on the screen will then reflect the interference pattern. This entails that, on the one hand, particles traveling through the lower slit will be affected by that part of the wave function going through the upper slit and vice versa; on the other, as particles' trajectories depend on the wave function, these bend in a non-Newtonian way (for there are no external forces acting upon particles bending their trajectories). Notably, though particles' trajectories may bend, they don't cross one another, so one can as a matter of fact infer which slit each particle passed through from its final position on the screen.

Section 2. The Arrow of Time and BM: an asymmetry in the initial conditions

The question of the arrow of time within BM boils down to the question about whether any of the above-listed elements manifests a temporal asymmetry (either structural or non-structural). It's often assumed that BM is paradigmatically a *structural time symmetric* quantum theory. This is likely due to, on the one hand, the allegedly time symmetric nature of its dynamics and, on the other, to the fact that the theory only deals with deterministic trajectories. The fact that BM is a non-collapse theory also feeds this idea: as already shown, collapse theories are often assumed to support a structural time asymmetry. In any case, as long as BM makes the *same* predictions as SQM-CP does and brings about the same time asymmetric predictions as SQM-CP does (for example, the sort of conditional probabilities expected in an experiment are not equal in backward direction of time and in the forward direction of time), BM should provide some explanation about them. A collapse mechanism is a no longer available choice, so there should be an alternative explanation. And this gives rise to the problem of the two realms as presented before: provided that the Schrödinger equation (I've been assuming this from the beginning of this Part 3) and the guidance equation (which I'll examine more closely afterwards) are time-reversal invariant, the question is: *where* do these temporal asymmetries come from?

Let's start examining a simple case. Consider the sort of stories that Bohmians tell about a Mach-Zehnder-type experiment. Imagine two sources of electrons (S1 and S2) and two electron detectors after each source (D1 and D2). When an electron is fired out by S1, the electron detector D1 will emit a flash of light. And the same goes for S2 and D2. Both detectors let each electron to pass through, when gets it detected, towards a half-silvered mirror placed in the middle of the experimental arrangement. When an electron reaches the half-silvered mirror, it has 50% of chances of bouncing off and 50% of chances of passing through it. At the other extreme of the experimental arrangement, there are two electron detectors (F1 and F2), which emit a flash of light when an electron has either been reflected by the half-silvered mirror or when pass through it (see Fig. 9.1)

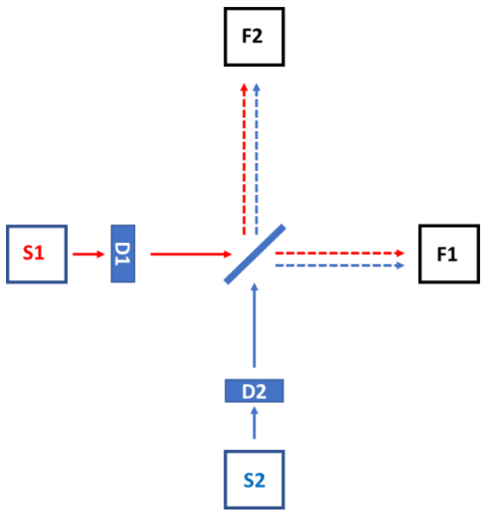

Fig 9.1 (adapted from Arntzenius 1997)

According to the standard formalism (SQM-CP), there are some FORWARD claims (as "what is the probability that F1 detects given that D2 detected?") regarding certain conditional probabilities of some *later* state given an *earlier* state that are true, whereas their BACKWARD counterpart ("what is the probability that D2 detects given that F1 detected?") aren't. Such an asymmetry, within SQM+CP, came from the dynamics in itself as the quantum state underwent an induced collapse when measured or observed (this was, for instance, Penrose's argument). BM should be able to only make true the same sort of FORWARD claims, but not their BACKWARD counterpart. Otherwise, as explained before, the theory would be plainly wrong.

Putting aside my analysis therein (and my caveats for considering this FORWARD-BACKWARD asymmetry as genuine), let's see a bit closer how it looks from a Bohmian perspective. One possible FORWARD question can be posed as following: what is the probability of getting an electron detected at F2 given that it was detected by D1? As BM yields the same frequencies as SQM, the answer will be 'a half'. But the story told about it varies significantly.

Consider first the future-headed case. At t_0, an electron is fired by S1. Its quantum state will be given by a wave-function $|\psi(t_0)\rangle_e$, whose value is not zero in a region covering the vicinity of the source. This wave function can be written down as a superposition of "detected by F1" ("$|F1\rangle$") and "detected by F2" ("$|F2\rangle$")

$$|\psi(t_0)\rangle_e = \sqrt{\frac{1}{2}}(|F1\rangle + |F2\rangle) \qquad (9.4\ )$$

Yet, as a matter of fact, the initial state is not only determined by the electron's wave function at t_0, but also by its position (q_i). According to BM, the final position of the electron, that is, whether it will get eventually detected by either F1 or F2, depends exclusively on which its exact position at $t_0$was. For simplicity, imagine that this may be either slightly on the right side or slightly on the left side within the region where the wave function is non-zero. It happens that if the electron started off on the left side (let's say, whether it has the rough position q_L), it will surely be reflected by the half-silvered mirror, and that if it rather started off on the right side (let's say whether it has the rough position q_R), it will surely get through the half-silvered mirror. One, of course, doesn't know exactly which the position of the electron at t_0 is but, as a matter of fact, it is in either of them. What one knows is the initial wave function of the electron and the frequency with which the electron begins in some either region, which is equal to $\int_q \ |\psi(q, t_0|^2$.

So, this is how the possible sequences of the experiments may run. The electron started off its trip being fired by S1 at t_0. Next, it gets detected by the first electron detector, D1. When it reaches the half-silvered mirror, it can either be reflected or pass through it. BM's predictions say that roughly half electrons will be reflected, and the other half will pass through. Interestingly, when the electron reaches the half-silvered mirror, its wave function splits into two regions where it is non-zero: one branch is reflected and the other is let passed through. Hence, only one of these branches (either $|F1\rangle$ or $|F2\rangle$) will "follow" the actual trajectory that the electron went along after reaching the half-silvered mirror. Finally, the electron is detected by, say, the detector F2 (what entails that it was reflected by the half-silvered mirror). And further, from here one can straightforwardly infer that the initial position of the electron was somewhere slightly on the left side of the wave function's original region: those electrons reflected started off their trips on the left side of the S1's region at t_0, whereas the rest started off on the right side of the S1's region at t_0. It is worth remembering that there is *no* collapse of the wave-function in BM, so the electron's wave function will still be in a superposition after being reflected or letting passed through. The point is that only one of its branches will remain close to the actual electron's position.

This produces the following temporal sequences

Future-headed sequences $S1\ e(q_R) \rightarrow D1 \rightarrow M \rightarrow\ F1\ /\ S1\ e(q_L) \rightarrow D1 \rightarrow M \rightarrow F2$

And, the usual FORWARD transitions probabilities. For instance,

FORWARD "the probability that F2 detects given that D1 detected is a half"

It is worth stressing that these probabilities are based on agent's ignorance about the initial position of the electron. If one knew the initial conditions, say the electron's initial position is $e(q_L)$, then the right sequence would be $S1\ e(q_L) \rightarrow D1 \rightarrow M \rightarrow F2$.

It's easy to see that this Bohmian story doesn't remain invariant when told backward in time. To put it simply, BM, as well as SQM, renders invariant forward transitions chances but non-invariant backward transitions chances. Setting aside any caveats concerning whether time is being reversed adequately, it is easy to see that an electron travelling backward from F2 has only a half of chances of being detected by D1, when in the forward case it was one. Furthermore, in the forward case in BM one assumed that the initial wave-function was around the source S1, and in a superposition of $\sqrt{\frac{1}{2}}(|F1\rangle + |F2\rangle)$. Now, in the backward case one starts with a precise position at F2 (either slightly on the left side or slightly on the right side), and a different wave-function being in a superposition of $\sqrt{\frac{1}{2}}(|D1\rangle + |D2\rangle)$. So, one ends up with the following (backward) sequence and BACKWARD claim

Past-headed sequences $F2\ e(q_R) \rightarrow M \rightarrow D1\ or\ F2\ e(q_L) \rightarrow M \rightarrow D2$

BACKWARD "the probability that D1 detects given that F2 detected is one half"

The crucial question at this point is: What is the source of this asymmetry between FORWARD and BACKWARD within BM? If BM is a quantum theory about particles and their trajectories, and each particles' history is invariably and completely determined by the Bohmian laws of motion and particles' initial state, the source of the asymmetry between FORWARD and BACKWAR must come from either of them. Moreover, the sort of asymmetry one gets exclusively depends on the source of it: an asymmetry at the level of the dynamics governing particles' behavior will yield a structural asymmetry. But, if the asymmetry will rather depend only on non-dynamics ingredients, then the asymmetry will be non-structural.

It was detailed above that BM involves two dynamical laws. Begins by considering whether the Schrödinger equation is time-reversal invariant. Clearly, if one gets to the conclusion that the Schrödinger equation is non-time-reversal invariant, then so is BM altogether. One should then take the charge of explaining whether the temporal asymmetry between FORWARD-BACKWARD come out from there, but in principle, and theoretically, the theory would indeed treat the past-to-future direction and the future-to-past direction differently. More interesting is the case in which the Schrödinger equation is time-reversal invariant. It is often taken that the guidance equation is time-reversal invariant if the Schrödinger equation is (see Callender 2000: 260, for instance). It follows from this that if one assumes for a moment that the Schrödinger equation is time-reversal invariant, then the guidance equation is time-reversal invariant as well. In this way, the whole Bohmian dynamics turns out to be fully time-reversal invariant, so the source of the aforementioned temporal asymmetry must come out from elsewhere[46].

[46] It would be interesting to evaluate this argumentation from the *Causal Symmetric Bohm Model* proposed by Roderick Sutherland (2008). The overall aim of Sutherland's model is to deal with Bell's theorem and to provide a natural explanation of Bell non-locality in a three-dimensional space. And that's why he proposes a Bohm model involving retrocausaton. Though Sutherland himself separates the notion of causal symmetry from the notion of time symmetry, his model introduces the notion of *backward causation*, which has been related to the direction of time in other approaches. To build the causal symmetric model, he proposes that the initial wave-function ψ_i (which specifies the initial boundary conditions) should be complemented with a *final* wave-function ψ_f (which specifies the *final* boundary conditions). Retro causation comes in the picture because, while ψ_i influences future events, ψ_f influences past events. Importantly, ψ_f is independent of ψ_i and at any time there are two wave-functions, ψ_i and ψ_f. What would be interesting to look at is (a) whether ψ_f makes true the same probability predictions as BACKWARD, (b) what is the relation between ψ_f and the time-reversed ψ_i.

The only option left at this point is that such a source be the initial state of the particle. The view that in BM the FORWARD-BACKWARD asymmetry is due to an asymmetry in the *initial conditions* has already been pointed out elsewhere. Frank Arntzenius (1997), for instance, claims that a *constraint on the initial conditions* is what implies the asymmetry that is to be accounted for in BM. The constraint consists in assuming that one initially had a wave function roughly near to the region S1 and D1, and in a superposition of being detected by F1 and F2. And when taking the scenario backward, one assumed a *different* initial (final) condition, starting with a wave-function roughly near to F2 and in a superposition of D1 and D2. This latter assumption is what explains the usual frequencies obtained at the lab. Arntzenius puts it in this way:

> "This is an assumption that we always make in Bohmian theories, and rarely even notice. Of course, it is a good thing that we do this, for if we did the opposite, we would have an empirically inadequate theory, for it would predict non-invariant forward transitions frequencies and invariant backward one, and that is just wrong" (Arntzenius 1997: S219)

The same is held by Callender (2000: 260).

And all this leads us to the following: there would be no reason to conclude that time *in itself* is temporally asymmetric in BM according to this argument. An asymmetry in the initial conditions of an electron is not the right kind of source capable of *producing* a structural arrow of time. This doesn't mean that BM is a timeless theory or that the theory is unable to draw any distinction between the two directions of time. It looks rather like one is dealing with a case of a *non-structural distinction* between the past-to-future and the future-to-past directions, which reduces to a non-structural distinction between forward transition probabilities and backward transitions probabilities. Though the FORWARD and the BACKWARD stories are equally possible according to BM's dynamics, and thereby the structure of BM's solutions is like $W = W^f + W^b$, a constraint on the initial conditions is required so as to pick up the right sort of solutions that makes true the right sort of transition probabilities experimentally obtained within *this* world. Each set of solutions instantiates unalike extrinsic properties, particularly, each of them starts off from *different* initial conditions, on which the observed asymmetry can be grounded. Or, to put it differently, each of them *is conditionalized* over different initial conditions. In this sense, the true of both FORWARD (or the future-headed sequence) and BACKWARD (or the past-headed sequence) must be related to the initial conditions one starts with, specifying particles' position and initial wave-function adequately. For all this, BM renders a *non-structural* arrow of time, which is relative to a constraint on the initial conditions.

Section 3. Time reversal and its special role in BM

The idea that time is not structurally asymmetric according to BM holds because the Bohmian dynamics is fully time-reversal invariant. It was also mentioned in passing before that, in general, the guidance equation is time-reversal invariant *if* the Schrödinger equation is time-reversal invariant. This is a natural presumption since the guidance equation requires the wave function in order to describe the evolution of the particles' positions. As wave function's behavior is ruled by the Schrödinger equation, and it is reasonable to think that if it remains invariant under time reversal, so will also be the guidance equation.

Importantly, the notion of time reversal has played a highly relevant role in building up a standard version of BM. In fn. 45, I noted that there are many workable candidates for the guidance equation, leading to the fact that there are thus many BM-like theories. If BM is an attempt to solve the measurement problem that requires the guidance equation, why shouldn't any of the alternative BM-like theories be taken instead? Dürr, Goldstein and Zanghi have put forward an argument for favoring BM and the official guidance equation instead of the other candidates. In a nutshell, the argument is based on the "theoretical virtues" of BM, in particular, that BM, as it is presented by them, and involving the guidance equation as is exactly introduced by them, is the simplest BM-like theory endowed with the right space-time invariances.

I won't get into details about this problem of the underdetermination of the guidance equation and about why one should prefer BM instead of any of the other BM-like candidates. The point I want to mention here is that in the argument used for this, Dürr, Goldstein and Zangui make a few noteworthy claims on time reversal within BM. That is, in *justifying* their version of the guidance equation and why this is the simplest Galilean-invariant *and* time-reversal invariant version, Dürr, Goldstein and Zangui make some noteworthy assumptions about time reversal and time-reversal invariance.

The justification runs as follows. The aim is to find the simplest Galilean-invariant version of a guidance equation. The equation mainly aims to explain how the particles move. And the simplest possibility is that it only prescribes velocities instead of higher derivatives (Dürr and Teufel 2009: 147). So, the starting point will consist in the wave-function determining a *velocity vector field* $v_1^\psi, \ldots\ v_N^\psi$ on configuration space for all the particles. This is how the wave-function will govern the evolution of particles' positions.

Next, they assume two things:

(a) on the one hand, that any candidate law has to guarantee that the particle's velocity at t depends on the wave-function, its spatial derivative at t, the particle's mass and $\hbar$;

(b) on the other hand, that such a candidate law must be Galilean invariant and time-reversal invariant. The former requirement follows from assuming that a Galilean space-time is the right setting for any BM-like theory, so that whatever the form of the guidance equation be, it must have a Galilean structure to fit into such setting[47].

In order to accomplish these conditions and for the equation to make sense, one needs to characterize the vector field $\boldsymbol{v}^\psi$ carefully. As the sought equation ought to be Galilean invariant, it should have the following form as the simplest one

$$\boldsymbol{v}^\psi = \alpha \frac{\nabla\psi}{\psi} \tag{9.5 }$$

Where α stands for $\frac{\hbar}{m}$ and the use of $\nabla\psi$ comes from the fact that the equation must be invariant under rotation, if it's Galilean invariant.

However, this equation could be either real or imaginary. Whether it is to be real or imaginary is something that Dürr, Goldstein and Zanghi leave to time reversal to decide. And that's why time reversal plays a special role within BM: Since the left side of the equation (9.5) represents a velocity, it will flip sign under time reversal. It follows because if $\boldsymbol{v}^\psi$ is roughly something like $\frac{dQ}{dt}$, a time-reversal operation that, at least, transforms t into $-t$ will transform $\frac{dQ}{dt}$ into $-\frac{dQ}{dt}$. If time-reversal invariance must hold (as Dürr, Goldstein and Zanghi assume), then the right side of the equation ought to transform in such a way that also produces a minus sign. How will the right side of the equation transform under time reversal? By *assuming* that the right time-reversal operator also takes the complex conjugate over the wave-function, transforming $T: \psi(\boldsymbol{q}, t) \rightarrow \psi^*(\boldsymbol{q}, -t)$, one obtains that the real part transforms as

[47] See Valentini 1997 for a different argument concluding that BM's underlying space-time is not Galilean but Aristotelian.

$$TRe\left(\alpha\frac{\nabla\psi}{\psi}\right) \to Re\left(\alpha\frac{\nabla\psi^*}{\psi^*}\right) \tag{9.6 }$$

while the imaginary part transforms as

$$TIm\left(\alpha\frac{\nabla\psi}{\psi}\right) \to -Im\left(\alpha\frac{\nabla\psi^*}{\psi^*}\right) \tag{9.7 }$$

Therefore, if time-reversal invariance is demanded, one should choose the imaginary part because is the only way of keeping the theory time-reversal invariant.

$$\boldsymbol{Tv}^{\psi} = \alpha T\mathrm{Im}\frac{\mathrm{T}(\nabla\psi)}{T(\psi)} \to \boldsymbol{-v}^{\psi} = \alpha - \mathrm{Im}\frac{\nabla\psi^*}{\psi^*} \tag{9.9 }$$

And this contributes to accept that the simplest version of the guidance equation is that proposed by BM.

To put it clearly, the form of the argument is as follows:

1. The underlying space-time is Galilean (assumption) and, in accordance with this, the guiding equation must be Galilean invariant (assumption).
2. From 1, if the guiding equation is Galilean invariant (assumption), it must have the form of eq. (9.5), which still leave open whether it takes either the real or the imaginary part.
3. The Schrödinger equation must be time-reversal invariant (assumption)
4. If the Schrödinger equation is time-reversal invariant, then so is the guidance equation
5. The Schrödinger equation is time-reversal invariant if and only if T_A formally represents time reversal (what was showed in Part 2)
6. From 4. And 5., the guidance equation must be invariant under the application of T_A
7. In order to accomplish 6., the simplest form of the guidance equation is that proposed by Dürr, Goldstein and Zangui (eq. 9.3).

The first premise is assumed without any further justification. In general, physics textbooks prescribe that the underlying space-time of any non-relativistic quantum theory is Galilean and, consequently, that the theory must be Galilean-invariant. Notwithstanding this, some interpretations of SQM might lead to a violation of such a prescription: as mentioned in a previous chapter, Valentini thinks that Bohmian Mechanics' underlying space-time is

Aristotelian –rather than Galilean–, and wave function realism, which conceives the wave-function as a scalar field, makes the theory non-Galilean-invariant as it is non-invariant under Galilean boosts (see Allori 2018). In any case, as the above-sketched argument shows, there is no way to get to this *particular* form of the guidance equation without assuming Galilean invariance. This leaves one with the following question: on which grounds does one justify that the underlying space-time is Galilean, and consequently, that the theory must be Galilean invariant?

The third premise is equally troublesome. As shown in Part 2, the assumption can be seriously challenged by resting upon a different view on symmetries: one could claim that symmetries shouldn't be a priori prescribed, but they should rather be considered as contingent properties of the dynamics one discovers a posteriori by dynamically-independent means. Clearly, if the Schrödinger equation is not supposed to be time-reversal invariant, then T_A is unmotivated as a genuine formal representation of time-reversal invariance (Premise 5). This would break the whole argument down as there wouldn't be, at first glance, any reason for choosing the imaginary part instead of the real part for eq. 9.5. In some sense, the argumentation seems to be running in circles. One desires to justify BM's dynamics (particularly, that the guidance equation is precisely as BM says that it is). For doing this, one relies on a symmetry argument prescribing, on the one hand, that a dynamical equation must remain time-reversal invariant and, on the other, that certain formal objects (as wave functions) must transform in a certain way. For this, the time-reversal transformation must be T_A rather than any other one. But the justification for taking T_A rather than any other time-reversal transformation (as T_U) depends upon prior facts about the dynamics one wanted to justify, in particular, that such a dynamic meets time-reversal invariance.

Let me be more precise about what is, at first sight, the problem with the argument. BM's defenders employ a "symmetry argument" to justify the official guidance equation. But the argument crucially rests on assumptions about the symmetry transformation that strongly depends on the dynamics one wants to justify in the first place. Bradford Skow (2010) runs a similar argument in saying that Dürr, Goldstein and Zanghi's strategy is ill-conceived since it fails to be "independently grounded", which means that it strongly ties up the form of the symmetry operator to the theory's dynamics to be justified. In the case of time reversal, he says:

> "But which operation is the time reversal operation is not so up for grabs. To determine whether a given equation is time reversal invariant, we have to first

> identify, in some independent way, how time reversal (and other symmetries) act on states, and only then check whether the equation is invariant under those symmetries. And what is true here about time reversal is also true for boosts" (Skow 2010: 405)

There are two ways out coming to my mind. The first consists in arguing that one actually wants to justify a particular formulation of the guidance equation (one of the elements of BM's dynamics) through time-reversal invariance, but that the justification of the time-reversal transformation (T_A) doesn't depend on any assumption about the guidance equation but on assumptions about the Schrödinger equation (the other element of BM's dynamics). Hence, any motivation for T_A is independently grounded in so far as it doesn't rely on the part of the dynamics one wanted to justify.

The other way out comes about from the fact that *only if* one takes the imaginary part one gets an empirically adequate formulation of BM. In this way, the justification of this version of the guidance equation ultimately arises from empirically-grounded reasons, avoiding circularity. This latter way out could be indeed a very good strategy to break the circle and to justify why one must choose the imaginary part for eq. 9.5, and in consequence, providing some indirect support for choosing T_A. This, of course, should be firstly investigated in deep within a BM framework; but, anyway, I'm not so sure that it succeeds in justifying T_A in any relevant sense. The reason is the same analyzed before: any successful justification of T_A *crucially* depends upon assuming time-reversal invariance beforehand. So empirical successfulness is not enough.

I think the former way out doesn't succeed either. Though it's true that the justification for T_A ultimately relies on the time-reversal invariance of the Schrödinger equation, and on how the wave function transforms under it, the guidance equation is so closely-tied to them that both laws cannot really be considered independent from one another. This manifests clearly in the Premise 4, which says that the guidance equation is time-reversal invariant *if* the Schrödinger equation is time-reversal invariant. But, by assuming that the Schrödinger equation is time reversal invariant and by grounding T_A on the fact that it's the unique time-reversal operator that leaves the equation invariant, there is no much more to say about the guidance equation: this *logically* entails that guidance equation is time-reversal invariant as well (for the conditional). So, there is no much room for any independence here.

But, besides this, the guidance equation is also assumed to be time-reversal invariant! (see Dürr, Goldstein and Zanghi 1992, for instance). Putting aside any justification of T_A based on the Schrödinger equation, taking T_A as the right formal representation of time reversal seems a bit unmotivated beyond the fact that it's the transformation that leaves *that version* of the guidance equation invariant. So, if the Schrödinger equation is putting aside, it seems that the form of the time-reversal operator depends on assuming that the guidance equation must be invariant under so-defined time-reversal operator, which at the same time justifies the form of the equation that motivates it. And this looks very much like a petitio principii.

Now look at this a bit differently. They say that the Schrödinger equation is time-reversal invariant because it is T_A-invariant. This already *implies* the time-reversal invariance of the guidance equation. But *which* guidance equation? Well, the one justified by using T_A for reversing time. Could it be differently? It couldn't, because the whole reasoning is grounded on prior facts on time reversal that logically imply, on the one side, that the guidance equation is time-reversal invariant and, on the other, that *that* is the right way to write the guidance equation down. This circular-like reasoning is not bad per se, and in fact it might look quite appealing so as to get an elegant and simple theory as BM intends to be. But I think one should be careful when doing this within a symmetry argument: there seems not to be any independent reason that yields T_A as the right time-reversal operator without relying either on the Schrödinger equation or on the fact that it helps to construct BM in its official looking.

There are two further points I'd like to bring out. First, the strategy runs the risk of trivializing the entire task of posing a symmetry argument. The point is that one can always come up with a symmetry transformation that changes the adequate variables, that transforms the state in such a precise way that keeps a given equation invariant. And if this at the same time logically implies the symmetry of the equation under construction, the entire task seems to be unwarranted. I have already pointed this out several times previously, and BM is here just incurring in the same sin, so to speak.

Second, this leads one nowhere when discussing about a *structural* arrow of time within BM. The Bohmian dynamics comes out time-reversal invariant because T_A. But this decision is explicitly grounded on the assumption that the Schrödinger equation, *and* the guidance equation consequently, *must* be time-reversal invariant. As mentioned above, this premise appears both in the justification of T_A and in the very process of constructing the theory. But there are no independent grounds to suppose that the theory must be time-reversal invariant simpliciter. One could well arrive at the conclusion that the BM's underlying space-time is a Galilean space-time, and thereby, that BM's dynamics should be built accordingly. But time-reversal invariance doesn't belong to the set of symmetries of a Galilean space-time. One's metaphysics could indeed play the role of grounding such an assumption, but then the right way to pose the argument should make this explicit. And, of course, the theory shouldn't be time-reversal invariant simpliciter, but in the light, and conditionalized over, some previously-taken metaphysical assumptions.

Section 4. Relationalist time symmetry and BM's PO

As I've argued along Part 2, one's metaphysics of time crucially affects the way in which one is to formally characterize the notion of time reversal. A *substantivalist stance* would naturally yield the Schrödinger equation non-time-reversal invariant, and thereby, a non-time-reversal invariant BM (see Albert 2000 and Callender 2000 for akin claims). This doesn't automatically imply that the guidance equation (whatever its form comes to be) is also time-reversal invariant for the substantivalist, though one is somewhat left clueless about how it should then transform under time reversal. Beyond this, it is easy to see that if one by any reason arrives at the conclusion that the Schrödinger equation is non-time-reversal invariant, then so is BM regardless whether the guidance equation is.

More interesting is the case of a relationalist stance with respect to time reversal, since it must dig into the "geographical" details of the theory to figure out *how* time reversal implements motion reversal. As mentioned early, BM is a quantum theory about *particles* and *their positions*, where wave function's role is, at least in the version of Dürr, Goldstein, and Zangui, preeminently nomological. So, a relationalist justification of how time reversal might be implemented within BM could reasonably consist in drawing attention to its PO. Interestingly, this would be an alternative justification of that presented above: it would mostly rely on the ontological aspects of the theory instead of purely on its dynamics.

Bryan Roberts (2010) has presented a rationale along this path in the framework of BM. He begins by pointing to the twofold dynamics of BM and to the fact that the guidance equation is the law that truly governs the stuff of the world. That is, the law that truly governs the PO of BM: particles and their trajectories. As a consequence of this, the relevant notion of time reversal for Bohmians is the one related to the guidance equation and to the stuff governed by it and not to the Schrödinger equation. This seems to suggest that the form of the "Bohmian time-reversal operator" should be worked out from how it should act upon the PO of BM. In fact, Roberts claims that a Bohmian time-reversal operator is demanded to act fundamentally on the positions living in a Bohmian configuration space, since this is its PO. In accordance with this, a Bohmian time-reversal operator is merely a bijection from a particle's position to a particle's position plus a reparameterization of $t \to -t$.

But, why is one entitled to call this transformation 'time reversal'? At first glance, this Bohmian time-reversal operator is just an "empty" characterization of what time reversal is supposed to do. A *time-reversal* operator cannot be such a bijection from particles' positions to particles' positions, if one wants 'time reversal' to preserve any relevant meaning. One is not whereby reversing anything dynamically relevant by mapping particles' positions. Furthermore, position is exactly the right sort of thing one expects to be left unaltered under time reversal. Though the value of position changes with time, position in itself is not a time-derivative magnitude: it's primitive. Therefore, appealing to the PO to set out the form of a Bohmian time-reversal operator seems to be superfluous: the ontology on its own is unable to give any information about how time reversal must be implemented. And this is for the simple fact that time has not even entered the picture yet: the ontology on its own remains changeless in time if a dynamic is not beforehand introduced in the picture (even more, from a strict relationalist stance, there is no even anything like time in such a static picture!). And, as I've argued before, a *relationalist* time-reversal operator should aim to generate the right sort of transformations that reverses physical systems' *motion*. And nothing like this can be ever reached by acting upon particles' positions, without determining how time reversal is to act upon the dynamics.

Roberts seems to note it in mentioning that there is a subtlety to deal with in determining the form of the Bohmian time-reversal operator. And the subtlety is precisely that the guidance equation also depends on the wave function. However, I don't think that this be a subtlety at all but quite the opposite: it's what is crucial to get some grasp on how a relationalist time(motion)-reversal operator must be formally implemented within BM. As many times mentioned, particles' trajectories are governed by the wave function (as it appears within the guidance equation), which determines particles' velocities. So, *these* are the relevant dynamical variables to reverse if one wants to get a backward-headed motion.

In dealing with the "subtlety", Roberts gives a rationale about how a Bohmian time-reversal operator should be further characterized. In contrast with Dürr, Goldstein, and Zangui's characterization, he relies on three conditions, seemingly independent of SQM:

(a) The Bohmian time-reversal operator is an involution in configuration space

(b) The action of the Bohmian time-reversal operator is such that if φ_{x0} is a position eigenfunction, then so is its image under the action of the Bohmian time-reversal operator.

(c) The Bohmian guidance equation is invariant under the Bohmian time-reversal operator for the free particle Hamiltonian.

The first condition attempts to capture the idea that the Bohmian time-reversal operator intends to reverse *motion*. The second one is required because, in BM, a symmetry operator should give special treatment to those vectors forming the position basis. The last condition seems to be a bit poorly justified, but let's set it aside for the time being. Thus-specified Bohmian time-reversal operator, though not acting upon velocities directly, allows one to figure out how velocities should transform under it. As positions remain unchanged

$$T\colon q \to q \tag{9.10 }$$

and time changes sign under time reversal,

$$T\colon t \to -t \tag{9.11 }$$

one gets that

$$\frac{dq}{dt} \to \frac{dTq}{dTt} = -\frac{dq}{dt} \tag{9.12 }$$

This is helpful to find out how the left side on the guiding equation transform under the Bohmian time-reversal operator, but what about the right side of the equation where a wave function appears? The point I want to make is that if one doesn't assume that the Bohmian time-reversal operator transforms the wave function by taking the complex conjugation as $T\colon \psi(\boldsymbol{q},t) \to \psi^*(\boldsymbol{q},-t)$, there is no way to get the guidance equation invariant. But the justification of why a Bohmian time-reversal operator must transform the wave-function into its complex conjugate mainly lies on the spectrum condition (see Chapter 4, Section 2.2), that is, within SQM's principles. In fact, this is somehow hidden in Robert's condition (c): so-specified Bohamian time-reversal operator says nothing about how the Hamiltonian should transform. So, when Roberts demands the guidance equation to keep the free particle

Hamiltonian invariant, this can be only achieved by introducing further requirements coming from SQM[48].

Therefore, it is not quite right to claim that this approach to a Bohmian time-reversal operator "doesn't presuppose any facts about time reversal in ordinary quantum mechanics" because it does it by specifying that time reversal within BM should turn the wave-function into its complex conjugate within the guidance equation. *Nothing* within BM alone may give us a hint about this. In fact, Dürr, Goldstein, and Zangui rely on the fact that the Schrödinger equation (a central piece of SQM) turns out to be invariant under T_A to take it as the adequate time-reversal operator for the guidance equation as well.

Let's leave Robert's argument aside and change the angle slightly. I mentioned earlier that Albert distinguishes between time-reversal invariance properly and *partial*-time-reversal invariance: whereas mostly all fundamental physical theories that have been taken seriously up to the presete are non-time-reversal invariant (according to Albert's thought), some of them *do* turn out to be *partial*-time-reversal invariant. Moreover, I've identified partial-time-reversal invariance as a case of a relationalist time-reversal transformation. In this case, a relationalist time-reversal operator concerns *only* the positions and trajectories of particles, that is, the Bohmian PO. Hence, it seems reasonable that if one holds a relationalist stance with respect to time (and then one commits oneself to formally implement time reversal in terms of motion reversal) and also holds that the fundamental stuff of the world consists of particles and their positions, partial-time-reversal invariance, in Albert's sense, is what one should mainly pay attention to.

But this takes one back to a point I've argued against before: there is almost always some straightforward transformation linking every state S_k with other state $\tilde{S}_k$ such that if there is a sequence going from S_i to S_f in accordance to the theory, there is then a time-reversed sequence going from $\tilde{S}_f$ to $\tilde{S}_i$ also in accordance to the theory. One problem with it: such a transformation is too weak on the demands. If it only aims at linking any state with other, which chiefly means

[48] One might come to think that the argument could succeed doing without condition (c). In fact, so as to pick out the imaginary part in the genesis of the guidance equation, one was in need of taking, no matter what, an anti-unitary operator (it's the only one that produces a minus sign to the imaginary part after applied), so conditions (a) and (b) seem to be enough to justify an anti-unitary Bohmian time-reversal operator that doesn't presuppose any fact about time reversal in ordinary quantum mechanics. Nevertheless, I'm not completely sure about this way out. At the end of the day, one also needs to know how the wave function transforms on the right side of the equation, whose behavior is given at the same time by the Schrödinger equation. So, I feel that there is nothing to hold on to justify that the wave function transforms as usual doing without information coming from what knows about the Schrödinger equation, that is, from SQM.

to linking any particle's position with a time-reversed particle's position (whatever this means) and keeping them invariant, then the transformed scenario will be whatever it needs to be to keep invariant the trajectories.

I'm not claiming that this is completely pointless. What I want to point to is that the question for a structural arrow of time is, again, *whether* a theory's dynamics (under the constraints I've already specified before) turns out to be invariant under time reversal, and not *in which way* it may be so. Trivially, there will be always a way to get your theory invariant under a transformation that changes the dynamics in such a precise way that keeps particles' position invariants. Or, to put it differently, there will always be a transformation capable of generating a backward movement keeping invariant the trajectories. Furthermore, as I've mentioned not long ago, *almost always* such a straightforward transformation is completely dependent of the dynamics at issue, that is, it isn't based on an independent way of justifying how the transformation is to act upon state without appealing to the dynamics.

Let me now change the angle radically. Suppose now that one declines any relationalist commitment with time and turns into a substantivalist framework. One thus knows two things:

(a) The Schrödinger equation comes out non-time-reversal invariant in SQM
(b) If the Schrödinger equation comes out non-time-reversal invariant, so does guidance equation.

Therefore, one is forced to conclude that BM is structurally and radically as non-time-reversal invariant as SQM, regardless its PO of particles and positions. Though there will always be a way to devise a transformation that links particles' positions to particles' positions, and thereby, that maps a trajectory to a transformed trajectory, the substantivalist will reasonably claim that one is not allowed to call that "artifice" *time-reversal*. And this is because one also has to specify how the wave-function transforms within the guidance equation: if one goes with a substantivalist time-reversal operator, one knows that time reversal doesn't take the complex conjugation over the wave function. Therefore, the guidance equation is non-time-reversal-invariant: the right side of the equation is not equal to the left one. This could be shocking at first glance within a deterministic theory. But, if you stop for a while and think things out, this is quite reasonably from the substantivalist's perspective: the time-reversal transformation acts upon first-time derivatives changing their sign. And, so long as the wave function's behavior is governed by the Schrödinger equation, and it turns out to be non-time-reversal invariant under T_U, there seems no way to keep the guidance equation invariant in this setting. To put it simpler: for the right side of the guidance equation to transform in the right way to keep it invariant, the time-reversal transformation must transform the wave function into its complex conjugate. But T_U is not the sort of transformation that does that job[49]. It follows from this that T_U will ruin the right side of the guidance equation. And this justifies the point (b). There is no much to add here.

Final Remarks

Along this chapter I've analyzed, on the one hand, whether Bohmian Mechanics *structurally* exhibits a time asymmetry and, on the other, the role of time reversal within the theory.

As to the former, the conventional wisdom on the matter is that it doesn't: any temporal asymmetry depends on an asymmetry in the initial conditions as a last resort. In this way, BM is able to explain any temporal asymmetry in the theory (for instance, the difference between forward and backward transitions) though is clear that such an explanation is not enough to manifest a structural temporal asymmetry. In fact, BM turns out to be, for the conventional wisdom, structurally time symmetric as both the Schrödinger equation and the guidance equation are time-reversal invariant, if T_A is taken. Nonetheless, as discussed previously, whether T_A is taken depends on previously-assumed metaphysical commitments with respect to time, so the claim should be conditionalized over them: if a relationalist stance with respect to time is assumed, then BM turns out structurally time symmetric because their fundamental equations of motion are T_A-invariant, and thereby, time-reversal invariant.

Unsurprisingly, BM turns out non-time-reversal invariant if a substantivalist stance with respect to time is rather assumed. The point is that the Schrödinger equation comes out non-time-reversal invariant under T_U. This implies that guidance equation will be non-time-reversal invariant too. This becomes evident when one focuses on the right side of the guidance equation and how it would transform under T_U. However, this result comes out *mainly* from the Schrödinger equation and how it relates to the guidance equation. In this sense, BM is as structurally time asymmetric as SQM.

I've also analyzed the role of time reversal in BM. Its defenders claim that time reversal plays a special role within the theory as it helps underpinning the simplest Galilean-invariant

[49] Neither would it even take the imaginary part for 9.5 to begin with! And if T_U is for any reason the right time-reversal transformation, why would one use one transformation for reversing time and a different one in the genesis of written down the guidance equation?

version of the guidance equation (the "official" guidance equation, so to speak). Whether the strategy succeeds or not has been criticized elsewhere (Skow 2010). I've rather focused on the role of time reversal in the argument. The conclusion is that, once again, there is no means to judge whether guidance equation is time-reversal invariance or not independently of whether the Schrödinger equation is.

In this sense, BM doesn't add any new argument for or against time-reversal invariance. Neither for or against a structural arrow of time. On the one side, whether the guidance equation is time-reversal invariant heavily depends on the time reversal transformation one goes with. And this, at the same time, depends on facts about the Schrödinger equation and its time-reversal invariance.

On the other, any strategy relying on BM's PO is somewhat hopeless. First, the PO *on its own* is unable to give us a hint about how time reversal should be implemented. It's the dynamics what one should look at so as to figure out how the relevant magnitudes should transform. And this leads one again to work out how the wave-function should transform. And any recipe for this will come from SQM, not from BM's PO.

ω.

Final Remarks

Or what all this was about…

And yet … and yet . . . Denying temporal succession, denying the self, denying the astronomical universe, are apparent desperations and secret consolations. Our destiny is not frightful by being unreal; it is frightful because it is irreversible and iron-clad. Time is the substance I am made of. Time is a river which sweeps me along, but I am the river; it is a tiger which destroys me, but I am the tiger; it is a fire which consumes me, but I am the fire.

(Jorge Luis Borges)

"And yet…and yet…Negar la sucesión temporal, negar el yo, negar el universo astronómico, son desesperaciones aparentes y consuelos secretos. Nuestro destino no es espantoso por irreal; es espantoso porque es irreversible y de hierro. El tiempo es la sustancia de que estoy hecho. El tiempo es un río que me arrebata, pero yo soy el río; es un tigre que me destroza, pero yo soy el tigre; es un fuego que me consume, pero yo soy el fuego"

(Jorge Luis Borges)

The main idea underlying this research has been that philosophers of physics and philosophers of time have, despite what it has been largely said, well-based reasons to hold that there is something like a *quantum* arrow of time, which doesn't exclusively and solely depend upon contingent facts about matter or initial conditions, but rather may be grounded in fundamental laws. This conclusion, as well as the opposite, is not of course free of assumptions. Hence, the other guiding idea driving this work is that our claims about the arrow of time, time reversal, time-reversal invariance, and so forth, intrinsically depend upon a panoply of previously-taken commitments. Some of them, surprisingly, have been largely overlooked. As it was mentioned at the end of the Introduction, my intention wasn't to coercively force others to take a heterodox view on the matter, or to affirm that the way in which things have been thought of so far was deeply and simply mistaken. Broadly speaking, the motivation was to point the way towards an (alternative) understanding of the problem of the arrow of time and of some of the concepts it involves, which have been gravitating around philosophy and physics since long ago. Naturally, I do think that this is a noteworthy path to set forth on, and that's what I hope I have justified convincingly.

As each part and chapter finished with a summary, there is no much need for extensive final remarks, but it will suffice to put all the pieces together briefly and to put on view the overall argumentation. I started off this research by pointing to the fact that the very formulation of the problem of the arrow of time (as typically understood in the literature) is already a philosophical problem in itself calling for clarification. The undertaken task mainly consisted in *unpacking* the problem, unthreading closely-related, though differentiable, notions and sorting out some problems commonly put under the same label. From this unpacking task, I arrived at two main conclusions: First, that the notion of *irreversibility* and *(non)-time-reversal invariance* ought to be properly distinguished. Second, that the general problem of the arrow of time encloses two different problems frequently mixed up: the problem of *a structural arrow of time* and the problem of *the two realms*. Whereas the latter greatly relies on irreversibility, the former on the time reversal. In addition, I've shown in which way the latter conceptually presupposes the former. Their modal scope and metaphysical implications were also stemmed from such a distinction. I finally highlighted that the notion of time-reversal invariance played a paramount role in asserting whether time is *structurally* directed or not, backed up by the symmetry-to-reality inference and the connections between dynamics' symmetries and space-time's symmetries.

With all this conceptual machinery under our belt, I stepped into standard non-relativistic quantum mechanics. I pointed out there that the problem to be primarily addressed was whether the Schrödinger equation turns out to be time-reversal invariant (Chapter IV). At this point, I argued that two approaches to time reversal, and thereby to time-reversal invariance, may be rightfully and soundly defended in this framework: and *orthodox approach* and a *heterodox approach*. After exposing conceptually and formally in detail the notion of time reversal according to the orthodox approach, I opposed it to the heterodox one. Notably, whereas the former concludes that the Schrödinger equation is time-reversal invariant, the latter arrives at the opposite result. The argumentation was then overall based on two pillars: (a) showing that both approaches don't really return contradictory conclusions, because in fact they understand time reversal differently (both in its formal and conceptual aspects), (b) pointing to the (overlooked) fact that both approaches rely heavily upon a bunch of metaphysical, epistemic and methodological statements on time and symmetries. Therefore, both approaches' lawfulness takes root in such a broader bunch of statements. This rendered two sides in the debate:

(i) The orthodox approach to time reversal is based on a *relationalist* metaphysics of time (as long as it understands time reversal as *motion* reversal and represent the symmetry transformation accordingly) and on a view that takes symmetries as *theory relative* and as *guides to theory construction*.

(ii) The heterodox approach to time reversal rather resorts on a *substantivalist* metaphysics of time (as long as it understands time reversal as metaphysically and physically distinct from motion reversal) and on a view that takes symmetries as *theory independent* and as *contingent properties of the dynamics.*

The (partial) upshot of all this was that the declaration of time-reversal invariance or non-time-reversal invariance should be therefore relativized to these backgrounds. Notwithstanding, I argued next that the orthodox approach, though defendable and workable in many contexts, might not be quite appropriate when addressing some metaphysical problems as the problem of a structural arrow of time (Chapter V). Then, it was shown that OA runs the risk of violating some of the conditions that make the problem of the arrow of time (like contingency and fundamentality) philosophically interesting. This led philosophers of physics and time to an uneasy scenario where something has to be given updropped: either the problem of the arrow of time is reformulated, or the orthodox approach is abandoned. The overall conclusion in Part 2 was that, even though two legitimate approaches to time reversal co-exist and can be rightfully held (under some relevant assumptions), the orthodox approach doesn’t seem adequate to address the problem of a structural arrow of time, which left us with the following bold conclusion: *if* these issues for the orthodox approach cannot be overcome, the heterodox approach should be rather taken when addressing the problem of a structural arrow of time within standard non-relativistic quantum mechanics. Therefore, the Schrödinger equation turns out non-time-reversal invariance. Therefore, it *does* exhibit a structural arrow of time.

Part 3 introduced some interpretations of standard non-relativistic quantum mechanics available to the present. Setting aside the above-mentioned conclusion, and assuming as usual that the Schrödinger equation is time-reversal invariant, it was explored whether, and in which sense, various interpretations of the standard formalism deliver new theoretical elements that might drive to a different treatment of the two directions of time. Those interpretations that modify the standard dynamics were exclusively taken into account: Orthodox Quantum Mechanics (which introduces the Collapse Postulate), Spontaneous Collapse Theories (paradigmatically, GRW-type theories) and Bohmian Mechanics. Though all these interpretations are able to objectively distinguish between the past-to-future and the future-to-past direction, only Spontaneous Collapse Theories are able to do it in a *structural* way, that is, based purely on intrinsic properties of the theory's dynamics. Therefore, GRW-type theories are the most promissory candidates for a structural arrow of time in non-relativistic quantum mechanics, even in the case that the Schrödinger equation be time-reversal invariant. Orthodox Quantum Mechanics and Bohmian Mechanics can just account for time asymmetry in terms of either dynamics' external properties or boundary conditions, yielding thereby a non-structural arrow of time.

After such a long road, it is worth coming back to the question this thesis started with:

> Does time (whatever it comes to be) *objectively* instantiate the property of "having a privileged direction" *in non-relativistic quantum mechanics*?

And the answer is, in general, "yes, it does but…". Where "but" must be unwrapped taken into consideration the distinctions I drew along this thesis: what sort of temporal asymmetry is being given, what one's metaphysical, epistemic and methodological commitments are with respect to the terms involved, and so forth. Even the most radical form of temporal asymmetry (a structural arrow of time grounded in the fact that the standard formalism, in the simplest cases, comes out non-time-reversal invariance) may be defendable by deconstructing the widely-extended orthodox view.

Philosophers and physicist have tirelessly disputed about the nature of time for centuries and it is highly likely that things will remain the same in the future. However, as mentioned in the Introduction, progress is sometimes made by clarifying what a problem is really about, by exploring what the alternative paths are and by exposing some deeply-ingrained assumptions. In such a philosophical exercise, one can, not only, discover new dead ends as well as unseen way outs, but also get to a far-reaching understanding of the problem and to deeper considerations of its current challenges as well as of its eventual answers. However, the philosophical task should not only push one's convictions beyond, but also allow one to occasionally come back to them on firmer grounds. In any case, all this drives our knowledge of the world farther.

References

Part 1

Aiello, M., Castagnino, M., and Lombardi, O. (2008). "The arrow of time: from universe time-symmetry to local irreversible processes". *Foundations of Physics*, 38: 257-292.

Albert, D. Z. (2000). *Time and Chance*. Cambridge, MA: Harvard University Press.

Arntzenius, F. (1997). "Mirrors and the direction of time". *Philosophy of Science,* 64: 213-222.

Ayer, A. J. (1954). *Philosophical Essays*, London: Macmillan.

Baker, D. (2010). "Symmetry and the metaphysics of physics". *Philosophy Compass* 5: 1157-1166

Barbour, J., and Bertotti, B. (1982). Mach's principle and the structure of dynamical theories. *Proceedings of the Royal Society A, 382*, 295–306.

Baron, S. (2015). "Feel the flow". *Synthese,* 194 (2): 609-630.

Biedenharn, L. and Sudarshan, E. (1994). "A basis-free approach to time reversal for symmetry group". *Pramana,* 43 (4): 255-272.

Belot, G. (2013). "Symmetry and equivalence", in R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics,* Oxford: Oxford University Press.

Benovsky, J. (2010). "The relationalist and substantivalist theories of time: foes or friends?" *European Journal of Philosophy,* 19 (4): 491-506.

Black, M. (1952). "The identity of indiscernibles". *Mind,* 61: 153-164.

Boltzmann, L. (1872). "Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen", in *Wissenschaftliche Abhandlungen*, F. Hasen¨ohrl (ed.), New York: Chelsea Publishing Company, 1968.

Brading, K. and Castellani, E. (2003). *Symmetries in physics: Philosophical Reflections*. Cambridge: Cambridge University Press.

Brading, K. and E. Castellani. (2007). "Symmetries and Invariances in Classical Physics". *Handbook of the Philosophy of Science, Philosophy of Physics, Part B*. Eds. J. Butterfield and J. Earman. The Netherlands: Elsevier, 1331–1367.

Bridgman, P. (1941) *The Nature of Thermodynamics*. Cambridge, MA: Harvard University Press.

Brighouse, S. (1994). "Spacetime and Holes". *PSA*: *Proceedings of the Biennial Meeting of the Philosophy of Science Association,* 1: 117-125.

Broad, C.D., (1927). *Scientific Thought*, London: Kegan Paul.

Callender, C. (1995). “The metaphysics of time reversal: Hutchison on classical mechanics”. *The British Journal for the Philosophy of Science*, 46: 331-340.

Callender, C. (1997). “What is “the problem of the direction of time’?”. *Philosophy of Science* (Supplement), 64: S223-S234.

Callender, C. (2004). “Measures, explanations and the past: should ‘special’ initial conditions be explained?”, *The British Journal for the Philosophy of Science*, 55: 195-217.

Callender, C. (2016). “Thermodynamic asymmetry in time”, in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, URL=<http://plato.stanford.edu/archives/win2016/entries /time-thermo/>.

Callender, C. (2004a). “There is No Puzzle about the Low Entropy Past”, in Christopher Hitchcock (ed.), *Contemporary Debates in the Philosophy of Science*, Oxford: Blackwell, pp. 240–256.

Callender, C. (2004b). “Measures, Explanation and the Past: Should ‘Special’ Initial Conditions Be Explained?”. *British Journal for the Philosophy of Science*, 55(2): 195–217.

Carnot, S. (1890), *Réflexions sur la Puissance Motrice du Feu et sur les Machines Propres à Développer cette Puissance* (1824). Trad. R. H. Thurston, *Reflections on the Motive Power of Heat and on Machines Fitted to Develop that Power*. New York: Steam Engine Library New York.

Carroll, J. (1994). *Laws of Nature*. Cambridge: Cambridge University Press.

Castagnino, M. and Lombardi, O. (2005). “A global and non-entropic approach to the problem of the arrow of time”, pp. 74-108, in A. Reimer (ed.), *Spacetime Physics Research Trends. Horizons in World Physics*. New York: Nova Science.

Castagnino, M. and Lombardi, O. (2009). “The global non-entropic arrow of time: from global geometrical asymmetry to local energy flow”. *Synthese*, 169: 1-25.

Castagnino, L. Lara and Lombardi, O. (2003). “The direction of time: From the global arrow to the local arrow”, *International Journal of Theoretical Physics*, 42: 2487-2504.

Caulton, A. (2015). “The role of symmetry in interpretation of physical theories”. *Studies in History and Philosophy of Modern Physics*, 52: 153-162.

Caulton, A. and J. Butterfield. “Symmetries and paraparticles as a motivation for structuralism”. *British Journal for the Philosophy of Science* 63.2 (2012): 233–285.

Clausius, R. (1864). *Abhandlungungen über die mechanische Wärmetheorie*. Vol. 1, Braunschweig: F. Vieweg. (English translation in Uffink 2001).

Craig, W. (2000). *The Tensed Theory of Time: A Crítical Examination*. Dordrecht: Kluwer Academic Publishers.

Dasgupta, S. (2015). “Substantivalism vs. Relationalism about space in classical physics”. *Philosophy Compass*, 10/9: 601-624.

Dasgupta, S. (2016). “Symmetry as an epistemic notion (twice over)”. *British Journal for Philosophy of Science*, **67**: 837-878.

Davies, P. (1995). *About time: Einstein’s unfinished revolution*. Harmondsworth: Penguin.

Deasy, D. (2017). “What is presentism?”. *Nous,* vol. 51, 2: 378-397.

Earman, J. (1974). “An attempt to add a little direction to ‘The Problem of the Direction of Time’”, *Philosophy of Science,* 41: 15-47.

Earman, J. (1986a). “The problem of Irreversibility”. *PSA: Proceedings of the Biennial Meeting for the Philosophy of Science Association,* Vol. 2: 226-233.

Earman, J. (1986b). *A Primer on Determinism*. Boson: Dordrecht.

Earman, J. (1989). *World enough and space-time. Absolute versus relational theories of space-time*. Cambridge, Massachusetts: MIT Press.

Earman, J. (2006). “The ‘past hypothesis’: not even false”, *Studies in History and Philosophy of Science,* 37: 399-430.

Eddington, A. (1928). *The Nature of the Physical World*. Cambridge: Cambridge University Press.

Zermelo, E. (1896). “Uber einen Satz der Dynamik un die mechanische Warmetheorie”**,** *Annalen der Physik,* 57: 485-494. Translation in *Kinetic Theory,* ed. S. G. Brush, 2 vols., II, 208-217. Pergamon: Oxford.

Esfeld, M. and Deckert, D. (2018). *A Minimalist Ontology of the Natural World.* New York: Routledge.

Farr, M. (2012). “On A- and B-theoretic elements of branching spacetimes”, *Synthese,* 188 (1): 86-116.

Frischhut, A. (2013). “What experience cannot teach us about time”. *Topoi,* 34 (1): 143-155.

Yehezkel, G. (2013). “The illusion of the experience of the passage of time”. *Disputatio,* V, No. 35: 67-80.

Goldstein, S., Tumulka, R., and Zanghì, N. (2016). “Is the Hypothesis about a Low Entropy Initial State of the Universe Necessary for Explaining the Arrow of Time?”. *Physical Review D*, 94(2): 023520.

Grib, Sean and Karim Thébault. 2016. “Time remains”. *British Journal for Philosophy of Science,* 67: 663-705.

Grünbaum, A. (1963). *Philosophical Problems of Space and Time*. New York: Alfred A. Knopf.

Hacking, I. (1975). “The identity of indiscernibles”. *The Journal of Philosophy*, 72: 249-256.

Henderson, J. (2010). “Classes of Copenhagen interpretations: Mechanisms of collapse as typologically determinative”. *Studies in History and Philosophy of Modern Physics,* 41: 1-8.

Hoefer, C. (1996). ‘The Metaphysics of Space-Time Substantivalism.’ The Journal of Philosophy 93: 5–27.

Hoerl, C. (2013). “Do we (seem to) perceive passage?” *Philosophical Explorations,* 17 (2): 188-202.

Horwich, P. (1987). *Asymmetries in Time*. Cambridge, MA: MIT Press.

Huggett, N., Vistarini, T., and Wüthrich, C. (2012). “Time in quantum gravity”, in Adriand Bardon and Heather Dyke (Eds.), *A Companion to the Philosophy of Time*, Wiley-Blackwell, pp. 242-260.

Hutchison, K. (1993). “Is classical mechanics really time-reversible and deterministic?” *The British Journal for the Philosophy of Science*, 44: 307-323.

McDaniel, K. (2016). “John M. E. McTaggart”, in Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, URL= https://plato.stanford.edu/archives/win2016/ entries/mctaggart/

Landsberg, P. (1982). *The Enigma of Time*. Bristol: Hilger.

Lange, M. (2000). *Natural Laws in Scientific Practice*. Oxford: Oxford University Press.

Le Poidevin, R. (2007). *The Images of Time*: *An Essay on Temporal Representation*. Oxford: Oxford University Press.

Lehmkuhl, D. (2018). “The Metaphysics of Super-Substantivalism”. *Nous*, 52 (1): 24-46.

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell Publishers.

Lewis, D. (1986). Collected papers (Vol. II). Oxford: Oxford University Press.

Lewis, D. (1994). “Humean supervenience debugged”. Mind, 103, 473–490.

Loewer, B (2012). “Two accounts of laws and time”. *Philosophical Studies*, 160: 115-137.

Loschmidt J. (1876). “Über den Zustand des Wärmegleichgewichtes eines Systems von Kőrpern mit Rücksicht auf die Schwerkraft”, *I, Sitzungsber. Kais. Akad. Wiss. Wien, Math. Naturwiss. Classe*, II. Abteilung **73**, 128-142, 1876. Translation: “On the state of thermal equilibrium in a system of bodies with considerations of gravity”, online version at www.loschmidt.cz.

Maudlin, T. (1989). “The essence of space-time”, in A. Fine and J. Leplin (eds.), *PSA 1988*, vol. 2. East Lansing: Philosophy of Science Association, 82-91.

Maudlin, T. (1993). Buckets of water and waves of space: why space-time is probably a substance. *Philosophy of Science*, *60*, 183–203.

Maudlin, T. (2002). “Remarks on the passing of time”, *Proceedings of the Aristotelian Society*, 102: 237-252.

Maudlin, T. (2007). *The Metaphysics within Physics.* New York: Oxford University Press

McTaggart, J.M.E. (1908). “The Unreality of Time”, *Mind*, 17: 457–73.

Mehlberg, H. (1961) “Physical Laws and time’s arrow”, in *Current Issues in the Philosophy of Science,* New York: Holt, Rinehart and Winston.

North, J. (2009). “The structure of Physical theories”. *Journal of Philosophy*, 106: 57-88.

North, J. (2008). “Two views on time reversal”. *Philosophy of Science,* 75: 201-223.

Norton, J. (2016). “The impossible process: Thermodynamics reversibility”. *Studies in History and Philosophy of Modern Physics*, 55: 43-61.

Norton, J. (2012). “Approximation and idealization: Why the difference matters”. *Philosophy of Science*, 79, 207–232.

Nozick, R. (2001). *Invariances: The Structure of the Objective World*. Cambridge, MA: Harvard University Press.

Paul, L. A. (2010). Temporal experience. *Journal of Philosophy, 107,* 333–359.

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

Planck, M. (1897.) *Vorlesungen ¨uber Thermodynamik*. Leipzig: Veit & Comp.

Pooley, O. (2013). Substantivalist and relationalist approaches to spacetime, In Batterman, R. (Ed.) *The oxford handbook of philosophy of physics* (pp. 522-586). Oxford: Oxford University Press.

Price, H. (1996). *Time's Arrow and Archimedes' point: New Directions for the Physics of Time*. New York: Oxford University Press.

Price, H. (2012). "The flow of time", in Craig Callender (ed.*), The Oxford Handbook of Time*, Oxford University Press.

Price, H. (2010). "Time's arrow and Eddington's challenge". *Séminaire Poincaré XV Le Temps:* 115-140.

Prigogine, I. (1980). *From Being to Becoming. Time and Complexity in the Physical Sciences.* New York: W. H. Freeman.

Prosser, S. (2013). "Passage and Perception". *Nous*, 47 (1): 69-84.

Putnam, H. (1967). "Time and Physical Geometry". *Journal of Philosophy*, 64: 240–247

Rasmussen, J. (2012). "Presentists May Say Goodbye to A-Properties". *Analysis*, 72(2): 270–276.

Raven, M. J. (2010). "Can time pass at the rate of 1 second per second?". *Australasian Journal of Philosophy*, 88: 1-7.

Rechel, E. (1947). "The reversible process in thermodynamics". *Journal of Chemical Education*, 24: 298-301.

Reichanbach, H. (1956). *The Direction of Time.* Berkeley: University of California Press.

Rietdijk, C.W. (1966). "A Rigorous Proof of Determinism Derived from the Special Theory of Relativity". *Philosophy of Science*, 33: 341–344.

Roberts, J. (2008). "A puzzle about laws, symmetries and measurability". *The British Journal for Philosophy of Science*, **59**: 143-168.

Rodriguez-Pereira, G. (2006). "How not to trivialise the identity of indiscernibles". In P. F. Strawson and A. Chakrabarti (eds.) *Concepts, Properties and Qualities*, Ashgate, pp. 205-223.

Rodriguez-Pereira, G. (2004). "The bundle theory is compatible with distinct but indiscernible particulars". *Analysis*, 64 (1): 72-81.

Rovelli, C. (2002). "Partial Observables". Physical Review, D65: 124013.

Rovelli, C. (2004). *Quantum gravity*. Cambridge: Cambridge University Press

Ruetsche, L. (2013). *Interpreting Quantum Theories*. Oxford: Oxford University Press.

Schaffer, J. (2009). "Spacetime the one substance". *Philosophical Studies*, 145(1): 131-148.

Schuster, M.M. (1986). "Is the flow of time subjective?" The *Review of Metaphysics*, 39 (4): 695-714.

Sider, T. (2001). *Four-dimensionalism.* Oxford: Oxford University Press.

Sklar, L. (1974). *Space, Time and Spacetime.* Berkeley: University of California Press.

Sklar, L. (1993). *Physics and Chance*, Cambridge: Cambridge University Press.

Swartz, N. (1985). *The concept of physical law*. Cambridge: Cambridge University Press.

Tallant, J. (2012). "(Existence) Presentism and the A-theory", *Analysis*, 72(4): 673–681.

Thomson, W. (Lord Kelvin). (1852). "On a universal tendency in nature to the dissipation of mechanical Energy", in *Mathematical and Physics Papers, Vol.1, Cambridge: Cambridge University Press*, pp. 511–514.

Uffink, J. (2001). "Bluff your way in the second law of thermodynamics", *Studies in History and Philosophy of Science*, 32: 305-394.

Uffink, J. (2007). "Compendium to the foundations of classical statistical physics", pp. 924-1074, in J. Butterfield and J. Earman (eds), *Handbook for the Philosophy of Physics*. Amsterdam: Elsevier.

Valente, M. B. (2012). "The relativity of simultaneity and presentism", at https://arxiv.org/abs/1302.2603.

Wallace, D. (2011). "The logic of the past hypothesis", in B. Loewer, E. Winsberg y B. Weslake (Eds.), currently-untitled volume on themes from David Albert's *Time and Chance*, Harvard.

Wallace, D. (2019). "Observability, redundancy and modality for dynamical symmetry transformations". Forthcoming.

Weyl, H. (1952). *Symmetry*. Princeton, NJ: Princeton University Press.

Zimmerman, D. (2008). "The Privileged Present: Defending an 'A-theory' of Time", in *Contemporary Debates in Metaphysics*, Theodore Sider, John Hawthorne, & Dean W. Zimmerman (eds), Oxford: Blackwell: 211–225.

Part 2

Abraham, R. and Marsden, J. E. (1978). *Foundations of Mechanics, 2nd edn,* Addison- Wesley Publishing Company, Inc.

Acuña, P. (2016). "Minkowski Spacetime and Lorentz Invariance: the cart and the horse or two sides of a single coin?". *Studies in History and Philosophy of Modern Physics*, 55: 1-12.

Arntzenius, F. and Greaves, H. (2009). "Time reversal in classical electromagnetism". *The British Journal for the Philosophy of Science*, 60: 557-584

Ballentine, L. (1998). *Quantum Mechanics. A modern Development*. Singapur: World Scientific.

Belot, G. (2001). "The Principle of Sufficient Reason." *Journal of Philosophy*, 98: 55–74.

Brown, H. (2005). *Physical Relativity*. Oxford: Oxford University Press.

Busch, P. (1990a). "On the energy-time uncertainty relation. Part I: Dynamical time and time indeterminacy". *Foundations of Physics*, Vol. 20, N°1: 1-32.

Busch, P. (1990b). "On the energy-time uncertainty relation. Part II: Pragmatic time versus energy indeterminacy". *Foundations of Physics*, Vol. 20, N°1: 33-43.

Callender, C. (2000). "Is time 'handed' in a quantum world?" *Proceedings of the Aristotelian Society*, 100: 247-269.

Chevallier, G. (2007). "Wigner's theorem and its generalizations". In K. Engesser, D. M. Gabbay and D. Lehmann (eds.), *Handbook of Quantum Logic and Quantum Structures*, New York: Elsevier Publisher, pp. *429-475*.

Costa de Beauregard, O. (1980). "CPT invariance and interpretation of quantum mechanics", *Foundations of Physics*, 10: 513-530.

Curiel, E. (2014). "Classical mechanics is Lagrangian: It is not Hamiltonian". *British Journal for Philosophy of Science*, 65: 269-321

Davies, P. (1974). *The Physics of Time Asymmetry*. Berkeley: University of California Press.

Dürr, D. and Teufel, S. (2009). *Bohmian Mechanics: The Physics and Mathematics of Quantum Theory*, Berlin: Springer-Verlag.

Earman, J. (2004). "Laws, symmetry, and symmetry breaking: invariance, conservation principles, and objectivity." *Philosophy of Science*, **71**: 1227-1241.

Gasiororowicz, S. (1966). *Elementary Particle Physics.* New York: John Wiley and Sons.

Gibson, W.M. and Pollard, B.R. (1976). *Symmetry Principles in elementary particle physics*. Cambrige: Cambridge University Press.

Haag, R. (1992). *Local Quantum Physics: Fields, Particles, Algebras.* Heidelberg: Springer-Verlag

Hicks, M. (2018). "What everyone should say about symmetries (and why Humeans get to say it)". http://philsci-archive.pitt.edu/14473/

Janssen, M. (2009). "Drawing the Line between Kinematics and Dynamics in Special Relativity", *Studies in History and Philosophy of Modern Physics*, 40: 26-52.

Jauch, J. and Rohrlich, F. (1955). *The Theory of Photons and Electrons*. London: Addison-Wesley Publishing Company.

Lange, M. (2007). "Laws and meta-laws of nature: conservation laws and symmetries". *Studies in History and Philosophy of Modern Physics*, 38: 457-481.

Lombardi, O., Castagnino, M., and Ardenghi, J. S. (2010). "The modal-Hamiltonian interpretation and the Galilean covariance of quantum mechanics," *Studies in History and Philosophy of Modern Physics*, **41**: 93-103.

Lopez, C. (2019). "Roads to the past: how to go and *not* to go backward in time in quantum theories". *European Journal for Philosophy of Science*, 9: 27.

Lopez, C. and Lombardi, O. (2019). "Space-time symmetries in quantum mechanics". In Lombardi et al. (eds), *Quantum Worlds: Perspectives on the Ontology of Quantum Mechanics*, Cambridge: Cambridge University Press, pp. 263-293.

Mach, E. (1919). *The science of mechanics: a critical and historical account of its development, 4th ed.., Translation by Thomas J. McCormack*. Chicago: Open Court.

Messiah, A. (1966). *Quantum Mechanics*. New York: John Wiley and Sons.

Pauli, W. (1980) [1958]. *General Principles of Quantum Mechanics*. Berlin: Springer-Verlag.

Peterson, D. (2015). "Prospect for a new account of time reversal". *Studies in History and Philosophy of Modern Physics*, 49: 42-56.

Rédei M. (1998). "Observables and states in the Hilbert space formalism of quantum mechanics". In: *Quantum Logic in Algebraic Approach. Fundamental Theories of Physics (An International Book Series on The Fundamental Theories of Physics: Their Clarification, Development and Application)*, vol 91. Dordrecht: Springer, pp. 11-28.

Roberts, B. (2017). "Three myths about time reversal invariance". *Philosophy of Science*, 84, 2: 315-334.

Roberts, B. (2018). "Time reversal" http://philsci-archive.pitt.edu/15033/1/Roberts2018-Time Reversal.pdf.

Roman, P. (1961). *Theory of Elementary Particles*. Amsterdam: North Holland Publishing Company.

Sachs, R. (1987). *The Physics of Time Reversal*. London: University Chicago Press.

Savitt, S. (1996). "The direction of time". *The British Journal for the Philosophy of Science*, 47: 347-370.

Schlosshauer, M. (2006). "Experimental motivation and empirical consistency in minimal no- collapse quantum mechanics". *Annals of Physics*, 321: 112-149.

Stone, M.H. (1930). "Linear transformations in Hilbert space. III. Operational methods and group theory." *Proceedings of the National Academy of Sciences of the United States of America*, 16: 172–175.

Styer, D. et al. (2002). "Nine formulations of quantum mechanics". *American Journal of Physics*, 70: 288-297.

Uhlhorn, U. (1961). "Representation of Symmetry Transformations in Quantum Mechanics." *Arkiv för Fysik* 23:307–40.

von Neumann, J. (1931). "Die Eindeutigkeit der Schr¨odingerschen Operatoren." *Mathematische Annalen*, 104: 570–578.

Von Neumann, J. (1932). *Mathematische Grundlagen dder Quantenmechanik*. Berlin: Springer-Verglag. English version: *Mathematical Foundations of Quantum Mechanics* (1955). Berlin: Princeton University Press.

Wallace, D. (2008). "Quantum Mechanics", in Rickles (ed), *The Ashgate Companion to the New Philosophy of Physics* (Ashgate, 2008). Published online under the title: "The Measurement Problem: State of Play".

Watanabe, S. (1951). "Reversibility of Quantum Electrodynamics". *Physical Review*, 84, 1008.

Wigner, E. (1932). *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra.* New York: Academic Press (1959).

Wigner, E. (1967). *Symmetries and Reflections*. Bloomington, IN: Indiana University Press.

Part 3

Aaserud, F. and Heilbron, J. (2013). "Love, literature and the quantum atom (Niels Bohr's 1913 Trilogy Revisitied)". *Isis*, 106 (4): 972-973.

Adler, S. (2007). "Lower and Upper Bounds on CSL Parameters from Latent Image Formation and IGM Heating". *Journal of Physics*, A40: 2935.

Aharonov, Y. and Rohrlich, D. (2005). *Quantum paradoxes: Quantum theory for the perplexed*. Weinheim: Wiley-VCH.

Aharonov, Y., Bergmann, P., and Lebowitz, J. (1964). "Time symmetry in the quantum process of measurement". *Physical Review B,* 134(6): 1410-1416.

Albert, D. (1990). "On the collapse of the wave-function". In Miller, A. (ed.), *Sixty-two years of Uncertainty*, New York: Plenum Press, pp. 153-166.

Albert, D. (1992). *Quantum mechanics and experience.* Cambridge, MA: Harvard University Press.

Albert, D. (1994). "The Foundations of Quantum Mechanics and the Approach to Thermodynamic Equilibrium." *British Journal for the Philosophy of Science,* 45, 669-677.

Albert, D. (1996). "Elementary quantum metaphysics". In J. Cushing, A. Fine, & S. Goldstein (Eds.), *Bohmian Mechanics and quantum theory*: An Appraisal, Kluwer, p. 277-284.

Albert, D. (2013). "Wave function realism". In D. Z. Albert, & A. Ney (Eds.), *The wave function: Essays in the Metaphysics of quantum mechanics*, New York: Oxford University Press, pp. 52-57.

Albert, D. and Loewer, B. (1990). "Wanted dead or alive: two attempts to solve Schrödinger's paradox". In Fine, Forbes and Wessels (eds.)., *Proceedings of PSA 1990*, Vol. 1, pp. 277-285.

Albert, D. and Loewer, B. (1996). "Tails of Schrödinger's cat". In Clifton, R. (ed), *Perspectives on quantum reality: non-relativistic, relativist and field theoretic,* Dordrecht: Kluwer, pp. 81-92.

Albert, D.Z. and Vaidman, L. (1989). "On a proposed postulate of state reduction", *Physics Letters*, A139: 1.

Allori, V. (2013). "Primitive ontology and the structure of fundamental physical theories". In D. Z. Albert, & A. Ney (Eds.), *The wave function: Essays in the Metaphysics of quantum mechanics*, New York: Oxford University Press, pp. 58-75.

Allori, V. (2018). "A new argument for the nomological interpretation of the wave-function: The Galilean group and the classical limit of non-relativistic quantum mechanics." *International Studies in the Philosophy of Science,* 31 (2): 177-188.

Allori, V. (2019). "Quantum mechanics, time and ontology". *Studies in History and Philosophy of Modern Physics,* forthcoming.

Allori, V. et al. (2014). "Predictions and primitive ontology in quantum foundations: A study of examples". *The British Journal for the Philosophy of Science*, 65 (2): 323-352.

Allori, V., Goldstein, S., Tumulka, R., and Zanghi, N., (2008). "On the Common Structure of Bohmian Mechanics and the Ghirardi-Rimini-Weber Theory". *British Journal for the Philosophy of Science*, 59: 353–389.

Atkinson, D. (2006). "Does quantum electrodynamics have an arrow of time?". *Studies in History and Philosophy of Modern Physics*, 37 (3): 528-541.

Bacciagalupi, G. (2007). "Probability, decoherence and the arrow of time". *Studies in History and Philosophy of Modern Physics*, 38 (2): 439-456.

Bacciagalupi, G. and Valentini, A. (2009). *Quantum Theory at the Crossroads: Reconsidering the 1927 Solvay Conference*. Cambridge: Cambridge University Press.

Bassi, A. and Ghirardi, G. (2003). "Dynamical Reduction Models", *Physics Reports*, 379: 257.

Bassi, A., Lochan, K., Satin, S., Singh, T.P., and Ulbricht, H. (2013). "Models of Wave-function Collapse, Underlying Theories, and Experimental Tests". *Review of Modern Physics*, 85: 47141.

Becker, L. (2004). "That von Neumann did not believe in a physical collapse". *British Journal for the Philosophy of Science*, 55 (1): 121-135.

Bell, J. (1987). *Speakable and unspeakable in quantum mechanics*. Cambridge: Cambridge University Press.

Bohm, D. (1952). "A suggested interpretation of the quantum theory in terms of 'hidden' variables. I and II". *Physical Review*, 85, 166-193.

Bohr, N. (1935). "Quantum mechanics and physical reality". *Nature*, 136: 1025-1026.

Bricmont, J. (2016). *Making Sense of Quantum Mechanics*, Cham: Springer.

Brock, S. (2003). *Niels Bohr's Philosophy of Quantum Physics*. Logos Verlag.

Bub, J. (2016). *Bananaworld: Quantum mechanics for primates*. Oxford: Oxford University Press.

Bub, J. (2018). "In defense of a "single-world" interpretation of quantum mechanics." *Studies in History and Philosophy of Modern Physics*, on line first.

Callender, C. (2018). *What makes time special?* Oxford: Oxford University Press.

Caves, C., Fuchs, C., and Schack, R. (2002), "Quantum Probabilities as Bayesian Probabilities". *Physical Review A*, 65, 022305, 1–6.

Cushing, J. (1994). *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*, Chicago: University of Chicago Press.

Deotto, E. and Ghirardi, G. (1998). "Bohmian Mechanics Revisited". *Foundations of Physics*, 28(1): 1–30.

Deutsch, D. (1986). ‘Three experimental implications of the Everett interpretation”. In R. Penrose and C.J. Isham (eds.), *Quantum Concepts of Space and Time,* Oxford: The Clarendon Press, pp. 204–214.

Deutsch, D. (2002). “The Structure of the Multiverse”, *Proceedings of the Royal Society of London* A, 458: 2911–2923.

Dickson, M. (2007). “Non-relativistic quantum mechanics”, in Jeremy Butterfield and John Earman (eds), *Philosophy of Physics*. Carolina: Elsevier.

Dieks, D. (2019). “Quantum mechanics and perspectivalism”. In Lombardi et al. (eds), *Quantum Worlds: Perspectives on the Ontology of Quantum Mechanics*, Cambridge: Cambridge University Press, pp. 51-70.

Dirac, P. (1935). *Principles of Quantum Mechanics.* Oxford: Oxford University Press.

Dürr, D., Goldstein, S. and Zanghi, N. (2004). “Bohmian mechanics and the meaning of the wave-function, contribution to experimental metaphysics”. In Cohen et al. (eds.), *Quantum Mechanical Studies for Abner Shimony*, vol. 1, Dordrecht: Kluwer – Boston Studies in the Philosophy of Science 193, pp. 25-38.

Dürr, D., Goldstein, S. and Zangui, N. (1996). “Bohmian mechanics as the foundation of quantum mechanics”. In J. T. Cushing, A. Fine, and S. Goldstein (eds), *Bohmian mechanics and quantum theory: An appraisal.* Dordrecht: Kluwer Academic Publishers.

Dürr, D., Goldstein, S. and Zangui, N. (2013). *Quantum Physics without Quantum Philosophy*. Dordrecht: Springer-Verlag.

Dürr, D., Goldstein, S., and Zanghi, N. (1992). “Quantum equilibrium and the origin of absolute uncertainty”. *Journal of Statistical Physics*, 67: 843-907.

Dürr, D., Goldstein, S., and Zanghi, N. (1997). “Bohmian mechanics and the meaning of the wave function”. In R. S. Cohen, M. Horne, & J. Stachel (Eds.), *Experimental metaphysics - quantum mechanical Studies for Abner Shimony*, Volume one; Boston studies in the philosophy of science 193. Kluwer Academic Publishers.

Earman, J. (2002). “What time-reversal invariance is and why it matters”. *International Studies in the Philosophy of Science*, 16: 245-264.

Ellis, G. F. R (2013b). “The arrow of time and the nature of spacetime”, *Studies in History and Philosophy of Modern Physics*, 44: 242-262.

Ellis, G. F. R. (2013a). “Space time and the passage of time”, en V. Petkov (ed.), *Springer Handbook of Spacetime*. Heidelberg: Springer. arXiv:1208.2611.

Esfeld, M. (2014). “Quantum humeanism, or: physicalism without properties”. *The Philosophical Quarterly*, vol. 64, n° 256: 453-470.

Esfeld, M. and Gisin, N. (2014). “The GRW flash theory: a relativistic quantum ontology of matter in space-time?”. *Philosophy of Science,* 81 (2): 248-264.

Esfeld, M. and Sachse, C. (2011). *Conservative Reductionism*. London: Routledge.

Everett, H. (1957). “Relative state formulation of quantum mechanics.” *Reviews of Modern Physics,* 29: 454-462.

Faye, J. (1991). *Niels Bohr, his Heritage and Legacy*. Kluwer: Springer.

Faye, J. (2014). “Copenhagen Interpretation of Quantum Mechanics”. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. URL = http://plato.stanford.edu/archives/fall2014/entries/qm-copenhagen/

Friederich, S. and Evans, P. (2019) “Retrocausality in quantum mechanics”. In: *The Stanford Encyclopedia of Philosophy*, E. Zalta (ed), URL= https://plato.stanford.edu/entries/qm-retrocausality/

Fuchs, C. (2010). “Qbism, the Perimeter of Quantum Bayesianism.” https://arxiv.org/abs/1003.5209.

Fuchs, C., Mermin, N.D., and Schack, R. (2014). “An Introduction to QBism with an Application to the Locality of Quantum Mechanics”. *American Journal of Physics*, 82(8): 749–54.

Ghirardi, G. (2016). “Collapse theories”, in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, URL =http://plato.stanford.edu/archives/win2016/entries/qm-collapse/

Ghirardi, G. and Rimini, A. (1990). “Old and New Ideas in the theory of quantum measurement”. In Miller, A. (ed.), *Sixty-two years of Uncertainty*, New York: Plenum Press, pp. 167-192.

Ghirardi, G.C., Grassi, R., and Benatti, F., (1995). “Describing the macroscopic world—Closing the circle within the dynamical reduction program”, *Foundations of Physics*, 25: 5-38

Ghirardi, G.C., Rimini, A., and Weber, T. (1986). “Unified dynamics for microscopic and macroscopic systems”, *Physical Review*, D34: 470.

Gisin, N., 1984, “Quantum measurements and stochastic processes”, *Physical Review Letters*, 52: 1657, and “Reply”, *ibid*., 53: 1776.

Goldstein, S. (2017). “Bohmian Mechanics”. In Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/sum2017/entries/qm-bohm/>

Goldstein, S. and Zanghi, N. (2013). “Reality and the role of the wave-function in quantum theory”. In D. Z. Albert, & A. Ney (Eds.), *The wave function: Essays in the Metaphysics of quantum mechanics*, New York: Oxford University Press, p. 91-109.

Harrigan, N. and Spekkens, R. (2010). “Einstein, incompleteness, and the epistemic view of quantum states”. *Foundations of Physics*, 40(2): 125-157.

Healey, R. (1981). “Statistical theories, quantum mechanics and the directedness of time”. In Healey, R. (ed.), *Reduction, Time and Reality*, Cambridge: Cambridge University Press, pp. 99-128.

Healey, R. (2002). “Can physics coherently deny the reality of time?”, in Callender, C. (ed.), *Time, Reality and Experience*, Cambridge: Cambridge University Press, pp. 293-316.

Heisenberg, W. (1958). *Physics and Philosophy*. New York: Harper and Brothers.

Henderson, J. (2010). “Classes of Copenhagen interpretations: Mechanisms of collapse as typologically determinative”. *Studies in History and Philosophy of Modern Physics*, 41: 1-8.

Kent, A. (1989). “’Quantum Jumps’ and Indistinguishability’, *Modern Physics Letters A*, **4(19)**: 1839–1845.

Lazarovici, D., Oldofredi, A. and Esfeld, M. (2018). "Observables and unobservables in quantum mechanics: how the no-hidden-variables theorems support the Bohmian ontology". Entropy, 20: 381.

Lombardi, O., Fortin, S., Castagnino M. y Ardenghi, J. S. (2011). "Compatibility Between Environment-Induced Decoherence and the Modal-Hamiltonian Interpretation of Quantum Mechanics". *Philosophy of Science*, **78**: 1024-1036.

London, F. and Bauer, E. (1939). "The theory of observation in quantum mechanics". In Wheeler and Zurek, *Quantum Theory and Measurement*, Princeton, NJ: Princeton University Press, pp. 217-259.

Lucas, J. (1999). "A century of time". In Butterfield, J (ed.), *The Arguments of Time*, Oxford: Oxford University Press, pp. 1-20.

Maudlin, T. (1995). "Three measurement problems". *Topoi*, 14: 7-15.

Ney, A. (2012). "The status of our ordinary three dimensions in a quantum universe". *Noûs*, 46(3): 525–560.

Ney, A. (2013). "Ontological Reduction and the wave function ontology". In D. Z. Albert, & A. Ney (Eds.), *The wave function: Essays in the Metaphysics of quantum mechanics*, New York: Oxford University Press, p. 168-183.

North, J. (2011). "Time in Thermodynamics", in C. Callender (ed.), *The Oxford Handbook of Philosophy of Time*. Oxford: Oxford University Press, pp. 312-350.

Pearl, P. (1989). "Combining stochastic dynamical state-vector Reduction with spontaneous localization". *Physical Review A*, 39(5), 2277-2284.

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

Peruzzi, G. and Rimini, A. (2000). "Compoundation invariance and Bohmian mechanics". *Foundations of Physics*, 30: 1445-1472.

Petersen, A. (1963). "The Philosophy of Niels Bohr". *Bulletin of the Atomic Scientist*, vol. 19: 8-14.

Popper, K. (1982). *Quantum Theory and the Schism in Physics.* London: Hutchinson.

Price, H. (2002). "Burbury's last case: the mystery of the entropic arrow". In Callender, C. (ed), *Time, Reality and Experience*, New York: Cambridge University Press, pp. 19-56.

Roberts, B. (2010). "How to time reverse a quantum system", at http://jamesowenweatherall.com/SCPPRG/RobertsBryan2011Man_TimeReversal_Irvine.pdf.

Saunders, S., Barret, J. and Wallace, D. (2010). *Many Worlds? Everett, Quantum Theory and Reality*. Oxford: Oxford University Press.

Savitt, S. (1995). *Time's arrow today*. New York: Cambridge University Press.

Shimony, A. (1990). "Desiderata for modified quantum dynamics", in *PSA 1990*, Volume 2, A. Fine, M. Forbes and L. Wessels (eds), Philosophy of Science Association, East Lansing, Michigan.

Skow, B. (2010). "On a symmetry argument for the guidance equation in Bohmian Mechanics". *International Studies in the Philosophy of Science*, 24: 393-410.

Spekkens, R. (2007). “Evidence for the epistemic view of quantum states: a toy theory”. *Physical Review A*, 75, 032110.

Stone, A. D. (1994). “Does the Bohm Theory Solve the Measurement Problem?” *Philosophy of Science,* 61(2): 250–266.

Sutherland, R. (2008). “Causally Symmetric Bohm Model”. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics,* 39(4): 782–805.

Tumulka, R. (2006). “A Relativistic Version of the Ghirardi–Rimini–Weber Model’, *Journal of Statistical Physics,* **125**: 821–840.

Valentini, A. (1997). “On galilean and lorentz invariance in pilot-wave dynamics”. *Physics Letters A* 228: 215-222.

Wallace, D. (2012). *The emergent multiverse*. Oxford: Oxford University Press.

Watanabe, S. (1965). “Conditional Probability in Physics”. *Progress of Theoretical Physics Supplement, Extra Number:* 135-167.

Zeilinger, A. (2005). “The Message of the Quantum.” *Nature*, 438(7069): 743.