

Faculté de philosophie, arts et lettres

Common Sense or Quantum Eraser: What Is the Direction of Causality?

Alignment of Arrows of time, Causality and
Planification

Auteur : Doan Vu Duc
Promoteur(s) : Alexandre Guay
Lecteur(s) : Peter Verdée, Charles Pence
Année académique 2018-2019
Master de Philosophie, finalité approfondie

Common Sense or Quantum Eraser: What Is the Direction of Causality?

ALIGNMENT OF ARROWS OF TIME, CAUSALITY AND PLANIFICATION

Doan VU DUC

Promoter: Alexandre GUAY

Master program of Philosophy of science

Université Catholique de Louvain

2018-2019

We trust that time is linear. That it proceeds eternally, uniformly. Into infinity. But the distinction between past, present and future is nothing but an illusion. Yesterday, today and tomorrow are not consecutive, they are connected in a never-ending circle. Everything is connected.

—The Stranger, *Dark*, Season 1: Secrets

Introduction

To start a philosophical discussion, one interesting —and above all, funny— way to proceed is to begin with fiction and argue whether what is at stake in the said movie is coherent or not. This method often leads to metaphysical questions that can be related to a philosophical literature and hence bring satisfying answers (in the best case); so, to approach the questions concerning the direction of causality, let us consider a largely exploited subject in famous movies, namely time travel.

In movies such as *Star Trek XI*, *Back to the Future*, *Terminator* or the series *Dark*, the idea of travelling back in the past in order to change something in the present has been used in different ways. Those stories dealt with the question of “what can be changed in the past?”, some claiming that we can change anything—thus encountering the well-known Grandfather Paradox— while others claim there is nothing that can be done in the past that can change the present—hence adopting a sort of determinism. One could discuss those answers and relate them to metaphysical theories that have been elaborated¹, hence following the method. Other questions like “what is time?” or “does future already exist?” can be approached as well.

Among the numerous potential questions that can be asked when watching a science-fiction movie, there is one that I wish to consider here: “what makes time travel *in the past* so strange?” The question may look trivial and people could think there is nothing bizarre in this kind of time travel; after all, fiction is evidence that we can imagine such a thing. However, it is no surprise that movies often fail to give a consistent account of what time travel implies, which is not sufficient to satisfy a philosopher’s reflection. Furthermore, there is a striking feature of time travel that cannot be denied and yet can be hardly explained: it is the “simple” fact that we cannot change something that has already happened. Or so it seems. For, answering the question of “why can we not bring something about in the past?” requires much more than referring to the obviousness of it. And this is the philosophical starting point of the reasoning that we will explore.

More precisely, I aim to show that we have an intuition about the direction of causality, namely that it can only but go from past to future—and never the other way around. This commonsensical intuition, I defend, is due to the way we learn causality and does not qualify as an objective feature of causality. Furthermore, an experiment of fundamental physics seems to support the idea that backwards causality is possible. I will defend that this experiment—namely the quantum eraser experiment— fits at

¹ A good overview of those theories can be found in Barry Dainton’s *Time and Space* (Dainton, 2010).

least partly in our commonsensical notion of causality, but requires us to abandon the temporal asymmetry of it. The discussion calls for many concepts and philosophical tools, which I will present when needed. The plan goes as follows:

1. Discuss, with Wilfrid Sellars' philosophy, what we mean by common sense, how it can be opposed to science and how it is constitutive of our experience of the world;
2. Give a state of the art of philosophy of causality with two tendencies, the theory of propagation-interaction and the theory of manipulation;
3. Give the manifest image of causality, explain what it consists in, where it comes from and give one striking feature of it, namely the fact that the direction of causality is aligned with the temporal direction in common sense;
4. Explain the link relation between the direction of causality and the direction of planification;
5. Describe the quantum eraser experiment, and show that it counts as a case of manipulation;
6. Defend that, despite the intrinsic connection between the causal arrow and the planification arrow, common sense about the direction of causality should be abandoned; the causal arrow does not necessarily follow the temporal arrow, but other features of causality in common sense continue to apply.

Chapter 1

Manifest and Scientific Images or the Clash between Common Sense and Science

1. To Begin with

This chapter provides a distinction between the manifest image and the scientific image of “man-in-the-world” as Wilfrid Sellars calls it. The presentation of this duality of images will allow us to explore what common sense means and what it consists of, as opposed to what science says. We shall end up with two images, which we should consider as two different types of *structures*, and we will see that despite their potential incompatibility, it is hard to get rid of the manifest or the scientific image, for different reasons.

First of all, consider the notions that are said to be “commonsensical”. In philosophy, common sense refers to what people understand intuitively and naturally. More precisely, the definition of common sense I adopt refers to a structure of the mind that encompasses propositions or beliefs, recognised as self-evident by most people and that are not due to esoteric knowledge or research, but to intuitive knowledge. Intuitive knowledge is at the level of beliefs or knowledge stemming from daily life experiences, which can structure intuition. For instance, think about the fact that people who have their driver’s licence know how to drive a car, but do not know how a car works; or people living near the sea know that after a high tide follows a low tide, but again do not know why; or the obviousness that anything in the world always *falls*, except very light things such as feathers; and more generally, think about those things of everyday life that people consider obvious but still do not know why. This is what I mean by common sense.

As humans, we have notions about things in the world that we hardly question because they seem obvious—which is confirmed by the fact that (almost) all humans share the same conception. Common sense is this set of description of things which shapes how we conceive things. That is, common sense does not only consist of sentences we make about the world (i.e. “all heavy objects fall on the ground”); but also of “felt” conceptions. For example, it is very hard to describe how to ride a bicycle: it is rather a sort of sensation that we learn by trying to keep our balance on it and eventually succeed at keeping it. What happens for a driver is very comparable—this explains why a driver does not keep thinking about the intensity with which she

should press the foot pedal. And in respect to the property of “falling down”, people could hardly explain why a heavy body falls and not a light one, other than because of their heaviness or lightness, respectively. These explanations are, of course, circular and simply indicate how obvious commonsensical ideas are; and how difficult they are to explain.

Next to the sphere of commonsensical conceptions, is the sphere of scientific conceptions. At a certain moment in the history of humanity (at least, in the West), some people thought it was time to give more satisfying explanations about the world. Over the course of time, science played a great role in how humanity grew and claimed to give an accurate account of the world *as it is*. So, science could explain precisely why a low tide follows a high tide; or enabled the invention of vehicles as cars, explaining how combustion can turn a motor for instance. Somehow, science gave and gives explanations in addition to how we perceive things in common sense, only from another point of view.

However, science can sometimes be problematic; or rather, common sense can. That is, there are certain cases in which common sense and science disagree and give opposite descriptions or explanations, resulting in some incompatibility. For instance, it is a very commonsensical idea that a heavy body drops *faster* than a lighter one. So, from the same height, a piano will fall faster than a pen and hence reach the ground earlier. Of course, this idea is mistaken and there is a huge amount of formal theories about free fall in vacuum. That is, that idea is mistaken because of air that slows down the objects (depending on their contact area). The speed of a free fall in vacuum does not depend on the mass of a body, and so a piano and pen fall at the same speed, despite what is thought in common sense.

“This case is not problematic, though!” one could say. Indeed, it does not have strong implications in everyday life: it will not technically change our way of living (how we speak, or how we walk for instance). But it should change a certain way of seeing the world; and this is sufficient for what I intend to show. For if incompatibilities of this kind result in necessity of abandoning common sense, then it means that there are potentially other incompatibilities; so, other commonsensical conceptions about the world that we should replace. And these might be more fundamental conceptions than the one of the example, namely the speed of free fall.

This distinction between common sense and science has been theorised by the American philosopher Wilfrid Sellars in his paper “Philosophy and the scientific image of man”. We will now explore his ideas in order to have a more precise overview

of these two ways of perceiving the world, but also how delicate is our position if we have to transform common sense.

2. “Philosophy and the Scientific Image of Man”

In “Philosophy and the scientific image of man” (Sellars, 1963), Sellars distinguishes between the manifest image and the scientific image of man-in-the-world. These so-called images refer to two different types of conceptions which, broadly speaking, refer respectively to common sense and to the scientifically elaborated concepts. This chapter consists of the reasoning that led Sellars to this distinction; later on shall we define more precisely both images and reply to a few objections against them.

To begin with, let us come back to the distinction we proposed roughly in the previous section. For Sellars, the difference between the manifest and scientific image comes from a rather naïve one, making a difference between practical and theoretical “knowledge”. To give a basic example, consider the difference there is between a self-taught violinist and a violinist who learnt music in a musical academy in how they perceive music.² It is probable that, when listening to a piece of music, the self-taught violinist “sees” fingers travelling on the strings, or “feels” where to put her fingers to reproduce the sounds; while the academic-taught violinist “sees” music notes on a score. However, both musicians listen to the same piece and both can accurately reproduce the piece —both representations are adequate.

HOW DID SELLARS COME TO THIS DISTINCTION?

To understand more precisely the duality between the images, we need to start from Sellars’ reflection of philosophy. In “Philosophy and the scientific image of man”, Sellars tries to determine *the* aim of philosophy. And at first sight, the answer is not clear at all, and for good reasons. Sellars observes that philosophy has been diversified in many “philosophies of” specific disciplines, which is confusing if we want to find the goal of philosophy in the big picture. For example, we deal nowadays with philosophy of biology, of physics, of politics, etc. That is, philosophy doesn’t look like it has a *predilection area* of investigation. Rather, it seems philosophy catches other disciplines’ subjects and investigates them. Therefore, Sellars gives the following broad definition of the aim of philosophy:

² This example is provided for it illustrates better the distinction between the two kinds of knowledge and what is meant by “perceiving”, than the examples of the previous section (with the tides, the free fall or the car). However, these were illustrative enough of what is meant by common sense.

“THE aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang on together in the broadest possible sense of the term.” (Sellars, 1963, p. 1)

What Sellars means here, is that philosophy aims to build “bridges” between specific disciplines in order to connect the dots of a bigger network: a *landscape*.³ Everything we consider as being part of the world is a part of that landscape, but no discipline has taken the role of building it and developing a general view that connects all the specific disciplines —this is precisely why we call them *specific* disciplines, namely because they aim at specialising, not connecting. In that respect, philosophy has no “special” subject-matter in itself, but all the subject-matters of the disciplines which do not —or cannot— look at *all* the other subject-matters surrounding them. Furthermore, history has taught us that when philosophy, at some point, has a special subject-matter, it is quickly taken over by some group of specialists —as was the case for mathematics around 2500 years ago. So, the role of philosophy rises from the necessity of the other disciplines to know how their subject-matters, their methods and principles of their thinking “fit into the intellectual landscape” (Sellars, 1963, p. 2). That is, while the special disciplines *do* know their way around in their subject-matters, they do not have a sense of their role in the broader image they form with all the other subject-matters.

For instance, physicists deal with physical events and physical phenomena —these are their subject-matters. But at some point, they might also face questions that are not approached in physics itself, for instance about what it means to look at the world as a physicist. So, physicists might reflect on non-physical questions that appear when trying to answer questions of physics —and these questions form a part of the subject-matter of philosophy.

A SYNOPSIS VIEW

“It is therefore the ‘eye on the whole’ which distinguishes the philosophical enterprise. Otherwise, there is little to distinguish the philosopher from the persistently reflective specialist; the philosopher of history from the persistently historian.” (Sellars, 1963, p. 3)

So, instead of creating new knowledge like the other specific disciplines⁴, philosophy starts from what is already known to form “a synoptic vision”, one that

³ A landscape corresponds to an image. Sellars used the word for pedagogical reason as a *landscape* is more easily understood than an image, which is a rather vague word. We shall come back to Sellars’ terminological choice later.

⁴ The terms “creating new knowledge” could be vague, for philosophy somehow creates knowledge as well; however, they refer to the fact that specific disciplines aim at specialising in their subject-matter.

builds a large-scale map with all the special subject-matters. And such a map (or landscape) is, to Sellars, properly an *image*. All the specific subject-matters can sometimes contradict each other, and this is, of course, something philosophy has to deal with.

TWO LITTLE OBJECTIONS

However, one could disagree for, up to now, it seems Sellars makes two problematic assumptions. First, one could say, Sellars presupposes that philosophy is necessary to the special disciplines considered chaotic, blurred, even *lost* without it. But the point is not that special disciplines need to wait for philosophical clarification of their subject-matter; they are, in fact, totally independent.

“What we must rather say is that the specialist knows his way around in his own neighbourhood, as his neighbourhood, but doesn’t know his way around in it in the same way as a part of the landscape as a whole.” (Sellars, 1963, p. 4)

Secondly, one could object Sellars supports the idea that philosophy has only one job: namely, catching the whole landscape in which stand the “details” (the subject-matters) discovered by all the special disciplines. Nevertheless, that landscape does not stand there as a unity, waiting to be caught by a philosopher. Rather, “the unity of the reflection vision is a task” (Sellars, 1963, p. 4). To be clear, philosophy needs previously discovered subject-matters in order to join the dots and capture the landscape which, as a matter of fact, does not weaken the relevance of special disciplines: they allow us to discover new things about the world. And *the* aim of philosophy is therefore to grasp the intellectual landscape in which all the researches of the special disciplines take place. They are like the fundamental details of a landscape within which connections are to be discovered to form a *united* picture; and this is what Sellars call an image.

THE TERM “IMAGE”

I draw the lecturers’ attention to the choice of the term “image”. It appears one could wonder to what it refers. Sellars uses this vague word on purpose and wants us to understand it in the broad sense of what an image can be. He says, about the choice of the term:

“On the one hand it suggests the contrast between an object, e.g. a tree, and a projection of the object on a plane, or its shadow on a wall. In this sense, an image is as much an existent as the object imaged, though, of course, it has a dependent status.

In the other sense, an 'image' is something imagined, and that which is imagined may well not exist, although the imagining of it does—in which case we can speak of the image as merely imaginary or unreal. But the imagined can exist; as when one imagines that someone is dancing in the next room, and someone is. This ambiguity enables me to imply that the philosopher is confronted by two projections of man-in-the-world on the human understanding. One of these projections I will call the manifest image, the other the scientific image.” (Sellars, 1963, p. 5)

The last two sentences are crucial. Sellars actually points out that there is no such thing as the *one* image. That is, philosophers are actually confronted by *two* such images⁵ (in principle; and in fact, by many). And those are the manifest image and the scientific image of man-in-the-world, a duality that should be fused by philosophy in what Sellars calls the *stereoscopic view* of philosophy. An illustration of the duality can be found in the example of the two violinists, or in an example of the previous section (the difference between the “knowing how” to drive a car, and the “knowing why” a car advances by knowing the mechanics and all the theory about cars). Both kinds of knowledge are accurate and complete, though different by nature.

3. The Manifest Image

Sellars describes the manifest image as the intuitive conception of the way humans realised themselves as being in the world, namely when humans first encountered *themselves*—when they came to be human. And that conception of themselves as being in the world did not come randomly: that humans have a conception of themselves is a kind of “necessity”, just as it is a necessity that humans would be a very different kind of human if they had a different conception of themselves. The conception of themselves *only* arises within a framework that was itself *non-conceptual*.

THE “ORIGINAL IMAGE”

“[A] diversified conceptual framework is a whole which, however sketchy, is prior to its parts, and cannot be construed as a coming together of parts which are already conceptual in character.” (Sellars, 1963, p. 6)

⁵ Precisely, philosophers are confronted by two pictures in principle, and by many in fact. Cf. (Sellars, 1963, p. 4)

So, the manifest image as a human conception of themselves appeared on a previous background that is itself non-conceptual and which Sellars calls the “original image”. It should be highlighted that the original image came to humans before any conception: the original image refers to a “deep” experience as humans—which is precisely what phenomenologists try to capture. And it is on that basis that the manifest image arose. The *manifest image* has nothing to do with a representation of man-in-the-world that would be pre-scientific, uncritical or even naïve (in the sense of a representation that would be waiting for the reflected, disciplined and critical look of science). The right way to understand the *manifest image* is to consider it as a “refinement” or “sophistication” of that original image—though a non-scientific one.

THE TWO PATHS OF REFINEMENT

“The conclusion is difficult to avoid that the transition from pre-conceptual patterns of behaviour to conceptual thinking was a holistic one, a jump to a level of awareness which is irreducibly new, a jump which was the coming into being of man.” (Sellars, 1963, p. 6)

The *refinement* of the original image follows two paths: the empirical one and the categorial one. The *empirical* refinement applies to the way humans encounter the world and themselves as living into it. Humans approach the world empirically, that is they *conceptualise* it with the “canons of inductive inference” as defined by John Stuart Mill⁶, and the ones of statistical inference. Those two methods are how humans experience the world, and are therefore the basis of the manifest image. In other words, the manifest image is an image abstracted from the inductive inference-framework and from correlations humans think they obtain in this framework. This point is worth highlighting, since it shows the manifest image is definitely not a *naïve* conception of the world. In this respect, the manifest image has been relevant even in the development of science, in which correlational and inductive methods have played an undisputable role. The postulational hypotheses presuppose “correlations to be explained, and [suggest] possible correlations to be investigated” (Sellars, 1963, p. 7). The manifest image gives us “glasses” to look through at the world. Now, whether or not Sellars has a realist position towards the manifest image is not clear; however, the manifest image aims at being complete and a philosophical tradition, the “perennial

⁶ In a nutshell, the inductive inference in John Stuart Mill epistemology is the method which generalises from experience in the wide sense. I.e., we *infer* from one or a few situations in which a phenomenon is observed, that this phenomenon will occur in all other situations of the same class, so the situations that resemble the one(s) in which the phenomenon was observed. The resemblance resides in the material characteristics of the situations.

Cf. (Mill, 1963-1991, p. 306)

philosophy”, has endorsed a realist position in which the manifest image as the *real* image.

Regarding the categorial refinement, Sellars says that in the first stages of human relations with the world, humans had to deal with a reality and did it by making *categories*. These categories arose when humans first tried to understand of what kind were the basic objects of the world. That is, humans had to deal with a non-conceptual image, the original image, and when dealing with it, they came to refine it following two ways; the refinement forms the manifest image.

“The categories of theoretical physics are not essences distilled from the framework of perceptual experience, yet, if the human mind can conceive of new categories, it can also refine the old; .and it is just as important not to over-estimate the role of creativity in the development of the framework in terms of which you and I experience the world, as it is not to underestimate its role in the scientific enterprise.” (Sellars, 1963, p. 10)

EXPERIENCES AS DETAILS OF A LANDSCAPE

Now, what is concretely the conceptualisation of primary experiences as a human-being? The human encounter of themselves should be understood in the broad sense of having an experience: the manifest image refers to the basic ones (the experience of being surrounded by others, the experiences of time, of thinking, of three-dimensional space, etc.) as well as the more complex ones (the aesthetical, logical, ethical aspects of human experience). All these conceptions are part of a large-scale picture, the *manifest image*, that can be compared to an intuitive representation of the world. To come back to the analogy of a landscape, we can consider all those experiences (ethics, aesthetics, experience of time, of others, etc.) as being *details* of the same landscape. Sellars adds, about the manifest image:

“There is man as he appears to the biochemist, to the physiologist, to the behaviourist, to the social scientist; and all of these images are to be contrasted with man as he appears to himself in sophisticated common sense, the manifest image which even today contains most of what he knows about himself at the properly human level.” (Sellars, 1963, p. 20)

That common sense (also called the “ordinary usage”) is the manifest image is crucial in this thesis. Even though there might be a discussion about whether Sellars actually thinks common sense is the manifest image or not, we will simply take it as an assumption.⁷ Common sense is described as a sophisticated image that refers to how we conceptualise things and experience everyday life. It follows common sense

⁷ However, the assumption is justified. Cf. (Sellars, 1963, pp. 8; 19-20; 22)

is structural to how humans perceive and experience the world, and comes from the original image. This, in turn, provides an essential argument for the conservation of the manifest image: the way we perceive the world in common sense is *dear* to us. This is why some incompatibility between common sense and science is problematic: we cannot easily get rid of the manifest image, nor can we escape it at will. Common sense is *constitutive* and *structural* to our experience of the world and in that respect, it is shared by everyone. In a sense, Sellars' idea is close to Immanuel Kant's categories of understanding, which are structural of how we perceive the phenomenal world.⁸ Because of its structural role, it is very hard to abandon, replace or change common sense; and this challenges the task of philosophy.

4. The Scientific Image

Two objections could be presented against Sellars' manifest image, but they will be approached in the next section, for they rest on the difference with the scientific image; so, more details should be given first.

PRIORITY OF THE MANIFEST IMAGE

As previously said, the manifest image is not the only existing image at the disposal of humans. Next to the prior experience of the world that led to the manifest image, the history of humanity has seen the rise of scientific approaches in different forms at different times. Those began with the first scientific reflections in Ancient Greece and keep going with the current developments in the different sciences—physics, chemistry, biology, etc.⁹ One notable characteristic of the scientific image is that it came *after* the manifest image. That is, the manifest was (and is) inevitable for humans needed to simply live in the world, hence have a representation of it following the empirical and categorial paths. Only then, and via other paths, would humans construct the scientific image.

Consequently, it is somehow connected to common sense, for the manifest image is structural to how we perceive the world—and it is upon such a perception that scientific approaches first arose (indeed, it seems difficult to elaborate a science without previous perception of the world, at least at the very beginning of history of

⁸ The closeness with Kant has been approached by Eric Watkins. Cf. (Watkins, 2012)

⁹ We can consider human sciences as sciences in Sellars' approach, though he is not clear about it. However, it is not an important matter for my purpose.

It should be noted that mathematics as a science does not qualify in the discussion, for it seems difficult to consider the manifest image corresponding to it – as such, there is no commonsensical image of mathematical objects.

science). Over the course of time, this led to another characteristic: namely that the manifest image often sticks to the things we can perceive (though certain things structuring the manifest image are perceived indirectly, as we can see in phenomenology with the notion of intentionality for instance), while the scientific image *inter alia* makes hypotheses of the existence of imperceptible things in order to explain the perceptible ones. That is, the scientific image results from postulational theory construction and following which correlations are predicted by theories elaborated in order to explain established correlations.

THE SCIENTIFIC IMAGE?

It must be specified here that Sellars is aware that there is no such thing as *the* scientific image, in the sense that there is no commonly shared view of the world among sciences: the scientific image is not easily unified. So, if we want to unify biology and physics, there are unavoidable difficulties for we need to determine the correspondent objects in biology and physics. This amounts to making a meticulous equation, as

“(...) to do this requires only an appreciation of the sense in which the objects of biochemical discourse can be equated with complex patterns of the objects of theoretical physics.” (Sellars, 1963, p. 21)

However, despite the difficulties due to the complex differences between sciences—which we can consider as forming a sort of heterogenous mosaic—, there is a way to understand them as participating to the scientific landscape. Each science elaborates its own image of the world and in that respect, there are as many scientific images as there are sciences—one for the neuroscientists, one for the astrophysicists, one for the biologists, etc.¹⁰

But Sellars makes of the scientific image an idealisation (just as the manifest image is one such) of the many particular images of sciences, which share the fact that they contrast with the manifest image. That is, even though the different scientific disciplines might use different methods, it is not misguided to consider them as having in common the fact that they are broadly parts of a same *picture*. In other words, the heterogeneity of sciences forms the one scientific image. So, each science participates to the *scientific image* as a landscape, just as the human experiences of time, moral, aesthetics participate to the manifest image.

¹⁰ A discussion about how we should consider this kaleidoscopic view of science is provided in (Brillard, Guay, Imbert, & Pradeu, 2011).

ONTOLOGICAL NEUTRALITY, AND RIVALRY

The last point that shall be approached in this section is an idea given above: the fact that the manifest image is *prior* to the scientific image. Any theoretical conception needs a ground, and that ground —according to Sellars— is the manifest image of human-in-the-world; humans conceived the scientific approach of the world way *after* they had conceptualised their very first and direct experience of it.¹¹

“(...) the categories of a theoretical science are logically dependent on categories pertaining to its methodological foundation in the manifest world of sophisticated common sense in such a way that there would be an absurdity in the notion of a world which illustrated its theoretical principles *without also illustrating the categories and principles of the manifest world.*” (Sellars, 1963, p. 20) (Italics original)

So, one could say, can we not simply weaken the scientific image and say it only provides useful fiction? In other words, do the so-called “objective” entities only have subjective experiences (the manifest image) as a basis? The answer consists of two points.

For one thing, we should point out that Sellars defines a framework in which he is *neutral* in respect to realism or antirealism about the images. That is, he does not stipulate whether the manifest or the scientific image is true about the world. They might or might not be, and the answer does not matter. So, when Sellars says that an image is real, he does not refer to any sort of truth or reality—in doing so, his neutrality entitles an image all there is in the world. What matters is that both images *can* be true, and both give a complete account of the world. They both aim to be true and as a matter of fact the scientific one aims to be a complete image of “the whole truth”, despite its methodological dependence on the manifest image of the world. Let us note that this leads to a rivalry between them: they capture two different pictures of the same reality. It is actually in that respect that Sellars suggests to view both of them as *whole* visions of the total amount of things in the world.¹² Again, the manifest and scientific images do not differ by their relation to truth or reality, there is no hierarchy between them in that respect. Rather, they differ in their structures —how they were elaborated, their bases, their methods, etc.

The second point concerns directly the said rivalry. In effect, one could be tempted to opt for the manifest image over the scientific one, as the latter came

¹¹ The priority we are talking about is *temporal*. The conceptual priority of the manifest image is defended by the perennial philosophy, but I do not intend to endorse this position.

¹² This is the base for Sellars project to articulate both images in a stereoscopic view. He suggests to see them as two pieces of an unfinished puzzle, and the task of philosophy as an attempt to fit them together. Cf. (Sellars, 1963, p. 19)

afterwards and has the former as a basis; again, it would be absurd to think humans could develop sciences while being virgin of any previous representation of the world. Precisely, says Sellars, the categories of a theoretical science are *logically* dependent on categories which relate to the manifest world. The manifest image is the one in which the scientific theories find their methodological ground. However, there are very good reasons to keep the scientific image and consider it as relevant for sciences are very *powerful*: they provide precise predictions about celestial bodies, climate, chemical reactions, etc. And, at least in physical sciences (which are the sciences to which I refer the most), the capacity of predicting accurately is central. The scientific image aims to be independent from the manifest image—which is part of the aim at being “the whole truth”. So, it would be bad faith to claim the scientific image is trivial and should be rejected. Furthermore, science is founded on a solid basis, both empirical and formal; it follows that in cases where the manifest image and the scientific image disagree, we will be in a delicate situation. On the one hand, we are naturally drawn to the scientific image for we take what science says for granted; on the other hand, it is very difficult to get rid of the common sense, as it is constitutive of how we perceive the world.

Consider the following example, which shows the problematic situation in which we are when there is an incompatibility between science and common sense. Common sense about the notion of time is that we all share the same one; that is, we intuitively think that time is shared by everyone *simultaneously* (so, we are all simultaneous to each other). We are induced to think so by the fact that in everyday life, this simultaneity is obvious: we do not think time passes differently for us when we are in an aeroplane than for someone who stayed home. We can even call that person, and we will observe that we are simultaneous to each other. Yet, one of the consequences of Albert Einstein’s special theory of relativity (1905) is that time depends on the speed of the frame of reference. Consequently, two persons moving at different speeds—say, one in an aeroplane and one staying home—have a different time. They “get older at different speeds”, to say it very roughly.¹³ As a matter of fact, the special theory of relativity is incompatible with the commonsensical notion of time and calls for a revision of the latter.

So, we have seen that Sellars is neutral about the true nature of the manifest and the scientific images. What matters to him is that they *can* be true and aim at being so by providing a complete account of the world. Hence, they differ from each other

¹³ The purpose of the sentence is more to illustrate than to give an accurate account of the STR, which it does not.

in the nature of their *structures*. The fact that they intend to be complete and to be true results in a rivalry between them. On the one hand, the manifest image is structural of how we perceive the world and served as a basis for science; on the other hand, we should listen to what science says for (1) it aims at being independent from the manifest image, (2) it is based on empirical methods and can give proof (what common sense hardly can), and more generally (3) it is quite powerful thanks to the precision of its predictions (and this is, to me, sufficient to justify that science must be taken into account, for the capacity to predict accurately is what makes physical sciences—which are central in this study—powerful). Thus, whenever the scientific image contradicts common sense, this puts us in a delicate situation: albeit we should change our manifest image, we struggle to drop it precisely because of its structural role—as we have seen in the example concerning time, we have not changed common sense about it, notwithstanding the incompatibility with the (proven) special theory of relativity.

5. An objection against the Manifest Image

The notions of manifest and scientific image and how they differ might encounter objections which will be discussed here, as they form the basis of my reasoning and there should not be disagreement about the assumptions of the thesis.

Sellars anticipated an objection against his assumption that there would be *the* scientific image. But what about talking of *the* manifest image? We have reason to think that there is no such thing as a unified manifest image; that there could be an infinity of images that are “manifest”. For instance, if we entertain the manifest image of the human body, Sellars should actually talk about the manifest *images* of it. Indeed, the manifest conception of it in Asia is probably different from the occidental one (while we are the heirs of a dualist tradition, Buddhist traditions are not). So, maybe we should circumscribe the manifest image of the human body, say, to the western manifest image of it.

A first reply is that the manifest image does not refer to linguistic uses. That is, we should pay attention to not consider a difference in languages for a difference of manifest images. Consider the case of colour, which is more illustrative: in other parts of the world, people refer to it in different manners. For instance, the literal translation of the English word “green” in Vietnamese is “the blue colour of the tree leaf”; if we follow the objection, this would mean that Vietnamese people perceive colours differently—namely, they would see “our” colour green as blue. Of course, assuming that they perceive the same qualia, this is absurd: this difference in language indicates

nothing of how they see colours, and they probably see them as we do (except for colour-blind people, which is the case all around the world). Furthermore, some readers have probably spotted the analogy of the manifest image with the phenomenology. Although we will not tackle this branch of philosophy, it is worth noting that phenomenologists make the same assumption as we do; namely, that there is something structurally common to how humans, all over the world, have certain experiences.¹⁴ The complexity of their task is then to catch this *something*, so hard to circumscribe with the language. It is actually this complexity that led Sellars to use the voluntarily vague term “image”, which is meant to capture the whole set of human conceptions (which are themselves divided into two structurally different images).

As a last remark, let us note that one does not need to adhere completely to Sellars’ assumption: as a reminder, the central point of this thesis is the direction of causality. Therefore, I will provide arguments in the next chapter in order to defend that there is something structurally common in the way humans perceive causality. Hence, it is sufficient to accept these arguments albeit not Sellars’ thesis.

Now we have clarified what the manifest and the scientific images are, we will approach in the next chapter the question of the direction of causality, in respect to the manifest and the scientific image. Precisely, we will discuss what the manifest image of causality consists in, and try to circumscribe its origin. In a second phase, we will confront it with the scientific image of causality; as we will see, there is an incompatibility between the manifest and scientific images of the direction of causality, and a choice will have to be made. For now, let us just remember that there is a difference of structure between common sense and science that can lead to a problematic contradiction: while it is tempting to follow science, common sense is structural of our perception of the world and is therefore hardly replaced or transformed.

¹⁴ The manifest image is universal to a certain extent. That is, there are certainly people who do not share it; just like colour-blind people do not see colours as most people do. However, this does not make any discussions about the human perception of colour irrelevant; and the same applies to the manifest image.

Chapter 2

The nature of causality

In the previous chapter, we have seen that Sellars distinguishes between the manifest and the scientific image. This distinction is important, as I aim to discuss how it applies on the concept of causality and what we can learn from the manifest image of causality. As we will see, the manifest image encompasses a characteristic that might be incompatible with a scientific experiment, and this is what we will discuss later on. But before beginning the debate, more must be said about causality. So, this chapter aims to provide a state of the art of the philosophical theories concerning the nature of causality.

1. Salmon's Theory of Propagation

There have been many debates among philosophers in order to determine what causality actually is. Is it physically into the world (ontic theories), or just some human way of talking about the world? Philosophers defending the theory of propagation answer to the question with the former, and in this section, we will see how they do so. The first theory we approach is Salmon's theory of propagation-interaction.

PROPAGATION THEORY

Wesley Salmon's theory of propagation (Galavotti, 2018; Salmon, 1979, 1980, 1984, 1994, 1997, 1998, 2002, 2010) is one of the most famous theories of causality. His main intuition is that causality is an *ontic* entity of the world. That is, Salmon believes that there are real causal processes in the world—in other words, that causality is not some human projection of the mind, as the conventionalist thesis suggests (we will come back to the conventionalist thesis later). According to Salmon's ontic thesis, events are explained by the mechanisms into which they take place, and the *causal processes* are physical entities.

What characterises a so-called “causal process” is that something *propagates* from one event (the cause) to another (the effect); all cases that do not involve the propagation or transfer of something are therefore non-causal. This might seem crystal-clear, and yet some cases are not so evident. Of course, we could say, pouring boiling water causes an ice cube to melt because heat *propagates* from the water to the ice cube. And it seems that the notion of propagation applies to a lot of cases that we call causal. Yet, imagine one forgets to water one's plant for a week: it would be counterintuitive to say that forgetting to water the plant did not cause its death. In other

words, it is likely that the nonchalance of the owner caused the plant's death. However, there is no propagation at all in that case, for there is nothing physically transmitted from the owner's absence of action (he did not water the plant) to the plant.

FACING SPURIOUS CAUSES

The starting point of Salmon's theory is his statistical-relevance model. At the beginning, he intended to propose a theory strictly based on statistics; broadly, the idea is that an event causes another if it increases the probability of its occurrence.

"According to the S-R model, in order to explain an event one must exhibit all the factors that are statistically relevant to its happening, without mentioning irrelevant elements, so that an explanatory account should include only information that is genuinely explanatory." (Galavotti, 2018)

The real power of the statistical-relevance model is explanatory. That is, the cause of an event consists of all the informative factors of it, which in turn are those that increase its probability.

However, one could say, there could be events that fulfil the requirement to be a cause (increasing the probability of another) without being actual causes. Therefore, we need to distinguish between real causes and *spurious causes*. It is no mystery that sometimes, we find correlations that have no explanatory dimension at all. Suppose we observe a drop in a barometer (*B*) and a few minutes later, a storm (*S*). While there is indeed a statistical relevance of the barometer to the storm—every time there is a drop in the barometer, a storm follows so that we can predict it—, no one would say that there is a direct *causal relevance* from the storm on the barometer.

CONJUNCTIVE FORKS

The notion Salmon appeals to is the *screening off*. Let us take the barometer example again, but add another event in the equation: the event *P*, which is the drop in atmospheric pressure. Given that there is a statistical dependence from *S* and from *B* on *P*, it follows that the event *P* screens off *B*, for *B* becomes irrelevant to *S* when we take *P* into account (even if *B* is statistically relevant to *S*, without *P*).

The *screening off* property of causality is connected to the principle of common causes which Salmon borrowed from Hans Reichenbach. This principle says that "if an improbable coincidence has occurred, there must exist a common cause" (Reichenbach, 1956, p. 157). That is, just like in the preceding example, we should find the common cause of two correlated but non-causal events. The principle of

common cause is based on the *conjunctive forks*, which allow us to determine spurious causes. The conjunctive forks in their formal version, work as follows: suppose that there are two events B and S whose occurrences happen simultaneously¹⁵ “more frequently than would be expected on the basis of pure chance” (Galavotti, 2018) (take, for instance, the barometer and the storm). These events are statistically dependent; but the principle says that, in the presence of another event P , if the correlation of B and S is absorbed—in other words, if they become reciprocally *independent*—, we have a conjunctive fork. The conjunctive fork is formally given as follows:

$$\begin{aligned} p(A\&B|C) &= p(A|C) \times p(B|C) \\ p(A\&B|\sim C) &= p(A|\sim C) \times p(B|\sim C) \\ p(A|C) &> p(A|\sim C) \\ p(B|C) &> p(B|\sim C). \end{aligned}$$

CAUSAL PROCESSES

Up to now, we have seen how Salmon determines causal explanations, that is via statistical relevance. The statistical-relevance model presented a weakness, namely when spurious causes are at stake. Therefore, he appealed to conjunctive forks, which allow to distinguish real causes, screening off spurious causes.

We have not talked yet about what Salmon means by causality, and this is what we approach here. Earlier, we mentioned the terms “causal process”, and these are central for Salmon.

“[A causal process is a] spatio-temporal continuous entity having the capacity to transmit information, structure, and causal influence.” (Salmon, 1998, p. 253)

Causal processes are responsible for causal propagation, so as to ensure the connection between a cause and its effect. Causal processes are continuous in Salmon’s theory¹⁶, and are characterised by the concepts of *mark transmission* and the *at-at theory*. The point is, causal processes have the capacity of transmitting marks: those marks can be information, signals, causal influence, etc. The mark must be found at both the cause and the effect. Only if there is a mark can we talk about causality.

To illustrate the mark transmission, Salmon gives the example of the spotlight: imagine a moving spotlight that projects a spot of light on a wall. It is uncontroversial that the spotlight causes the spot of light and hence, to use Salmon’s words, that there

¹⁵ The events do not need to be necessarily simultaneous. But their occurrence of the joint events must happen more than frequently than would be expected on the basis of pure chance.

¹⁶ As opposed to authors like Suppes, Good or Reichenbach who put forward the idea of causal chains as a collection of events.

is a causal process from the spotlight to the spot of light —namely, the moving light beam. As a matter of fact, the moving spot of light (that we see as a round shade on the wall) is no process at all —it is not the cause of the light beam; therefore, it is a *pseudo-process*. How do we know there is the transmission of a mark? Salmon replies that, in order to prove that the light beam is an actual process, we can test it by stretching a red veil just in front of the spotlight: in doing so, the light beam turns red (and so, too, does the spot of light on the wall). This shows that a mark has been introduced between the spot of light and the light beam, as it is transmitted along the beam. The same does not happen if we stretch the red veil on the wall —the spot of light on the wall turns red, but loses the red colour as soon as the source moves— so that the spot of light on the wall is itself no causal process, but results from a causal process. It should be noted that a causal process is characterised by a *capacity* to transmit marks, and not the fact that it *actually* does so. This can be observed in the spotlight example: the light beam has the capacity to transmit the mark of the red veil, if we introduce that mark (namely, if we stretch a red veil in front of the light source).

But the mark transmission is not sufficient to escape Hume's criticism, which would consist in accusing Salmon of introducing occult causal powers (Galavotti, 2018). Therefore, inspired by Russell's causal lines (Russell, 1948), he applies the *at-at* theory on the causal processes.

“ The transmission of a mark from a given point *A* in a process to some point *B* in the same process simply consists in the fact that it appears at each point between *A* and *B*.” (Galavotti, 2018)

He even added a counterfactual clause to the *at-at* theory (Galavotti, 2018), so that even if we need counterfactuals to characterise causal processes, they are always determined by what sciences tell us. Hence, he uses counterfactuals not in terms of possible worlds, but instead in those terms:

“[a counterfactual is] an experimental interpretation according to which counterfactual statements make reference to observed statistical relations which can be tested by controlled experiment.” (Salmon, 1984, pp. 149-150)

This means that, for Salmon, counterfactuals do not refer to possible worlds as, for instance, David Lewis (1973) defines it. Instead, a counterfactual statement describes what would happen given what science tells us. That is, science benefits from experiments of which the observation provides statistical relations. Those enable us to know what would happen if a mark were transmitted, thanks to other scientific experiments.

CONSERVED QUANTITIES

The last point we approach here is the notion of *conserved quantity*. We know now that a causal process is characterised by its continuity and the transmission of a mark. Without causal mark, no causality. The conserved quantities characterise precisely what a mark consists in, and are a notion that Salmon borrowed from Phil Dowe (Dowe, 1992a, 1992b, 1995).

Examples of quantities are the mass-energy, the linear momentum, the charge, etc. The notion of quantity is voluntarily flexible enough so that it can potentially cover all scientific cases. So, in order to be the transmission of a causal mark, there must be a transmission of a significant conserved quantity. Only, we should know what a significant quantity is for each case; and this must be determined by science itself. Salmon says:

“A process is causal if it is capable of transmitting a mark, whether or not it is actually transmitting one. The fact that it has the capacity to transmit a mark is merely a symptom of the fact that it is actually transmitting something else. That other something I described as information, structure, and causal influence.” (Salmon, 1997, p. 462) (Italics original)

Conserved quantities are actually possessed by causal processes. Therefore, a process is causal if and only if it transmits a conserved quantity, while pseudo-processes cannot and are therefore easily distinguishable. The principle of conserved quantity does not rule out the mark transmission or the at-at theory: the mark transmission helps detecting causal relationships (the red veil), and the at-at theory stipulates that a conserved quantity “is at the appropriate place at the appropriate stage in the process.” (Salmon, 2010, p. 10)

CONCLUSION

As a reminder, Salmon’s theory consists in an ontic approach of causality. According to him, causality itself physically exists in the world, so that causal processes are real physical entities and “explain how the world works” (Salmon, 1984, p. 133).

“The complete causal structure is a fact of nature that exists quite independently of our knowledge or interests.” (Salmon, 2002, p. 126)

But those causal processes are established by scientific theories. Science indicates what a significant quantity is—or is not. Hence, Salmon is *not* interested in catching the manifest image of causality and this is the reason why I have to abandon this theory. That is, what I intend to do is to determine what is the manifest image of

causality —what it means in common sense— and discuss whether it is compatible with science or not. And even if the conjunctive forks, as well as the statistical-relevance model, are shared by other theories that could interest us; Salmon’s theory of propagation is not adequate for my purpose. That is, I aim to bring face to face the manifest image of causality and science in order to detect and discuss the incompatibilities, and that is not possible with a theory that assumes that what is causal must be determined by science. Notwithstanding, the theory of propagation-interaction is illuminating for many cases and gives a satisfying account of causality from a scientific point of view.

2. Menzies and Price’s Agency Theory

Another theory of the nature of causality is the theory of manipulation. Just as there are different theories of propagation-interaction, there are different theories within the group of the manipulationist theories; but all share the idea that to cause something is to manipulate something. We will study here Menzies and Price’s agency theory, which is certainly one of the most famous manipulationist theories. But more importantly, I defend that the two authors succeeded at reaching what they claimed: defining the common sense of causality.

MAKING A DIFFERENCE

The Agency theory of Menzies and Price is in line with other *difference making* theories, and is a version of the manipulation theory. In their article “Causation as a Secondary Quality”, Menzies and Price aim to open up a new direction to theories of manipulability, which were until then rather unpopular. This is why they call it the “agency theory” which, as the name suggests, gives a central role to the notion of *agent* in the definition of causality. The central statement of the agency theory goes as follows:

“an event *A* is a cause of a distinct event *B* just in case bringing about the occurrence of *A* would be an effective means by which a free agent could bring about the occurrence of *B*.” (Menzies & Price, 1993, p. 187)

The paper is articulated so as to provide a defence of the theory, for the objections against it have been—to the authors—overrated. We shall see here the four objections they deal with in order to explain why causality should be considered as analogous to the concept of colour in the dispositional theory, namely as a *secondary quality*. For that matter, they use analogies with the dispositional theory of colours all along their reasoning, and some of their examples will be approached here. The point

is, few would disagree with the fact that an adequate account of colour needs to be given with references to the human perception, with terms as “perceptual states” or “perceptual capacities” —which actually amounts to referring to colours as secondary qualities.¹⁷ And this is why Menzies and Price use the definition of colour in the dispositional theory: namely, the idea that “to be red is to be disposed to look red to a normal observer under standard conditions” (Menzies & Price, 1993, p. 188). However, it should be noted that the Menzies and Price thesis is *not* that causality is a secondary quality, but that the use of the dispositional theory of colour provides an adequate portrait of causality in the agency theory.

The very notion of secondary quality refers to human perception, and this is not innocent, for they would not elaborate their theory if they did not think the common sense of causality covers, at least partially, the truth about causality. Of course, the authors proceed to a philosophical initiative which is to theorise something in a coherent system; but that “something” is the commonsensical idea of causality.¹⁸

CONFUSION BETWEEN EPISTEMOLOGY AND METAPHYSICS

Now, let’s discuss a first objection considered by Menzies and Price: the confusion between epistemology and metaphysics. That objection can be expressed in those terms: while Menzies and Price pretend that they provide a conceptual (metaphysical) analysis, they actually give an empirical methodology (epistemology) for testing causal relations. That is, the authors claim that the notion of agency plays a constitutive role in the analysis of causality —it seems they adopt the verificationist thesis, following which the means used to test a statement determine its meaning.

Menzies and Price’s reply is that they do not adopt a verificationist position at all, and it is no part of their theory. The reason is, the notion of agency does not have the role of an empirical method. Rather, this notion has been introduced because the concept of causality is metaphysically rooted in the notion of manipulation. In other

¹⁷ In John Locke’s philosophy, secondary qualities are opposed to primary qualities. Those last ones are intrinsic qualities of an object, in the sense that they exist within its body; in other words, they exist outside of the human perception. These qualities are for instance bulk, number, figure or motion. The secondary qualities are qualities that do not exist within the object, but rather are “powers” caused by the object on the human perception so that we have ideas of colours, smells or sounds. Locke claimed the primary qualities are objective while the secondary qualities are contingent on perception; of course, more has been said about secondary qualities by authors like Berkeley and the notion as evolved ever since. Cf. (Kaijo, 2008).

¹⁸ And referring to the human perception in order to give an account of a concept can be understood as referring to the manifest image of it. We shall return to this point in the next chapter.

words, the very idea of “manipulating things”¹⁹ (which is linked to the notion of agency) endorses the very nature of causality.

The argument is best understood with the analogy of colour. If we adopt the dispositional theory—or any theory that makes the notion of colour dependent on human capacities—we will unavoidably face epistemological implications. The reason is, the best evidence for an object’s “looking red” is to look so to a *normal observer* and under *right conditions*. However, the theory makes no confusion of methodological considerations with conceptual analysis of colour. Now, there is no doubt such an analysis of the concept of causality deals with epistemological issues, as the best kind of evidence is provided by experimental manipulation: that is, the agency approach defines the concept of causality in terms of the way an agent’s bringing about, producing or manipulating an event modifies the probability of another event. However, thinking that there is here a confusion of a verificationist order, would be a mistake. The central idea of the agency theory is that causality is explained by relation to our experience as agents, just as the concept of colours (as a second quality) is explained by relation to our experience as observers. In other words, causality is an *extrinsic* or *relational* concept.

UNMANIPULABLE CAUSES

A second objection is the one of unmanipulable causes. Even if we accept Menzies and Price’s theory after their answer to the last objection, there is another that easily comes to mind: the concept of causality is bound to an agent’s bringing about an event, but it looks like the theory works on the assumption that *all* events are manipulable, which is obviously not the case. Consider the example of the meteorite that is thought to have caused the extinction of the dinosaurs: no human being, were it living at that time, could have manipulated the meteorite. Is it to say that this example does not contain a causal relation? Saying that the meteorite indeed caused the extinction of dinosaurs, even though no normal (human) agent would have been capable of intervening on the meteorite, is not surprising.

And the same can be said about colours: an object that looks red can look another colour under certain specific conditions, so that an observer would not be able to see it red. But that problem is inherent to secondary qualities and the reference to human perception. So, the answer Menzies and Price give is that, for unmanipulable causes, we should imagine what would happen if an agent had superhuman powers. In the meteorite example, a “super-agent” would be able to deflect the meteorite’s

¹⁹ Which can be understood as a particular case of « making a difference ».

trajectory and hence manipulate the cause of the dinosaurs extinction. In case the notion of superhuman power is too blurred, the authors refer to the use of *models*. Namely, scientists (seismologists for example) simulate events thanks to virtual models, and hence simulate the manipulation of events.²⁰ The point is, when we face an unmanipulable event, we should refer to another pair of events that are in the same causal relation and where the supposed cause *is* manipulable—in other words, we should use analogies to determine whether a supposed cause is *actually* a cause.

However, James Woodward did not agree with the reference to analogies, for he believes Menzies and Price are not precise enough (Woodward, 2004). For example, they do not specify what must be selected in order to provide an accurate model. That is, we do not know how to choose the right features to model, what the criteria are for a good model, and whether we manipulate *involuntarily* events that are not represented in the model—which, *in fine*, results in a failure to fulfil the condition of isolation²¹ for a good manipulation.²²

CIRCULARITY AND THE EXPERIENCE OF SUCCESS

The most important objection is the one of circularity and runs as follows: when the authors utilise the term “bringing about” in their definition of causality for the agency theory, it seems they define it *circularly*, for “bringing about” actually means something like “causing the occurrence of an event”, which in turn is a causal vocabulary. In other words, they define causality *with* causality, which is useless and irrelevant. Menzies and Price’s defence is that their use of the term “bringing about” does not refer to causal notions, but rather to a very primitive one. That is, the definition of causality lies in that first experience as an agent when we experience our power as an agent.

“The basic premise is that from an early age, we all have direct experience of acting as agents [...] we all have direct personal experience of doing one thing and thence achieving another.” (Menzies & Price, 1993, p. 194)

This is the experience of success. From Menzies and Price’s point of view, the concept of causality finds its roots in an experience that can barely be described in words, for it is analogous to the experience of colours—namely, the ostension of it. Precisely, they start their explanation of the non-circularity of the definition by making it clear what it means to *bring about*, as the terms seem circular. It is indeed very likely

²⁰ This, in turn, allows to avoid scale problems for instance.

²¹ The manipulation of an event must be isolated so that only *that* event is manipulated. We will come back to that notion in the next section.

²² Price replied to those objections in a paper (Price, 2014).

that “to bring about” does not mean anything more than “to be the cause” of something. However, to interpret their theory this way would be mistaken; rather, we should consider *how* we learn the concept of causality.

It is actually at a very young age that we learn the notion of causality. Only, this primitive process of learning does not happen the same way we learn mathematics for instance, but rather how we learn the name of colours; i.e., by *ostension*. For example, our parents or close family show us a tree leaf, and say “green!” a considerable number of times. And the more a baby is shown green objects, the quicker it learns that colour until it has no doubt anymore about what it means to be green. In a sense, what it means to be green is “to be green”; nevertheless, it would be bad faith to say that this merely amounts to a circular definition.

The process of learning causality is similar. At a very young age as well, we have the experience of success. The point is, humans are a very special kind of living being, for they are, and perceive themselves as, agents. And one main consequence of being an agent in the world, is that they can manipulate things in order to produce something else. So, a child has the experience of wishing the occurrence of an event *B*, and to use its “agent-power” to bring about an event *A* so that *B* actually occurs. In other words, the way we learn causality is by planning the occurrence of certain events by means of provoking other *earlier* ones, and actually *succeeding* at it. And this has to do with the experience of agent. Despite the lack of precision of Menzies and Price, we can understand their definition of causality in a counterfactual way, namely that causality is related to what we *could* do, given our capacities.

“The basic premise is that from an early age, we all have direct experience of acting as agents [...] we all have direct personal experience of doing one thing and thence achieving another.”
(Menzies & Price, 1993, p. 194)

This is not to say that we merely have the Humean experience of the succession of events. Instead, we experience successions of events where the earlier event is due to our action oriented to a goal and that goal is achieved (which is the later event). Menzies and Price call this the experience of *success* —or what I call *planification*. In doing so, the definition of causality via the experience of agent is not circular.

James Woodward attacked this argument, claiming that it is misguided. For him, we do not acquire the concept of causality via experience, but via complex neuronal processes and learning. He says:

“[...] the idea that our concept of causation is derived purely from experience (whether of agency or of anything else) is simply mistaken. As with other concepts, the acquisition of the concept of

causality involves a complicated interaction between prespecified neural mechanisms and ‘learning’.” (Woodward, 2004, p. 126)

Woodward claims that only very few learning processes are actually based on experience, as Menzies and Price mean it. It is not obvious that our experience as agents is empty of any causal concept and is a *primary experience*—which lets us intriguingly think about Kant’s *a priori* concept of causality. It must be noted that Woodward’s manipulation theory does not take common sense as central. But the point of Menzies and Price is precisely to argue that there is something like common sense in the notion of causality; Price himself replied (Price, 2014) that they do not deny there might be more complex processes behind the first experience as agent; but they say that this is what gives common sense. That Woodward pretends there is no learning from experiences like Menzies and Price describe it, is not convincing: it simply highlights the peculiarity of the experience as agents, a fundamental one.

ANTHROPOCENTRICITY

The last objection we will approach here is that scientists could say a central problem of Menzies and Price’s theory is the *unacceptable anthropocentricity*. That is, in a possible other world where agents have different capacities of perception and powers, Menzies and Price’s account means that different causal relations obtain.

For example, consider a possible world which is exactly like the actual world except that by an accident of natural selection agents have a much more restricted set of manipulative powers. At first sight, it might appear that the agency approach is committed to saying, quite implausibly, that the truncated manipulative powers of the agents imply a truncated set of causal relations.” (Menzies & Price, 1993, p. 198)

A first defence could be to say that the reference is the agents of our *actual* world; but this is not very convincing, for it does not seem to evacuate the anthropocentric problem. For that matter, the authors refer to an actualised criticism from David Lewis (1989, p. 199). It says that *if*—in the possible other world— we had fewer capacities of perception, but *still* applied a concept of causality relying on our capacities; *then* it would change the meaning of the clause saying that a causal relation is a “relation between events such that an actual agent could manipulate one event as a means to bringing about the other” (Menzies & Price, 1993, p. 199). And this difference would hold, despite the fact that those potential agents would have the same way of fixing references. It results from this argument that there is a problematic relativity of the concept of causality.

This is the moment where a scientist would feel comfortable in her idea that causality is some kind of human projection and cannot be objective. Indeed, the fact that there is a form of relativity is unavoidable, and the authors accept it. However, they reject the idea that this relativity is problematic. Their argument is that there is a difference of degree between the concept of causality and other concepts like colours or tastes. “[C]ausation is very much less sensitive than colour” (Menzies & Price, 1993, p. 199) they say; so, in a sense, causality is more objective than usual secondary qualities.

To illustrate this point, let us take their example: suppose that in a possible other world, the agents have the taste of sourness, like us. So, if they drink vinegar, they will say it tastes sour, just like we do in our actual world. Now, suppose that when they taste lemon, they do not find it sour at all. That means lemon is not sour in that world (hence the relativity of sourness), but let us assume these counterfactual agents have the exact same meaning for “sour” as us. We arrive at the conclusion that “sourness is in the eye of the beholder” (Menzies & Price, 1993, p. 200). Translated into causal terms, this means that even though the concept of causality is related to our capacities, the notion of causality remains connected to the notion of manipulation, whatever that is that we can manipulate—it differs because of our capacities. Furthermore, we have seen with the analogies that we do not need to *actually* perform the manipulation, as long as we have a good analogy with a manipulable cause—so, we can make the abstraction of the capacities we do not have. But in all cases, causality remains determined by manipulation.

Menzies and Price add that, in a pragmatic sense, the potential relativity of causality does not pose any problem, just as the notion of sourness does not change in science as well as in ordinary life. Furthermore, they say the notion of causality is far less sensitive to potential variations in the human conditions than taste or colour. That is, causality is a very peculiar concept, connected to a peculiar sort of experience. For that point, let us go back to the objection of relativity because of anthropocentricity.

We have already seen that the agency approach can be extended in counterfactual worlds in which agents can *actually* manipulate things of this counterfactual world (*cf.* the answer to unmanipulable causes). In other words, it is not obvious that

“any modification of mere *degree* in our powers as agent will issue in any modification in the causal relations we are thus inclined to ascribe. [...] On the contrary, it seems that agents with different capacities will nevertheless envisage the same range of possible causal relations,

provided that they employ the principle of analogical reasoning we noted earlier as licensing the extrapolations of their manipulative capacities.” (Menzies & Price, 1993, p. 200)

The only problematic case would be in a counterfactual world where the “cognitive beings” have no agency power at all; they could therefore not elaborate a notion of causality. So, in spite of the dependence on capacities, cognitive people of a counterfactual world have *either* a notion that works with the application of those capacities, *or* nothing at all.

Also, the agency causality does not contradict that causality is more objective than colour or taste, since there is a huge difference of degree between causality and secondary qualities. Of course, scientists would say that this difference of degree is not enough to settle the objectivity of causality. But, Menzies and Price answer, what is it for a property to be objective or subjective? Maybe it actually *is* a matter of degree. In conclusion, Menzies and Price defend the fact that the objections approached are, in the end, not more powerful than those against the orthodox account of colours.

CONCLUSION

In this section, we have provided a description of the agency theory. Menzies and Price’s version of the manipulationist theory is interesting in that they try to make sense of the commonsensical notion of causality, which they relate to agency. As I will defend in the next chapter, their theory captures the manifest image of causality, and this will be the starting point of my discussion.

Chapter 3

Manifest and Scientific Images of Causality: symmetry and asymmetry

In the first chapter, we have proposed a distinction between what falls under respectively common sense and science. This distinction had been proposed by Sellars, who referred to it as the manifest and the scientific images of man-in-the-world. Sellars' analysis is intriguing for he points out something fundamental about common sense; namely, that it is structural of the way we perceive the world that surrounds us. That is, we perceive things and have representations of them (via the empirical and categorial paths), and this is what common sense actually is. In other words, it is somehow stronger than one could have expected. On the other hand, the scientific image is very powerful and despite the fact that the manifest image is prior to it, it aims at being independent of it. The upshot of this divergence is that we face difficult situations when common sense and science contradict each other. On the one side, there are good reasons to listen to what science says—it succeeds very well at predicting and proving things, empirically and formally—and to reject common sense, which does not benefit from an observable basis; on the other side, the role of common sense makes it difficult to get rid of, which is problematic if we need to transform it (as it has been the case for the notion of time). In case of incompatibility, we had therefore better stay as close as possible to the manifest image.

In the second chapter, a state of the art in philosophy of causality was given, so as to have an overview of two main philosophical tendencies in respect to the nature of causality. I have claimed that, while Salmon's theory of propagation-interaction does not interest us here for it is not based on common sense but on science (though Salmon gives a definition of causality that is at least acceptable from a commonsensical point of view), Menzies and Price theory of agency is more promising for my purpose.

This chapter consists of two main points. The first is that Menzies and Price's theory of agency is precisely a theory that catches the manifest image of causality. As I defend, they succeed at doing what they claim. The second point is that a part of that manifest image is not really discussed by the authors: the direction of causality. As we will see, we should look closely at it, for this part of the manifest image might be incompatible with science, which forms the basis of the discussion.

1. Agency Theory and Manifest Image

A FIRST-PERSON EXPERIENCE

A first point to consider, is that Menzies and Price *try* to catch common sense. That is, they do not intend to show how causality should be understood or what it physically consists in. Rather, they want to give a satisfactory account of causality from a commonsensical point of view. They do not adopt an ontic thesis in respect to causality, as Salmon does. Instead, they show what causality is from our perspective as humans.

“What feature of our perspective could it be that manifests itself in the cause-effect distinction? The most plausible answer is that we acquire the notion of causation in virtue of our experience as agents.”(Price, 1996, p. 157)

For Menzies and Price, causality is properly human. This is why they compare it to secondary qualities; even if secondary qualities, such as colours, can be defined only with the human capacities of perception, there is an important sense in which we can define colours objectively. However, causality is a first-person experience, like Sellars defines the manifest image (experiences that structure common sense are subjective experiences).

AN EARLY-AGE EXPERIENCE

Now, the fact that they attempt to capture the manifest image does not necessarily mean that they succeed. But if we pay attention to their arguments, we can see that they are very similar to what Sellars meant with the terms “manifest image”. In that respect, the striking feature of the agency theory is that we learn the notion of causality from an early age, or more precisely from the moment we discover we are agents and succeed at bringing things about. The authors are not very precise about the experience of agent, but there is a coherent way to interpret it from a Sellarsian perspective.

As a reminder, the manifest image of man-in-the-world is a refinement of the original image, following two paths (the categorial and the empirical one). Sellars did not clarify at what age the refinement occurs, but we have good reasons to think it is at an early age. That is, the manifest image is structural of how we perceive the world, and that perception is possible only on such a structure. While the original image is somehow chaotic—in that it has not been refined, so the objects are not categorised yet—the manifest image is directly built. Without the refinement of the original

image, humans would be “blind”, and this is why it must happen at a young age. It is also the reason why the manifest image appears intuitive to us.

And this what Menzies and Price defend. At a young age, we experience the world and deal with a lot of different things. Then, we refine these experiences following the categorial and empirical paths: we experience success. That is, we experience that we succeed at producing the occurrence of an event by bringing about another one, which follows the two paths of refinement. From then on, the experience of success structures the way we perceive things, and this is what we commonsensically recognise as *causal*. Of course, this is not the end of the refinement: over the course of time, we extend the pattern observed in the experience of success to other cases, so that we recognise certain unmanipulable events as causal. But the analogy with the way Sellars defines the manifest image is striking, so that we can say that Menzies and Price have captured the manifest image of causality in their theory.

2. Manifest Image of Causality and the Causal Arrow

That Menzies and Price succeed at catching the manifest image of causality is one thing. However, it is not evident that they succeeded at catching the *whole* thing about causality (from a commonsensical point of view). That is, they define the nature of causality; but what about the direction of causality? In this section, we will see that there is a disagreement between the manifest and scientific images of causality in respect to its direction. We will then engage on a quest for a solution to this delicate position, whose first step will be to find where common sense of causality comes from. The next steps will be presented progressively.

COMMON SENSE AND CAUSAL ARROW

First of all, we should determine what common sense says about the direction of causality, that is, what we think a relation of cause-effect is in everyday life. This is what we call the direction of the *causal arrow*, which is a feature of the manifest image of causality.

The causal arrow, or the direction of causality, can mean two things. First, it may refer to the asymmetry of causality, namely that a causal relation goes from the cause to the effect, and never from the effect to the cause. Huw Price (Price, 1996) provides good arguments about it, though we will not approach them here. Second—and it is the meaning I aim to discuss—it may refer to the fact that the causal relation that goes from the cause to the effect is always (or so we think, in common sense)

oriented from the past towards the future. In other words, a cause temporally precedes its effect. This is what we call the *temporal asymmetry* of the causal arrow.

PRINCIPLE OF INNOCENCE

One could be surprised by the obviousness of the statement: it is indeed indubitable that a cause precedes its effect —how could it be otherwise? It seems uncontroversial that a cause does not follow its effect. And this is what Price calls the principle of innocence (Price, 1996, pp. 116-131). The principle of innocence says that

“ordinarily, we take it for granted that the properties of two physical systems will be independent of one another, unless and until they interact. We assume that when two systems encounter one another for the first time, innocence precedes experience: the two systems remain blissfully unaware of each other’s existence, until they first meet.” (Price, 1996, p. 118)

As a matter of fact, it is counterintuitive to imagine a case in which an effect happens before its cause. Therefore, says Price,

“[w]e have seen that this principle has been invoked at a number of points in an attempt to explain some of the observed temporal asymmetry in the world —though usually, as I pointed out, by writers who failed to see that the real puzzle is why the corresponding principle doesn’t hold in the opposite temporal direction.” (Price, 1996, p. 118)

What is important is that we usually do not think a future event can cause a past one, although it would sometimes give satisfying explanations of “magical” correlations. Imagine, for instance, that one day a friend of yours claims she can read the future (she says she has prophetic dreams); she then declares that something is going to happen, and it actually does happen. This is a disturbing correlation, and in such a situation, we will be more inclined to think this is pure luck (unless we already believe prophetic dreams exist), rather than trying to make sense of an explanation where the friend’s claim is dependent on a future occurrence. In other words, we take it for granted that causality is asymmetric and follows the time’s arrow. And this is how the causal arrow is conceived by the manifest image of causality.

3. Scientific Image of Causality

Now we have determined one of the features of the manifest image of causality (namely the temporal asymmetry of its direction), let us approach the scientific image of it. That is, what does science say about the direction of causality? What is the theoretical or empirical basis of the scientific claims about its direction? To approach the scientific conception of causality, we will have to see why physicists think that the

temporal asymmetry of interactions is accidental in a properly symmetric universe. However, if the asymmetry benefits from an observable basis in macrophysics, it is not the case of microphysics. It turns out physicists usually try to eliminate the concept of causality from physics—in other words, they adopt eliminativism.

THE WORLD OBEYS TO T-SYMMETRY

Before talking about causality, let us begin with what physicists say about the temporal direction. It is commonly accepted that we live in a world where things, phenomena, and events follow the arrow going from past to future. For instance, no one would disagree with the fact that we will never observe the parts of a broken glass assemble together and form the initial glass; neither would one disagree that we have never observed scattered pins on a bowling lane stand up straight and shoot out a bowling ball towards the hand of a player. These two examples seem rather to describe what we would observe if we watched *in reverse* a video of these events, but in the way they normally happen—so, a glass that breaks or a player shooting out a bowling ball towards pins. But we never have (and it is unlikely that we ever will) really observed such things.

However, according to physics, there is no less probability for events to happen in the opposite direction (as described at the beginning of the example). That is to say, the laws of physics are *symmetric* in respect to the temporal direction of physical interactions. In other words, the laws of physics are regardless of time. So, any interaction described by the laws of physics could actually happen in the opposite direction. “T-symmetry” is the name Price gives to this lawlike principle that the laws of physics satisfy.

PRINCIPLE OF INNOCENCE IN THE MACROWORLD

If the laws do not prefer either direction, how does it come that interactions seem to be *dependent* on the asymmetry of time direction—namely, the direction given by the increase in entropy?

Physicists had to face this question, for it shows a gap between everyday life and scientific statements. The answer has usually been that the asymmetry of interactions is due to the *boundary conditions*, namely the accidental conditions of the universe after the Big Bang. That is, the universe is asymmetric but this is not due to any lawlike principle—only to its contingent state at its beginning.

The asymmetry stemming from the boundary conditions gave rise to the *principle of innocence* that we have talked about in the previous section. As a reminder, the principle of innocence says that the properties of interacting systems are independent before they interact, but not afterwards. As a feature produced by the boundary conditions, the principle of innocence is compatible with T-symmetry and therefore presents no inherent problem in itself.

Concerning this last point, we should specify that the principle of innocence benefits from an observable basis. Namely, we find mostly correlations *after* interaction of macroscopic systems—and not the opposite. Price says:

“This provides one sense in which real-world physical systems are uncorrelated before interactions, then: they don’t display entropy-reducing correlations.” (Price, 1996, p. 120)

In other words, the manifestations of the second law of thermodynamics counts as evidence that there is an asymmetry in respect to post-interactive correlations: we observe more correlations following one temporal direction (from past to future) than in the other. So, we do have an empirical ground for the principle of innocence at the macroscopic level.

EXTENSION OF PRINCIPLE OF INNOCENCE IN THE MICROWORLD

The problem arises when we extend the principle of innocence to the microworld, for there is actually no observable ground in the microworld that is to its advantage. On the face of it, the similarity with the principle of innocence of the macroworld has been assumed in physics. Again, the principle of innocence in the microworld has been thought to come from boundary conditions as well.

“Rather, we rely on the asymmetric principle that interaction *produces* correlations only in one temporal direction —”towards the future”, and not “towards the past.” (Price, 1996, p. 122)

The absence of observation of μ Innocence (for microworld innocence) means that assuming it amounts to make it a fundamental principle in its own right. So, physicists who follow this path presuppose that the asymmetry required by μ Innocence is given in the boundary conditions; this, in turn, means that there is no “initial” asymmetry at all. One of the objections Price provides, is that:

“ μ Innocence does not need observational support, because it can be ruled out on the grounds that its failure would require ‘miraculous’ correlations between systems which are about to interact.” (Price, 1996, p. 122)

However, he says, the problem of this objection is that it takes μ Innocence for granted; that is, it amounts to say that there is an asymmetry for, if there were not, there would not be asymmetry—and this would allow miraculous things to happen.

What this objection shows—and, more generally, the assumption of μ Innocence without observable basis—is that we stick to the manifest image, as Sellars described it. Even if T-symmetry is stipulated by physics, physicists themselves tried to make sense of the manifest image at the microlevel, showing that, as a commonsensical notion, it is constitutive of how we perceive things (and therefore, we can hardly get rid of it).

Furthermore, as we will see, assuming μ Innocence commits to consider quantum phenomena as mysterious.²³ We haven't said anything yet about causality in science, but the discussion about symmetry, asymmetry and innocence provided a good overview of how physics deals with the notions underlying causality. To sum up, science has to deal with the following dilemma: on the one hand, physics of the past decades seems to show that the laws of nature do not distinguish between past and future, and are therefore symmetric in respect to time (this is what Price calls T-symmetry); furthermore, the idea that systems can interact before they meet makes quantum physics bizarre (Bell's Theorem, Copenhagen Interpretation). On the other hand, it seems obvious that innocence is part of the world (even if it is produced by the asymmetry given by the boundary conditions), and that it should be extended from the large scale (macroworld) where observable basis can be found, to the small scale (microworld). It follows that one of the two ideas should be put aside, and this context provides the basis of the scientific image of causality.

SCIENTIFIC IMAGE OF CAUSALITY AND ELIMINATIVISM

The reason why we tackled the notion of innocence is because the notion of innocence is strikingly related to causality. If μ Innocence does not amount to the same thing—innocence is about the previous “ignorance” of two systems that have not interacted *yet*—, the scientific position about the former implies huge consequences for the latter. If there is indeed symmetry, so that a relation of dependence can precede the encounter of two systems, then it is hard to see how to make sense of causality as we have seen it in the manifest image. So, although we would naturally tend to defend

²³ Bell's Theorem has been considered as an evidence that at the quantum level, there is some kind of nonlocality (a term which we will come back to later); Bohr's “Copenhagen Interpretation” implies that reality is somehow “indeterminate” *until* a measurement is made (which involves a distinction between past and future and an asymmetry). Cf. (Price, 1996, pp. 124-125)

that causality is aligned with the time's arrow (as we think commonsensically), the claims about innocence and symmetry favour a different position concerning causality.

The situation from the point of view of physics appears as follows: on the one hand, there are good reasons to think that the world at the microlevel is symmetric in time, as the laws predict (events happening following one temporal direction can actually happen following the opposite direction). On the other hand, it seems reasonable to maintain that what is happening at the microlevel are causes and effects; and that these causes are previous to their effects, so that any event depends on an earlier state of affairs rather than a later one.

This situation leads to the eliminativist position with regard to causality, and much adopted by physicists (and advocated by Bertrand Russell (Russell, 1963, p. 132), but also by Ruth Kastner (Kastner, 2017, pp. 164-167)). Eliminativism says that the concept of causality should be rejected from physics or at least from microphysics. This actually amounts to much the same: physicists tend to think according to a reductionist position, namely that the things of the macroworld are composed by smaller things from the microworld. So, concerning causality, it seems causal relations between macro-things rest on similar causal relations between smaller micro-things. In other words, rejecting causality at the microlevel seems to imply rejecting it at the macrolevel as well; and abandoning causality means abandoning common sense of it, the asymmetry of dependence but also the commonsensical idea that we can affect the future but certainly not the past. The issue of the choice is therefore more important than it looks. Be that as it may, we will see in the next chapter that a scientific experiment might force our hand in the choice.

PHYSICISTS DO NOT DO WHAT THEY PREACH

Before the next chapter, which will talk about the quantum eraser experiment, a last remark should be highlighted as concerns eliminativism. We have seen that physicists usually banish causality from physics as the consequences coming from the symmetry of dependence would be too problematic. And two comments should be given here.

First, the fact that eliminativism does not consider that causality could actually be different from what is thought commonsensically (namely, what we have highlighted about the manifest image of its asymmetric temporal direction). In other words, causality might not necessarily need to follow the time's arrow, and this is an option we will consider later.

Second, the fact that physicists do not do what they preach. That is, even if they agree to say that causality should be put aside, some actually continue to work this commonsensical idea. For instance, John Wheeler who introduced one of the first delayed-choice experiments in physics (which will play an important role in the next chapter), asks —about the results of the experiment— whether “this result means that present choice influences past dynamics, in contravention of every principle of causality” (Wheeler, 1978, p. 41). Notwithstanding, Wheeler does not clarify to what such “principle” he refers. What I aim to show is how strong are manifest images; this notion is so deeply structural of our way of perceiving the world, even *thinking*, that we assume causality can only but happen in one direction —from past towards the future. And this is the core of a problematic tension between the manifest and scientific images of causality.

CONCLUSION

Summing up, this chapter aimed to summarise the content of the manifest and scientific images of causality. On the one hand, we have seen that an important feature of causality in common sense is its direction; that is, we think that when two events are related by a causal connection, the cause must precede the effect. In other words, the causal arrow follows the arrow of the temporal asymmetry, namely the direction from past towards the future.

On the other side, we have seen that scientists are mute as regards the direction of causality. Their traditional view of the concept of causality is eliminativism, which claims that causality should be rejected from physics (or at least, microphysics, which in turn amounts to the same thing) —usually, they consider causality as some subjective human projection on the world. However, physicists *do* position themselves concerning the symmetry or asymmetry of dependence. That is, physical laws are symmetric in time and the usual explanation of the observed asymmetries at the macrolevel —namely, the principle of innocence— is that they come from boundary conditions. So, if the world *should* be symmetric, it is *not* —and there is sufficient observable evidence for that point. The problem is, at the microlevel, even though we would be tempted to extend the principle of innocence, there is no observable basis. In other words, asserting that two systems are not correlated *before* they interact benefits from no observable evidence. Therefore, most physicists claim we should stop considering the world as asymmetric (at least in the microworld) and begin to see it as the laws predict. Yet, the assumption of innocence at that level —which Price calls μ Innocence— is common in quantum physics, so that physicists do not actually do

what they preach. And it is no surprise that considering that there is no innocence, impacts on causality.

It should be noted that whether the concepts of law or interaction are connected to causality depends on previous assumptions: for instance, from Salmon's point of view they are intimately related for causality is an ontic entity of the world, while from the point of view of the manipulation theory, this link is not necessary. However, from the point of view of scientists, the relation between the principle of innocence and causality is assumed so that the former constraints the latter.

In conclusion, it seems that there is an incompatibility between the manifest and scientific image of causality. If the asymmetry of dependence does have an observable basis at the macrolevel —and, with that in mind, is compatible with common sense—, it is not the case at the microlevel. Notwithstanding, the question of the symmetry or asymmetry of dependence has not yet been solved among physicists. But in the next chapter, we will discover a quantum experiment that might force their hand.

Chapter 4

The Quantum Eraser Experiment: Against the Manifest Image of Causality?

1. Quantum Physics and Double-Slit Experiment

In the recent years, research in quantum mechanics has led to very interesting experiments, the results of which surprised physicists. These quantum experiments, as we call them, concern subatomic entities, like photons or electrons (even though they may have implications at the macrolevel). There is no completely accepted interpretation of quantum physics, though there is an orthodox interpretation based on Niels Bohr's Copenhagen Interpretation (Heisenberg, 2007).

Quantum physics has provided a large and complex development in science, which we cannot tackle in this thesis because of its complexity. Rather, we will focus on one of the strange quantum phenomena that still surprises physicists: the *quantum eraser experiment*. The quantum eraser experiment is intriguing in that it presents correlations that depend on events that have not yet happened. On the face of it, it would be tempting to claim that this counts as a proof of backwards in time causality; but this is a claim that cannot be taken lightly. For that reason, we will approach the questions concerning causality in the next chapter.

Quantum mechanics has been elaborated experimentally as well as mathematically thanks to the famous physicists Niels Bohr, Louis de Broglie or Erwin Schrödinger, amongst many others. But in this chapter, what will interest us is the experimental development of quantum mechanics, precisely the experiments derived from the previous double-slit experiment. Therefore, we will approach Thomas Young's double-slit experiment, and then the quantum eraser experiment.

THOMAS YOUNG AND THE DOUBLE-SLIT EXPERIMENT

The origins of the quantum experiments are to be found in Thomas Young's experiments, 19th century. Young wanted to prove that light behaves as a wave rather than as a particle —contrary to what most scientists since Newton thought, namely that light behaved as a particle (as announced, quantum physics concerns the entities of the microworld, so the objects quantum experiments deal with are photons or electrons). For that matter, Young set up an experiment proving, at the time, the wave-behaviour of photons. Over the course of time, this experiment led to the “double-slit experiment” in quantum mechanics.

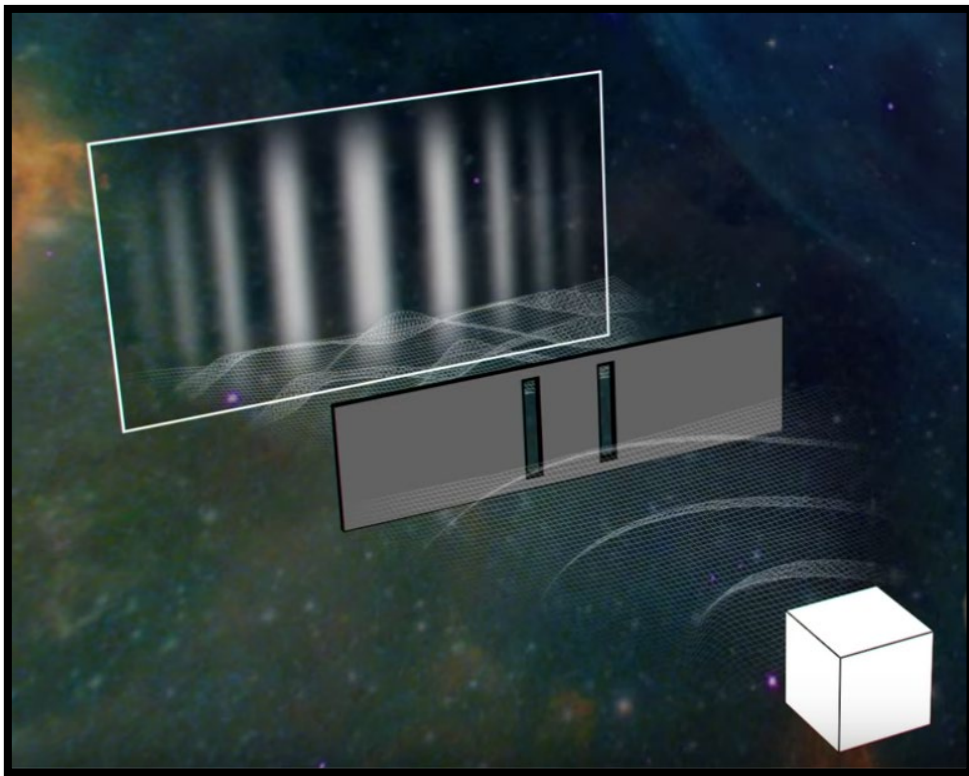


Figure 1 — A beam of light is directed towards an interference screen. The white cube is the source of light. The interference screen presents an interference pattern. (source: PBS Digital studio, “how the quantum eraser rewrites the past”, https://www.youtube.com/watch?v=8ORLN_KwAgs&t=428s)

What is common to the quantum experiments we will study is that a source of photons (or electrons), such as a beam of light, sends photons towards a screen on which each photon is observed; the screen is called an “interference screen”. Another common feature of those experiments is the use of slits: namely, in the middle of the photon’s run towards the screen, a plate is placed on which there is one or two slits, so that the photon can pass through it and reach the screen.

Now, let us describe the double-slit experiment in quantum mechanics (Figure 1). A beam of light is directed towards an interference screen, and between the source of light and the screen, a plate with two parallel slits is placed. When the experiment begins, light passes through the slits and the photons are observed on the interference screen, on which they appear as points. So, after several photons have passed through the slits, they are piled up on the screen. The expectation, if light behaved as a particle, would be that we should obtain *two* piles on the interference screen of the size of the slits. That is, we would expect photons to behave just like tennis balls would behave if they were shot (from the same standpoint) through two slits, and hit a wall beyond the slits.

But this is not what is observed. Instead, the result on the screen presented a pattern of interference: and a pattern of interference could only prove the hypothesis

that light behaves as waves. The reason is, if photons behave as waves, then any photon passes through both slits at the same time. Only, after the photon has passed through the slits, it interferes with itself —strictly speaking, the wave interferes with itself. Interference patterns can actually be found in many different contexts, such as in music, when two frequencies interfere with each other, or when we observe waves at sea. In the double-slit experiment, the interference is due to the slits and appears on the screen as an alternation of darker and brighter vertical bands, as different densities of points that appear as those bands.²⁴

DOUBLE-SLIT EXPERIMENT AND DETECTORS

Now, the double-slit experiment is definitely not the strangest quantum experiment that has been discovered in quantum physics. As we said, interference patterns can take many different forms; the double-slit experiment was proof that one of those forms is light, which hence has a wave-nature. But a later version of the experiment would really confuse physicists (Figure 2).

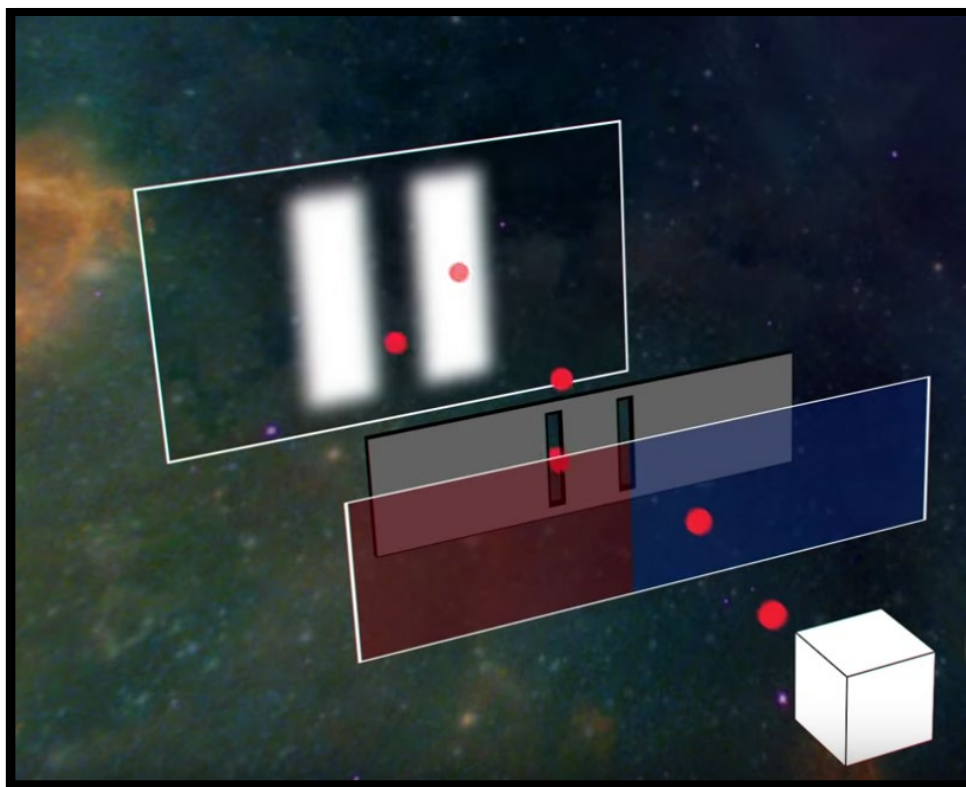


Figure 2 — In that new version with detectors (in blue and red), which can be placed before or after the two slits, the photons form two piles on the interference screen. (source: PBS Digital studio, “how the quantum eraser rewrites the past”, https://www.youtube.com/watch?v=8ORLN_KwAgs&t=428s)

²⁴ It should be noted that there is a version of the experiment with only one slit. In this case as well, the pattern on the screen is not a pile of the size of the slit; but a pattern larger than the slit —since the photon spreads as a wave. In a sense, we could say that the smaller the orifice, the broader the angle.

In that new version, *detectors* are placed after each slit, whose role is to light up when the photon passes through the slits. The purpose of this new setup is to verify that each photon passes through both slits at the same time, as would be expected given their wave-nature. In other words, the two detectors should light up simultaneously. Yet, the result of the observation is that when a photon is sent, only *one* detector lights up; in other words, each photon detected passes through *one* slit, indicating that it behaves as a particle. To describe this phenomenon, we say that the detectors “destroy” the interference pattern that is expected from a wave-nature entity and that this demonstrates the principle of the wave-particle duality of the nature of a photon. Richard Feynman said, about this version of the double-slit experiment, that it is

“a phenomenon which is impossible (...) to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the only mystery [of quantum mechanics].” (Feynman, 2010, pp. 1-1)

COMPLEMENTARITY PRINCIPLE

The wave-particle duality of photons is actually the manifestation of a more general principle, namely the *complementarity principle*. This principle, say Kim, Yu, Kulik and Shih is

“perhaps the most basic principle of quantum mechanics, [it] distinguishes the world of quantum phenomena from the realm of classical physics. Quantum mechanically, one can never expect to measure both precise position and momentum of a quantum at the same time.” (Kim, R. Yu, Kulik, & Shih, 2000, p. 1)

In another paper, Kim and Shih say about the complementarity principle:

“Another feature of this principle is that we cannot measure from which slit a photon passed through, and at the same time observe an interference pattern on the interference screen; and that is exactly what was observed in the double-slit experiment version with detectors. It is prohibited. We say that the quantum observables ‘position’ and ‘momentum’ are ‘complementary’ because the precise knowledge of the position (momentum) implies that all possible outcomes of measuring the momentum (position) are equally probable.” (Kim & Shih, 1999, p. 1)

It turns out that scientists are, to use an illustrative expression, somehow “fooled” by photons. That is, what they want to know is the “which-path information” of photons behaving as waves. But the presence or absence of detectors reveals

respectively either the which-path information (with destroyed interference) or an interference pattern (without which-path information).

In that perspective, other versions of the double-slit experiment took place in order to find what makes the presence of detectors so peculiar. And one of those consisted in delaying the uncovering of the which-path information. That is, given that each photon behaves as a particle when being measured, an alternative is to delay the measurement so as to let it reach the interference screen *before*. In so doing, scientists let the photon “think” that it will not be measured (that is, given the initial setup, there is no way one could know whether a measurement will take place or not). But even in those versions where the which-path information is measured *after* the photon has reached the interference screen (Ma et al., 2013; Ma, Kofler, & Zeilinger, 2016), the complementarity principle applies —whether or not the measurement is performed later is correlated with the interference or no-interference pattern on the screen.

2. Quantum Eraser Experiment

As announced, the quantum eraser experiment is derived from the double-slit experiment we have described in the previous section. We said that despite the first version of the double-slit experiment proposed by Young, that was thought to prove the unique wave-nature of light, physicists proposed a new version of it with detectors. Those detectors were supposed to light up when measuring a photon passing through one or the other slit; and against the expectations, every time a photon is sent, only one detector lights up, resulting in two piles of photons on the interference screen —instead of an interference pattern. This is the context that led scientists to elaborate the idea of the quantum eraser experiment.

THE TWO PURPOSES OF THE QUANTUM ERASER EXPERIMENT

Scully and Drühl proposed one of the first versions of the quantum eraser (Scully & Drühl, 1982), followed by Wheeler (1983), and Kim, Yu, Kulik and Shih (Kim et al., 2000). This last one will be the subject of our study in this chapter.

There are two purposes to the quantum eraser experiment. The first one consists in setting up an experiment so as to delay the measurement of the photon. That is, as the result of the experiment differs according to the presence or absence of detectors, it should be verified whether the measurement influences in any way the behaviour of the photon. It could be thought, indeed, that the presence of detectors impacts the photon so as to make it behave as a particle rather than a wave because it is in interaction with the whole system before being sent through the slits. So, the delay of

the measurement which takes place *after* the photon has reached the interference screen allows to check this possibility. In other words, the purpose is to see whether correlations hold even when the measurement happens *later* than the observation on the interference screen.

The second purpose of the experiment is to test to what extent the occurrence of the measurement influences the photon's behaviour. More precisely, the point is to observe what happens if the photon is measured but the information of that measurement —namely, the *which-path* information— is erased before being revealed. And to make the experiment more explicit, the erasure of the which-path information takes place unpredictably (the occurrence or non-occurrence of the erasure is independent of any previous event).

KIM ET AL.'S EXPERIMENT

Now, let us describe in detail Kim et al.'s experiment (Figure 3, Figure 5). A first feature of it, is the *non-linear optical crystal*. As we have said, the measurement of the photon must be delayed, and the crystal is the object that allows it to fulfil this condition. The crystal is placed after the two slits and generates two new twin photons, one going towards the interference screen, the other towards the quantum eraser setup.

Before a photon reaches the double-slit (and, right afterwards, the crystal), it must be sent by a source of photons: a *laser beam*. Then, there is a plate with a double-slit through which pass each photon. Behind each slit, there is a non-linear optical crystal.

The non-linear optical crystal is meant to generate “twin” photons from the incoming photon, which form an *entangled pair*. That is, any incoming photon is divided into two photons, each having half of the energy of the original one. Entanglement is a phenomenon observed in quantum mechanics, in which particles are connected in such a way that the description of the quantum state of one of them is dependent on the quantum state of the others, even if the particles are separated by a large distance and are not in each other's light cone (which implies nonlocality, a term that will be explained later). As we will see, the entanglement is central to the quantum eraser experiment. So, those entangled photons are sent in two different directions: one towards the interference screen, the other towards the setup containing the quantum eraser. This way, the screen-directed photon allows us to observe whether there is an interference pattern; and the *detector-directed* photon is used to know which slit the original photon passed through —the which-slit information.

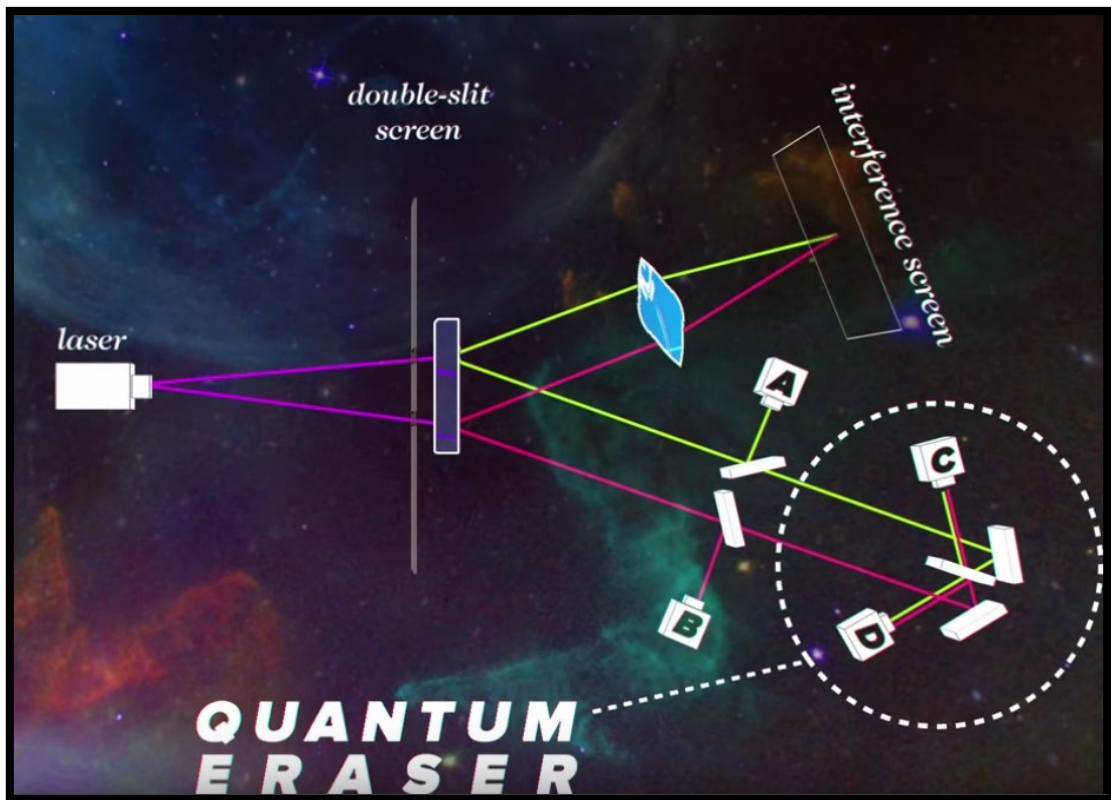


Figure 3 — The quantum eraser experiment. The dark-blue bloc after the two slits is the nonlinear optical crystal. The 5 white blocs are the beam splitters A, B, C and D are detectors. If we consider the green travel, the photon has a 50% chance of being reflected towards detector A, and a 50% chance of continuing towards the others beam splitters and hence to be reflected either in detector C or in detector D.

It should be noted that the screen-directed photon reaches the interference screen *before* (7.7 nanoseconds) the detector-directed photon reaches the setup containing the quantum eraser – the distance between the crystal and detectors A and B is 2.3 metres longer than the distance between the crystal and the interference screen. This fulfils the first purpose we shed light on in the previous section. So, any sent photon passes through either *slit A* or *slit B*, and is divided into two entangled photons (screen-directed photons A or B, entangled with detector-directed photons A or B).

SIMILARITY WITH THE DOUBLE-SLIT EXPERIMENT

Now, let us talk about the detector-directed photons. A twin photon whose original photon has passed through *slit A* activates the *detector A* and, as a matter of fact, a photon whose original photon has passed through *slit B* activates the *detector B*. Up to now, the setup of the experiment is similar to the double-slit experiment with detectors. Consequently, even when photons that have passed through *slit A* or *slit B* are measured *after* their screen-directed twin reached the screen, there is no interference pattern on the screen—it has been destroyed.

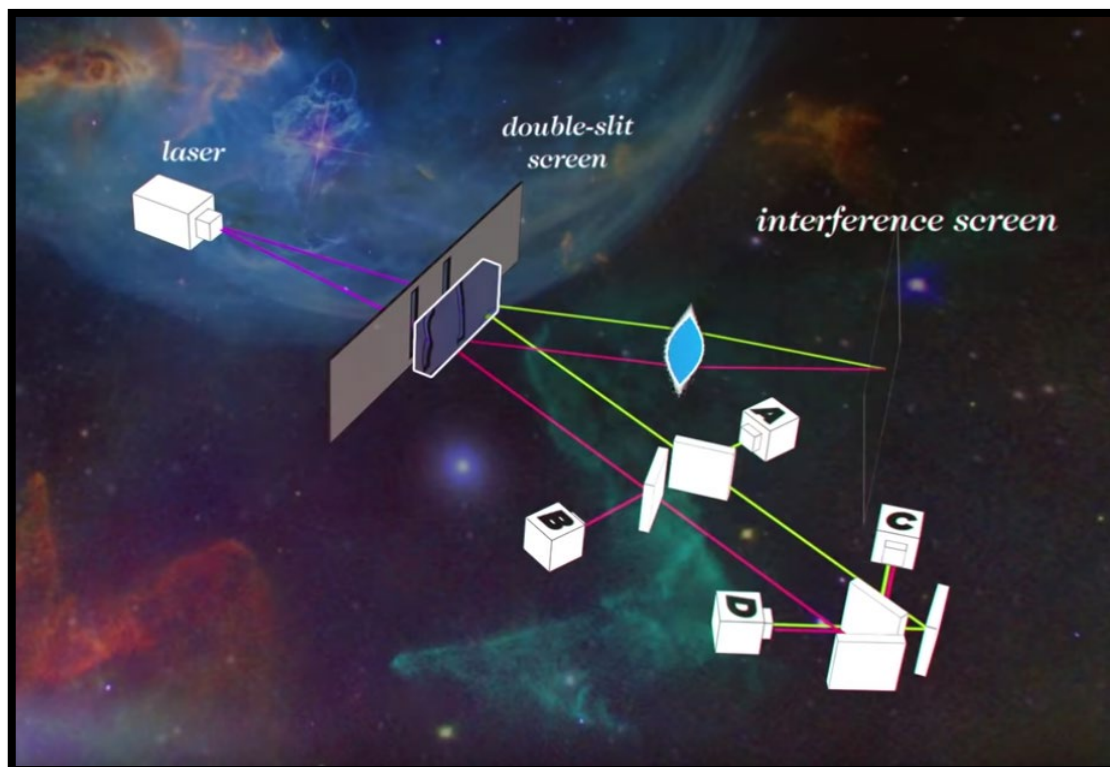


Figure 5 — The quantum eraser under another angle.

THE PECULIARITY OF THE ERASER

Now, let us describe the quantum eraser itself, which is what makes the experiment peculiar. Two detectors, *C* and *D*, are added to the setup. Any photon *A* (i.e. having passed through *slit A*) passes through the beam splitter α . The role of a beam splitter is to reflect 50% of the incoming photons towards the *detector A*; and to let the other 50% through. Those last ones continue their run towards another beam splitter γ . Again, γ reflects 50% of the incoming photons towards the *detector C*, and lets 50% of them through; these last ones continue their run towards *detector D*.

Exactly the same applies to any photon *B*. It passes through a beam splitter β that reflects 50% of them towards *detector B*, and lets the other 50% through. These last ones reach the beam splitter γ (the one we talked about earlier) that reflects 50% of them towards *detector D*, and lets 50% of them through, towards *detector C* (again, detectors *C* and *D* are the same ones we talked about earlier).

WHY IS IT CALLED AN ERASER?

The reason why it is called a quantum eraser is that, for any photon passing through *slit A* or *slit B*, there is a 50% chance that its which-path information will be erased. This 50% chance corresponds to the 50% chance that *detector C* or *detector D*

lights up. The reason is, when one of those two detectors lights up, there is no way to know from which slit the incoming photon comes, as the beam splitter γ lets 50% of the incoming photons through.

That is, the which-path information is conserved only when *detector A* or *detector B* are activated. So, the originality of the experiment of Kim and his colleagues, as compared to John Wheeler, is that we do not know whether a photon is going to be measured, nor if its which-path information will be conserved—or erased. It must be noted that, in the experiment, “we ‘see’ both wave (interference) and which-path (particle-like) with the same apparatus” (Kim et al., 2000, p. 2).

3. Results of the Experiment

The intriguing part of the experiment consists in the results of it. In effect, every time *detector C* or *detector D* light up, what is observed on the interference screen is an interference pattern. And every time *detector A* or *detector B* light up, the interference screen presents a destroyed interference pattern. That is to say, there is a correlation between the photon’s behaviour on the interference screen and the (later) unpredictable occurrence of the erasure of its measurement.

Therefore, the correlations show that there is an important sense in which relations of dependence hold between the twin-photons. Whenever the detector-directed photon is measured, the screen-directed photon behaved as a particle. Only, these correlations violate μ Innocence; in effect, the pattern on the interference screen is observed *before* the erasure or conservation of the which-path information.

BACKWARDS IN TIME?

This is the disconcerting part of the experiment: scientists tried as much as possible to somehow “confuse the issue”. That is, there is no way to know whether the which-path information will be erased before the twin has reached the screen. Therefore, even a bizarre hypothesis where the original photon “knew” that it would be measured later does not work. More precisely, in *detector C* and *detector D*, the photon *is* measured; but scientists “fooled” themselves so that the lighting up of *detector C* or *detector D* is uninformative—the measure is lost. In Xiao-Song Ma’s words, the aleatory erasure

“rules out the possibility that the photon knew the configuration beforehand and adapted its behaviour accordingly.” (Ma et al., 2013, p. 1221)

So, the experience taught us that the erasure or revealing of the which-path information is correlated with the *preceding* behaviour of the other twin photon on the interference screen. It seems therefore that the experiment is a case where an earlier event depends on a later one.

Wheeler said about the quantum eraser experiment :

“In this sense, we have a strange inversion of the normal order of time. We, now, by moving the mirror in or out have an unavoidable effect on what we have a right to say about the *already* past history of that photon. (...) Thus, one decides whether the photon ‘shall have come by one route or by both routes’ after it has *already done* its travel.”²⁵

ELIMINATIVISM AND COMMON SENSE

In most articles, scientists are very cautious in their description of the experiment, particularly in respect to causality—as we said in chapter 2, physicists adopt traditionally eliminativism. Werbos, for instance, said:

“(...) as Einstein emphasised over and over again, we need to pull ourselves away from our everyday intuitive attachment to time as a kind of magical ‘special’ dimension.” (Werbos & Dolmatova, 2000, p. 7)

There is, he admits, a “common-sense belief that causality can only operate forwards in time”(Werbos & Dolmatova, 2000, p. 7), but we need to put it aside. He adds up that Von Neumann “expressed the view that quantum mechanics appears to violate everyday notions about causality” (Werbos & Dolmatova, 2000, p. 3). Again, to shed light on the eliminativist tradition, Scarcelli et al. declare:

“Any attempt to interpret the physics of the quantum eraser in terms of complementarity examination on a single photon leads to counterintuitive results and paradoxical conclusions.” (Scarcelli, Zhou, & Shih, 2007, pp. 1-2)

So, these authors concede that the results given by quantum experiments—including the quantum eraser experiment—are counterintuitive and challenge common sense.

²⁵ Wheeler cited by (Ma et al., 2016, p. 6)

It should be noted that, in this quote, the expression “normal order of time” puts forward the argument given in the previous chapter, namely that the temporal asymmetry of dependence is important for experimentalists.

4. Nonlocality and eliminativism

This section will provide an explanation of nonlocality by way of another quantum experiment. Amongst the many different versions of the double-slit experiments²⁶, we will focus on the one of Ma, Kofler and Zeilinger (Ma et al., 2016). They set up an experiment whose aim was to prove that the correlations observed in the delayed choice experiment²⁷ are *nonlocal*. Precisely, their aim was to show the non-causal nature of the relation between the two entangled twin photons. That is, the moment where one of the twins is observed on the interference screen is *not* in the light cone of the event of the erasure (or conservation) of the which-path information. In other words, they say, any causal relation going from one to the other would violate locality; hence, the relation cannot be causal.

WHAT IS NONLOCALITY?

Nonlocality is actually a consequence of Albert Einstein's special theory of relativity. The theory of relativity is based on the principles that the speed of light is invariant in all inertial frames of reference, and independent of its source, and that the laws of nature are invariant in all inertial frames of reference. A consequence of the invariance of the speed of light is that two things that interact must be in each other's *light cone*. Precisely, it stipulates that, for any event e , we can determine the set S of all events that could be causally connected to e ; hence, S includes all events that would be able to reach e if they had a speed equal to (or slower than) the speed of light. This is commonly known as the principle of locality, and says that the interaction between two physical systems must be local.

In Ma et al.'s experiment, the proof is given that if there is any sort of communication between the two entangled photons, the information at the basis of the correlations should go *faster* than the speed of light. This, in turn, is interpreted by the authors as a proof of the noncausal nature of the correlation. Practically, Ma et al. set up detectors that were placed at 55 metres from the non-linear optical crystal, right after the double slit (in Vienna, 2007, through an optical fibre), and at 144 kilometres in another one (in the Canary Islands, 2008, via a free-space link). So, the choice of measurement (revealing or erasing the which-path information) was delayed long after one of the twin photons had reached the interference screen.

²⁶ Cf. (Ionicioiu & Terno, 2011; Peruzzo, Shadbolt, Brunner, Poposecu, & O'Brien, 2012; Scarcelli et al., 2007; Walborn, Cunha, Padua, & Monken, 2002)

²⁷ The delayed choice experiment is similar to the quantum eraser experiment, but does not contain an eraser.

“(...) the erasure event of which-path information had to be space-like separated from the passage of the interfering system through the interferometer as well as its detection event. Based on the special theory of relativity, the event of quantum erasure was therefore causally disconnected from all relevant interference events. (...) one was able to either acquire which-path information of the system photon and observe no interference, or erase which-path information and observe interference.” (Ma et al., 2016, p. 18)

IS ELIMINATIVISM THE SOLUTION?

Considering causality cannot hold in nonlocal cases, eliminativism seems a reasonable position for scientists, which explains this habit and Ma et al.’s position. They claim that “we should not have any naïve realistic picture for interpreting quantum phenomena” (Ma et al., 2013, p. 1226), and we should completely give up causality in the interpretation of quantum experiments.

“No naïve realistic picture is compatible with our results because whether a quantum could be seen as showing particle- or wave-like behavior would depend on a causally disconnected choice. It is therefore suggestive to abandon such pictures altogether.” (Ma et al., 2013, p. 1221)

Only, little has been said about what it means to be causal. That is, it is not evident that causality is a physical phenomenon—in which case there are better reasons to think that causality must obey special relativity. Ma et al.’s argument lacks a precise definition of what they mean by a cause—and by causality—which, in turn, appeals to a philosophical investigation. Their concept of a causality is both unclear and of minimal influence, for how we consider causality has a huge impact on scientific (especially physical) explanations or interpretations of the quantum eraser experiment. This is why a philosophical discussion is necessary to elaborate a non-trivial discussion about the causal or acausal nature of the experiment.

CONCLUSION

In this chapter, we have seen that since Young, quantum experiments have shown strange results²⁸, leading to different interpretations²⁹ of quantum mechanics. One element is nevertheless common to them: the dual nature wave-particle of quantum particles. We have studied Kim et al.’s quantum eraser experiment, which presents correlations whose relation of dependence does not follow the past-future direction of time; in other words, it contradicts the principle of μ Innocence that we

²⁸ Gaasbeek proposed to “demystify” the delayed choice experiments (Gaasbeek, 2010).

²⁹ The orthodox interpretation of quantum mechanics is Niels Bohr’s “Copenhagen Interpretation”. Cf. (Heisenberg, 2007, pp. 14-25)

discussed in the previous chapter (in which we have seen that it was already endangered as it does not benefit from an observational basis).

The fact that this kind of correlations obtains strongly suggests that we are dealing with a case of backwards causality. As we already discussed in the previous chapter, the absence of μ Innocence implies most probably that causality is symmetric and does not necessarily follow the arrow of time.

Most physicists follow the eliminativist tradition, even if the argument that causality is necessarily local lacks clarity, and hence appeals to a philosophical investigation. Nevertheless, it shows that there is a real incompatibility between common sense and science.

In the next two chapters, we will approach two questions. The first concerns clarification as regards the manifest image. Science suggests that common sense of causality should be rejected, but before opting for such a radical option, we should first thoroughly study our subject. So, we will try to find where common sense of causality comes from and why we intuitively think that the causal arrow should follow the temporal arrow. It might be useful in our project to accommodate the manifest image so as to be compatible with science. In chapter 5, we will determine what causality is and whether the quantum eraser experiment can be considered as a causal case.

Chapter 5

Looking for the origin of the manifest image of causality

1. Success as Planification

The quantum eraser experiment showed us that we are dealing with a striking incompatibility with common sense. The experiment presents a case where the principle of μ Innocence is not respected and so, strongly suggests to abandon the common sense of causality. But before considering adopting an eliminativist position and rejecting the concept of causality, we should try to find a less radical path. For that matter, we should determine where the common sense of the direction of causality comes from, and why it considers the causal arrow should be aligned with the time's arrow.

EXPERIENCE OF SUCCESS

So, where does the idea that the causal arrow should be aligned with the time's arrow come from? To this question, one could appeal to Humean regularities³⁰ and answer that the arrows are aligned thanks to the repetition of causal relations following the arrow of time. However, I do not think this is sufficiently satisfying to explain the origin of common sense. Rather, I find it more convincing to refer to the experience of planification.

The idea that causality stems from the experience of planification is to be found in Menzies and Price's paper, "Causation as a Secondary Quality" (Menzies & Price, 1993, pp. 193-195). As a reminder, a manifest image consists of conceptions of the world that surrounds us stemming from primitive experiences (the original image), and proceeds from the subjective experience we have of that world. And this is precisely what Menzies and Price have in mind: namely an experience we have at a young age, the experience of *success*, that is structural of how we perceive the world—in this case, agency gives us the causal structure of the world. And I point out that what matters in the manifest image of the direction of causality is precisely the experience of success as a planification.

As we have seen in chapter 2, the notion of causality stems from our condition of agent. Only, this notion is commonsensical in that it comes from the mere

³⁰ David Hume proposed that the concept of causality comes from regularities. That is, we know that two things are related by a causal connection thanks to the repetition of their correlations. Cf. (Hume, 1999)

experience of succeeding at bringing things about. To succeed is actually to plan an action towards a goal, and this consists in the refinement of the original image, that structures the way we relate to the world. That is, were we unable to manipulate things, we would not have the experience of manipulating things. Hence, we would have a different conception of causality —especially, we would not have structured our commonsensical perception of the direction of causality in such a way that it follows the temporal direction.³¹

WHAT DOES IT MEAN TO “PLAN”?

What I mean by planification is, strictly speaking, the deliberation that a free agent³² can make about its own acts to come. Planification is properly subjective, and this is the whole point, as we try to solve the origin of the manifest image of causality. In everyday life, we plan the occurrence of events by means of planning our future actions in order to reach our goal. For instance, when bowling, we plan the success of a “strike” —namely a shoot that hits all the pins. In order to do so, right before the shoot, we observe the pins and plan the means to succeed —the power, direction and spin of the bowling ball. In a planification, there must therefore be an agent capable of making free choices, and who thinks up an end.

It turns out that it is this experience that teaches us causality. When considering the means by which we can reach a purpose, we must unavoidably assume that there is some sort of relationship of dependence between the means and the end; and hence do we assume causality. Therefore, there is an important sense in which causality is a *tool* that allows us to make a success of a plan. Causality is, to say it roughly, a *side effect* of planification (which itself stems from agency).

PLANIFICATION ARROW AND TIME’S ARROW

Now, we should clarify the relation between planification and the temporal asymmetry of causality in common sense, as defined in chapter 2. And answers can be

³¹ In that respect, Craig Callender (2008) explains in his paper *The Common Now* why the (human) manifest image of time assumes that time follows the direction from past to future and is asymmetric. According to him, humans are presentist (cf. (Dainton, 2010, pp. 13-41)) because of our cognitive mechanisms.

³² We do not use the term “human” here, in order to avoid problematic anthropocentrism. A free agent could be an alien, though the word “free” could be disturbing. I use it here with reference to the fact that a human being is capable of making free choices and deliberate about those choices. James Woodward criticises the use of the terms “free agent” or “free choice” (Woodward, 2004, 2016), and Price proposed a reply (Price, 2014).

found in Frank Ramsey's "General Propositions and Causality" (Ramsey, 1978), which inspired Huw Price.

Ramsey thought that common sense of causality originates in the deliberation humans make about their own actions (in this sense, he is in accordance with the notion of planification). Yet, he says, we can only plan *future* things. There is nothing metaphysical in that feature of planification, as expected; all that we can plan lies ineluctably in the future. This has to do with our condition of human, which is made so that the past seems fixed, while the future is not. As a matter of fact, everything we are able to plan lies in the open future—as it cannot be fixed yet. Even though a planification cannot be reduced to a mere manipulation³³, the principle of planification is that something must be manipulated in order to reach the desired end. And this is why the planification arrow follows the time's arrow.

The argument of Ramsey explains the alignment of the causal arrow on the time's arrow by the observation that we deliberate counterfactually about our possible actions; that is, we wonder what is going to happen if we do such things. In the perspective of planification, we ask ourselves whether our goal is going to be achieved if we perform such action. Now, as we noticed in chapter 2, the fact that future depends on the present but not the past seems to be an essential feature of planification. We are therefore tempted to generalise the distinction between *before* and *after*, as the past seems unsettled as opposed to the future. In the end, we assume this difference is *a priori*. Hence, it turns out we think that "any present event is irrelevant to the probability for us of any past event" (Ramsey, 1978, p. 146). But this is unwise.

"What is true is this, that any possible present volition of ours is (for us) irrelevant to any past event. To another (or to ourselves in the future) it can serve as a sign of the past, but to us, now what we do affects only the probability of the future. (...) that I cannot affect the past, is a way of saying something quite clearly true about my degrees of belief. (...) We are then engaged not on disinterested knowledge or classification (to which this difference is utterly foreign), but on tracing the different consequences of our possible actions, which we naturally do in sequences forward in time, proceeding from cause to effect not from effect to cause." (Ramsey, 1978, p. 146)

Summing up, the planification arrow is aligned with the time's arrow. As human beings capable of deliberating about our actions, we can only affect the future. From this experience—the experience of planification—we deduce a generalisation about causality, for in common sense "we think explicitly in general terms" (Ramsey, 1978, p. 141). It follows that the alignment of the causal arrow with the temporal arrow

³³ More details about manipulation will be given in the next chapter.

is contingent, and that it would be possible to conceive counterfactually humans with capacities that allow to break that alignment.

One could notice that the fact that we have personally the experience of planification does not specifically indicate that other people can only affect the future as well; yet, it seems they do. Ramsey presents two replies; first, we proceed by analogy and hence think that other people, just like us, from their point of view can only affect the future. Secondly, we regard people's actions through the causal generalisation we talked about. In so doing, their actions must either be either a cause or an effect; and of course, it can only be a cause of what comes later (indeed, it never happens that we plan something in the past; and causality stems from the capability of planning). Therefore, given other people's action,

“we can only affect indirectly (in our calculation) events than it. In a sense my present action is an ultimate and the only ultimate contingency” (Ramsey, 1978, p. 146).

RECAPITULATION

Up to now, I have argued that the origin of the manifest image of causality lies in the experience of planification. As common sense is endangered by the scientific image in which, it seems, causality can go backwards in time, it was urgent to find out how common sense came to conceive causality as following the time's arrow. The experience of planification is the basis of the temporal asymmetry of causality in the manifest image; at a very young age, we experience our capacity as an agent and especially, of planification. Precisely, we have the experience of planning the occurrence of a desired event *B* and succeed at producing its occurrence by means of provoking the occurrence of another event *A*.

One could object that this is not sufficient to explicate how the notion of causality arises, and why it is aligned with the time's arrow. My reply contains two parts: one concerning the conception of causal connections, and the other based on the works of Ramsey. First, humans would be ignorant if they figured out that they had the power of acting and planning, but did not deduce causal relations between their actions and the desired end. That is, it would be very pointless to be able to succeed at producing certain events by means of others, and still not conceive that there is something more than mere correlations between these events. Again, we are talking about the manifest image; there is something intuitive in conceiving connections between our producing events and the success of the occurrence of a planned one. The notion of causality is therefore a *tool* supporting these connections, and hence being

structural of the way planification is perceived. This first argument was meant to defend that the notion of causality arose from the experience of planification.

My second reply is simpler and focuses on the arguments given in the last section. As we have seen with Ramsey, humans tend to think in general terms. It is thence no surprise that we generalise an experience, especially if the experience is a primitive one; which is the case of the experience of planification, intimately related to agency (and we have seen in the first chapter that the manifest image structures the commonsensical conceptions we have of things, and is based on primary experiences—the original image). We could say that this is what happens with colours. That is, we never see the colour blue itself, but objects that appear to us with a certain colour. That we can see objects in colours is a primitive human experience; and all we do is to generalise the colour blue from the objects appearing with shades, indeed different, but close enough to be named after the same word—“blue”.

There should therefore be no doubt about the fact that we generalise the experience of planification, whose main tool is causality. It appears that a desired event obtains thanks to the success of a planification, in which events produced by our own share a relation with the desired (and actualised) end; and these relations are what we call *causality*. Besides, these relations need to be asymmetric in time so as to lead to the occurrence of the desired event. In other words, any planification can only be oriented towards the open future, never to the fixed past.

2. Scientific Experiments, Causality's and Time's Arrow

One could feel that the gap between the manifest and the scientific images is too wide; that there is a void that we will never be able to fill. The only choice left is thence to distinguish two types of causalities. There would therefore be, on the one hand, causality as understood in common sense C_1 , aligned with the planification arrow; and on the other hand, causality as defined according to scientific claims about the relations of dependence—namely that it is asymmetric and that there cannot be μ Innocence, and hence no principle of innocence at all—, C_2 . In other words, C_1 matches what we said about common sense in chapter 3, while C_2 corresponds to the observation of the quantum eraser experiment in chapter 4.

Nonetheless, let us reconsider what Price said about those physicists who do not practise what they preach. Namely, he accuses some physicists of defining causality as asymmetric in time while there is nothing that supports μ Innocence in the microworld. They should, Price says, simply drop their commonsensical notion of

causality. But there is an important sense in which the temporal asymmetry of causality is *essential* for physics.

One of the important features of physics and of science in general, is empirical experimentation. Physicists as well as chemists proceed to experiments and these have an important role to play in the development of their sciences. And this is where my argument intervenes: without temporal asymmetry of causality, these experiments are irrelevant. That is, an experimenter has to take the temporal direction into account, for an experiment consists of a setup and a result. Or, to be more explicit, it consists of an *initial*, previous state of the setup and a *final*, later one. Were an experiment considered symmetric in time, then it would not make sense, for scientists would have no clue to know what is to be interpreted—that is to say, the initial or final states. In other words, science needs the asymmetry of causality.

Be it a human projection or not, the principle of innocence—and, with it, the direction of causality—is not purely subjective, it is necessary for a correct interpretation of scientific experiments. It might well be that at the microlevel, things happen differently than at the macrolevel and have no observational basis to assume μ Innocence. But there is an important sense in which the principle of innocence is essential to the experimental dimension of science and therefore, there cannot be two forms of causalities (one commonsensical, the other scientific).

3. Causal Arrow, Planification Arrow and Time's Arrow

In conclusion, it turns out that the reason why common sense aligns causality on the time's arrow, stems from the fact that it is itself aligned with the planification arrow. And the planification arrow is aligned with the time's arrow.

Only, the diagram is unbalanced. The quantum eraser experiment—as far as it is causal—implies that the causal arrow cannot be aligned with the time's arrow. The arrow of planification is the reason why common sense is incompatible with the scientific image given by physics. Causality can therefore not be considered asymmetric in time, as science contradicts it and common sense cannot defend it, as it comes from the generalisation of planification.

The outcome is therefore unavoidable: causality cannot be considered fundamentally asymmetric. However, two points should be taken into account before closing the subject. First, we have not decided yet whether the quantum eraser experiment is causal. Secondly, as we said in chapter 1, common sense is structural of the way we perceive the world. Such as it was for the case of time, it will be difficult to get rid of the commonsensical notion of causality; it would thence be an economic

move to make sense of causality in the light of the manifest image. These are those two points that we will tackle in the next chapter.

Chapter 6

The Quantum Eraser Experiment Is a Manipulation, Not a Planification

In chapter 3, we have explored the manifest image of causality, and ended up with the assumption of principle of innocence (and μ Innocence). This principle underlies a temporal asymmetry of causality. Yet, in the same chapter (and in chapter 4 especially), we have seen that science does not agree with this temporal asymmetry. Not only is there no observable basis available for μ Innocence, but the quantum eraser experiment strongly suggests that causality is symmetric in time —hence, that backwards causality is possible. At least, the experiment is troubling for it is a case that violates the principle of μ Innocence, which we should definitely get rid of. This led us to reflect about common sense itself, and it turns out that our commonsensical notion of causality stems from the experience of planification (or success, as Menzies and Price say).

Now, I draw attention to what we said in chapter 1 about manifest images. Namely, that it is precisely difficult to replace a manifest image, for it is constitutive of how we perceive the world —which is similar of how we perceive colours, we cannot replace the way we see the colour blue. Replacing common sense of causality is not necessarily the best option, for what is problematic is one of its features (the temporal asymmetry). There might be a way to make sense of the direction of causality in a way that is acceptable by common sense and yet compatible with the scientific image; a less radical solution than mere eliminativism, in fact.

In chapter 2, I have argued that Menzies and Price's theory of agency catches the manifest image of causality. Only, it is not precise concerning the direction of causality, and I have defended that in common sense, the causal arrow is aligned with the time's arrow. This feature of the manifest image is endangered for it seems to be incompatible with the quantum eraser experiment and hence. But it does not necessarily imply that we reject the manifest image of the nature of causality —which, I defended, is captured by Menzies and Price. Hence, our task is to find out whether the agency theory can be accommodated to the quantum eraser experiment. If it can, then only the assumption of the alignment of the causal arrow with the temporal arrow, stemming from the direction of planification, must be rejected and replaced. In this case, common sense of causality continues to make sense while being compatible with physics (which stipulates that the world is essentially symmetric in time —but asymmetric for contingent factors) and the quantum eraser experiment. Only, we have to show that the quantum eraser experiment is causal from an agentist point of view.

In this chapter, we will see what a manipulation consists in and analyse the quantum eraser experiment from a manipulationist perspective. If this is the case, we have proven that it qualifies as a case of backwards causality, considering backwards causality makes sense metaphysically (which is a point we will approach); but at the same time, we have proven that the manifest image of causality, as defined by Menzies and Price, is consistent. Therefore, we do not have to abandon the manifest image but only the assumption of the alignment of the causal and temporal arrows, and the upshot is the less radical possible.

1. Quantum Eraser Experiment Is a Manipulation

Our task is to determine if the quantum eraser experiment “matches” Menzies and Price’s theory; in other words, we will see if the experiment qualifies as a causal case if we follow Menzies and Price’s steps. In this section, we have two purposes: one is to determine *what* conditions must be fulfilled in order to be causal, from the manipulationist point of view; for that matter, we will approach the theory of intervention of James Woodward. The other is to analyse the quantum eraser experiment according to these conditions.

JAMES WOODWARD AND INTERVENTIONS

In chapter 2, we studied Menzies and Price’s theory of agency as a manipulationist theory of causality. Of course, this is not the only theory of manipulation available, and even if the positions amongst the theories of manipulation diverge, they agree about the idea that to cause something is to make a difference. Now, I defended in chapter 3 that Menzies and Price’s theory of agency catches the manifest image of causality. However, their theory lacks precision concerning the criteria that allow to distinguish what is a proper manipulation and what is not. Therefore, we turn to James Woodward.

James Woodward’s version of manipulation is different from Menzies and Price’s agency theory, especially in that he does not try to give an account of common sense. To him, there is a way to define causality objectively by referring to science, and common sense should be taken into account in the definition of causality. Woodward did not intend to make sense of common sense, and rather presented a version of manipulation which got rid of agency and its inherent subjectivity. That Menzies and Price shed light on the origin of the concept is one thing; but, according to Woodward, they did not capture what causality is. In other words, he separates the concept from its origin. As a reminder, they expressed formally their theory as follows:

“an event *A* is a cause of a distinct event *B* just in case bringing about the occurrence of *A* would be an effective means by which a free agent could bring about the occurrence of *B*.” (Menzies & Price, 1993, p. 187)

Woodward meant to keep the idea that the nature of a causal connection is a manipulation, which he calls an *intervention*. As he puts it, the relation between a cause and an effect should be expressed this way:

“an intervention on some variable *X* with respect to some second variable *Y* is a causal process that changes *X* in an appropriately exogenous way, so that if a change in *Y* occurs, it occurs only in virtue of the change in *X* and not as a result of some other set of causal factors.” (Woodward, 2000, pp. 199-200)

There is an important sense in which Woodward aims at providing an objective theory, that does not take common sense into account (this is why he accommodates his theory so that it is compatible with science). In other words, he does not pay attention to the manifest image of causality in his description of an intervention.

However, the disagreement between Woodward and Menzies and Price concerns more the interpretation of the nature of causality, than the theory itself. That is, despite what Woodward claims, his theory and Menzies and Price’s are part of the same family. My point is that, despite the obvious differences between the theories, these are close enough about their account of the concept of causality. Therefore, even if Menzies and Price have captured the manifest image of causality, their lack of precision can be completed by Woodward’s criteria for a manipulation which are compatible with the agency theory. In a sense, Woodward provided the good formal account for a manipulation to be a cause.

WOODWARD’S CONDITIONS

Woodward’s version intends to refine manipulationist theories (such as the agency theory, for instance) by considering the criteria that should be fulfilled in order to qualify as a genuine manipulation—that is, an intervention. Consider the following situation, where:

- *I* is an intervention on the variable *X*;
- *X* is a property of individual *i*;
- (G) is the postulated relationship holding between *X* and *Y*. That is, according to (G), any change in *X* causes or explains changes in the other variable *Y*—to do so, we consider the prediction given by (G) about the changes in *Y* produced by the intervention on *X*, and observe this intervention.

- Before the intervention, the value of the variable X possessed by i is called x_0 . After the intervention, its value becomes x_I .

Then, for I to be a genuine intervention, it should satisfy the four conditions given by (M). These conditions are:

“M1) I changes the value of X possessed by i from what it would have been in the absence of the intervention (i.e. $x_I \neq x_0$) and this change in X is entirely due to I .

M2) The change in X produced by I is claimed by (G) to change the value of Y . That is, according to (G), the value, y_0 , that Y takes when $X = x_0$, is different from the value, y_I , that Y takes when $X = x_I$.

M3) I changes Y , if at all, only through X and not directly or through some other route. That is, I does not directly cause Y and does not change any causes of Y that are distinct from X except, of course, for those causes of Y , if any, that are built into the I – X – Y connection itself; that is, except for (a) any causes of Y that are effects of X (i.e. variables that are causally between X and Y) and (b) any causes of Y that are between I and X and have no effect on Y independently of X . In addition, I does not change the causal relationships between Y and its other causes besides X . Moreover, a similar point holds for any cause Z of I itself —i.e. Z must change Y , if at all, only through X and not through some other route.

M4) I is not correlated with other causes of Y besides X (either via a common cause of I and Y or for some other reason) except for those falling under (M3a) and (M3b) above.” (Woodward, 2000, p. 201)

In respect to what we said earlier, these conditions do not necessarily involve human actions (e.g. agency). That is, as long as a causal process satisfies the conditions M1 to M4, it is an intervention —even though no human being is involved.³⁴ It is worth noting that these conditions guarantee the invariance of a manipulated system. That is, given Menzies and Price’s definition of a manipulation, one legitimate criticism is that it allows certain correlations to qualify as causal processes even if, for instance, the manipulation of the event A involves changes in event B but not via A . In that respect, conditions M1—M4 guarantee that what is at stake in the intervention —the chain I – X – Y — remains isolated.

IS THE QUANTUM ERASER EXPERIMENT A MANIPULATION?

Now we have chosen how we will determine whether the quantum eraser experiment consists in a manipulation —and whether that manipulation is explanatory

³⁴ Other features of the conditions M1—M4, such as the problem of circularity, are discussed in (Woodward, 2000, pp. 201-204).

or not—, we should apply the criteria proposed by Woodward. Before analysing the quantum eraser experiment, one point should be highlighted: the fact that two entangled photons are actually two separated entities. That is, we assume that two twin photons are not the same entity, like two parts of the same thing (which would explain more easily the mysterious correlations we observed); rather, we consider that, after the original photon has been divided into two twin photons, these are separated entities.

Even though the erasure occurs automatically, I consider it as an intervention for it is very similar to the double-slit experiment where detectors are set up in the experiment. Only, in the quantum eraser experiment, instead of humans putting detectors and make decisions that could interfere with the correlations, the measure is produced unpredictably, and automatically. So, the experiment is very similar to the other version of the double-slit experiment —where humans actually intervene by including detectors. Only, instead of a human intervention consisting in including or removing the detectors, the activation of detectors is automatic. Or, to be more precise, the conservation or erasure of the which-path information is automatic.

So, in the quantum eraser experiment, the intervention I is the erasure of the which-path information. That is, any photon passing through the slits will be *measured* in that experiment; what happens is the 50% chance of its erasure (when the photon is measured in *detectors C or D*), which hence has the role of the intervention. There is no need to say that the intervention does not involve human action. The variable X on which acts the intervention I , can take two values represented by the couple A-B and C-D, and is possessed by the detector-directed photon. Either the which-path information of the photon is *revealed* (x_0); or it is *erased* (x_1). The other variable, Y , can take two values as well. It is possessed by the entangled photon, directed towards the interference screen, and either behaves as a particle (y_0) or as a wave (y_1).

(G) is the postulated relationship holding between X and Y . In that case, (G) is supported by the entanglement holding between the two twin photons. That is, the entanglement predicts that any change in X in the detector-directed photon and predicts that any change on X produces change on Y : any change on the detector-directed photon (measurement revealed or erased) implies change on the screen-directed photon (particle or wave behaviour). This is also supported by standard quantum mechanical calculation (Kim et al., 2000, pp. 2-3). Then, we observe the intervention according to (G). In the quantum eraser experiment, X and Y can take only two values. Now, let us focus on M1—M4.

There is no doubt that condition M1 is fulfilled. The intervention of the erasure changes the value of X from what it would have been without it; and we know that by

observation. In effect, the peculiarity of the quantum eraser experiment is that the occurrence of the eraser is aleatory and has a 50% chance of happening. In other words, we observe that when the photon is not observed in detectors C or D, it is measured in detectors A and B—hence, the which-path information is revealed. That is, when the erasure does not take place, the value of X is x_0 . Hence, the intervention changes the value of X from what it would have been without it.

M2 is satisfied as well. The supposed relationship holding between the two entangled photons says that the screen-directed photon behaves as a wave when there is no quantum erasure; and this behaviour is different from when an erasure takes place—namely, the screen-directed photon behaves as a particle. We know counterfactually that if the which-path information is erased, the behaviour of the photon on the screen is wave-like (and if it is not, we have the which-path information and the photon in the screen is particle-like).

As regards M3, the erasure of the which-path information of the detector-directed photon does change the behaviour of the screen-directed photon, and this only through its twin photon. Two things should be highlighted: first of all, we are dealing with an experiment, not with some natural phenomena we could encounter in a random context. That is, the purpose of such an experiment is to shed light on the—supposed—causes while all other variables remain invariant. The quantum eraser experiment is set up so that the occurrence of the erasure does not affect other variables than those of the twin photons. Of course, this does not mean that the experiment is indubitably invariant; but it strongly suggests that the causal relationship holding between I and Y in the experiment, passes only through X or any effect of X affecting Y . Secondly, we should look closer at the correlations at stake. In the experiment, we observe that Y never takes the value y_I (wave behaviour), besides when X takes the value x_I . Furthermore, on the basis of other versions of the double-slit experiment, we know that the behaviour of the photon on the interference screen is correlated with the measurement of it (or of its entangled twin). Therefore, there is no way the erasure of the which-path information could be a direct cause of the patterns observed on the interference screen. A last argument concerns the nature of quantum mechanics itself. One of the important features of the quantum eraser experiment, is that the pattern on the interference screen appears *earlier* (8 nanoseconds) than the lighting up of the detectors.

“The experiment is designed in such a way that L_0 , the optical distance between atoms A, B and detector D_0 , is much shorter than L_i , which is the optical distance between atoms A, B and

detectors D_1 , D_2 , D_3 , and D_4 , respectively. So that D_0 will be triggered much earlier by photon 1.” (Kim et al., 2000, pp. 1-2)

So, even if we consider that backwards causality is possible, it would be surprising that an intervention such as the quantum erasure could affect *directly* the behaviour of the twin photon on the interference screen. Rather, it seems to me that the only possible way for the intervention to affect it, is through the entanglement of the two photons. And entanglement is a special characteristic of quantum mechanics—just as the fact that observation does affect the phenomena at stake, as opposed to classical mechanics where observation is neutral. Hence, I causes Y only through X . One could object that we take it for granted that a measurement can be considered as an intervention, which is not evident. Only, I reply that one cannot perform a measurement on an entangled photon without disturbing the system, because this is how quantum mechanics works. And this is why Woodward accepts that, at the quantum level, a phenomenon that does not follow the laws of quantum mechanics which are unmanipulable, is *causal* in principle. That is, we can make a counterfactual based on scientific experiments, which is very similar to what is accepted as a manipulation (Woodward, 2016). So, that is not to say that all counterfactuals are manipulations, but in the case of quantum mechanics, it is convenient to consider a measurement as a manipulation. Hence, in the quantum eraser experiment, we know counterfactually that if the erasure did not occur, then the screen-directed photon would behave as a particle (which correlated with the fact that we conserve its which-path information), and so the erasure of the which-path information is an intervention (if it fulfils the conditions M1—M4).

Concerning M4, the experiment presents no correlations between the erasure and other causes of the screen-directed photon—besides the directed-photon, whose which-path information is either revealed or erased by the non-occurrence or occurrence of the intervention.

CONCLUSION

In this section, we have seen the conditions Woodward believes are relevant for an event to be an intervention. I have argued that we have good reasons to add these conditions to Menzies and Price’s agency theory, given their imprecision (and the compatibility of the conditions with the theory). These conditions guarantee that the intervention on one variable X , provoking change in another variable Y , does it only through the variable X . Roughly speaking, Woodward’s criteria ensure that the whole

system remains invariant, besides the variables implied in the supposed causal relationship.

It turns out that those conditions are fulfilled by the quantum eraser experiment. In other words, we can say that it consists in a manipulation and hence, that there is a causal connection between the revealing or erasure of the which-path information, and the behaviour of the twin photon of the interference screen. The experiment is therefore causal, if we adopt the manipulation theory. Now, the puzzle concerning the backwards nature of this case of causality remains, and needs further investigation as regards the possibility of backwards causality.

2. Quantum Eraser Experiment and Bilking Argument

Our further investigation leads us to two famous arguments concerning backwards causality. Those were formulated by Max Black—the bilking argument—and by Michael Dummett, whose argument consists of a thought experiment very close to the quantum eraser experiment.

BILKING ARGUMENT

In 1956, Max Black discussed the paradoxes involved by backwards causality. One of them is similar to the paradox of time travel; it is commonplace that, when watching a movie such as *Back to the Future*, we are confused by the implications of travelling in the past. I refer to the *grandfather paradox*: if Marty travels back in the past and prevents his parents from meeting, he prevents himself from being born, hence he cannot exist and travel in the past.

Max Black defends that the same applies to backwards in time causality, thanks to his *bilking argument*. Consider that someone claims we can produce an event A at t_1 by bringing about an event B at t_3 . In other words, there is a positive correlation between A and B , so that B increases the probability of occurrence of A . The bilking argument says that the correlation can be reliably called “causal”. That is, there is a way to disturb the correlations between B and A , so that the claimed causal connection between them does not hold anymore.

Imagine we prepare a device that is able to detect the occurrence of A . When A happens, the device provokes an event C at t_2 so that a positive correlation holds between A and C . But this is not all: the device guarantees that B occurs if and only if C does *not* happen. That is, the device ensures that a negative correlation holds between B and C . Hence, we end up with a striking impossibility, and the claimed correlation between A and B is “bilked”.

To illustrate, consider the following example. Consider a magician who claims he is capable of determining which card from a 54-cards deck you are going to take. The game is simple: you ask him which card you are going to draw, he tells you the name of a card, you shuffle the cards in order to ensure he is not cheating, and then draw one randomly. And so, the game begins. The magician claims that the Queen of Hearts will be drawn (*A*); you shuffle the cards, pick one card and discover it is, indeed, the Queen of Hearts (*B*). But there can be no logical causal connection between the two events, despite the correlation (and the low probability for the magician to claim the name of the right card).

Of course, we must assume that the claim of the magician does not influence the drawing of the cards (for that matter, all cards are face down). In other words, the Queen of Hearts would have been picked, whether we had asked our question to the magician or not. We make the assumption also that the two events are not due to a common cause, and that the guess of the magician is due only to the future drawing of the Queen of Hearts—and not to some other earlier event. The cause of the drawing of the Queen of Hearts is a sufficient cause for its occurrence, but is causally independent from of the magician's guess.

In this situation, the bilking argument stipulates that there can be no causal link between the later event (drawing of the Queen of Hearts) and the earlier one (the magician's guess). That is, the device we have talked about consists in you, acting after the magician's guess; after the magician says "the Queen of Hearts will be drawn", you can prevent the occurrence of the drawing of the card (*C*). As the drawing (*B*) is independent of the guess (*A*), we are free to simply *not draw a card*.

One could object that *B* causes *A* only when *B* occurs—*A* is therefore due to another cause when *B* does not happen. That is, the drawing of the Queen of Hearts must happen in order to bring about the guess of the magician, who can claim "Queen of Hearts" for other reasons in situations where the drawing does not occur. But, Black replies, there is a way to contradict this point by learning how to draw a card, and yet managing to draw another card. For instance, when the magician claims "the Queen of Hearts will be drawn", you discretely look at the cards, find the Queen of Hearts and place it at the top of the deck; hence, you know you have to take any other card of the deck. This being done, the magician's claim is in disagreement with the drawing of the card. Furthermore, if we figure out that we cannot manage to find the Queen of Hearts every time we try to do so, this means that the drawing of the Queen of Hearts is *dependent* on the magician's guess—hence, one of the conditions we have listed is not satisfied.

Black claims that this impossibility applies to all cases of supposed backwards causality; it is a logical impossibility inherent to the direction of causality. The bilking argument is quite powerful and might seem difficult to disagree with it. However, Michael Dummett proposed a reply to Black's thesis that we will approach in the next section, and might offer a good defence of the quantum eraser experiment as a case of backwards causality.

DUMMETT'S TRIBE EXAMPLE

In 1964, Michael Dummett published his paper "Bringing about the past" (Dummett, 1964) where he provides an answer to the bilking argument. However, Dummett believes that the asymmetry holding between past and future, when we describe the causal relation of two events, actually reflects the objective asymmetry of causality. That is, were we mere *observers* instead of *agents*, we would still make an asymmetric account of causality. More precisely, Dummett claims that even though our conception of causality stems from agency (in this sense, he refers to Menzies and Price's theory of agency), we would build another concept of causality if we were mere observers. And this concept would encompass the same asymmetry.

But let us consider the fact that we are agents in the world. As we have discussed in chapter 3, backwards causality seems absurd from a commonsensical point of view, and this has to do with our condition of agents. When asked *why* we are not able to bring about a past event, a first answer that comes to mind is that we are not able to modify things that have already happened. As a matter of fact, we cannot make an event not to have happened, if it has already happened.³⁵ It is therefore insane to think an effect could precede its cause, or that we could bring about an event in the past.

But entertain the following situation. After I have taken all my exams, I have a few days free before the revealing of the results. Stress increases over the course of time, and on the D-day, I log into the website of the university and open the link that contains my results. Strangely, I find myself praying or asking that I passed the philosophy exam. This kind of situation is not atypical, but in a sense, is strange. That is, when praying or asking "Please, make it so that I passed the exam!", all I am doing is trying to influence a past event. For instance, I pray that I answered correctly to the questions, which is useless, of course: I do not have access anymore to that event, which is already settled.

³⁵ One of the consequences of this, is the direction of planification.

A similar example led Dummett to point out something important in the bilking argument: namely, that for a correlation to hold between A (at t_1 —the magician's guess) and C (at t_2 —the decision to not draw a card), the observation is fundamental. This might seem obvious, but it is central to Dummett's argument. For the correlation to exist, there must be an observation of A . Price points out that

“it is assumed that the presence of such an observational procedure could not provide a plausible explanation for the failure of the contentious A-B correlation.” (Price, 1996, p. 173)

Dummett exploited this point and proposed the “tribe” thought experiment, in order to show that backwards causality is coherent, as long as there is a way to know whether the supposed earlier effect A has occurred, before the supposed later cause B has happened. Imagine we visit a tribe, where some traditional ritual is to send teenagers hunting, far from the village, for two days. Only, they have to walk for two days to reach the hunting place, and it takes two days to come back. So, the whole trip duration is six days. The point of the ritual is that during the hunt, observers of the same tribe accompany the teenagers in order to report, back to the village, whether they have been brave or not —which is the whole point of the ritual. Surprisingly, the chief of the village dances a “courage dance” that is meant to make the teenagers be brave during the hunt. Only, the chief does not dance during the two days of the hunt; instead, he dances every day, for six days, until they come back. That is, he dances even when the hunt is already done and the teenagers are on their way back.

Surprised by the chief who thinks he can affect the past, we tell him that to keep dancing after the hunt has taken place, is absurd. That is, he would not think about doing it *after* the teenagers have returned. But the chief replies that it is precisely because he *knows* whether the teenagers have been courageous or not. We then tell the chief, in vain, that he can only affect the teenagers' behaviour during the first four days, not after. But he replies that experience is against that (every time he performed the dance for the six whole days, the teenagers had been courageous while the few times he only performed the dance during the first *four* days, things went badly for them and they acted cowardly).

But then, we argue this: *either* the teenagers have been brave, in which case they have been brave whatever the chief does next (a dance, for instance); the chief's dancing will not be effective for causing the occurrence of their having been brave. *Or* the teenagers have been cowardly, in which case they have been so whatever the chief does next. In other words, either the chief's dancing is superfluous, or it is fruitless. But the chief replies:

“If they have been brave, then indeed there is a sense in which it will be true to say that, even if I do not dance, they will have been brave; but this is not incompatible with its also being true to say that, if I do not dance, they will not have been brave. Now what saying that my continuing to dance is effective in causing them to have been brave amounts to is that it is true both that, if I go on dancing, they have been brave, and that, if I do not dance, they have not been brave. I have excellent empirical grounds for believing both these two statements to be true; and neither is incompatible with the truth of the statement that if I do not dance, they have been brave, although, indeed, I have no reason for believing *that* statement.” (Dummett, 1964, p. 350)

And so on and so forth. Every time we try to defend the incoherence of believing in backwards causality, the chief finds a satisfying answer. In the end, says Dummett, it turns out that the chief’s belief is not that his dancing brings about the brave behaviour of the young men. Rather, it is his performance combined with the chief’s ignorance of their behaviour during the hunt. And the experience makes this belief accurate and rational: there is a positive correlation between the chief’s dancing and the teenagers’ having been brave. So, Dummett says, the chief began with three beliefs:

“(i) that there was a positive correlation between his dancing and the previous brave behaviour of the young men; and (ii) that the dancing was something in his power to do as he chooses. We are tempted to think of these two beliefs as incompatible, and I described people attempting to devise a series of experiments to convince the chief of this . [...] But in the process, a third belief, which we naturally take for granted, has had to be abandoned in order to hang onto the first two: the belief, namely, that it is possible for me to find out what has happened (whether the young men have been brave or not) independently of my intentions.” (Dummett, 1964, p. 357)

What Dummett’s thought experiment provides in the debate, is that the bilking argument cannot apply in this case. From the chief’s perspective, there is no inconsistency in believing that he can perform an action that brings about an event in the past, as long as he has no way to know what has happened. Therefore, the chief cannot apply the bilking argument—namely, to bring about the event *C*.

My conclusion therefore is this. If anyone were to claim, of some type of action *A*, (i) that experience gave grounds for holding the performance of *A* as increasing the probability of the previous occurrence of a type of event *E*; and (ii) that experience gave no grounds for regarding *A* as an action which it was ever not in his power to perform—that is, for entertaining the possibility of his trying to perform it and failing—then we could either force him to abandon one or other of these beliefs, or else to abandon the belief (iii) that it was ever possible for him to have knowledge, independent of his intention to perform *A* or not, of whether an event *E* had occurred. Now doubtless most normal human beings would rather abandon either (i) or (ii) than

(iii), because we have the prejudice that (iii) must hold good for every type of event: but if someone were, in a particular case, more ready to give up (iii) than (i) or (ii), I cannot see any argument we could use to dissuade him. And so long as he was not dissuaded, he could sensibly speak of performing *A* in order that *E* should have occurred.” (Dummett, 1964, pp. 357-358)

ANALOGY BETWEEN TRIBE EXAMPLE AND QUANTUM ERASER EXPERIMENT

Price points out a weakness in Dummett’s defence: it takes for granted that the way we can observe the outcome of past events does not disturb anything. More precisely, Dummett defeats the bilking argument by showing that, in the mechanism linking *B*, *A* and *C*, the correlations holding between *A* and *C*, and those between *C* and *B*, disturb the circumstances necessary for *B* and *A* to be causally connected. In other words, it is not the absence of correlations needed for the triangle *B—A—C* to be closed that contradicts the bilking argument.

“So, a perfect parallel would be a case in which the past-directed bilking argument was defeated not by the nonexistence of a means to observe the relevant past events—as Dummett assumes—but by the fact that the mechanism of such observations inevitably disturbed the system in such a way that the claimed correlation failed to hold.” (Price, 1996, p. 174)

In this section, we will see that there is an important sense in which the quantum eraser experiment matches the condition highlighted by Dummett and the additional comment of Price. As we will see, the quantum eraser experiment is very close to the tribe example. Consider the following features of the bilking argument:

- *A* is the erasure of the which-path information, or the dancing of the chief;
- *B* is the behaviour of the screen-directed photon on the interference screen, or the brave behaviour of the young men who went hunting;
- the brave behaviour of the young men corresponds to the wave behaviour of the entangled twin photon;
- and the cowardly behaviour of the young men corresponds to the particle behaviour of the twin photon.

Just as for the chief, experience shows that every time an erasure is performed in the quantum eraser experiment, the previous behaviour of the entangled twin photon was wave-like. The chief sometimes could not manage to dance (say, for he was sick and could not move), which was correlated to the young men having acted cowardly; and the same applies to the experiment. When the erasure does not take place, for its occurrence is *aleatory*, the behaviour on the interference screen of the twin photon is particle-like. So, the real question here, is whether the quantum eraser experiment defeats the bilking argument—in other words, how analogous it is to the tribe

example. And I will defend that the quantum eraser experiment is a perfect case defeating the bilking argument, hence being a coherent and acceptable case of backwards causality.

The first point we should point out, is the fact that in Dummett's tribe example, from the chief's point of view, we do not know what was the young men's behaviour. Formally speaking, from *B*'s point of view, there is no way it knows anything about *A* (even though it is practically possible, from our external point of view, to inform *B* about *A*; namely, through a device that introduces *C*, as we have seen in the bilking argument). So, we must approach the case from the eraser point of view, whose action of erasure is analogous to the chief's dancing.

Secondly, we must assume that the supposed causal connection between the two events is not directed forwards in time. This is supported by the fact that the beam splitters unpredictably let 50% of the photons through; therefore, I do not see how one could defend that the previous photon behaviour on the screen influences the beam splitters, so as to let a twin photon being reflected in detectors *A* or *B*, or being let through towards detectors *C* or *D*.

Thirdly, and most importantly, we must consider whether the experiment could be bilked. In Dummett's example, the chief of the village would perform his dance while not knowing what had been the behaviour of the young men, and this was essential for it to be claimed a retrocausal case. Dummett provided one possibility to bilk the causal link between the dancing and the young men's behaviour, namely by imagining that the reporters who accompany the hunters inform the chief of their behaviour.

"Once he had the information, provided by the observers, about the behavior of the young men, then, under the old dispensation, his attitude changed: he no longer had a motive to perform the dances." (Dummett, 1964, p. 355)

As he says, the bilking is defeated for the attitude of the chief changes. The same applies to the quantum eraser experiment: it cannot be bilked. The reason is, in order to get the information of the screen-directed photon, a *measurement* must be performed. And, as a matter of fact, any measurement in a quantum experiment disturbs the whole experiment (more precisely, any measurement destroys the interference pattern, as in the double-slit experiment with detectors). And this is *precisely* what Price talks about, when he says that a perfect case contradicting the bilking argument, is one where the mechanism of the observations disturbs the system, so that the claimed correlation fails to hold.

“In Dummett’s terms, for example, the parallel case would be that in which the ritual was believed to be effective only under conditions which would be defeated by any attempt to discover the outcome of the hunt in advance.” (Price, 1996, p. 174)

CONCLUSION

In this chapter, we discussed what it means to manipulate something. We appealed to Woodward’s conditions in order to complete the agency theory. According to these conditions and Woodward’s position concerning quantum mechanics, we showed that, from the point of view of the manifest image, the quantum eraser experiment is causal. Hence, there is an inherent incompatibility, for one of the features of common sense is that backwards causality is not possible.

Therefore, we started a analyse of the coherence of backwards causality. Max Black provides a striking defence which deserves a consistent answer if we wish to defend the possibility of backwards causality. In cases such as the paradoxes encompassed by time travelling in the past, it is possible to bilk the causal connection holding between the later and the earlier events. However Michael Dummett admits the power of the bilking argument, he provides a kind of case where it does not work: namely, when—from the point of view of the later event—we cannot know what was the outcome of the earlier event. Precisely, it is not the fact that there is no, say, *practical* means to know what was that outcome. Rather, it is the mechanism of the causal system that makes any attempt to bilk it useless for it significantly *disturbs* the system. And this is precisely what happens with the quantum eraser experiment, which provides a case intriguingly close to Dummett’s tribe thought experiment.

Summing up, this chapter proved that the quantum eraser experiment qualifies as a case of backwards causality from a commonsensical point of view. The alignment of the causal arrow with the planification arrow must therefore be rejected.

Conclusion

In this thesis, we started with a presentation of the Sellarsian distinction between the manifest and scientific images of man-in-the-world, a distinction that is interesting in that it shows certain incompatibilities. The manifest image is actually what phenomenologists deal with, namely the structure of our perception of the world, which is common sense. Hence, common sense does not stem from ignorance or naïve ideas; it is properly constitutive of how we think and perceive the world. But incompatibilities can arise when confronting it with the scientific image.

When we applied the distinction to the notion of causality, I defended that Menzies and Price provided the best theory—namely, the agency theory—describing the manifest image of causality, despite the inherent difficulty to give an account of the structure of how we perceive the world. Their theory is all the more satisfactory because it is compatible with science—or, at least, with fundamental physics which I chose as the most representative of sciences given the purpose of my thesis, which deals with quantum mechanics. That is, the idea at the core of the agency theory is that something is a cause when it *makes a difference* (an idea shared by many other theories). The agency theory not only catches the manifest image of causality, but also is compatible with science; which is something that theories such as Salmon's theory of propagation-interaction does not (since it is not his purpose to capture common sense). Yet, the theory of propagation-interaction can have an explanatory role, in cases that the agency theory considers as causal—for there will be cases where the manipulation actually is brought about by a propagation.

Afterwards, we deepened our analysis, and pointed out a feature of the manifest image that is not taken into account in Menzies and Price's theory of agency: the direction of the causal arrow. It is indeed a commonsensical assumption that a cause always precedes its effect. Yet, in physics, the laws of physics are symmetric in respect to time, which implies some incompatibility between the manifest and the scientific images of causality. And even if one could think we should reject common sense, things are not so simple—again, common sense is structural of our perception. One point supporting my thesis is that, even amongst physicists, the assumption of the alignment of the causal arrow with the temporal arrow remains. As Price puts it, they do not do what they preach—to reject the commonsensical assumption of their claims—which is due to the importance of common sense. Furthermore, scientific experiments can only work *with* the assumption that the causal arrow is aligned with the temporal arrow. That is, it would be ridiculous to setup an experiment, and consider it as the result of a future state of the experiment. To interpret correctly a scientific

experiment, scientists must assume that the correlation holding between the final state and the initial state follows the time's arrow.

But a striking incompatibility between common sense and science arises when we faced the quantum eraser experiment, supporting the disagreement between the manifest and the scientific images of the direction of causality. Assuming that a causal relation holds in the experiment (which we showed later), the experiment shows a case of backwards causality, where the conservation or erasure of one photon influences the previous behaviour of its twin (considering, as we discussed in chapter 6, that the photons are separated entities). We know that the direction of the correlations does not go forward in time, for the occurrence of the erasure is unpredictable, assuming the past-to-future direction of time. In other words, we have to deal with correlations that are mysterious for we cannot find a common cause between the two photons (unless we look at the future).

The quantum eraser experiment has an important consequence: it implies a revision of the commonsensical concept of causality. Common sense is incompatible with physics, and there are few reasons to think we should prefer common sense to science. Furthermore, we must take into account the fact that, to modify common sense is a difficult project. As it is structural of our perception of the world, it is hard to change it; that is, it would be like asking someone to stop seeing blue objects as blue. This is why we cannot change our commonsensical notion of time, despite the revision that requires the special theory of relativity. And even if it is not our task to consider the feasibility of a modification of common sense, I think this is a sufficiently convincing argument to try to accommodate common sense, so as to "save it". That is, it would be a real advantage to make the commonsensical notion of causality compatible with science, and prevent it from being merely "a useful fiction". As Sellars puts it, the manifest image is an adequate conception of the world, and this was my guideline.

Therefore, we were left with two tasks. The first one was to analyse where the commonsensical notion of the direction of time stems from. The other was to analyse whether it is possible to make sense of the quantum eraser experiment as a causal case, from the point of view of the agency theory.

I showed, with Ramsey, Menzies, and Price, that humans have at a very young age the experience of planification (which is what Menzies and Price call the experience of *success*). And as humans use to generalise their experiences (which, from a Sellarsian point of view, can be interpreted as the construction of the manifest

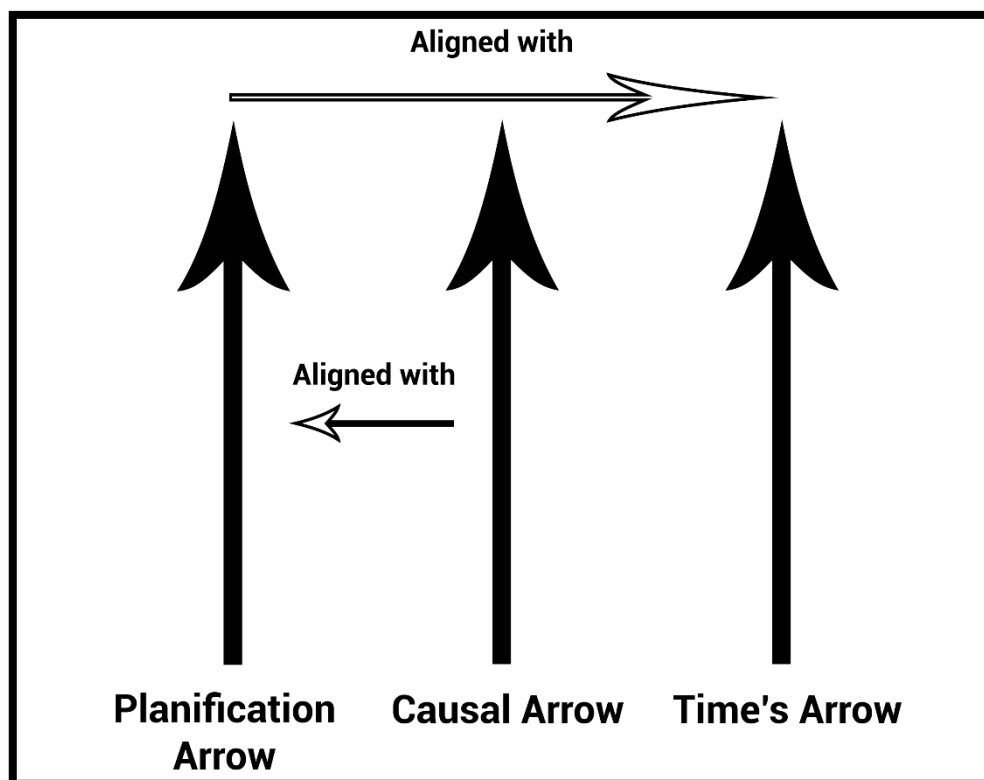


Figure 6 — Alignments of the three arrows.

image of the world), they generalise the direction of planification. Consequently, as causality is a “side effect” of planification³⁶, the causal arrow is aligned on the planification arrow. This implies that causality is related to our capacities, i.e. what we are able to do; but also what we would be able to do, if we had different capacities. Summing up, we ended up with the following alignments (Figure 6).

Our second task was to prove that the quantum eraser experiment was causal, from Menzies and Price’s perspective. For that matter, I appealed to Woodward’s criteria; even if Woodward does not agree with Menzies and Price, his criteria are illuminating if we want to find out whether an event is a manipulation (or an intervention, as Woodward puts it) or not. In other words, Woodward’s conditions for a manipulation helped to complete Menzies and Price’s theory.

Despite the fact that the Quantum eraser experiment fulfils the criteria M1—M4, one point remained in discussion: the fact that we have to consider a measurement as a manipulation. But as I said, Woodward himself ended up considering a measurement in quantum mechanics as an intervention, for we know counterfactually what would happen if the measurement were not performed (thanks

³⁶ Menzies and Price say that the experience of success teaches us that to cause something is similar to situations where we bring about an event *A* in order to provoke an event *B*. cf. (Menzies & Price, 1993, pp. 193-195).

to experiments). The last step was to show that, considering the quantum eraser experiment as a case of backwards causality is coherent, as it is very similar to Dummett's reply to Black's bilking argument.

Hence, I have proven that we can separate the alignment of the causal arrow with the planification arrow, from the alignment of the causal arrow with the time's arrow. That is, I separated agency from the experience of success, even if they are deeply related. That is, the experience of success is at the origin of the concept of causality (as I said, causality is a side effect of the experience of planification), but even if it must be rejected, that does not mean that we should throw the baby out with the bath water. Menzies and Price's agency theory makes sense of the manifest image, and is compatible with both commonsensical cases and with science, both in their definition of the nature of causality *and* the direction of causality. From an agentist point of view, it is possible to make sense of causality as being symmetric in respect to time.

It should be noted that I assume observing backwards causality in the microworld is sufficient to revise the whole concept of causality (or at least, the common sense of its direction), even at the macrolevel. The reason is, the point of physics is precisely to bring the macroworld and the microworld together, in order to have general laws —or general principles. The project of physics, and hence of quantum physics, is reductionist. Therefore, it makes sense not to differentiate the macroworld from the microworld, as regards causality. Therefore, I privilege a position where science is confronted to common sense, despite the differences of scales. It is also the reason why I rejected the position assuming that there are two types of causalities. As I interpret Sellars, even if the manifest and scientific images are two different representations, they are related to the same things —hence, the manifest image and the scientific image of causality are two ways to relate to the same concept.

Summing up, I have proven that despite the incompatibility between common sense and the quantum eraser experiment, there is a way to keep the manifest image of causality as making sense, in respect to science. Common sense has to be rejected, for the alignment of the causal arrow with the time's arrow cannot hold. But the manifest image of causality can be safe, and remains an adequate conception of the world. The commonsensical notion of causality can be thence compatible with science.

Bibliography

- Braillard, P.-A., Guay, A., Imbert, C., & Pradeu, T. (2011). Une objectivité kaléidoscopique : construire l'image scientifique du monde. *Philosophie*(110), 46-71.
- Callender, C. (2008). The Common Now. *Philosophical Issues*, 18(1), 339-361.
- Dainton, B. (2010). *Time and Space* (2nd edition ed.). Durham (UK): Taylor & Francis Ltd.
- Dowe, P. (1992a). Process Causality and Asymmetry. *Erkenntnis* (1975-), 37(2), 179-192.
- Dowe, P. (1992b). Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory. *Philosophy of Science*, 59(2), 195-216.
- Dowe, P. (1995). Causality and Conserved Quantities: A Reply to Salmon. *Philosophy of Science*, 62(2), 321-333.
- Dummett, M. (1964). Bringing about the past. *Philosophical Review*, 73(3), 338-359.
- Einstein, A. (1905). Zur Elektrodynamik bewegter Körper. *Annalen der Physik*, 17, 891-921.
- Feynman, R. (2010). *The Feynman Lectures on Physics. Volume III*. New York: Perseus Book Group.
- Gaasbeek, B. (2010). *Demystifying the Delayed Choice Experiments*. Retrieved from arXiv:1007.3977
- Galavotti, M. C. (2018). Wesley Salmon. *The Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/archives/fall2018/entries/wesley-salmon/>
- Heisenberg, W. (2007). *Physics and Philosophy: The Revolution in Modern Science*. New York: Penguin Books.
- Hume, D. (1999). *An Enquiry Concerning Human Understanding* (T. L. Beauchamp Ed.): Oxford University Press.
- Ionicioiu, R., & Terno, D. R. (2011). Proposal for a Quantum Delayed-Choice Experiment. *Physical Review Letters*, 107(230406), 1-5.
- Kaijo, C. (2008). Primary and Secondary Quality. *California Undergraduate Philosophy Review*, 1, 49-54.
- Kastner, R. E. (2017). *The Transactional Interpretation of Quantum Mechanics*. Cambridge: Cambridge University Press.
- Kim, Y.-H., R. Yu, Kulik, S. P., & Shih, Y. H. (2000). A Delayed Choice Quantum Eraser. *Physical Review Letters*, 84, 1-5.
- Kim, Y.-H., & Shih, Y. (1999). Experimental realization of Popper's Experiment: Violation of the Uncertainty principle? *Conference Proceedings*, 1-6.
- Lewis, D. (1973). *Counterfactuals*. Malden (USA): Blackwell Publishers.
- Lewis, D. (1989). Dispositional Theories of Value. *Proceedings of the Aristotelian Society*, 63, 113-137.

- Ma, X.-S., Kofler, J., Qarry, A., Tetik, N., Scheidl, T., Ursin, R., & Ramelow, S. (2013). Quantum erasure with causally disconnected choice. *PNAS*, 110(4), 1221-1226.
- Ma, X.-S., Kofler, J., & Zeilinger, A. (2016). Delayed-choice gedanken experiments and their realizations. *Reviews of Modern Physics*, 88, 1-28.
- Menzies, P., & Price, H. (1993). Causation as a Secondary Quality. *British Journal of Philosophy of Science*, 44, 187-203.
- Mill, J. S. (1963-1991). *The Collected Works of John Stuart Mill* (University of Toronto Press / Routledge and Kegan Paul ed.). Toronto / London: John M. Robson.
- Peruzzo, A., Shadbolt, P., Brunner, N., Poposecu, S., & O'Brien, J. L. (2012). A quantum delayed choice experiment. *Science*, 338(6107), 634-637.
- Price, H. (1996). *Time's Arrow and Archimedes' Point. New Directions for the Physics of Time*. New York: Oxford University Press.
- Price, H. (2014). Causation, Intervention and Agency – Woodward on Menzies and Price. In H. Beebe, C. Hitchcock, & H. Price (Eds.), *Making a Difference* (pp. 1-29). Oxford (UK): Oxford University Press.
- Ramsey, F. P. (1978). General Propositions and Causality. In D. H. Mellor (Ed.), *Foundations: Essays in Philosophy, Logic, Mathematics and Economics* (pp. 133-151). London: Routledge & Kegan Paul.
- Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.
- Russell, B. (1948). *Human Knowledge: its scope and limits*. New York: Simon and Schuster.
- Russell, B. (1963). *Mysticism and Logic*. London: Allen & Unwin.
- Salmon, W. (1979). Alternative Models of Scientific Explanation. *American Anthropologist*, 81(1), 61-74.
- Salmon, W. (1980). Probabilistic Causality. *Pacific Philosophical Quarterly*, 61(1-2), 50-74.
- Salmon, W. (1984). *Scientific Explanaiton and The Causal Structure of the World*. Princeton: Princeton University Press.
- Salmon, W. (1994). Causality without counterfactuals. *Philosophy of Science*, 61, 297-312.
- Salmon, W. (1997). Causality and Explanation: A Reply to Two Critiques. *Philosophy of Science*, 64(3), 461-477.
- Salmon, W. (1998). *Causality and Explanation*. New York: Oxford University Press.
- Salmon, W. (2002). A Realistic Account of Causation. In M. Marsonet (Ed.), *The Problem of Realism* (pp. 106-134). Aldershot and Burlington: Ashgate Publishing.
- Salmon, W. (2010). The Causal Structure of the World. *Metatheoria*, 1(1), 1-13.
- Scarcelli, G., Zhou, Y., & Shih, Y. (2007). Random Delayed-Choice Quantum Etraser via Two-Photon Imaging. *European Physical Journal D*, 1-10.

- Scully, M. O., & Drühl, K. (1982). Quantum eraser: A proposed photon correlation experiment concerning observation and "delayed choice" in quantum mechanics *Physical Review A*, 25(4), 2208-2213.
- Sellars, W. (1963). Philosophy and the Scientific Image of Man. In *Empiricism and the Philosophy of Mind* (pp. 1-40). London: Routledge & Kegan Paul Ltd.
- Walborn, S. P., Cunha, M. O. T., Padua, S., & Monken, C. H. (2002). A double-slit quantum eraser. *Physical Review A*, 65(3), 1-6.
- Watkins, E. (2012). Kant, Sellars, and the Myth of the Given. *The Philosophical Forum*, 43(12), 311-326.
- Werbos, P., & Dolmatova, L. (2000). The Backwards-time interpretation of quantum mechanics: Revisited with experiment. *Arxiv.org*, 1-29. Retrieved from <https://arxiv.org/abs/quant-ph/0008036>
- Wheeler, J. (1978). The 'Past' and the 'Delayed Choice' Double-Slit Experiment. In A. Marlow (Ed.), *Mathematical Foundations of Quantum Theory* (pp. 9-48). New York: Academic Press.
- Wheeler, J. (1983). Delayed Choice. In J. Wheeler & W. H. Zurek (Eds.), *Quantum Theory and Measurement*. Princeton: Princeton University Press.
- Woodward, J. (2000). Explanation and Invariance in the Special Sciences. *British Journal of Philosophy of Science*, 51, 197-254.
- Woodward, J. (2004). *Making Things Happen: A Theory of Causal Explanation*. Oxford and New York: Oxford University Press.
- Woodward, J. (2016). Causation and Manipulability. *The Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/archives/win2016/entries/causation-mani/>

Table of Contents

Introduction	7
Chapter 1 Manifest and Scientific Images or the Clash between Common Sense and Science	9
1. To Begin with	9
2. “Philosophy and the Scientific Image of Man”	11
How did Sellars Come to This Distinction?	11
A Synoptic View	12
Two Little Objections	13
The term “image”	13
3. The Manifest Image	14
The “original image”	14
The Two Paths of Refinement	15
Experiences as Details of a Landscape	16
4. The Scientific Image	17
Priority of the Manifest Image	17
<i>The Scientific Image?</i>	18
Ontological Neutrality, and Rivalry	19
5. An objection against the Manifest Image	21
Chapter 2 The nature of causality	23
1. Salmon’s Theory of Propagation	23
Propagation Theory	23
Facing Spurious Causes	24
Conjunctive Forks	24
Causal Processes	25
Conserved Quantities	27
Conclusion	27
2. Menzies and Price’s Agency Theory	28
Making a Difference	28

Confusion Between Epistemology and Metaphysics.....	29
Unmanipulable Causes.....	30
Circularity and the Experience of Success.....	31
Anthropocentricity	33
Conclusion	35
Chapter 3 Manifest and Scientific Images of Causality: symmetry and asymmetry	36
1. Agency Theory and Manifest Image.....	37
A First-Person Experience	37
An Early-Age Experience	37
2. Manifest Image of Causality and the Causal Arrow	38
Common Sense and Causal Arrow	38
Principle of Innocence	39
3. Scientific Image of Causality	39
The World Obeys to T-Symmetry	40
Principle of Innocence in the Macroworld.....	40
Extension of Principle of Innocence in the Microworld	41
Scientific Image of Causality and Eliminativism	42
Physicists Do Not Do What They Preach	43
Conclusion	44
Chapter 4 The Quantum Eraser Experiment: Against the Manifest Image of Causality?.....	46
1. Quantum Physics and Double-Slit Experiment.....	46
Thomas Young and the Double-Slit Experiment.....	46
Double-Slit Experiment and Detectors	48
Complementarity Principle	49
2. Quantum Eraser Experiment	50
The Two Purposes of the Quantum Eraser Experiment.....	50
Kim et al.'s Experiment	51
Similarity With the Double-Slit Experiment	52

The Peculiarity of the Eraser	53
Why Is It Called an Eraser?	53
3. Results of the Experiment.....	54
Backwards in Time?	54
Eliminativism and Common Sense.....	55
4. Nonlocality and eliminativism.....	56
What Is Nonlocality?	56
Is Eliminativism the Solution?.....	57
Conclusion	57
Chapter 5 Looking for the origin of the manifest image of causality	59
1. Success as Planification	59
Experience of Success	59
What Does It Mean to “plan”?.....	60
Planification Arrow and Time’s Arrow	60
Recapitulation	62
2. Scientific Experiments, Causality’s and Time’s Arrow	63
3. Causal Arrow, Planification Arrow and Time’s Arrow	64
Chapter 6 The Quantum Eraser Experiment Is a Manipulation, Not a Planification.....	66
1. Quantum Eraser Experiment Is a Manipulation	67
James Woodward and Interventions	67
Woodward’s Conditions	68
Is the Quantum Eraser Experiment a Manipulation?.....	69
Conclusion	72
2. Quantum Eraser Experiment and Bilking Argument.....	73
Bilking Argument	73
Dummett’s Tribe Example	75
Analogy Between Tribe Example and Quantum Eraser Experiment	78
Conclusion	80

Conclusion	81
Bibliography	85

Acknowledgements

I would like to thank Eléonore le Hardÿ de Beaulieu, Alessandro Cierro, Alexandra Syskova, Florian Marion and Gwenaël Laurent. Your daily support and presence, as well as our discussions and your advices provided me confidence, courage and support. That means a lot to me.

Special thanks to my godparents, Crispin Sompo and Jean Knightson, as well as Liz Ross. Not only your revisions, but also your daily care were amazing and really helped.

Finally, I would like to thank Alexandre Guay, who guided me through the elaboration of the thesis. This is the fruit of three years of discussions and advices which contributed to help me grow in philosophy.

