

# Roads to the past: how to go and *not* to go backward in time in quantum theories

Cristian López

## Affiliations

University of Buenos Aires (Argentina), Department of Philosophy - National Scientific and Technical Research Council (CONICET, Argentina) -University of Lausanne (Switzerland), Faculty of Humanities  
2224 Perito Moreno, 5501, Mendoza, Argentina  
E-mail address: [lopez.cristian1987@gmail.com](mailto:lopez.cristian1987@gmail.com)  
ORCID: [0000-0002-2883-4037](https://orcid.org/0000-0002-2883-4037)

## Abstract

In this article I shall defend, against the conventional understanding of the matter, that two coherent and tenable approaches to time reversal can be suitably introduced in standard quantum mechanics: an “orthodox” approach that demands time reversal to be represented in terms of an anti-unitary and anti-linear time-reversal operator, and a “heterodox” approach that represents time reversal in terms of an unitary, linear time-reversal operator. The rationale shall be that the orthodox approach in quantum theories assumes a *relationalist* metaphysics of time, according to which time reversal is nothing but motion reversal. But, when one shifts gears and turn to a *substantivalist* metaphysics of time the heterodox approach to time reversal in quantum mechanics comes up in a more natural way.

## Key words

Time reversal – metaphysics of time – substantivalism – relationalism – quantum theories

## 1. Introduction

There is a well-seated tradition tracking back to the work of Eugene Wigner (1932) that has formally characterized the notion of time reversal in quantum theories in terms of an anti-unitary and anti-linear time-reversal operator. According to such an approach, time reversal must be mathematically represented by an operator  $T_A$  that not only transforms the variable  $t$  as  $t \rightarrow -t$ , but also performs a complex conjugation ( $\psi \rightarrow \psi^*$ ) and changes momentum’s sign ( $P \rightarrow -P$ ). This view has been supported by the overwhelming majority of physicists and philosophers of physics, both in its formal aspect (Wigner 1932, Gibson and Pollard 1976 and the majority of specialized textbooks) as well as in its conceptual bases (Sachs 1987, Roberts 2017). I shall call it “the orthodox approach” to time reversal in quantum mechanics (OA thereafter).

Despite this widely-extended consensus, there is an alternative view on time reversal that characterizes it in terms of a unitary and linear time-reversal operator,  $T_U$ . This approach typically comes up in the literature in two quite different ways: (a) as an unacceptable way to formally represent time reversal in quantum theories (see for instance Gasiorowicz 1966: 27); and (b) as a

more genuine and broader way to formally represent time reversal in physics (see Albert 2000, Callender 2000). Overall, this approach claims that time reversal must be mathematically represented by  $t \rightarrow -t$ , and that other magnitudes' behavior under time reversal should be worked out by way of examining whether they are time-derivative or not. I shall call this approach the "heterodox approach" (HA from then on).

Which approach is the correct one? At first glance, this is not a purely philosophical question inasmuch as the notion of time reversal involves scientific and empirical research on symmetries in fundamental physics. But it is not an exclusively scientific question either, inasmuch as the notion of time reversal has been conceptually insightful to face genuine metaphysical concerns as that of whether time has a privileged direction. However, there is a further and deeper sense in which the question is eminently philosophical: not only is time reversal understood differently in each approach, but they also endorse unlike metaphysical commitments with respect to the nature of time. I think that the literature has largely overlooked this strongly philosophical side of the above-mentioned question, and has rather hinged on its scientific aspects solely.

In this paper, I shall develop further this overlooked philosophical side so as to show that *both* approaches can be suitably defended, although on different metaphysical grounds. In particular, I shall on the one side argue that OA involves a *relationalist metaphysics* of time, according to which time reversal is *nothing but* motion reversal. Within this underlying metaphysical framework,  $T_A$  arises as the only way to correctly represent time reversal in quantum theories. On the other side, I shall claim that a proper defense of HA can be reliably pursued on an alternative metaphysical background: as long as one holds a *substantialist metaphysics* of time, one can properly endorse HA and neutralize most of the arguments favoring  $T_A$  against  $T_U$ . The article overall aims to show, against the conventional understanding on the matter, that there is not one single way to represent time reversal in physical theories to the extent that there is not univocal understanding of what time reversal is, metaphysically speaking.

The structure of the article is as follows. In Section 2, I will give a general presentation of time reversal and time-reversal invariance in physics. In Section 3, I will introduce OA and HA. In Section 4, I will put forward some physical arguments to support OA, and then I will show its underlying metaphysical component (relationalism). In Section 5, I will argue that HA can be

successfully defended from a substantivalist metaphysics of time, and I will show how OA's arguments against HA can be overcome. In Section 6, final remarks.

## 2. Time reversal and time-reversal invariance in physics

Symmetries, as invariances with respect to a transformation  $X$ , have increasingly gained relevance in physics (see Brading and Castellani 2007, Baker 2010, Caulton 2015). In particular, the symmetries that physical theories have or lack shed light on the type of structure required by those theories. Time reversal and time-reversal invariance (or time symmetry) are just an instance of this: whether a physical theory is time-reversal invariant or not would illuminate the type of temporal structure assumed by the theory. Therefore, if a physical theory fails to be time-reversal invariant, then the theory *structurally* distinguishes the past-to-future direction from the future-to-past one. This is not only interesting from a purely scientific view, but also from a philosophical one: time-reversal invariance has, for instance, profound implications for the arrow of time debate (Horwich 1987: 52-55, Price 1996: 116, Wallace 2012).

To the extent that the time symmetry is an invariance under an operator that stands for a change of direction of time, a good deal of what time symmetry actually means is encoded in how one characterizes such a *time-reversal operator* (T-operator henceforth). First, it is relatively uncontroversial in the physicist's and philosophers of physics' community that time reversal is a theory-dependent notion as long as the proper characterization of the time-reversal operator changes across theories (but see Savitt 1996 and Peterson 2015 for discussion<sup>1</sup>). Secondly, time reversal is often understood as a transformation acting upon a *dynamical* law by means of an operator  $T$  that reverses time (see Castagnino and Lombardi 2009 for a defense of why time reversal applies to dynamical laws). According to this view, *time-reversal invariance* is a property that dynamical laws may instantiate or not: the property of 'being invariant under the *action* of reversing the direction of time by means of a *T-operator*'. When a dynamical law is  $T$ -invariant, it is said that the equation has a pair of temporally-mirrored physical evolutions (i.e. solutions of the

---

<sup>1</sup> Despite this theory-dependency, some attempt to coarsely characterize time reversal in a theory-independent way. Steven Savitt (1996: 12-14) has inventoried three kinds of time transformations that might be fairly called 'time reversal'; some of them boil down to a proper characterization within a physical theory, but others intend to be broader. Recently, Daniel Peterson (2015) argued that some accounts of time reversal (which he calls 'intuitives') start out by characterizing a time-reversal operator from a theory-independent perspective.

equation): one being future-directed with  $+t$ , and the other one past-directed with  $-t$ . The property of ‘being  $T$ -invariant’ can therefore be defined as following:

A dynamical law  $L$  is  $T$ -invariant iff, if  $e_t$  is a solution of  $L$ , then  $Te_t$  exists and is also a solution of  $L$ .

Naturally, the non-existence of  $Te_t$  means that the dynamical law lacks the property of being  $T$ -invariant, that is, the law is asymmetric under the action of inverting the direction of time in it.

### 3. The orthodox and the heterodox approaches to time reversal in quantum theories

One needs to put some flesh on the bones of the  $T$ -operator. So, how is the  $T$ -operator formally characterized in the quantum context? Most physics textbooks (see for instance Gibson and Pollard 1976, or Ballentine 1998) commonly start out by warning against, so to speak, classical expectations that time reversal merely maps  $t \rightarrow -t$  by means of a *unitary*  $T$ -operator,  $T_U$ . Taking for instance Schrödinger’s equation<sup>2</sup>, textbooks typically bring up an *anti-unitary* operator ( $T_A$ ) that not only does it flip the  $t$ ’s sign,  $t \rightarrow -t$ .

$$(1) TH|\psi\rangle = i\hbar \frac{\partial T|\psi\rangle}{\partial Tt} = H|\psi\rangle = -i\hbar \frac{\partial |\psi\rangle}{\partial t}$$

but also takes the complex conjugate (K) on (2).

$$(2) KH|\psi\rangle = -i\hbar \frac{\partial K|\psi\rangle}{\partial t}$$

Allegedly, this is how one gets a time-reversed Schrödinger’s equation, wherein  $H = H^*$

$$(3) H|\psi^*\rangle = i\hbar \frac{\partial |\psi^*\rangle}{\partial t}$$

Evidently, Schrödinger’s equation is  $T$ -invariant under  $T_A$ <sup>3</sup>. Strictly speaking,  $T_A$  is a combination of a unitary operator and an anti-unitary one:  $T_A = UK$ , where  $U$  is the unitary operator and  $KzK^{-1} = z^*$ . Notably,  $T_A$  is also demanded *by definition* (see, for instance, Ballentine 1998: 377-

<sup>2</sup> I will completely circumscribe myself to Schrödinger’s equation. Some interpretation-dependent dynamical equations can also be considered as part of the formal apparatus of quantum mechanics, for instance in GRW or Bohmian Mechanics.

<sup>3</sup> To be clear: ‘being  $T$ -invariant under  $T_A$ ’ means ‘being  $T_A$ -invariant’ as  $T_A$  specifies the form of the  $T$ -operator.

378) to change the sign of momentum,  $TPT^{-1} = -\mathbf{P}$ . One should bear in mind that such stipulation establishes a smooth continuity between time reversal in classical mechanics and in quantum mechanics: not only does time reversal transform the same magnitudes similarly, but it also keeps the theory's fundamental equation invariant. Indeed, the features of the unitary operator  $U$  are a consequence of the classical conditions for time reversal, namely,  $TxT^{-1} = x$ ,  $TPT^{-1} = -\mathbf{P}$  and  $T\sigma T^{-1} = -\sigma$  (see Sachs 1987: 34).

So, I will take OA as claiming that

**OA** Time reversal in quantum theories must be represented by a  $T$ -operator whose form is given by the  $T_A$ -operator

Either by strengthening OA or by pursuing a positive defense, time reversal has been represented differently. Some authors have argued that a  $T$ -operator should not involve any other property over and above that of turning  $t$  around (see, iconically, Callender 2000, Albert 2000; Costa de Beauregard 1980 also defends such view in quantum field theory). So, when one writes Schrödinger's equation down and then applies a time-reversal operator  $T$  that merely changes the sign of  $t$  and of all those quantities expressed in function of time (or non-basic magnitudes, in Albert's vocabulary), one obtains the equation (1) again

$$(1) TH|\psi\rangle = i\hbar \frac{\partial T|\psi\rangle}{\partial Tt} = H|\psi\rangle = -i\hbar \frac{\partial |\psi\rangle}{\partial t}$$

And that is all what one should expect time reversal to carry out. Period. This is essentially HA's spirit.

A couple of clarifications are in order here. To begin with, as opposed to OA, this heterodox way to define a  $T$ -operator does not take into account whether it must change the sign of momentum or not. In fact, it should not. In the second place,  $T$  is now a unitary and linear operator,  $T_U$ , as it does not take the complex conjugation on states. This feature is particularly troublesome here for it is what produces a minus sign on the right side of the equation. In other words, as  $T$  is a unitary-linear operator, then  $THT^{-1} = -H$ , entailing that if  $|\psi\rangle$  is an eigenstate of the Hamiltonian with energy  $E$ , then the temporally mirrored eigenstate  $T^{-1}|\psi\rangle$  should involve negative energies  $-E$ . It is quite easy to see that there is an obvious asymmetry in (1) (indeed, a

very deep and radical one) as one obtains no solution at all when time direction is inverted this way. In this respect, the Schrödinger equation is non-T-invariant under  $T_U$ <sup>4</sup>.

I will thus take HA as claiming

**HA** Time reversal in quantum theories must be represented by a T-operator whose form is given by the  $T_U$ -operator

#### **4. Underpinning OA: three physical arguments and a *relationalist* metaphysics of time**

There are at least three closely-related arguments to underpin OA, and to thereby cast HA aside. Indeed, the three are built on 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 “the involution-based 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), and I will call it “the Hamiltonian-based argument”; and the third one establishes that momentum (in this case, the momentum operator,  $\mathbf{P}$ ) must change its sign under time reversal, and I will call this third argument “the momentum-based argument”. I will spell them out before coming to set the relationalist metaphysical background that supports them.

##### **4.1 Three physical arguments**

**I.** Let us start with the *involution-based argument*. Wigner’s introduction to time reversal begins by claiming that time reversal is a transformation such that, when the following operations are sequentially performed, one obtains the identity:

$$(4a) \text{ time displacement by } t \times \text{time reversal} \times \text{time displacement by } t \times \text{time reversal} = \mathbf{I}$$

Taking  $T$  as any  $T$ -operator that meets (4a), one obtains that:

$$(4b) T[U_{\Delta t_2} T(U_{\Delta t_1} s_0)] = s_0$$

---

<sup>4</sup> To be clear: ‘being non-T-invariant under  $T_U$ ’ means ‘being non- $T_U$ -invariant’ as  $T_U$  specifies the form of the T-operator

Where  $s_0$  is the initial state, and  $\Delta t_1 = t_1 - t_2 = t_2 - t_1 = \Delta t_2$ . What this means is that time reversal represents an *involution*, where after applying time reversal twice, one should obtain the initial state again. So, an appropriate  $T$ -operator in quantum mechanics must meet conditions (4a) and (4b): to be a  $T$ -operator *is* to be an operator that *generates* an involution as specified above. Next, Wigner establishes that  $T$  has to preserve transition probabilities as well (otherwise, an involution is no longer possible)

$$(4c) |\langle \psi | \varphi \rangle| = |\langle T\psi | T\varphi \rangle|$$

Beyond this characterization of the  $T$ -operator in quantum theories, Wigner has so far remained silent about the specific form of  $T$ . His famous theorem claims that  $T$  must be either unitary ( $T_U$ ) or anti-unitary ( $T_A$ ). As  $T_U$  meets neither (4a), (4b) nor (4c), 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 straightforwardly sketched as follows:

- (1) Assume that  $T_U$  fairly represents time reversal
- (2a) If  $T_U$  represents time reversal, the  $T_U$  generates an involution (meeting 4a, 4b and 4c)
- (3a) As matter of fact,  $T_U$  does not generate an involution as it does not meet above-mentioned conditions
- (C)  $T_U$  does not represent time reversal

As there are only two games in town, it directly follows that  $T_A$  represents time reversal.

**II.** Let us move on to the *Hamiltonian-based argument*. Schematically represented,

- (1) Assume that  $T_U$  fairly represents time reversal
- (2b) If  $T_U$  represents time reversal,  $T_U$  keeps the Hamiltonian invariant
- (3b) As matter of fact,  $T_U$  does not keep the Hamiltonian invariant as  $T_U H T_U^{-1} = -H$
- (C)  $T_U$  does not represent time reversal

Let me spell premises (2b) and (3b) out. The reason of why the Hamiltonian must remain invariant under time reversal is that, otherwise, an involution 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). Quantum states are said to be “physically meaningless” (in the light of non-relativistic quantum mechanics itself) when their Hamiltonian’s spectrum features negative energies. Putting it drastically, physical systems with negative energies must *not* be considered as *quantum mechanics* systems any longer. I will take fully for granted that a Hamiltonian unbounded from below is physically meaningless, from the non-standard quantum mechanics perspective<sup>5</sup>. 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.

**III.** Finally, the *momentum-based argument* can be sketched as following

- (1) Assume that  $T_U$  fairly represents time reversal
- (2c) If  $T_U$  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 represents time reversal

The reasons for momentum’s sign to change under time reversal turns out to be somewhat unclear in the literature (though Roberts 2017 introduces a clearer and purely quantum-based argument to ground this premise). 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, time reversal flips the sign of momentum in quantum mechanics *by definition* (1998: 377-378), and, in the same vein, Albert Messiah (1966)

---

<sup>5</sup> The predicates “positive” or “negative” for the energy spectrum, or “unbounded from below/from above” for Hamiltonians are actually matter of convention. So, the argument could not hinge on which predicate one adopts to describe the system properly. The real problem is not whether or not the Hamiltonian is unbounded from below. The problem is if one starts with a Hamiltonian unbounded from above (but bounded from below) *and* one ends up with a Hamiltonian unbounded from below (but bounded from above) after a transformation. In some sense, the problem is if there is in general a bound *at all*. More precisely, a *specific* Hamiltonian must be bounded (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.



simply defines time reversal as transforming  $r$  (position) and  $p$  (momentum) into  $r$  and  $-p$  respectively (see Messiah 1966: 667, see Davies 1974: 24-25 for an akin definition). 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).

#### 4.2 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 is widely known, there are metaphysically two views about the nature of time. *Substantivalists* with respect to time claim that time is an entity that exists independently of events and things placed within it. On the contrary, *relationalists* support the idea that time intrinsically depends on events and on things in it. My point here is that the ways in which time-reversal operators are formally characterized follow from being engaged in one of the two sides in the metaphysical debate about the nature of time. Particularly, OA relies on a relationalist metaphysics of time, while HA may be soundly defended from a substantivalist one.

There are many different sorts of relationalist-like views 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 what it is 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, Barbour and Bertotti (1982) have argued for a Machian relationalism in physics according to which an absolute temporal-ordering structure is assumed for classical mechanics (see Gryb and Thébault 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 seems to be defended by Michael Esfeld et al. (2018) as change not only exhibits a temporal order but also a *direction* (Esfeld et al. 2018: 31).

Be that as it may, I will henceforth take the 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 (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.

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. However, I am particularly interested in how this metaphysical background underlies the defense and characterization of time reversal as  $T_A$  in quantum theories.

To start with, the  $T: t \rightarrow -t$  transformation must not be taken too seriously. It would be naïve to take  $T$  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 equation of *motions*, the physical meaning of time reversal is entirely exhausted by the dynamically relevant transformations that relate 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  $\Delta t = t_2 - t_1$ . To say “I will time reverse the system by applying  $T$ ” is simply a shortcut for “I will motion reverse the system by applying a bunch of dynamically relevant transformations to take the system back to the initial state”. In this toy example, the  $T$ -operator should be properly spelt out into the moment transformation  $P: p \rightarrow -p$ , and the position transformation  $X: x \rightarrow x$ . The transformation  $T: t \rightarrow -t$  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 moment transformation and the position transformation as specified above.

In the light of this, the theory-dependent nature of the  $T$ -operator is quite clear. 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 time translation, and reversing them properly so as to get to the state one began with. The defense of  $T_A$  in quantum theories is just an instance of this more general maneuver. Thus, a  $T$ -operator from a relationalist metaphysics of time is defined as follows

- $T_{Rel}$
- (a) A physically meaningless re-parametrization of  $t$  by  $T: t \rightarrow -t$
  - (b) A change of all dynamically relevant magnitudes so as to generate a moving-backward system (*involution*), which is expressed by extensionally specifying the bunch of dynamically relevant transformations.

This relationalist metaphysical background undoubtedly underlies Wigner’s definition (4a) of time reversal as an *involution*. According to definition (4a), 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 assures that we are 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 takes this point for granted in claiming that time reversal is nothing but *motion* reversal (Wigner 1932: 325, Sakurai 2011: 266), warning us that “no metaphysical notion of reversal of the direction of the flow of time is involved” by means of the  $T$ -operator (Gibson and Pollard 1976: 177) or that the notion of “time reversal is misleading” and one should rather refer to motion reversal (Ballentine 1998: 377). Strictly speaking, the notion of time reversal is neither misleading, nor metaphysically misguided: as time is just an abstraction at which we arrive by means of motion (paraphrasing Mach’s expression, Mach 1919: 224), time reversal is simply an abstraction of motion reversal.

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 outwardly regards whether physical theories imply that the change 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 prescribes *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.

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

Metaphysics comes first in the sense that determines not only what *time* reversal actually *is* but also *upon what* it is supposed to act. There is no shadow of doubt that on a relationalist basis, time reversal *must* be represented by  $T_A$ . But when one shifts gears, OA's arguments lose its dispositive force. Suppose now that one has sound reasons to decline such a relationalist basis for time and to rather turn to a substantivalist one. How should time reversal be now metaphysically and formally characterized?

As in the case of the relationalism, there are also many versions of substantivalist views on time (see Sklar 1974, Earman 1989, Pooley 2013 for comprehensive overviews of different kinds of substantivalisms). 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, the 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 which lies outside the scope of dynamics<sup>6</sup>.

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.

Naturally, **S1** and **S2** 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, irrespectively of changes in the world, and thus the structure of change supervenes upon the structure of time. In this sense, time can be said to be “substantial”.

In the light of this view, how must time reversal, thus, be characterized? For the one thing, as time stands by itself, there are no metaphysical reasons for time 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 keep with respect to time within a specific theory (e.g. whether they are first-time derivative, and so on). For another thing, as time is now a physically meaningful external parameter, an inversion of the direction of time must outwardly mean an inversion of the external parameter itself. This view of time reversal seems to be in Jill North’s 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)

---

<sup>6</sup> Certainly, I would be naïve to infer 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.

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 have not to be considered as an unphysical re-parametrization. Then, a  $T$ -operator from **S1** and **S2** is 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

To begin with, the  $T$ -operator is no longer demanded to produce an *involution*. Metaphysically, this is clear from a substantialist 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 horses. Under specific circumstances, an inversion of the direction of time *may* lead to an involution; for instance, when the dynamically relevant magnitudes that define the state and/or the direction of motion are first-time derivative (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 an involution, 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 non-relativistic quantum mechanics from  $T_{Sub}$ ). As a result, the involution-based argument no longer runs. Particularly, because the premise (2a) does not hold to be true in the light of  $T_{Sub}$ .

Let me make the same point slightly differently. According to Wigner's definition (4a) of time reversal, the existence of the second time translation must be guaranteed to set the form of  $T$  properly.  $T_U$  was precisely discarded for not being able to meet (4a), but this rationale no longer runs when a substantialist view on time is rather assumed. In some sense,  $T_{Sub}$  is at odds with Wigner's definition of time reversal because definition (4a) demands further actions to represent time reversal of which are actually necessary conforming to S1, S2 and  $T_{Sub}$ . Under this view, time reversal looks much more like a *reflection* than like an involution. Hence,  $T_U$  can be re-established on the proper metaphysical ground to the extent that it fairly represents such a reflection (see Arntzenius 1997 and Savitt 1996: chapter 1 for time reversal as a reflection).

Besides the involution-based argument and the metaphysical reasons to support HA, the rest of the physical arguments raised by OA against HA must also be overcome. Let us see in which way HA along with  $T_{Sub}$  cope with the momentum-based and the Hamiltonian-based argument.

### **(a) Overcoming momentum-based argument**

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: if time reversal aims at representing something like a backward movement, then reversing the sign of momentum seems mandatory as what one is in need of restoring a past-headed physically possible evolution. And this partially explains 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).

Craig Callender (2000) and David Albert (2000) have singly introduced non-standard approaches to time reversal wherein the transformation rule for momentum does not hold in some cases. Let me take a slightly different path to overcome the momentum-based argument, and claim that there is a reasonable way from  $T_{Sub}$  in which the transformation rule for momentum could not hold in some cases, specially, in quantum mechanics.

Recall that the  $t \rightarrow -t$  transformation is not a simple physically meaningless shortcut according to  $T_{Sub}$ , but it concentrates the whole physical meaning of what time reversal is. Hence, the  $T$ -operator does not need to be reduced to or explicated as a bunch of dynamically relevant transformations. Whether momentum’s sign must flip under time reversal or not mainly depends on the sort of relation that the magnitude holds with time according to the physical theory at issue. Let me spell it out by comparing Newtonian classical mechanics and non-relativistic quantum mechanics. In the former, to the extent that velocity involves the first derivative with respect to  $t$ , one *logically* obtains that particles’ velocities must also be transformed under  $T$ . Then, 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  $T$ -operator flips the sign of velocity because velocity is a first-time derivative magnitude. Thus, as long as momentum is defined within the theory as

$$(5) p = mv = m \frac{\Delta d}{\Delta t}$$

it follows *logically* (taking Callender's expression, Callender 2000) that momentum must also change its sign under  $T$ . To be clear: momentum changes under time reversal because, firstly, the time-reversal operator was defined as performing  $t \rightarrow -t$  *and* changing all time-derivative magnitude, and secondly, because momentum is defined as a first-time-derivative magnitude in Newtonian classical mechanics. There is no need to wonder whether time reversal must in general involve a reversion of the direction of motion, or if it must feature further properties by definition:  $T_{Sub}$  only commits to such a minimum  $T$ -operator which acts just like a *reflection*. Obviously, to claim that such a reflection is physically meaningless is to accept neither **S1** nor **S2**.

It is worth noticing that in Newtonian classical mechanics, a time-reversed evolution (provided that external forces have been ruled out) in fact *looks like* an inversion of the direction of motion, but this is a side effect of both  $T_{Sub}$  and the mathematical-physical structure of Newtonian classical mechanics (since it involves certain laws and not others, momentum is defined as time-derivative and rather than in some other way, and so forth). That a time-reversed evolution looks like a going-backward movement is an *effect* or *consequence of* and not a *definition* of time reversal.

So, according to  $T_{Sub}$ , when one goes on to apply  $T$  in the realm of quantum mechanics, one should not *beforehand* expect momentum to change under time reversal. First of all, as the definitions of quantities change across theories, one should firstly draw the attention to how momentum is formally represented in standard quantum mechanics. As it is well-known, momentum is now an operator (**P**) defined as

$$(6) \mathbf{P}(\psi) = -i\hbar \frac{\partial \psi}{\partial x}$$

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. So, let us look at the facts from  $T_{Sub}$ 's viewpoint. On the one hand,  $T$  need not feature any reference to any particular magnitude, so why should one beforehand expect that a certain magnitude behaves in a specific way? On the other hand,  $T_{Sub}$ 's definition refers *intensionally* to how other magnitudes behave under time reversal, a point implicitly suggested by Callender (2000): wherever you find a time-derivative magnitude in a theory, invert its sign accordingly. So, why should one expect a space-



derivative magnitude to change sign under  $T$ ? That is the reason why it looks quite unnatural to reverse the sign of momentum in quantum mechanics from  $T_{Sub}$  and HA: it is not a time-derivative magnitude and it is thus out of the range of the  $T$ -operator.

**(b) Overcoming 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 Schrödinger’s equation must render a physically meaningful solution upon change in the direction of time. So, it is a fundamental feature of  $T_A$  in non-relativistic quantum mechanics that it prevents negative energy solutions from coming up. It would be no possible to generate the second time translation in Wigner’s (4a) if the  $T$ -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. Non-relativistic quantum mechanics undoubtedly sets up a strong relation between time and the system’s energy (its Hamiltonian). In its simplest form, the Hamiltonian operator can be written down as

$$(7) H = i\hbar \frac{\partial}{\partial t}$$

As in the case of momentum in Newtonian classical mechanics, such definition points out to a tied-up relation between the Hamiltonian and time. From  $T_{Sub}$ , the Hamiltonian’s sign is *logically* inverted under time reversal because it is defined in function of time. Arguing that the  $T$ -operator should leave invariant certain quantities in quantum mechanics misses the point, since the proper definition of  $T_{Sub}$  and the metaphysical background of HA do not reflect such concern. To be clear:  $T_U$  changing of the 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.

However, one can go deeper into the reasoning. One of the motivations not to regard  $T_U$  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 representing an involution. As mentioned above, the generation of such an involution is a sine-

qua-non condition for representing time reversal properly from  $T_{Rel}$ ; 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 represent an involution, so the argument is neutralized.

It is worth stressing that non-relativistic quantum mechanics is non-invariant under  $T_U$ , that is, non-time-reversal invariant from  $T_{Sub}$  and HA. At first glance, this seems to be scandalously problematic, and as a compelling reason to throw the whole package away. However, I think that things must be approached from the appropriate angle. The fact that non-relativistic quantum mechanics 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. 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 a physical theory), *and* time is nothing but motion, then this directly entails that theories must be time-reversal invariant. But the implication heavily relies on taking the conjunction as true, and clearly  $T_{Sub}$  and HA might remain reluctant to accept it. Under HA, it makes full sense that a theory is simultaneously motion-reversal invariant and non-time-reversal invariant.

## 6. Final remarks

Let me summarize the points I've made along the article and put all pieces together. I have argued in favor of the legitimacy of two different approaches to represent time reversal in non-relativistic quantum mechanics against the conventional understanding that only one can be correct. The grounds for doing so was that OA, which promotes  $T_A$  as the right way to represent time reversal (defined in terms of  $T_{Rel}$ ) in standard quantum mechanics assumes a *relationalist* metaphysics of time (which I've summarized in the tenets **R1** and **R2**), and the physical arguments to support OA

are largely dependent on such a metaphysical stance. But when one turns to a *substantivalist* metaphysics of time (summarized in the tenets **S1** and **S2**), an alternative approach, HA, comes up more naturally in terms of  $T_U$ . Many of the arguments to throw HA away can be neutralized by defining time reversal in terms of  $T_{Sub}$ . When methodological and physical considerations are set alongside metaphysical commitments, the two approaches to time reversal emerge as coherent and tenable.

This conclusion paves the way to re-think some of the metaphysical problems that have largely depended upon the notion of time reversal in physics. For instance, the problem of the arrow of time is usually posed in terms of whether our fundamental dynamical equations are time-reversal invariant. Typically, such laws are taken to be time-reversal invariant and, thereby, blind to the direction of time. However, my analysis in this article offers some grounds for caution regarding such claims. Much of the philosophical meaning of the notion of ‘time reversal’ lies on our understanding of the nature of time, upon which time reversal is supposed to act. Not only are the answers to those questions non-univocal, but, furthermore, both ways to represent time reversal in quantum theories would enable quite different metaphysical scenarios with respect to the arrow of time debate. From time substantivalism holding  $T_{Sub}$ , one has sound reasons to claim that at least that one fundamental dynamical law (Schrödinger’s equation) is non-time-reversal invariant as it is non- $T_U$ -invariant. From a relationalist metaphysics of time, one has instead strong reason to believe the opposite and to remain sympathetic to the mainstream view.

It is worth re-stressing that nothing here implies that what physicists and philosophers of physics have extensively taken as a symmetry of the theory was never really a symmetry of the theory. In fact, bracketing the names ‘time reversal’ and ‘motion reversal’, standard quantum mechanics is as non- $T_U$ -invariant as it is  $T_A$ -invariant under any respect. The point is that one side of the debate comes to regard  $T_U$  as non-physical and to identify motion with time under the same transformation ( $T_A$ ). The other side of the debate regards  $T_U$  as a physical transformation acting upon time itself, where motion and time are metaphysically different. For this reason, nothing of the empirical or theoretical gain of  $T_A$ -symmetry disappears. From HA the situation just becomes ‘unfolded’: therein a metaphysical identification was carried out, one should break it down into two transformations: motion reversal on the one side, and time reversal on the other.

Which metaphysical background should be abandoned and which should be endorsed is still an ongoing discussion in metaphysics and philosophy of physics. Although plenty of empirical and theoretical research in physics seems to favor relationalism of (space)-time, substantivalism has not been abandoned yet, and many substantivalists still lie in wait.

## References

- 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.
- Baker, D. (2010). "Symmetry and the metaphysics of physics". *Philosophy Compass* 5: 1157-1166
- Ballentine, L. (1998). *Quantum Mechanics. A modern Development*. Singapur: World Scientific.
- Barbour, J., and Bertotti, B. (1982). Mach's principle and the structure of dynamical theories. *Proceedings of the Royal Society A*, 382, 295–306.
- 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.
- Brighouse, S. (1994). "Spacetime and Holes". *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1: 117-125.
- Callender, C. (2000). "Is time 'handed' in a quantum world?" *Proceedings of the Aristotelian Society*, 100: 247-269.
- 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.
- 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.
- Costa de Beauregard, O. (1980). "CPT invariance and interpretation of quantum mechanics", *Foundations of Physics*, 10: 513-530.
- Davies, P. (1974). *The Physics of Time Asymmetry*. Berkeley: University of California Press.

- Dasgupta, S. (2015). "Substantivalism vs. Relationalism about space in classical physics". *Philosophy Compass*, 10/9: 601-624
- Earman, J. (1989). *World enough and space-time. Absolute versus relational theories of space-time*. Cambridge, Massachusetts: MIT Press.
- Esfeld, M. and Deckert, D. (2018). *A Minimalist Ontology of the Natural World*. New York: Routledge.
- Gasiorowicz, 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*. Cambridge: Cambridge University Press.
- Gryb, Sean and Karim Thébault. 2016. "Time remains". *British Journal for Philosophy of Science*, 67: 663-705.
- Hofer, C. (1996). 'The Metaphysics of Space-Time Substantivalism.' *The Journal of Philosophy* 93: 5–27.
- Horwich, P. (1987). *Asymmetries in Time*. Cambridge, MA: MIT Press.
- Huggett, N., Vistarini, T., and Wütrich, 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.
- 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.
- 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.
- Messiah, A. (1966). *Quantum Mechanics*. New York: John Wiley and Sons,
- North, J. (2009). "Two views on time reversal". *Philosophy of Science*, 75: 201-223.
- Peterson, D. (2015). "Prospect for a new account of time reversal". *Studies in History and Philosophy of Modern Physics*, 49: 42-56
- 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

- Roberts, B. (2017). "Three myths about time reversal invariance". *Philosophy of Science*, 84, 2: 315-331
- Rovelli, C. (2004). *Quantum gravity*. Cambridge: Cambridge University Press.
- Sachs, R. (1987). *The Physics of Time Reversal*. London: University Chicago Press.
- Sakurai, J.J. and Napolitano, J. (2011). *Modern Quantum Mechanics*. San Francisco: Adison-Wesley
- Savitt, S. (1996). "The direction of time". *The British Journal for the Philosophy of Science*, 47: 347-370.
- Sklar, L. (1974). *Space, Time and Spacetime*. Berkeley: University of California Press.
- Wallace, D. (2012). "The arrow of time in physics", in Adrian Bardon and Heather Dyke (Eds.), *A Companion to the Philosophy of Time*, Wiley-Blackwell, pp. 262-281.
- Wigner, E. P. (1932). *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra*. New York: Academic Press (1959).