UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FILOSOFÍA

Departamento de Lógica y Filosofía de la Ciencia

TESIS DOCTORAL

Explanatory depth and statistical mechanical interventionism

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Fernanda Samaniego Bañuelos

Director

Mauricio Suárez Aller

Madrid
Ed. electrónica 2019

Explanatory Depth and Statistical Mechanical Interventionism

Phd thesis
Fernanda Samaniego Bañuelos
Logic and Philosophy of Science Department
Complutense University

Thesis supervisor: Mauricio Suárez Aller

To Lucio, my parents, my sister, and my friends.

Acknowledgments

I would like to thank José Marquina for awaking my interest in philosophy of physics and to Ana Rosa Pérez Ransanz for being my friendliest guide throughout my career. I am also deeply grateful to Mauricio Suárez for directing this research with dedication and patience. The conversations we had in several occasions about the content of this thesis had been very fruitful and elucidating. I thank the members and ex-members of our research group "Methods of Causal Inference and Representation in Science": Iñaki, Pedro, Albert, Isabel and Carlos, and to Maria for coming to the reading group about Woodward's book. The advice I received from Jos Uffink, Carl Hoefer, Henrik Zinkernagel and Olimpia Lobardi during the EPSA'09 Conference in Amsterdam and at the Workshop in Utrecht pointed at crucial considerations that helped me to define the problem in which my research is focused. I particularly thank Henrik for the conversations we had about the past hypothesis, and Carl Hoefer for driving my attention to the approach developed by Meir Hemmo and Orly Shenker. In November of 2010 I had the opportunity to meet Orly Shenker and Meir Hemmo personally at the University of Haifa and the University of Jerusalem. Thank you both for such a warm welcome and for the constructive discussions. All your comments were of great importance for this work. Also thanks to Mauricio for his support on making this encounter possible. I am very grateful to Miklos Rédei, Roman Frigg and Luis de la Peña Auerbach for teaching me foundations of statistical mechanics and philosophy of quantum mechanics and to Federica Russo and Isabelle Drouet for our profitable conversations during the Workshops on Causality in Madrid and Paris in March and June 2011. The feedback on the last draft of this thesis, provided by Federica Russo, Miklos Rédei and Iñaki San Pedro, was particularly helpful. I greatly appreciate the wisdom and support provided by Rocío Carretero and Carla Boyer during the writing process. I would finally like to thank to the members of our new Seminar for the Philosophy of Science in Madrid: María, Javier, Alex and Manolo and to my colleges Emilio, Guillermo, Bladimir, Pompeya, Isabel Gamero and David Teira. I owe more than words can tell to Nalliely, Laura, Blanca, Nelson, José Ricardo, Ekai, Federico, Esther, Aarón, Ángel, Adriana, Enrique, Rocío, Patricia, Arturo, Ahmed, Salomón, Maximiliano, Abraham, Mariana, Luis, Manuel, José Luis, Verónica, Natalia, Valentín, Martha, Manola and Lucio. Thank you all. This research has been carried out thanks to the scholarship number 190963 (2008-2011) of the Mexican Council of Science and Technology (Consejo Nacional de Ciencia y Tecnología CONACyT).

Fernanda Samaniego, Madrid, July 2011.

Contents

Introduction

The incompatibility between the asymmetrical second law of thermodynamics and the symmetry of the underlying statistical mechanical dynamics is one of the long-standing problems in the foundations of statistical mechanics and it is usually referred to as "the irreversibility problem". A common solution to this problem posits that thermodynamic evolutions are the result of an initially low-entropy state of the system. The so called *interventionists* (Blatt, 1959, [9]; Ridderbos & Redhead, 1998, [82]; Hemmo & Shenker, 2005, [55]), by contrast, believe that this low-entropy initial condition is not sufficient to account for the thermodynamic behavior. They propose an alternative (or an additional) solution to the irreversibility problem. In their view, the thermodynamic approach to equilibrium is ultimately produced by the environmental perturbations acting upon the system.

Soon after the interventionist approach was first proposed (in the 1950's) a specific kind of experiments, known as the spin-echo experiments, became a case of great interest in the debate about interventionism for two reasons. Firstly, in the spin-echo experiments it appears that the relevant system is completely isolated, and thus it approaches equilibrium without any influence of the environment. If that is the case, interventionism would not be able to account for the behaviour of the system in this experiment. Secondly, the entropy of the system appears to increase and decrease several times over the course of the experiment. In other words, the evolution of the system apparently violates the second law of thermodynamics. For both reasons the spin-echo experiments have been fertile ground for debates regarding the irreversibility problem, and constitute a particularly challenging case for interventionism.

The aim of this thesis is to analyze and compare several interventionist explanations of the spin-echo experiments. Both classical and quantum-based interventionist explanations are assessed by means of James Woodward's manipulability theory of causal explanation (Woodward, 2003, [106]). The application of the manipulability theory is not only relevant for understanding in detail the irreversible process that takes place during the spin-echo experiments, but also for clarifying the difficulties that a satisfactory version of the interventionist approach must overcome. The analysis additionally reveals some interesting features of the manipulability theory.

Based on the results of the analysis developed in this thesis, I will argue that the correlation between the causes postulated by interventionists, renders their explanations 'shallow' in accordance with the criteria of the manipulability theory. I will argue that this lack of 'explanatory depth', rather than reveling a disadvantage of interventionism, points to a weakness of Woodward's manipulability theory of causal explanation.

The thesis is organized as follows. In chapter 1 the irreversibility problem is explained and distinguished from other problems that are commonly related to it. In chapter 2 statistical mechanical interventionism is introduced and the spin-echo experiments are described in detail. Three philosophically relevant questions about the experiments are then answered from the interventionist perspective. The main objections to such answers are discussed in chapter 3. This discussion motivates the application of the manipulability theory to various explanations of the spin-echo experiments. The manipulability theory itself is presented in chapter 4. Chapters 5 and 6 develop the full analysis of various explanations of the spin-echo experiments in terms of the manipulability theory. This is the main original contribution of the thesis since the manipulability theory has never been applied to statistical mechanical interventionism before. The results of this analysis are discussed at the end of chapter 6. Finally, chapter 7 contains the overall conclusions of the thesis.

Chapter 1

The Irreversibility Problem

In this chapter we introduce the irreversibility problem. Some central notions in foundations of statistical mechanics must be previously defined. As we will see, the second law of thermodynamics and the notion of entropy can be formulated in several different ways. Also the notions of 'time asymmetry', 'time reversal invariance', 'irreversibility' and 'the arrow of time' have diverse uses in the literature and the relations that are said to hold among them depend on the philosophical view we adopt. We must define each of these notions carefully.

The first part of this chapter is devoted to such definitions. In the second part of the chapter the irreversibility problem is defined and it is distinguished form the problem of justifying the second law of thermodynamics more generally. Section 1.5 particularly stresses the distinction between 'the problem of the arrow of time' and 'the irreversibility problem'.

1.1 Statistical Mechanics

Statistical Mechanics (SM) is a theory that was born from modifications of the kinetic theory of gases, and constitutes the physical framework in which the irreversibility problem emerged. In the kinetic theory (fist formulated by Daniel Bernoulli in 1738, then by Nicolas Carnot (early 19th century) and afterwards (mid 19 century) developed by Rudolf Clausius, William Thomson Kelvin, James Clerk Maxwell and Ludwig Boltzmann) it is assumed that gases constituted by molecules and the properties of the system are determined by the mechanical properties of such molecules. For example, in order to compute the pressure exerted upon the walls of a container full of gas, the kinetic model assumes that all the molecules have the same speed, one-third of them moving parallel to each one of the three edges of the container. The pressure is then computed by attaching probabilities to the states of motion of the molecules (see Ehrenfests, 1912, [37]:4). The statistical mechanical twist arrives when probabilities, previously predicated of the state of motion of a molecule, become a property of the state of the *entire gas system*. The principal consequence of this change is that the mechanical and statistical concepts, previously entangled in the kinetic theory, separate for the very first time in SM (see Uffink, 2006, [95]:932-933).

SM evolved in such a way that it eventually split in two essentially different traditions. The tradition represented by Ludwig Boltzmann, on the one hand, and the tradition represented by Josiah Willard Gibbs on the other. Although it is possible nowadays to identify within the scientific community defenders of each tradition, the most common scientific attitude is using one or the other depending on the feature that results more convenient for the problem that one is willing to solve. For our purposes it is important to clarify the differences between the Boltzmannian and the Gibbsian approaches to SM, mainly because each of them defines the concept of entropy – closely related to the irreversibility problem– in a different way. I turn now to explain (in sections 1.2 and 1.3) the main differences between these two traditions.

1.2 Boltzmann and Gibbs: two approaches to Statistical Mechanics

A 'dynamical system' is a mathematical representation of any evolution function describing the trajectory of one point –or a set of points– across time in an abstract space known as state space or phase space. How the phase space is defined and what the physical meaning of the evolution function is depends on the particular application.

In SM the systems are constituted by a very large number of elements ($\approx 10^{23}$), with a finite number of degrees of freedom. The phase space Γ for a system of n particles is normally defined as a $2n$-dimensional space in which a point $x \in \Gamma$ represents the state of the system at any time. So x specifies the total position-momentum and is given by $x = (q_i, p_i) = (q_1, q_2, q_3, \ldots q_n, p_1, p_2, p_3, \ldots p_n)$ where q_i are the generalized space coordinates and p_i the conjugated momentum. The multidimensional volume associated to any set A in Γ is then measured in units of position $\otimes$ momentum. Lebesgue and Liouville are common measures and they are usually normalized ($0 \leq \mu \leq 1$). For most applications the system's evolution is described by the Hamiltonian equations. The evolution function ϕ_t is called "flow" when $t \in \mathbb{R}$, and that is the case in SM as t=time. The flow is measure preserving on Γ, meaning that for all t and for every measurable set $A \subseteq \Gamma$, the volume of A equals the volume of $\phi_t(A)$. In short, in what follows we will deal with dynamical systems defined as a tuple (Γ, ϕ_t, μ) where:

(a) The phase space Γ.

(b) The Hamiltonian flow $\phi_t; \Gamma \longrightarrow \Gamma, t \in \mathbb{R}$ such that for every subset $A \subseteq \Gamma$

$\phi_0(A) = 0$
$\phi_{t+s}(A) = \phi_s(\phi_t(A))$
$\phi_{-t}(A) = {\phi_t}^{-1}$

(c) The measure μ normalized and invariant under the flow,
i.e. for any measurable set $A \subseteq \Gamma$, $\mu(\phi_t(A)) = \mu(A)$

It is worth noting that the system may be constrained into some hypersurface of Γ. If the system is energetically isolated, for example, the system will be confined to configurations of position and momentum with the same energy as the initial condition. The states of the system will lie within the constant energy hypersurface $\Gamma_E \subseteq \Gamma$. The so called microcanonical measure $\mu(x)$ is the uniform measure over the energy hypersurface.

Within the Boltzmannian approach, the instantaneous microstate of a statistical mechanical system is represented by a single point in the phase space Γ. The value of a thermodynamical observable (i.e. a property, e.g. pressure), is related to that state point in the phase space by a phase function $f(x); f : \Gamma \longrightarrow \mathbb{R}$ (where $x \subseteq \Gamma$). That is to say, phase functions $f(x)$ associate phase space points to values of a given physical quantity. Given a Lebesgue-integrable phase function $f(x)$ defined on Γ_E, the phase average of $f(x)$ is defined as:

$$< f(x) >:= \int_{\Gamma_E} f(x) d\mu(x)$$

where $$d\mu(x) = \frac{1}{\omega(E)} \delta(H(x) - E) dx$$

δ is the Dirac delta function, $H(x)$ is the Hamiltonian, E the energy parameter and $\omega(E)$ is the area of the energy hypersurface given by:

$$\omega(E) = \int_{\Gamma_E} \delta(H(x) - E) dx$$

In the Gibbsian approach, by contrast, the object of study is not a single system but an ensemble of several identically prepared systems. The state of each of the systems belonging to the ensemble ε is represented by one point in Γ. Thus, the situation of the ensemble as a whole is represented by a set of points in the phase space. When the number of systems is very high we have a positive density distribution $\rho(x)$ of representative points over the phase space. This density distribution ρ provides information about the probability that a representative point lies in some region of the phase space.

The integral of $\rho(q_i, p_i)$ over the phase space region delimited by $[q_a, q_b] \otimes [p_c, p_d], (a, b, c, d \leq n)$ with respect to the measure ν gives the amount of states that satisfy $q_i \in [q_a, q_b]$ and $p_i \in [p_c, p_d]$. If the system is confined to the constant energy hypersurface Γ_E the integral is computed over Γ_E and the ensemble is called *microcanonical ensemble*.

The average of any phase function over the ensemble ε is then computed by:

$$< f(x) >= \int_{\varepsilon} f(x)\rho(x)d\nu$$

where ν is the measure with respect to the density distribution ρ. The Gibbsian approach postulates that this average of the phase function $< f(x) >$ serves to compute the value of the thermodynamic observable (e.g. the pressure) associated to the phase function (see Frigg, 2007, [41]:138). This is the so called *Gibbsian average method*, and it has shown to be an extremely successful method for predicting values that coincide with the values obtained by means of experimental measurements.

1.3 The Second Law of Thermodynamics

1.3.1 The second law and phase space regions

Consider a container full of gas. Let us define the macrostate M as the state in which all the gas molecules are concentrated in the left hand side of the container. There is a region in the phase space Γ that corresponds to this macrostate M. Such a region contains all (and only) the points in Γ such that they are macroscopically manifested by M. Each point inside that region is however unique in the sense that it represents a particular microstate in which the positions and momenta of the gas molecules are different. So, roughly speaking, regardless the specific position and velocity of the molecules, if the state of the system lies in the region associated to M, we will always observe the same macroscopic state in the laboratory, namely, the gas confined in the left hand side of the container.

Suppose that we divide the phase space Γ in several regions of this kind. Under such a division of the phase space, the region corresponding to the equilibrium (regularly corresponding to the macrostate in which the gas is spread out all over the container) will occupy the greatest volume of the phase space (see Fig.1), while the region corresponding to rare macrostates (for instance, all gas molecules perfectly aligned forming a cube in the middle of the container), will have, by contrast, a very reduced volume.[1] This is

[1]It is worth noting that entropy is not necessary opposite to order. For more details see Landsberg (1984, [66]).

due to the fact that there are not that many possible micro-configurations associated to that rare macrostate.

The intuition behind the second law of thermodynamics (illustrated in Fig.1) is that an isolated system with an initial state lying in a small region of the phase space, will evolve naturally in such a way that the succession of states will cross through regions of increasingly greater volume until it finally reaches the equilibrium region with the greatest volume.

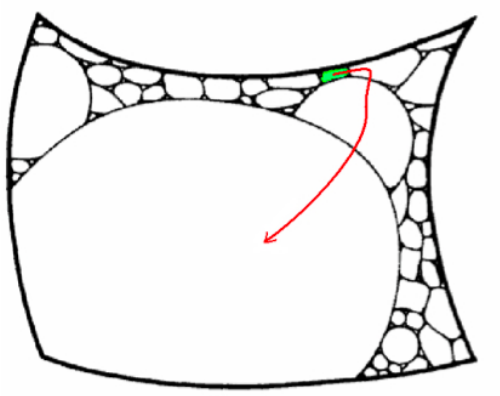


Fig.1. Evolution towards equilibrium.[2]

If we define a quantity proportional to the volume of the different regions of the phase space Γ we can then claim that, according to the second law of thermodynamics, such a quantity increases during the natural evolution of the system. Entropy is one such quantity. Given the huge differences between the volumes of the regions in Γ for systems with a large number of particles, the entropy is conveniently associated to the logarithm of the volume rather than to the volume of the region. Entropy is then defined as $S = k_B\ ln\Omega$, where k_B is the Boltzmann constant, and Ω is the number of microstates that correspond to the same macrostate (see Penrose, 1989, [78]: 400-406). Intuitively, the second law of thermodynamics states that isolated systems evolve in such a way that the entropy S either increases or remains equal, but it never decreases.

[2]Figure taken from Penrose, 1989, [78]:402, and then modified.

1.3.2 Three considerations about the second law

The formulation of the second law presented in the previous section leads to three interesting considerations (i, ii and iii). The first consideration (i) concerns how great the equilibrium region is compared with the rest of the regions in the phase space. Imagine, for instance, a rather simple system such that the number of different possible micro configurations compatible with a given macrostate is not very small compared with the number of micro states compatible with the equilibrium. In that case the union of all of the non-equilibrium regions may be larger in volume than the equilibrium region alone. How can we be sure then that the system will evolve into the equilibrium region?

A different issue (ii) concerns the domain of the law. What kind of systems behave in accordance with the law? Is behavior in accordance with the second law manifest only in isolated systems and not in open systems? Are we talking about an empirical and merely descriptive principle or, are we dealing with a normative statement that is supposed to hold universally?

And a third and delicate question (iii) concerns the meaning of 'entropy' and whether this meaning changes if we translate the situation to the Gibbsian framework. Within the Boltzmannian framework the initial state of the system is represented by a single point in the phase space. Within the Gibbsian framework, by contrast, the initial states of the systems in the ensemble are represented by a initial distribution $\rho(x)$. One may ask then whether "entropy' is a property of an individual system or a property of the ensemble. How do we define entropy within this framework, and what does its value depend on?

A common response to consideration (i) is to claim that statistical mechanics studies systems of a high number of degrees of freedom. If the system has a high number of degrees of freedom the volume of the equilibrium region is overwhelmingly greater than the volume of other regions together and the problem (i) is resolved.

To answer questions (ii) and (iii) we may consider different formulations of the second law of thermodynamics.[3] Probably the most empirically grounded

[3]Lieb-Yngvason, Carathéodory, Fourirer, Prigogine, Reichenbach and some other authors have been excluded because it is not my intention to provide an exhaustive review of every formulation of the second law of thermodynamics. For a scrupulous historical and philosophical analysis see Jos Uffink (2001, [94]). Formal and historical details of classical

formulations can be traced back to the very first works of Carnot, Kelvin and Clausius. Those formulations might be seen as the "the seeds of the law", since the concept of entropy was not yet explicitly involved. Clausius' general idea[4] was that in any machine unaided by external sources of energy, no transfer of heat from a body to another at a higher temperature is possible. Kelvin's principle, in turn, was that "a transformation whose sole final thermodynamic result is to transform into mechanical energy heat extracted from a source which is at the same temperature throughout is impossible" (see Adkins, 1987, [1]:104). These two principles were shown to be equivalent and are now known as prohibitions of the *perpetuum mobile* of the second kind[5]. Isolated systems performing cyclic processes (i.e. processes where the final state of the system is equal to the initial state) comprise the domain of Clausius' and Kelvin's principles. Since the kinetic theory had a very practical dimension related with heat engines it is not surprising that, in these early principles, no intention of describing the behaviour of the entire universe is explicitly manifest.

Later formulations by Clausius himself differ on just this point. In 1862 he extends the principle to no-cyclic processes and soon after he formulates the law as applying to the universe as a whole:

> There is "a generally prevailing tendency in Nature towards changes in a definite sense. If one applies this to the universe in total, one reaches a remarkable conclusion.[...]Namely, if, in the universe, heat always shows the endeavour to change its distribution in such a way that existing temperature differences are thereby smoothened, then the universe must continually get closer and closer to the state, where the forces cannot produce any new motions, and no further temperature differences exist." (Claussius,1864, quoted in Uffink [94]:36)

In 1879 Clausius proposes 'The Entropy Principle' according to which 'The entropy of the universe tends to a maximum'. Here the concept of *entropy* appears for the very first time in a formulation of the second law. In

thermodynamics can be found in James Serrin's (1979, [87]).

[4]For the original formulations see Jesudason (2003, [60]), Jaynes (1984, [59]) or Albert (2000, [4]:28-30).

[5]The prohibition of *perpetuum mobile* of the first kind corresponds to violations of the conservation of energy. For more details see Fink (2009, [39]:306).

the same year Max Planck proposes a formulation of the second law also in terms of entropy: "For all processes in the universe, the total entropy of the systems involved never decreases, and therefore all processes are irreversible" (Plank quoted in Uffink [94]:92).

The famous 'Minimum Theorem', published in 1872 by Boltzmann and now known as the **H-theorem**, was another attempt to solve the mysteries behind previous formulations of the second law and the phenomenon of irreversibility. Boltzmann analyzed the case of a dilute gas, constituted by spherical molecules inside a container with elastic walls. The theorem proves that, if such a system is energetically isolated over infinite time, it will spend the overwhelming part of the time in microstates near the equilibrium. And the equilibrium microstate is to be uniquely described by the Maxwell distribution (for more details see Sklar, 1993, [91]:177 and Brown, Myrvold and Uffink, 2009, [14]:175).

Boltzmann's H-theorem was criticized for relying on an unjustified probabilistic assumption (*Stosszahlansatz*) about the number of collisions between molecules as well as for the incompatibility between the theorem and the underlying dynamics[6]. In response to his critics, Boltzmann reformulated his conception of the second law of thermodynamics and proposed that "corresponding to any description of the gas in terms of its 'macroscopic' observable properties, there are many possible 'microstates' –many possible configurations of molecules which all give the same macrostate." (Boltzmann quoted in Price (2004, [80]:222-225)). Under this renewed Boltzmannian conception, also described at the beginning of this section, entropy is defined as $S = k_B ln\mu(M)$ and the second law may be formulated as follows:

Second Law of Thermodynamics (2LTD): The entropy of an isolated system tends to increase over time, approaching a maximum value at the equilibrium state.

This statement has become in many contexts the standard formulation of the second law of thermodynamics.

Sometimes this law is also articulated in terms of probabilities. In order to do so probabilities are associated to the measure μ operating over the phase space Γ. Intuitively this means that the greater the volume of a region in Γ is, the higher is the probability that a state point x lies in that region.

[6]This deserves to be treated in detail so the following section will focus on it.

This idea has been formulated more precisely in the following principle:

Proportionality principle: If M is the macro-state of a system at time t, the probability at that time that the micro state of the system lies in a subset A of Γ_M is $\mu(A)/\mu(\Gamma_M)$ where μ is the standard Lebesgue measure.

If this principle holds, the microstates of a given region in Γ that correspond to the same macrostate M (see Fig.1 in page 8) can also be defined as *equally likely micro states*. Using this terms yet another alternative formulation of the second law of thermodynamics is possible:

Second Law assuming the Proportionality Principle (2L-PP): Isolated systems evolve form *unlikely* to *likely* states.

We may now provide a conclusion regarding consideration (ii) concerning the domain of the law. The above presented discussion of different versions of the second law of thermodynamics shows how the domain of that law has changed across its history. It was firstly conformed by cyclic processes and was later extended to non-cyclic processes; in 'The Entropy Principle' the domain is the entire universe; and in Planck's formulation the domain includes all kinds of physical processes. Finally in the currently most widely accepted version of the law (2LTD) the domain of the law are energetically isolated systems. And –recalling the solution to issue (i)– it is also commonly assumed that those isolated systems have a high number of degrees of freedom.

Let us turn now to our third and last issue (iii) concerning whether or not the meaning of the entropy changes in the Gibbsian approach. Is it possible to describe the evolution of an *ensemble* of systems in non-equilibrium states toward an *ensemble* of systems in states of equilibrium? To put it slightly different, can the law be stated as 2LTD –or as an equivalent statement– also in the Gibbsian approach? To answer these questions let us inquire into how the Gibbsian entropy is defined and whether it increases in accordance with the second law.

The fine-grained Gibbsian entropy is denoted by S_G and is defined as:

$$S_G(\rho) := -k_B \int_\Gamma \rho \ \ log(\rho) d\Gamma$$

Where k_B is the Boltzmann constant and $\rho(q_i, p_i, t)$ is the (positive) fine-grained density distribution of the states of the ensemble.

Our question is whether this fine-grained entropy increases when a system approaches equilibrium, i.e., whether the non-equilibrium distribution of the system ρ approaches the microcanonical distribution –which characterizes equilibrium systems in the Gibbsian formalism. And the answer is that such a change in the distribution is impossible. The systems conforming the Gibbsian ensemble evolve in accordance with **Liouville's theorem** which states that the derivative of the density function equals zero:

$$\frac{d\rho}{dt} = 0$$

By definition, a distribution ρ is called 'stationary' iff Lioville's theorem holds (and thus measure-preservation holds) for all times t. A direct consequence of the theorem is that the measure μ is invariant under the Hamiltonian flow ϕ_t.

Intuitively Liouville's theorem means that, just alike an incompressible fluid, a density distribution in the phase space may change its shape but never its volume (see Fig.2 below). As a direct consequence, the fine-grained entropy, which is defined in terms of the density distribution, cannot increase. Non-stationary distributions always remain non-stationary and stationary distributions always remain stationary. This constitutes a serious problem because it is in conflict with the second law of thermodynamics.[7].

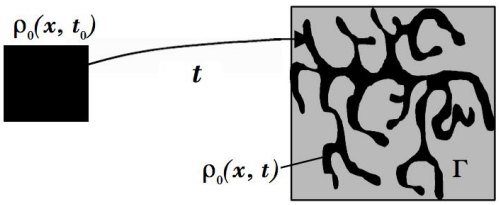


Fig.2. Ilustration of Liouville's Theorem.[8]

In the attempt to solve and explain the situation, Gibbs (1902) defined another kind of entropy called coarse-grained entropy. If the phase space is divided in disjointed cells ω_j $(j = 1, 2, 3, ...m)$ the so-called coarse-grained entropy is defined by:[9]

[7]For more details see Frigg (2007, [41]:138), Lombardi (2003, [68]: 17) or appendix A

[8]Taken from Lombardi, 2003, [68]:17.

[9]The definitions of entropy presented in this section follow Frigg (2007, [41]:151-153). Gibbs original book is (1902, [44]).

$$S_\omega(\rho) = S_G(\overline{\rho}_\omega) := -k_B \int_\Gamma \overline{\rho} \;\; log(\overline{\rho}) d\Gamma$$

where $\overline{\rho}_\omega$ is the uniform density in each cell, taking the average value of the density ρ in that cell i to:

$$\overline{\rho}_\omega(q_i, p_i, t) := \frac{1}{\delta\omega} \int_{\omega(q_i,p_i)} c(q_i', p_i', t) d\Gamma'$$

The following figures illustrate the conceptual difference between ρ and $\overline{\rho}_\omega$.

Both figures represent an equilibrium state: Fig.3 shows the picture of the fine-grained density distribution already spread along the phase space, while Fig.4 shows a different picture of exactly the same situation, but seen from the coarse-grained perspective. The difference between the two figures shows how coarse-grained takes the average value of the fine-grained distribution in each cell, disregarding the precise distribution inside it. Note that the density is evenly distributed across each cell -hence the tone of grey in Fig.4.

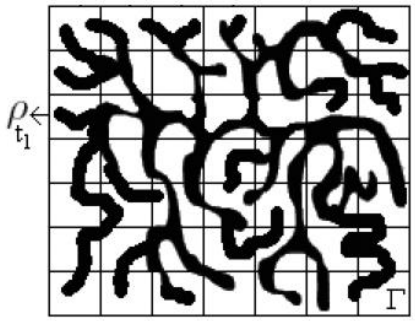


Fig.3. Fine-grained distribution

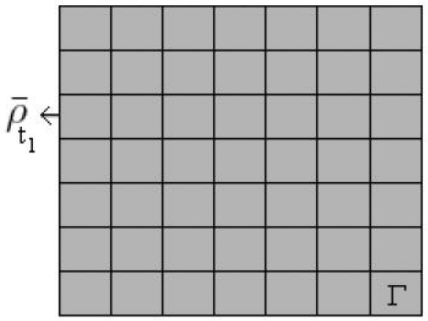


Fig.4.Coarse-grained distribution

If the evolution of the system is such that the initial fine-grained distribution $\rho(q_i, p_i, t_0)$ spreads over more and more cells ω_j as time increases, then, the coarse-grained entropy will also increase. Figures 2 and 3 represent

the system at a time $t_1(t_1 > t_0)$ close to the relaxation time, i.e., when the system has almost reached equilibrium.

The advantage of the coarse-grained entropy is that the uniform density $\overline{\rho}_\omega$) is not governed by Liouville's equations (Frigg, 2007, [41]:152) and so, it is not as restricted as the fine-grained entropy. In accordance with the second law of thermodynamics, the coarse-grained entropy increases while the system approaches equilibrium. This is also referred as the Gibbsian explanation of the irreversibility of the density distribution.

It is worth stressing, however, that full mixing –and subsequently full ergodicity– is a necessary condition for the desired increment of coarse-grained entropy.[10]

At least two other comments are in order regarding the Gibbsian description of the approach to equilibrium. Firstly, if the coarse grained entropy tends to its maximum value only as time tends to infinity then, it cannot be appropriate for describing real systems with finite relaxing times. (see Ehrenfests, 1912, [37]). Secondly, the consequence of defining both Gibbsian entropies in terms of the probability distribution ρ is that they become properties of an ensemble rather than an individual system. [11] Thus, "they are not properly speaking properties of individual systems which depend upon their changes in microstate at all. The Gibbsian entropies of a system are fixed by the constraints upon it, not by the actual state the gas takes on while so constrained." (Sklar, 1974, [90]:405-406).

This affects how the second law is to be conceived, because entropy is no longer considered as a property of an individual system. For instance, the statement 'the state of the system lies in the equilibrium region' not longer makes sense within the Gibbsian approach.[12] One could think that it is only a matter of changing from a Boltzmannian 'point state' to a Gibbsian density distribution ρ over the phase space, but the problem is not that simple. The main difficulty is how entropy is defined for non-equilibrium states – and this is precisely the essence of question (iii).

[10]Some arguments for mixing as necessary condition for the increment of coarse-grained entropy are based upon the geometrical interpretation of mixing and the convergence theorem. For more details see Frigg (2007, [41]:153) or Labarca (2005, [65]:60).

[11]This is the main difference between Boltzmann's and Gibb's approaches

[12]"This 'ensemble character' carries over to other physical quantities, most notably temperature, which are also properties of an ensemble and not of an individual system." (Frigg, 2007, [41]:170).

For example, in a box full of gas where a barrier is removed at t_1 (allowing the gas, previously trapped in the left half of the box, to expand over the whole volume of the box) the entropy will be defined for t_1 and also for t_2 when the gas is uniformly expanded all over the box. But whether the entropy of the intermediate non-equilibrium stages can be uniquely defined is not completely clear (see Uffink, 2001, [94]:94). In that sense, the following formulation of the second law seems more appropriate than 2LTD and 2L-PP:

Refined 2LTD (R-2LTD): If an isolated system evolves from a macrostate M_1 with entropy S_1 to the equilibrium macrostate M_2 with entropy S_2, then $S_2 \geq S_1$.

This refined version of the law is an improvement on previous formulations in at least two different respects. On the one hand, unlike 2L-PP, it makes no commitment to any relation between probabilities and entropy. This is an advantage because how probabilities are to be interpreted in SM is still a matter of debate. On the other hand the refined R-2LTD says nothing about the value of the entropy at intermediate stages during the evolution of the system. The idea underlying 2LTD is that the system evolves form a non-equilibrium state toward an equilibrium state, describing a sort of *continuum evolution*. The refined version R-2LTD, by contrast, only claims that the entropy of the final state is never less than the entropy of the initial state. And this move makes R-2LTD equally compatible both with the Boltzmannian and the Gibbsian definitions of entropy.

1.4 Loschmidt's and Zermelo's Objections

There are two particularly important objections that emerged as reactions to Bolztmann's formulation of the H-theorem. They are known as Loschmidt's Reversibility Objection and Zermelo's Recurrence Objection respectively. Both of them point to the incompatibility between the asymmetry built into the H-theorem and the symmetrical underlying dynamics of classical mechanics. In order to explain these objections let me first define the notion of 'time-reversal invariance'.

Definition of time-reversal invariance: Let $\mathbf{T}$ be the time-reversal operator that transforms t into $-t$. A dynamical equation (or a law-like statement) is said to be *time-reversal invariant* iff it is invariant under the application of $\mathbf{T}$.[13]

[13]See Castagnino et al (2005, [23]:3), or alternative definitions in Earman (1974, [33]:25

Using this definition the reversibility objection (Loschmidt, 1876, translated in Brush [16]) can be formulated as follows. Hamiltonian micro-dynamics are time-reversal invariant. Hence, for any SM dynamical system the sequence of states $Seq = x_1, x_2, x_3, \ldots, x_n$ is just as likely as the reverse sequence $\mathbf{T}(Seq)x_n, x_{n-1}, \ldots, x_3, X_2, x_1$. An evolution from a low-entropy initial state to a high-entropy final state is, therefore, just as likely as the reverse evolution.

The recurrence objection (Zermelo, 1896, translated in Brush [16]), raised some years later, is based on Poincaré's recurrence theorem. According to that theorem "any classical mechanical system, with a bounded phase space, returns to a state arbitrarily closely to its initial state, and indeed repeat this infinitely often" (Uffink, 2006, [95]:64). Zermelo noted that, as a consequence of Poincaré's theorem, for any dynamical system confined to a finite region of the phase space (for example, the hypersurfice Γ_E), it is impossible to define a continuous function such that it always increases for all initial states. In other words, this objection argues that the irreversible processes described in the H-theorem are impossible to obtain.

So both objections show, in short, that nothing in the dynamics governing the components of the statistical mechanical systems can account for the asymmetry displayed by thermodynamic behaviour.[14]

If these objections are true, it follows that, according to the dynamics in SM, every non-maximal entropy state represents a local entropy minimum of the possible trajectories coming from states with higher entropy. This means that, at any arbitrary intermediate state during the evolution of a system, one can retrodict by means of the SM formalism that the system had a higher entropy in the past (see Fig.5). But this is incompatible with the unidirectionality of the second law of thermodynamics, according to which entropy cannot ever decrease.

and 2002, [35]:246).

[14]For a detailed analysis of these objections see Brown *et al*, 2009, [14].

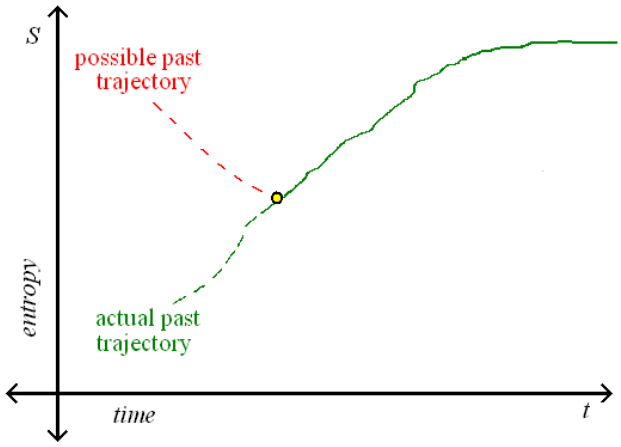


Fig.5. Thermodynamically abnormal retrodictions follow from SM dynamics.

The incompatibility between Hamiltonian SM and thermodynamics pointed by these objections is commonly defeated by postulating that the system's evolution depends upon its low-entropy initial condition. From this perspective, it is true that both thermodynamically normal and abnormal evolutions are possible according to SM, but, the *matter of fact* low-entropy initial conditions have been such that they lead to evolutions that obey the second law of thermodynamics. For example, if we observe a glass with water and some small ice-cubes inside it, SM dynamics tell us that past states where 'the ice-cubes were bigger' are equally likely as past states where 'ice-cubes were melted'. However, as a matter of fact, it has always been the case that initial states of 'glasses with water and ice-cubes' are such that ice-cubes liquefy and not solidify in water.

This solution, however, seems to simply 'displace' the problem from the present to some other state in the past (namely, the initial state). For whatever the election of the arbitrary *initial state* is, it will turn out (again) that many possible trajectories compatible with that state will come from higher entropies (see Fig.5 above). To avoid this prevalence of the anti-thermodynamical possible past, one posits that the low-entropy state holds at the very beginning of the universe (in some sense, at a moment "without past"). This is the so called **past hypothesis** and has been defended, among others, by Boltzmann (1895), David Albert (2000) and Huw Price (2004).

Many efforts have been devoted to make sense of the past hypothesis within modern cosmology models. These efforts appeal to several different features of the universe to explain the past hypothesis: inflation (P. Davies, 1983, critically analyzed by D. Page, 1983, and by H. Price 1996, 2002, 2003, 2004), black and white holes (S. Hawking, 1942), or the Weyl tensor (R. Penrose, 1989,2006).

Some philosophers of science, nonetheless, refuse to accept the idea that the initial conditions are sufficient to explain the approach to equilibrium. So they work in a different direction. They believe that some *additional* causal mechanism must be taken into account to provide an entirely satisfactory explanation of thermodynamic behaviour. Interventionism is one of those approaches that postulate some additional causal mechanism (and will be reviewed in chapter 2).

And yet another group of authors not only question the sufficiency of the past hypothesis (as interventionists do), but also question its necessity. Castagnino, Lombari and Lara (2003, [22]), for instance, propose to explain the temporal asymmetric behaviour of the universe without appealing to the past hypothesis – or to any other entropic consideration.

The main concern shared by many philosophers regarding the past hypothesis (among them Earman 2006, Winsberg 2004, and Callender 2010) is that, even if an acceptable definition of entropy could be proposed for the early universe (a goal that has not been reached), is not clear at all that the entropic behaviour of the universe as a whole is enough to justify the behaviour of small local systems.

1.5 Definition of the Irreversibility Problem

In the previous sections we have been using indistinctively the terms "irreversibility", "time-asymmetry", "direction of time" and "temporal reversal non-invariance". For the sake of clarity, it is now convenient to make explicit some assumptions regarding the differences and the relations among these concepts.[15]

[15]I'm very grateful with Olimpia Lombardi for clarifying the differences between these concepts both in conversations and her papers (2003, [68]; 2005, [23] and 2005, [65].

As mentioned before **time reversal invariance** (or 't-invariance' for short) is predicated of an equation or a law iff it remains invariant under the transformation **T**, which replaces the variable t with the variable $-t$ (see definition on p.16).

The terms 'reversibility' and 'irreversibility', by contrast, are not applied to equations or laws but only to physical processes. **Reversible processes** are governed by t-invariant laws. And a physical process is said to be **reversible** if its initial state can be completely restored. The paradigmatic example of reversibility is a very slow process in which the system always remains in states close to equilibrium. But a system's evolution may be more 'violent', beginning in a non-equilibrium state, and it may remain possible to restore the initial state by some auxiliary mechanism in such a way that the inverse sequence of states will take place. Reversible processes are temporally symmetric.

Irreversible processes, by contrast, are governed by non-t-invariant laws. A process is called **irreversible** if the sequence of events that make up the process occurs in only one temporal sequence and never in the inverse sequence. Irreversible processes are temporally asymmetric.

As a matter of fact there exist in nature many processes in which reversibility is impossible: gases expand, ice-cubes melt, and waves in lakes always propagate away from the point were a stone has generated the wave. Maybe because of this, irreversible processes have often been considered as the source of our intuitions about the flow of time. This brings about at least three philosophical problems that must be carefully distinguished:

(P-I) The problem of "the subjective arrow of time" concerns the psychological perception of a time flow.

(P-II) The problem of the arrow of time explores wether a distinction between past and future can be drawn. It aims at defining and justifying time's direction.

(P-III) The Irreversibility Problem concerns how to account for the *de facto* irreversible physical processes despite the fact that the underlying mechanics are t-invariant.

In this thesis P-I and P-II are not addressed. Only P-III will be examined in detail. More precisely, we will focus on a particular approach to P-III known as interventionism which will be explained in the following chapter.

Chapter 2

Statistical Mechanical Interventionism

In this chapter the interventionist approach to statistical mechanics is introduced. After describing the main features of this approach (in section 2.1) I will discuss the spin echo experiments and explain why are they relevant in the context of the irreversibility problem (in section 2.2). In section 2.3 I will explain the classical interventionist models, particularly Ridderbos and Redhead's model, in relation with the spin-echo experiments. At the end of this chapter (in section 2.4) I will explain a recent version of interventionism which is based on quantum mechanics.

2.1 Distinctive features of interventionism

The approach to statistical mechanics called 'interventionism' was originally advanced as a solution to both the problem of the arrow of time (P-II) and the irreversibility problem (P-III). It is however nowadays accepted that interventionism is not able to define a preferred direction of time; hence it does not solve P-II. For this reason interventionism is only considered in this thesis as an attempt to solve exclusively P-III. In other words, I understand interventionism as an approach such that, *once the direction of time is defined*, attempts to explain the irreversible behaviour of thermodynamic processes given that the underlying laws are t-invariant.[1]

The behaviour of thermodynamic systems is asymmetric. So it is natural to think that an appropriate explanation of such behaviour requires the

[1] Another relevant attempt to solve P-III is the ergodic theory. For some details about ergodic theory see appendix B.

presence of an asymmetry in the explanans. The most widely accepted view is that this asymmetry in the explanans is precisely the low entropy initial condition. However, interventionists do not accept that this initial condition explains irreversible processes. Whence they propose introducing a second asymmetry in the explanans, viz., a *causal mechanism* that ensures that entropy will not decrease (see Price, 2004, [80]:section 3). It is important to stress, however, that interventionism never suggests *replacing* the first asymmetry (low-entropy initial condition asymmetry) with the second one (causal asymmetry) but rather it is meant to *supplement* it.

According to interventionism, the random influence of the external environment acts as a source of perturbation over the system. And it is precisely this environmental disturbance that causes the irreversible increment of entropy. Hence, the essential idea posit by interventionism is that real systems are never isolated from their surrounding environment; On the contrary, they are in constant interaction with it. Isolation, if possible, is only achievable for finite and very short times. For any larger time range, for most physical systems, the interaction with the environment becomes crucial.

One may think that this kind of causal mechanism appeared for the first time in Boltzmann's H-theorem. After all, the H-theorem suggests that the entropy increment is caused by collisions between the gas particles and takes into account the causal interaction with the walls of the container. However, Boltzmann's reformulations of the H-theorem diverged importantly from what we call today 'interventionism'. More specifically, in his last version of the theorem there is no appeal to any causal mechanism but only to initial conditions. Therefore Boltzmann should not be considered 'the founder' of interventionism.

It is more appropriate to trace back the origins of interventionism to the mid 20th century, more precisely, to Peter Bergmann and Joel Lebowitz's paper 'New Approach to Nonequilibrium Processes' (1955, [7]) and to Reichenbach's book *The Direction of Time* (1956, [83]). Also the model that J.M. Blatt put forward a few years later (1959, [9]) is one of the pioneer interventionist proposals. These early interventionists of the 1950's were followed some decades later by Michael L.G Redhead and T.M. Ridderbros (1998, [82]). From now on I will refer to all these authors as "the classical interventionists". More recently, Meir Hemmo and Orly Shenker (2003, [54] and 2005, [55]) have put forward an explanation of irreversibility based on *quantum* decoherence for open systems. Despite the fact that Hemmo and Shenker's approach importantly differs from the classical interventionist ef-

forts in the sense that it appeals to quantum mechanics, the authors name themselves interventionists because the interaction between the system and the environment plays a crucial role in their explanation. We can fairly say that this is the list of the most relevant interventionist proposals.

The above mentioned interventionists appeal to different interacting elements to explain the irreversibility. The particles that make up the system may interact with the walls, with external particles or with themselves. In Bergmann's and Lebowitz's model (1955, [7]; and 1959, [64]), for instance, the particles interact with the container in which the gas is confined. The authors refer to the container as a 'driving reservoir' and impose on it a set of theoretical conditions. They assume that the reservoir has a fixed temperature and infinite heat capacity. Another remarkable condition assumed in this model is that the reservoir is composed out of an infinite number of parts and, each part of the reservoir interacts only once with the system:

> "If the reservoir interacts for brief spans of time only, by then strongly, it may be assumed that the net effect of such impulsive interaction will be to move the representative point in system's phase space a finite distance, that depends both on its original location and on the state of the reservoir just prior to interaction. If we further simplify the reservoir by assuming that it consist of a sensibly infinite number of disconnected and similar parts, that each such part interacts with the system but once, and that prior to interaction there is statistical independence between that reservoir component and the system, then we can average over the possible states of the reservoir."(Bergmann and Lebowitz, 1995, [7]:579)

Bergmann and Lebowitz impose such a condition (to divide the container in infinite parts) in order to make the mathematical model work.[2] The evolution of the joint system (system of particles + driving reservoir) is then described by a differential equation in the phase space. As shown below, the equation derived is equivalent to Liouville's equation but with an extra stochastic term in the right hand side. Liouville's original equation is:

[2]It is worth mentioning that interventionist models reject the idealization of considering 'completely isolated' systems. And they also reject other idealizations. For example, they criticise the ergodic program for using mixing properties because they require infinite times for systems to reach equilibrium. However, someone may argue, Bergmann and Lebowitz idealize the features of the reservoir by assuming that it has an infinite number of parts. Thus interventionism, at least this particular interventionist model, is not free of idealizations.

$$\frac{d\rho}{dt} + \{\rho, H\} = 0$$

where H is the Hamiltonian and $\{. \ , .\}$ is the Poisson bracket.

The modification on Liouville's equation proposed by classical interventionists is the following

$$\frac{d\rho}{dt} + \{\rho, H\} = \int_{x'} K(x, x')\rho(x') - K(x', x)\rho(x)dx'$$

where $K(x, x')$ is the probability that state x is displaced to state x' in Γ (see Fig.6 below)

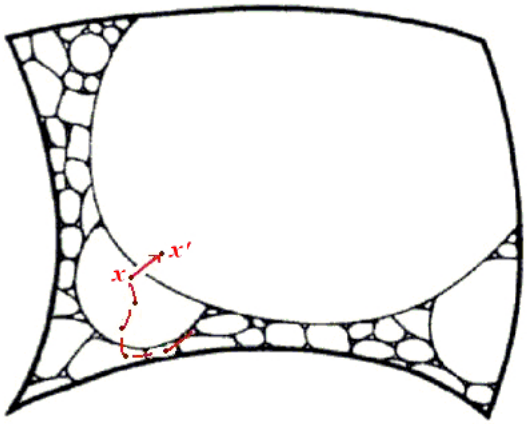


Fig.6. Displacement of the system from state x to state x'.

Modifying Liouville's equation is essential to Bergman and Lebowitz's proposal because, as the original formulation of Liouville's equation no longer holds, fine-grained entropy is no longer constant through the system evolution. Consequently, the approach to equilibrium can be now associated with the increment of fine-grained entropy without appealing to Gibbs' coarse-grained method.[3] It is shown that, for any arbitrary initial ensemble distribution, Bergmann and Lebowitz's model implies that the ensemble will approach equilibrium (to the microcanonical distribution) in the presence of a single reservoir.

Blatt's model (1959,[9]), in turn, explains the approach to equilibrium by appealing to the collisions between the molecules surrounding the system and

[3]Gibbs' coarse-grained method is explained above, see Fig.3 and Fig.4 on page 14.

the external side of the container. The interactions generate perturbations over the system that must be described in stochastic terms. The random or stochastic nature of these interactions is not ontological but, rather, it is due to our limited knowledge of the specific collisions. That is to say, the collisions are in principle deterministic, but we describe them as being stochastic because this is the only option available given our epistemic limitations. A peculiarity of Blatt's model is that the number of particles within the system, and the corresponding degrees of freedom, is irrelevant and need not be extremely large. This is interesting because, as mentioned before (1.3) the second law of thermodynamics was formulated for describing thermodynamic systems made up of a high number of particles. And, accordingly, the formulations of the law (either 2LTD, PP-2L or R-2LTD) assume that the system under cosideration has a high number of degrees os freedom. Hence it may be convenient to understand Blatt's model as an attempt to explain the *matter of fact* irreversibility of physical processes rather than an attempt to explain the second law of thermodynamics.

Just as Bergmann and Lebowitz, Blatt argues against the coarse-graining method proposed by Gibbs. In Blatt's view this method "ignores" the correlational information contained in the system. In other words, as the coarse graining method takes the average over the phase space cells, some information about the particular distribution in each cell is unavoidably lost. As a consequence, from the coarse-graining perspective we do not really know if the system is genuinely approaching to equilibrium or if, by contrast, it has kept a 'hidden order'. This makes coarse-graining, according to Blatt, a useless strategy for explaining a system's approach to equilibrium.

The same argument against the coarse-graining method is later defended (in 1998) by Ridderbos and Redhead. They suggested giving up coarse-graining in non-equilibrium SM and adopting more appropriate strategies instead (Ridderbos & Redhead, 1998, [82]:1273). Ridderbos and Redhead argue that the degrees of freedom of the system are entangled with the more numerous and sometimes unobservable degrees of freedom of the environment. This entanglement allows the authors to model the interaction between system and environment introducing a stochastic term, generating again a modified Liouville equation (different of Bergmann and Lebowitz because the walls of the container play no role here). Once Liouville's equation is modified, the increment of Gibbsian entropy is possible. In this way, assuming the stochastic interactions enable interventionist to derive the asymmetrical thermodynamic behaviour.

In sum, all the classical interventionist models mentioned above share a common *two-step plan* to solve the irreversibility problem. In a first step they reject Gibb's coarse-graining method. In a second step they propose a modification in the dynamics. More specifically, they *replace* Liouville's equation with a new equation of a stochastic nature.

To expand on the interventionist position we need to first consider the spin-echo experiments which constitute a case of particular relevance for our philosophical discussion. In the next section (2.2) I will describe those experiments and explain why are they relevant for interventionism. In section 2.3 I will explain the main features of classical interventionism focusing on the interventionist models of the spin-echo experiments. Section 2.4 addresses the quantum version of interventionism.

2.2 The Spin-Echo Experiments

In the so-called Spin-Echo Experiments ("SE experiments" from now on) a system of spins suffers several changes due to magnetic alterations. These experiments possess two interesting features relevant to the philosophical debate regarding the irreversibility problem. Firstly, during the experiment it is actually possible to control some of the microscopic variables of the system– something usually impossible for thermodynamic systems. And secondly, isolation from the external influence is highly controlled. This combination of features is rarely found in a single experimental device. Additionally, the SE experiments represent a challenge particularly for interventionists because the system of spins fluctuates from a state of high entropy to a state of low entropy (and vice versa) and the environment seems to play no role at all during the whole process. For this reason, studying the SE experiments in some detail will be useful for understanding the advantages and disadvantages of interventionist approaches.

The underlying mechanism of the SE experiments can be understood using an analogy offered by Erwin Hahn (Hahn, 1953, [50]: 5) who first carried out the experiment in the 1950's. The analogy consists of an imaginary Olympic race. The relevant system is the group of runners in this race. The runners are aligned along the starting line marked on the track (a very *ordered* state). The race takes place and when it finishes some runners are ahead of others so the positions of the runners are not longer aligned (representing a *disordered* state of the system). Now let us imagine that the runners are standing in their final positions and we ask them to turn around and start running back. Assuming (for the sake of the adequacy of the metaphor) that every runner will equal the velocity she had during the first phase of the race, the runners will reach their original positions at the starting line and recover the ordered alignment they had at the start of the race. If the duration of the original race is τ the runners will regain their original positions after a total time of 2τ.

In an SE experiment we have a set of nuclear spins (normally belonging to protons in a sample of glycerine) placed in a strong magnetic field. Through the application of a first radio-frequency pulse, the spins are initially aligned. In other words, the initial positions of the spins configure an ordered initial state. Let us say, for example, that the direction of the magnetic field is in the z-axis while the superposed spins lie over the xy-plane and they are all pointing in the same direction (as shown in Fig.7 below). Due to the presence of the magnetic field all the spins are initially precessing with the

same frequency and this produces the emission of an electromagnetic signal (represented in Fig.7 as the arrow). The spins are then left to evolve for a while and the discontinuities in the magnetic field cause slight differences in the precession rates of the spins (Fig.8). The precession rate of each spin is more or less affected depending on the strength of the field at each point. The evolution of the spins' system during this stage is analogous to the fist 'race' among the runners.

After an interval of time (τ), the spins reach a *disordered* state, in which they are not pointing in the same direction anymore and the electromagnetic macroscopic signal completely disappears (Fig.9).

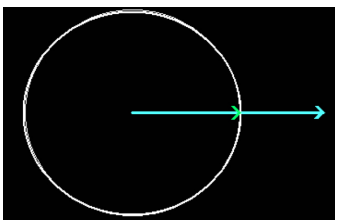


Fig.7. Initial signal

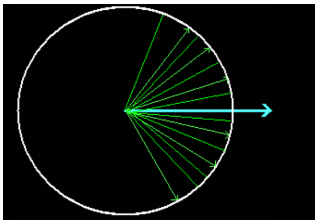


Fig.8. Defocusing

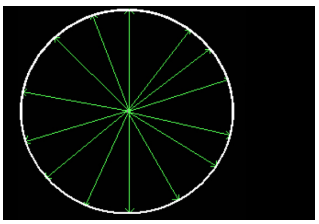


Fig.9. Total defocus

In a second part of the experiment (which corresponds to the runners running back) a reversal is induced by another radio-frequency pulse. After an interval of time (2τ) the spins are realigned and this causes the re-emission of the electromagnetic signal, which demonstrates that the spin's system has returned to an *ordered* state equal to its original state[4]

The repetition of the signal is the phenomenon that gives the name to the experiments. The evolution of the spin's system after the second r-f pulse would correspond to going from Fig.9, passing through Fig.8, and finally arriving, in Fig.7, to the original state and the emission of the echo-signal. In that moment the intensity of the electromagnetic signal reaches again a maximum point.

During the second part of the experiment it seems that "the spins go from an intuitively high entropy state to an intuitively low entropy state" (Ainsworth, 2005, [2]: 622). Refrigerators also produce this kind of anti-thermodynamic evolutions, but in that case an injection of energy explains

[4]Now it will be in the xz-plane instead of xy-plane but that is not relevant for our purposes. Other descriptions of spin-echo experiments can be found in Ridderbos & Redhead 1998 [82]; Albert 2000 [4]; Shenker 2001 [89]; Lavis 2003 [67] and Ainsworth 2005,[2].

the condensation of ice-cubes. In an SE experiment, by contrast, the echo system is supposedly in perfect isolation. So a natural question to ask is: Are the SE experiments a couter-example to the second law of thermodynamics?

We may of course repeat the procedure again on the system any number of times. We can control the radio-frequency pulses in such a way that we successively produce and destroy the alignment of the spins. As a result, every time that the spins are in phase the electromagnetic signal is emitted; and every time the spins are out of phase the signal disappears. So we have here a succession of signals in time. However, not all the echo signals are identical. Precisely as in an *echo*, the maximum intensity of the electromagnetic signals always decreases until the signal eventually disappears (see Fig.10.)

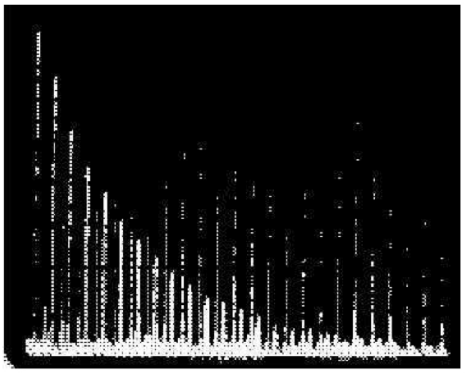


Fig.10. Echo-signal decay. [5]

This gives rise to a second question: Why does the intensity of the echo signals decrease?

As we mentioned before the SE experiments represent a challenge for interventionists. The reason is that the entropy apparently increases and decreases during the experiments, and the environment seems to play no role at all during the whole process. So, to summarize, at least three relevant questions have emerged from the SE experiments:

Q1: Is the second law of thermodynamics violated in a SE experiment?
Q2: How can we explain the decay and disappearance of the echo signal in successive repetitions of the radio-frequency pulses?
Q3: How do interventionist approaches account for the evolution of the spin's system?

[5]Image taken from the Massachusetts Institute of Technology website: http://ocw.mit.edu.

2.3 Classical Interventionism and the SE experiments

As regards question Q1 interventionists believe that the SE experiments do not violate the second law of thermodynamics. Let us assume that the violation of the second law requires a *decrease of entropy* to occur in a *closed* system *evolving spontaneously*. Then, in order to deny that the second law of thermodynamics is violated in the SE experiments, interventionists may follow three different strategies (or combinations thereof): They can deny that there is entropy decrease during the SE experiments; Alternatively, they can argue that in the SE experiments the system of spins is not closed; Or, finally, they can argue that the evolution of the spin's system in the SE experiments is not spontaneous but controlled. Let us consider these three strategies in turn.

2.3.1 Entropy in the spin-echo experiments

The first strategy has been adopted, for instance, by Blatt (1959, [9]) and Ridderbos and Redhead (1998, [82]). These classical interventionists argue that whether or not the second law of thermodynamics is violated by the in the SE experiments depends on our interpretation of the concept of entropy. If one interprets 'entropy' as the Gibbsian coarse-grained entropy one certainly arrives to the conclusion that *entropy* increases during the first part of the in the SE experiments, after the first pulse of radio-frequency is applied; and decreases after the second radio-frequency pulse is applied. However, one may adopt instead an "interpretation of entropy captured by the approach of counting on a logarithmic scale the number of accessible states of a system" (Ridderbos and Redhead, 1998,[82]:1237; see also Fig.1 on p.8). And, according to interventionists, interpreting entropy in this 'objective' way, one arrives to the correct conclusion that the spin system evolves in perfect accordance with the second law, i.e., that entropy in the SE experiments never decreases. Ridderbos and Redhead's reason for considering inadequate the description of the in the SE experiments provided by coarse-grained is that

> "...it [the coarse-graining method] amounts to ignoring the correlations which are built up in the system under dynamical evolution. [Thus,] the apparent anti-thermodynamic behaviour of the spin system according to the coarse graining approach arises precisely because this approach ignores the fact that the original order of the system is spread out into correlational information,

> transforming the order into a kind of 'hidden order'." (Ridderbos and Redhead, 1998,[82]:1251-1252).

This means that, although from the macroscopic point of view the system of spins seems to reach a state of equilibrium when the spins are out of phase, in fact the correlational information about the initial order is still contained in the system. That is the reason why the system of spins is still able to recover its initial order. Such a state, in which the spins 'remember' the information about their initial state, corresponds to what Blatt (1959, [9]:749) called a state of *quasi-equilibrium*. However, the state of *true equilibrium*, is only reached when the correlational information has been dissipated and, once a system arrives to such a state of *true equilibrium* it stays there forever.

It is worth noting that the concept of *irreversibility* itself is redefined by classical interventionists. Namely, they call a process 'irreversible' if and only if at the end of that process the system reaches a state of *true equilibrium*. Once the system is in this state of true equilibrium, the initial state of the system cannot be restored anymore because the correlational information has vanished on the microscopic level. (Ridderbos and Redhead 1998, [82]: 1256). In other words, when a system under study is in a quasi equilibrium state the order has disappeared at the macroscopical level, but it is still 'hidden' at the microscopic level in the form of correlational information. In states of true equilibrium, by contrast, there is no such a 'hidden order'. If the nuclear spins were in a state of true equilibrium, they would not be able to get back in phase again.

Thus, the argument shared by Blatt and Ridderbos-Redhead can be summarized as follows: The system of spins neither evolves toward true equilibrium after the first r-f pulse, nor does it evolve away from true equilibrium after the second r-f pulse. In the interval of time $[0, 2\tau]$ the system of spins only fluctuates from states closer or farther to *quasi-equilibrium*. The coarse graining method yields an incorrect interpretation of the concept of entropy. And this coarse-graning method leads us to the wrong impression that the second law of thermodynamics is violated in the SE experiments. However, the relevant question is not 'why does entropy decreases during the second part of the in the SE experiments? (which is, according to classical interventionists, a question based on a wrong concept of entropy) but rather how the usual descriptions of entropy in SM (in particular the coarse graining method) should be modified or replaced in order to correctly describe the *thermodynamically normal* evolution of the system of spins in the SE experiments.

Blatt additionally comments that the SE experiments provide a good reason to defend the fine-graining interpretation of entropy. He drives our attention to the fact that the concept of coarse-grained entropy is often preferred to the concept of fine-grained entropy because a macroscopic observer is supposed to be limited to "coarse-grained" experiments. In fact, the concept of fine-grained entropy is sometimes considered meaningless appealing to the fact that the detailed information about fine-grained entropy is impossible to obtain by experimental means. However, in Blatt's view, the SE experiments proof that information about fine-grained entropy can actually be obtained via the intensity of the electromagnetic signal. Therefore, the SE experiments render irrelevant this particular reason to prefer the coarse-grained concept of entropy (see Blatt, 1959, [9]:745-746).

Rejecting the concept of coarse-grained entropy, in any case, is not an easy decision to take. Let us remember that Gibbs introduced the coarse-grained method precisely in order to account for the increase in entropy; an increase that was impossible for fine-grained entropy. According to Liouville's theorem the volume of the distribution probability ρ is preserved all along the system's evolution and, as a direct consequence, fine-grained entropy[6] S_G remains constant too. So, in addition to the arguments against coarse-graining presented above, Ridderbos and Redhead need to provide an argument for fine-grained entropy that overcomes the difficulty related with Liouville's theorem. The difficulty generated by Liouville's theorem is solved by interventionists by means of the second strategy, which will be discussed in the next section.

2.3.2 The system of spins is open

The second strategy consists in denying that the system of spins is closed during a SE experiment. Since this strategy offers an explanation of the echo signal decay, it provides an answer to question Q2 (on p.29).

As a matter of fact, the intensity of the echo signal decreases and eventually disappears, even if the r-f pulses are still repeatedly applied on the spin-echo system.[7] Interventionism can explain this fact by appealing to the interaction between the system of nuclear spins and its environment. This is an advantage of interventionism compared with other approaches to the

[6] Defined in section 1.3.

[7] In the case we are assessing in this thesis, i.e. in glycerine samples, the decay is exponential.

irreversibility problem that fail to explain this aspect of the spin-echo experiments.

Let us see how exactly the system of spins is said to interact with the environment. We are hence looking for a progressive dissipation effect in the SE experiments that, in terms of the race analogy, helps us explain why not all the runners went back to the starting line (or at least not as fast as they had run in the original race). Then, what we need is a physical process that makes spins 'lose their memory of their initial state', or some kind of energy dissipation that affects the spin's frequency of precession, preventing them from regaining the spin-alignment.

Magnetic energy is exchanged during the experiment in several different ways. Sometimes the nuclear spins transfer their magnetic energy of precession to the sample molecules in the form of kinetic energy.[8] The magnetic energy of precession is some other times transferred to the neighboring spins[9]. Additionally, the system is affected by the Brownian motion of the glycerine molecules and the fluctuations in the local magnetic fields due to neighbouring moments. These two latter phenomena may drive the momentum of some molecules from the static magnetic field (chosen by the experimenter and controlled with the r-f pulses) into another randomly differing magnetic field (Hahn, 1953, [50]:6). As a consequence, the spins' frequency of precession is perturbed.

The presence of these effects is used by classical interventionists to defend that, during the SE experiments, the spins change their frequency, not only as a result of their interaction with the experimental set up (static magnetic field plus r-f pulses) but also because they transmit magnetic energy to the environment. The spins that relax their energy and are not in phase with the rest of the spins will not contribute to the next echo-signal. Due to this constant disturbance from the environment, the number of out-of-phase spins increases during the experiment and this explains why the intensity of the echo-signals is progressively reduced.

In Blatt's terminology (on p.31), completely isolated systems never reach true equilibrium, instead they always stay in a state of quasi-equilibrium. In this state the velocities of the system can still be reverted showing that the

[8]The time that it takes for this effect to occur is denoted by T_1 and is known as "spin-lattice relaxation time".

[9]The time that it takes this second effect to occur is denoted by T_2 and is known as "spin-spin relaxation time".

system 'remembers' the information about its initial state. This is the case for the system of spins at $t = \tau$ (i.e. when the spins are out of phase for the first time). However, once the r-f pulses are applied many times the echo-signal eventually disappears. The system of spins has then reached a state of *true equilibrium* and the process is not longer *reversible*. According to classical interventionists, the gradual decay in the intensity of the echo-signal proofs that the correlational information dissipates into the environment, and this only makes sense if we conceive the system of spins as an open system.

Before proceeding to discuss the third strategy, let me just remark that the compatibility between Liouville's theorem and the above mentioned interventionist description of the SE experiments is no longer problematic once we recognize that the system of spins is not closed. When the environmental degrees of freedom come into the picture in interventionist models, Liouville's equation is modified and the theorem no longer holds.

Let us sum up the evolution of fine-grained entropy during a SE experiment. In the first part of the experiment the effect of the environment is almost imperceptible and fine-grained entropy is conserved. Liouville's theorem approximately holds and, from the interventionist point of view, appealing to coarse-grained entropy is unnecessary. In the second part of the experiment, and during the consecutive r-f pulses and echo-signals, the environment increasingly affects the system, diffusing its correlational information. Then, the stochastic term becomes more important and Liouville's modified equation holds instead of Liouville's theorem.[10]

2.3.3 Controlled versus free or spontaneous evolutions

So far, we have said nothing about the third strategy to deny the violation of the second law of thermodynamics. This strategy consists in denying that the evolution away from equilibrium manifested in the spin-echo experiments is *free* or *spontaneous*. In other words, it is argued that the second law only applies to closed systems that evolve "on their own", i.e., free of external influences. A decrease of entropy in a system evolving under such circumstances would genuinely represent a violation of the second law. However, the evolution of the system of spins in an SE experiment, far from spontaneous, may be rather considered as a highly controlled evolution.

[10] See Liouville's equation and the interventionists' modified equation on p.24.

This strategy, however, is problematic in the sense that classifying evolutions as 'controlled' or 'spontaneous' seems to depend on our election of the relevant system under study (see Shenker, 2001, [89]). For example, if in the SE experiments we only take the system to be the set of nuclear spins, it seems that the evolution has been induced and controlled from the outside world (via the r-f pulses). By contrast, if the environment is taken to be a part of the system itself the evolution seems spontaneous.

A similar point holds for what we classify as 'open' or 'closed' system. If an ice-cube placed on the table in a warm room is melting, should we consider the whole room as the system of study? If so, we should also take into account the warm air interacting with the system through the windows and this depends on the weather which depends in turn on the atmospheric pressure, and so on. Should we consider the whole planet as the relevant system then? Or should we consider the whole universe to be the only really closed system? We will return to this issue in the next chapter. All we need to stress now is that the interventionist strategies may be controversial. And that, particularly the third strategy, presupposes a well-defined notion of 'closed' system and concomitantly 'spontaneous' or 'free' evolution of such system.

We have so far explicitly addressed questions Q1 and Q2 . However question Q3 has been addressed implicitly too in the previous sections. As mentioned, classical interventionism accounts for the SE experiments by appealing to the distinction between quasi and true equilibrium; redefining the notion of irreversibility; and rejecting Gibb's definition of coarse-grained entropy. We may turn now to a discussion of two quantum based approaches to the irreversibility problem, and how these approaches explain the evolution of the spin echo system.

2.4 Quantum-based approaches to the irreversibility problem

Meir Hemmo and Orly Shenker (2001,[53]; 2003,[54]; 2005,[55]) proposed an explanation of the irreversible thermodynamic behaviour based on the underlying quantum mechanical dynamics. The base of reduction is different in nature in this new proposal. Hemmo and Shenker's only predecessor in this enterprise is David Albert (2000, [4]).

The most relevant difference between Albert's and Hemmo-Shenker's approaches is that, while the former is based on the GRW theory of the quantum *collapse* of the wave function (Ghirardi, Rimini and Weber, 1986, [45]), the latter is developed within *no-collapse* interpretations of quantum mechanics. More specifically, Hemmo and Shenker appeal to quantum mechanical models of environmental decoherence (Zurek and Paz, 1994, [109]) to account for the interaction between the system and its environment at the quantum level. This interaction, combined with the stochastic nature of the quantum dynamics, leads to the irreversible increase of entropy (more details below). The environment and its influence on the system, by contrast, are completely irrelevant in Albert's proposal.

Classical interventionism aims to reduce thermodynamic processes to SM or, more precisely, to a modified version of SM –because Liouville's equation is replaced. Instead, the quantum-based approaches (both the GRW-based and decoherence-based approaches) propose solving the irreversibility problem by reducing thermodynamic processes using a *different* reduction base, namely, quantum mechanics. I turn now to explain the quantum-based approaches in detail.

2.4.1 GRW-based approach

Let me review the basic idea involved in Albert's theory, which is related to the following argument:[11]

The initial macrocondition M of a given system is compatible with a collection $\{L\}$ of both normal and abnormal microconditions mc[12] and there happens to be a *breathtaking straightforward measure* on the set $\{L\}$ of a system which has the following characteristics: (1) The measure counts the collection of normal points in $\{L\}$ as vastly larger than the collection of abnormal points in $\{L\}$; And (2) the measure is preserved[13] by the equations of motion (Albert, 2000, [4]:151).

The characteristics (1) and (2) imply that, under small perturbations of the system, the property of being a *normal mc* is stable, while the property of being an *abnormal mc* is extraordinarily unstable. In such circumstances, if the system was frequently, microscopically and randomly perturbed, the evolution of the system would tend to equilibrium independently of which particular microcondition mc in $\{L\}$ initially obtained.

Albert then suggests that the quantum collapses postulated by GRW theory "turn out to be just the sort of perturbations we need" (Albert 2000, [4]:151). Therefore, in Albert's view, the thermodynamic behaviour (i.e., the fact that every single mc in $\{L\}$ will be overwhelmingly likely to evolve towards equilibrium) is a consequence of the GRW dynamics. There is no need to postulate any other process to account for irreversibility.

Note that, in contrast with the *external* mechanism that introduces randomness in the classical interventionist approaches, GRW collapses in Albert's theory act as an *internal* source of perturbation.

[11] The original argument is in Albert, 2000, [4]:151-158.

[12] Note that a microcondition $mc \in \{L\}$ is considered *normal* if it belongs to a *normal thermodynamic trajectory*, and it is considered abnormal if it does not. Since these terms are defined relative to a trajectory, they do not concern only a specific moment but all the future and past states of the system. A consequence of defining *(ab)normality* in this way is the following: One cannot associate an *abnormal mc* with an ordered arrangement of particles M_o (gas in the corner of a box) and *normal mc* with a disordered arrangement of particles M_d (gas dispersed all around the box volume). In other words, in Albert's argument *normality* is not ascribed to particular states, but rather to trajectories conformed by a large succesion of states.

[13] Preservation in accordance to Liouville's theorem or to its quantum mechanical correlate, the Principle of Unitarity.

Let us apply Albert's model to describe the process of a gas spreading out. The wave function of such a system will evolve in accordance with the GRW dynamics. According to GRW, there is a high probability for a collapse to occur in a temporal interval which is reasonably short in the thermodynamic scale, but long enough in the quantum scale. This is guaranteed because in GRW the parameters of the dynamics are chosen in such a way that the quantum predictions for microscopic systems remain fully valid while the macroscopic superposition in measurement-like situations is suppressed in very short times. In fact, GRW dynamics are also referred as *unified* dynamics because GRW introduced a unique formalism for describing both microscopic and macroscopic systems. After choosing the most convenient parameters, it was computed that a collapse should occur for a microscopic system, on average, every hundred million years, while for a macroscopic one it should occur every 10^{-7} seconds.[14] Thus, in the gas example we are considering here, it is reasonable to suposse that a GRW collapse will occur in the experiment and the wave function of the gas will collapse into a specific state.

While the gas is spreading several collapses occur giving place to a sequence of states that conform a trajectory in the state space (see Hemmo & Shenker, 2003,[54]: 338-340). The crucial question is whether or not the trajectory of states traced over the state space is a *thermodynamic trajectory*, i.e., a trajectory obeying the laws of thermodynamics. To answer this question Albert proposes the following *dynamical hypothesis*:

"The GRW dynamics are such that the probabilities for the collapse transitions reproduce the probabilities of the trajectories calculated from the standard statistical mechanics measure for any given macrostate of the system." (Albert, 2000, [4]: 151-152)

Two direct consequences follow from Albert's dynamical hypothesis:
(a) The set of the thermodynamic abnormal evolutions has measure zero
(b) In every microscopic neighbourhood the thermodynamically abnormal states are uniformly distributed among the thermodynamically normal states.

[14]Commonly a *microscopic* system is taken to be of atomic dimensions or smaller, while a *macroscopic* system is large enough to be visible in the ordinary sense. A more exact definition relies on the number of particles in the system. Let us call N the number of particles. Then macroscopic systems are such that $N^{-1/2} << 1$. This definition is useful to determine that a system must contain at least about ten thousand particles in order for statistical arguments apply to reasonable accuracy (see Fitzpatrick, 2006, [40]).

The following image may be helpful to illustrate Albert's arguments.

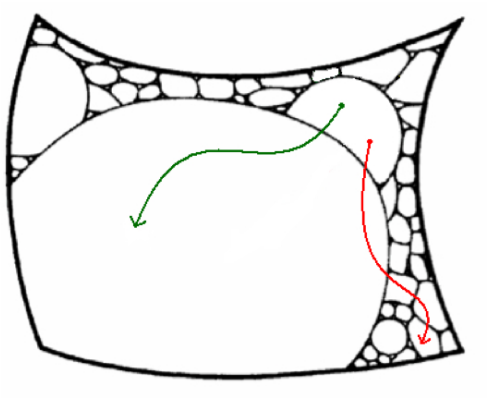


Fig.11. Normal and abnormal trajectories.

In every region of the phase space there are many states that belong to a "well behaved" trajectory, i.e., to a trajectory obeying the second law of thermodynamics. Let us call them the 'good' states. There are also a few states (a set of measure zero) that will evolve anti-thermodynamically and hence belong to abnormal trajectories. Those are the 'bad' states. Since both the "good" and "bad" states belong to the same region in the phase space, we are not able to distinguish them at the macroscopic level. What the GRW dynamics guarantee, so Albert's argument goes, is that the number of good states in each region is overwhelmingly larger than the number of bad states. So much larger in fact, that the white regions in Fig.10 would comprise good states, and the bad states would be hardly perceptible.

2.4.2 Decoherence-based interventionism

A different way to underpin thermodynamic regularities in quantum mechanical grounds has been proposed by Meir Hemmo and Orly Shenker (2001, [53]; 2003, [54]; 2005, [55]). According to their approach, quantum decoherence is the mechanism that "brings about an approach to equilibrium in the classical sense of, for example, an evolution towards the most probable macrostate." (Hemmo & Shenker, 2003, [54]: 348).

Even though some features of theories of decoherence are not completely recovered by Hemmo and Shenker, it is clear that some central ideas in their approach rely on those theories, particularly on results by Zurek and Paz (1994, [109]) and Joos and Zeh (1985, [61]). For instance, the interaction with the environment is used by Zurek and Paz to account for the increase in von Neumann's entropy. Hemmo and Shenker refer to this result and also consider the environment as key but they do not need to make use of von Neumann's entropy.[15]

In the stochastic no-collapse interpretations of quantum mechanics some additional dynamical laws are introduced in the quantum description of the system. These laws are said to produce the so-called *effective* collapses, which in contrast with the usual collapses of the quantum state, are not considered as real collapses. Hemmo and Shenker built their answer to the irreversibility problem on such no-collapse interpretations of quantum mechanics, assuming that "when macro-systems undergo decoherence interactions with their environment the extra dynamics results in *effective* collapses onto coherent states corresponding to what [the authors] have called quantum mechanically normal states." (Hemmo & Shenker, 2005, [55]: 632, original emphasis).[16]

From this perspective, the total quantum state does not evolve in accordance with GRW dynamics as Albert proposed, but evolves in accordance with the Schrödinger equation. Due to the decoherence, effective collapses of the quantum state will bring about transitions from one effective state to another one. In other words, the system seems to jump from one Schrödinger

[15]The relation between von Neumann's entropy and thermodynamic entropy is very intricate. Hemmo and Shenker carefully discuss this topic in [54] and [55]) but we will not detail that discussion here for two reasons: The first one is that, according to Hemmo and Shenker's proposal, it is possible to recover the thermodynamic regularities without appealing to von Neumann's entropy. The second is that, although Zurek and Paz's results give support to their decoherence-approach, the authors mention that the approach would stand even without those results.

[16]In particular, "decoherence ensures *effective* collapses onto the coherent states given by Gaussians in both position and momentum." (Hemmo & Shenker, 2005, [55]: 632).

trajectory to another one. This is the so-called *effective collapse*.

A system's effective state is not uniquely determined by its previous effective state. On the contrary, "transitions between effective states are genuinely stochastic [and] on this assumption the result is that the effective state of the thermodynamic system changes in a stochastic way in the course of decoherence" (Hemmo & Shenker, 2003, [54]:351). In this way, the stochastic nature of the system's evolution, according to this approach, is the result of decoherence together with the stochastic extra dynamic laws.

In order to recover the predictions of classical statistical mechanics, Hemmo and Shenker also need to postulate the following dynamical hypothesis; which is analogous to Albert's dynamical hypothesis.

"The quantum mechanical probabilities [produced by the extra dynamical laws] reproduce the quantitative predictions of classical statistical mechanics" (Hemmo & Shenker, 2003, [54]:633)

Although the authors recognize that they cannot provide a proof of their own version of the dynamical hypothesis, they argue that its plausibility can be defended on the basis of the characteristics of spontaneous effective collapses. More precisely, they argue that effective collapses are extremely frequent in macroscopic systems experimenting decoherence. And these collapses induce extremely small changes in position comparable with changes in position at statistical mechanical scales.

In sum, the irreversibility problem is solved in the decoherence-approach by offering a mechanism, viz., effective collapses in decoherence situations, which guarantee that the evolution of a macroscopic system has a high probability of being thermodynamically 'normal'. Let us see now how this applies specifically to the case of the spin-echo experiments.

According to decoherence-based approach, the system is not only affected by the external interaction with the environment (as in classical interventionist approaches) but it is also affected *internally* by the stochastic dynamics, i.e., by the effective collapses. This is also manifested in the double role played by decoherence in the SE experiments: On the one hand decoherence directly affects the spins through the influence of the environment. On the other hand, a diffusing effect is generated by the interaction between spins themselves, which influences their states. Let us call these decoherence processes *external* and *internal* decoherence respectively.

The internal decoherence, i.e. the interaction between spins, leads in turn

to the echo-signal decay by a process that possesses two different stages. In a first stage the spin-spin interaction induces stochastic effective collapses. In the second stage, each collapse of a spin produces a kind of "holistic" diffusing effect in the spin states of other particles. A direct consequence of this is that, if a decoherence based approach is correct, an experiment in which the spin-spin interaction is reduced (for example, by diluting the sample in the glycerine) should slow-down the decay of signal intensity. This opens the way to possible experimental tests for the decoherence-based approach to the problem of irreversibility.

2.4.3 Final remarks about quantum-based approaches

Another feature of GRW-based and decoherence-based approaches is worth commenting on since it distinguishes them from previous interventionist approaches. In both quantum approaches there is no need to appeal to the low entropy initial condition of the system in order to derive the thermodynamic regularities. The derivation of thermodynamic regularities from the underlying GRW dynamics in Albert's case, and from decoherence in Hemmo-Shenker's case, is said to be independent of the particular thermodynamic initial state. In other words, they solve the irreversibility problem by making the thermodynamic *normality* of the trajectories fully accountable for by the internal quantum dynamics.

It does not follow from this that Albert's and Hemmo-Shenker's quantum-based approaches are incompatible with a low entropy initial restriction. Each approach may be combined with the past hypothesis (i.e. with the low-entropy initial condition of the universe). The role that such hypothesis may play is setting the direction of time from past to future. Namely, the past direction is defined towards the low-entropy initial condition. But once the time direction is defined, the quantum-based approaches would leave the past hypothesis aside and appeal only to their respective dynamical hypotheses together with (either real or effective) collapses to recover the thermodynamic behaviour of the specific system under study. So although the quantum approaches are consistent with the past hypothesis, they *do not* require it to solve the irreversibility problem.

The weak point of these approaches may be their reliance upon their dynamical hypotheses. These hypotheses in essence express an intimate link between quantum dynamics and thermodynamics, which may be as hard to substantiate as the irreversibility problem itself.

Chapter 3

Objections to Interventionism

It has been claimed that interventionism cannot explain the increment of entropy in the universe as a whole – since this system lacks an environment. There are also philosophical debates about whether the randomness of the interaction between environment and system is genuine or simply related to our ignorance. And probably the most relevant question is this: is interventionism able to avoid the undesirable consequence that the system tends to equilibrium not only towards the future but also towards the past? (Davies, 1974, [30]; Sklar, 1993, [91]; Price, 1996 [79]; Callender, 1999, [17]). Ridderbos and Redhead attempted to provide some answers to these objections from the interventionist point of view. Nevertheless, new arguments have been recently raised against interventionism (Price, 2004, [80]; Callender, 2004, [18]; Ainsworth, 2005, [2]; Uffink, 2006, [95]). In this chapter I review some of these criticisms. In section 3.1 the "parity of reasoning problem" will be explained. In section 3.2 the problem regarding the ontological or epistemological nature of randomness will be discussed. Both (sections 3.1 and 3.2) include the counterarguments that interventionist have raised in retort to the objections. Section 3.3 adds a brief comment on idealizations.

3.1 The Parity of Reasoning Problem

Interventionists' arguments explaining how environmental perturbations lead the system to equilibrium *toward the future* can also apply, by parity of reasoning, *toward the past*. For example, if we observe a system at some arbitrary time t during its evolution we could say (using interventionist arguments) that, given that the system interacted with its environment *in the past*, it was then randomly perturbed. Hence, the previous states were more disordered and "closer to true equilibrium" than the present state. Hence, the system has been evolving from higher to lower entropy states. But this conclusion is in conflict with our empirical experience. And it is not what interventionists aim to defend. This problem is known as the "parity of reasoning problem" or the "double standard fallacy" and is frequently associated with interventionism in the literature (Sklar, 1993, [91]:254; Callender, 1999, [17]:363; Price, 1996, [79]:68 or 2004 [80]:223).

The parity of reasoning problem is, in some way, inherited by interventionism from older causal proposals. The mechanism of molecular interaction, proposed in Boltzmann's H-theorem to explain irreversibility, for example, faced the same difficulty as shown by Loschmidt's so-called reversibility objection (in section 1.4). For this reason, in criticisms to a wider set of proposals, it has been claimed (see Callender, 1999, [17] and Price, 2004, [80]) that the applicability of arguments in both temporal directions is a problem shared by all approaches that attempt to explain irreversibility by postulating causal mechanisms: "The point [interventionists need] to keep in mind is that, as in all such causal approaches, the mechanism needs to be time-asymmetric, if it is not to force entropy to be non-decreasing in both directions."(Price, 2004, [80]:223).

There are at least two possible reactions one may have to the parity of reasoning problem. One attempts to provide reasons for thinking that the statistical arguments do not apply with equal force in both temporal directions. I will refer to this option as "the optimistic reaction" as it still tries to save interventionism. The alternative is to acknowledge the difficulty and allege that tracing a preferred temporal direction (i.e. solving the problem of the arrow of time P-II) is out of the scope of interventionism. I will call it the "pessimistic reaction" because it marks a limit in the scope of interventionism.

3.1.1 The Optimistic Reaction

An example of the optimistic reaction is Ridderbos and Redhead's defence of the Bergmann-Lobowitz model against Sklar's parity of reasoning objection. Let us explain Ridderbos and Redhead's argument ([82]:1261) by applying it to the common example of a gas spreading. We consider a gas in a box divided in two halves. The gas is initially all contained in one half. Once we remove the barrier between both halves the gas spreads out.

Imagine that the system is in a non-equilibrium state at some point during its evolution. Interventionism tells us that the correlational information dissipates during the evolution. More precisely, after interacting with the system, every part of the reservoir interacts in turn with its own environment. Consequently, the information flows away from the system (*it is exported*) and any later interaction between the reservoir and the system is statistically independent from its past interactions. In other words, it is reasonable to assume statistical independence between the interactions because, even if some specific part of the reservoir interacted again with the system, the information produced in the first interaction would already be missing. Hence it makes sense to suppose that this part of the reservoir and system are interacting for the very first time.

However, if we invert the time direction and study the interaction between the system and its environment in reversed evolution, we find that the part of the system (e.g. a gas particle) that has interacted with the reservoir at a given time has no way to dissipate the information produced in that interaction. The gas particle containing the information of the collision does not interact with the external environment (as parts of the reservoir do). Therefore, the next interaction between that specific particle and the reservoir will be correlated with the previous one.

Thus, in contrast with the normal time-directed case, in the reversed case we cannot claim that interactions are statistically independent from each other. Therefore, Ridderbos and Redhead's conclusion is that interventionist arguments do not apply equally in either direction of time.

In the spin-echo experiments this idea can be expressed as follows. The glycerine molecules are able to dissipate the energy, for example, in the form of thermal energy. Let us imagine that we have filmed the experiment and now we are looking at the movie backwards. Although the spins can "absorb" the thermal energy from the molecules, they are unable to dissipate that energy. They can only keep it and use it to precess.

Let us now consider some objections to Ridderbos and Redhead's solution to the parity of reasoning argument. It has first been suggested that attempts like this to solve the parity of reasoning problem simply shift it rather than resolve it. The temporal asymmetry is presupposed rather than explained again and, therefore, the argument begs the question (Price, 1996, [79]:68). For instance, Ridderbos and Redhead assume that the exchange of energy towards the future dissipates the correlational information, but backward energetic exchange does not. Proving this assumption is tantamount to proving that entropy tends to increase and never decreases.

In retort, one could argue that Ridderbos and Redhead do not *just postulate* the asymmetric behaviour but they also provide us with a reason to believe that the past-future symmetry is broken. They do this by appealing to differences between the system (with no degrees of freedom to dissipate information) and the environment (with plenty of degrees of freedom to dissipate information).

One may then object that reliance on differences between system and environment may be compromised by the fact that a new system composed by the original system plus its environment can always be considered as our new system of study. For example, Ridderbos and Redhead's argument not only serves to explain why the fine-grained entropy of the gas has increased due to the dissipation of correlational information produced by the interaction with the box. It also serves to explain the entropic behaviour of the laboratory where the box is located; or even the entropy of the whole building. This reasoning can be applied to a chain of systems that ends with the system consisting of the universe as a whole.

However, the universe as a whole is a system that has no environment to interact with. So how can we explain the entropy increase of such a system from an interventionist point of view? Or are we to accept that the universe's entropy is constant? Ridderbos and Redhead explicitly accept this view when they claim that fine-grained entropy remains constant for the universe as a whole, but it increases for any of its subsystems (see Ridderbos & Redhead, 1998, [82]:1261-1262). They argue that this does not contradict the usual idea in cosmology that universal entropy grows, because cosmologists base their concept of entropy on the distribution of matter in the universe, which is best understood as a coarse-grained entropy rather than a fine-grainied entropy.

It is not hard to show that an increase in coarse-grained entropy is con-

sistent with constant fine-grained entropy. After all this was precisely the reason why Gibbs introduced the coarse-grained method aiming to explain the second law of thermodynamics (see p.13). It is harder to justify that the fine-grained entropy of every subsystem of the universe may be growing, while the fine-grained entropy of the universe as a whole remains constant. In other words, if the entropy of the universe is constant then where in the universe is entropy decreasing in order to compensate the fact that every subsystem of the universe has equal or growing but never decreasing entropy? This consideration may lead critics of interventionism to claim that the interventionists' optimistic reaction to the parity of reasoning problem is inconsistent.

It is worth emphasizing that, unlike the classical interventionist approach, both Albert's and Hemmo and Shenker's quantum-based approaches assume that the perturbations are (at least partially) generated *inside* the system. This enables both quantum approaches to account for the behaviour of the universe as a whole in the same way they account for that of any of the universe's subsystems. Classical interventionism, by contrast, is committed to all perturbations to the system coming from its interaction with the environment and therefore cannot apply this reasoning.

3.1.2 The Pessimistic Reaction

There are in fact many diverse pessimistic reactions to the parity of reasoning argument. What all of them have in common is their claim that perturbations exerted by the environment over the system are time-directed and, with it, they acknowledge that interventionism simply does not solve the problem of the arrow of time.[1]

Perhaps the first example of pessimistic reaction is the one Sklar himself mentions immediately after discussing the parity of reasoning problem:

> "I can only imagine one way in which the interventionist can block this argument from parity of reasoning. It would be to argue that the intervention from the outside is itself time-directed. Because intervention is causation, and because causation is from past to future, the intervention can only modify the ensemble toward the

[1] This is the reason why in this thesis we take interventionism as a solution to the irreversibility problem (P-III) but not as a solution to the problem of the arrow of time (P-II). Both P-II and P-III have been defined in section 1.5, on p.20.

> future direction of time. But without some deeper understanding of what is being used here, some understanding of how causation is playing some role over and above lawlike correlation of states, this sounds more like an a priori restrictive instruction on when to use statistical mechanics and when not to." (Sklar, 1993, [91]:254).

In this quote Sklar suggests that the only possible way out is to relate the time-directed environmental interventions to causality. We may then consider that perturbations of the environment cause entropy to rise. The past-future asymmetry is then broken by appealing to the temporally asymmetric relation between causes and effects. This reaction is pessimistic about interventionism being able to underpin temporal directionality, in the sense that the asymmetry is not really brought about by the environmental perturbations, but by the asymmetry built into causality. And there is no appeal (at least not explicitly) to this causal asymmetry in the interventionists' original models.

Another pessimistic reaction is due to Orly Shenker[2](Shenker, 2001, [89]: sections 2 and 3) who accepts the double standard problem in mathematical terms, but then offers an epistemological argument to defend interventionism:

> "Interventionism uses the environment to predict the unknown future, not to explain the known past (...) Awareness of the distinction between explanation and prediction prevents us from committing Price's (1996) double standard fallacy, for we consciously put in a time asymmetry originating in our experience, by hand, without claiming that this experience reflects a fundamental asymmetry in nature." (see Shenker, 2001, [89]:2-10)

So, basing her argument on the difference between 'explanation of the past' and 'prediction of the future', Shenker suggests that interventionism works only in one temporal direction (not for physical or mathematical reasons but) simply because of the way it is built.

The critics of interventionism (e.g. Callender 1999, [17]:363-364; Davies, 1974, [30]:74; Price, 2004, [80]:225; Uffink, 2006, [95]:sec 7.5.2) claim that as far as interventionism is applicable in both temporal directions, it is not able

[2]In fact, Shenker is optimistic about interventionism but not about interventionism being capable to provide answers to the problem of the arrow of time, because, form her point of view, that is not even a goal of interventionism.

to solve any problem related with time-asymmetries. Rather interventionism introduces asymmetry by hand just in order to solve the problem. And this is moreover a problem that has already been solved by the initial condition asymmetry –which, by the way, serves to solve both the arrow of time and the irreversibility problem. From their point of view, there is no satisfactory way out to the problem of parity of reasoning. Optimistic reactions (as the one defended by Ridderbos and Redhead) are problematic since they base explanations of irreversibility upon a "unwarranted temporally asymmetric assumption, analogous to Boltzmann's *Stosszahalansatz* [...that] merely ends up pushing the question back a step" (see Callender, 1999, [17]:363). Pessimistic reactions, on the other hand, lead to recognize that the interventionist approach "is wrong half of the time" (see Callender, 1999, [17]:364). Namely, interventionism is right when it is applied toward the future, and wrong when it is applied toward the past.

3.2 Is Randomness Ontological or Epistemological?

In order to guarantee the approach to equilibrium, interventionism assumes that the environmental perturbations are *random*. However, the notion of *randomness* itself is a philosophically intricate notion.
[3]

It is worth asking, then, whether this randomness is ontological or epistemological. Interpreting randomness ontologically entails considering it as a real and genuine feature of the environment itself; while interpreting randomness epistemologically entails considering it a consequence of our ignorance.

If randomness is interpreted as an ontological feature of the environment, it seems that the laws governing the environment are essentially different from the Hamiltonian deterministic laws that hold inside the system. Hence, the ontological view of randomness seems to commit classical interventionism

[3]Randomness is an enormously important topic in the philosophy of science. This section focuses on a particular criticism that has been raised against interventionism on grounds related to randomness. However, it is worth pointing out (as stressed to me by me internal examiner, Miklos Rédei) that a number of proposals have been put forward lately to shed light onto the concept of randomness and its role in statistical mechanics. Berkovitz, Frigg & Kronz (2006, [8]) and Uffink (2006, [95]:sec 7.5) are examples of this. For a more general review of the fundamental problems in the concept of randomness see Coffa (1974, [26]: part III), Bennet (2011, [6]), Eagle (2011, [32]) and references thereby.

to a patchwork understanding of laws. However it can be argued that, as the environment is constituted by the same kind of particles than the proper system under study (atoms, molecules, etc.) both system and environment should be treated in the same way (see Albert, 1994, [3]:672; Callender, 1999, [17]:363; Hagar, 2005, [48]:474; Ainsworth, 2005, [2]:628; or Frigg, 2007, [41]:161). Additionally, as mentioned before, the division between system and environment is often arbitrary.

Maybe for these reasons classical interventionists usually choose the other option, namely, interpreting randomness epistemologically. The epistemological interpretation is usually identified with a Laplacian picture of the world, according to which, matter interacts in a purely deterministic way but randomness is introduced in our descriptions merely as expression of our ignorance. This seems to be the interventionists' position. For example, according to Blatt, the interactions between the system and its environment are considered random only as a result of our limited knowledge (see Blatt, 1959, [9]). So the components of the environment are governed by exactly the same laws as the system. It is the complexity of the environment and its high number of degrees of freedom that forces interventionists to build stochastic models.

A direct consequence of this epistemological interpretation, pointed by Peter Ainsworth (2005, [2]), is that the increase of entropy or, more precisely, the increase in the volume of the probability distributionρ [4] due to environmental interactions "can only represent a diminution of what we know about the location of the system in phase space" (Ainsworth, 2005, [2]: 628-629). Ainsworth also claims that, as a consequence of the epistemological interpretation, real and quasi equilibriums are ontologically equivalent. But I believe that this later claim is objectionable, as quasi and true equilibrium seem to have objective physical differences.[5]

Let us imagine that someone gets into the laboratory when a spin-echo experiment is taking place. This "ignorant observer" arrives when the spins are out of phase for the very first time (that is to say, when $t = \tau$). I believe that Ainsworth correctly points at the fact that this ignorant observer would be unable to tell us if the system is in a quasi-equilibrium or a real-equilibrium state.[6]

[4]See last paragraph in section 2.3.1.

[5]See definitions of quasi equilibrium and true equilibrium on p.31.

[6]I thank Orly Shenler for proposing a similar example in a personal conversation.

Let us refer as "the demon" to the experimenter that has been inside the laboratory from the beginning of the experiment. This demon, unlike the ignorant observer, knows exactly which radio-frequency pulses must be applied to the system in order to bring the spins back to an ordered state.

After the dacay of the echo-sginal, however, even the demon is unable to reverse the spins to their original state, despite the fact s/he possess all the required knowledge required to do it. This, I believe, shows that an objective physical difference exists between what intrventionists call a quasi-equilibrium state and what they call a "true-equilibrium" sate. And this difference is independent of what the observer knows about the system. According to interventionism, as I understand it, ignorance plays no role in the actual system$\tilde{\text{O}}$s approach to equilibrium, but only in our descriptions of such approach.

3.3 Idealizations

Although interventionists reject ideal assumptions – such as the system being isolated or the system approaching to equilibrium when time tends to infinite – interventionist models themselves are not free of idealizations. Quite the contrary, classical interventionists introduced highly idealised models of the environment to account for irreversible processes (see Shenker 2001, [89]:7; or van Lith, 2001, [98]:168-169). Of course idealizations are not *per se* undesirable. The point to stress here is that interventionism aims to describe systems as we find them in Nature. It was put forward by their proponents as an alternative to other approaches that impose ideal conditions on the systems (e.g. the ergodic programme). Yet, it is hard to conceive the classical interventionist models as straightforwardly applicable to systems as we find them in Nature.

Chapter 4

The Manipulability Theory of Causal Explanation

The concept of *intervention* is conspicuous in the interventionist solution to the irreversibility problem that we have considered. As we saw Sklar even explicitly links the concept of intervention to causation. Hence, it makes sense to analyze the credentials of interventionism from the point of view of a philosophical theory of causation. Among the diverse theories and models available in the literature, the manipulability account of explanation and causation (as presented in Woodward, 2003, [106], and referred from now on as m-theory) provides the most convenient elements to analyse the approaches to irreversibility studied in this thesis. Interventionists explain the SE experiments by postulating causes and the m-theory is a theory of *causal explanations*. Besides, the m-theory is closely related to *controlling* causal factors and this perspective fits very well with the experimental context of the SE experiments. The m-theory provides us with a tool suitable to evaluate and compare different causal explanations of these experiments. Moreover, the concept of *intervention* itself is prominent and central to the m-theory.[1] For these reasons the m-theory is the right philosophical framework for this analysis. The aim of this chapter is to introduce the central ideas of the m-theory and formally define its key notions, including different types of causes in addition to the central notions of *intervention, invariance and depth*. All these notions are complex and the picture presented does not intend to be exhaustive.

[1] In statistical mechanics an *intervention* is an environmental disturbance exerted upon the system. In the m-theory, in turn, an *intervention* is a procedure that changes the value of a putative cause to test if the value of the putative effect correspondingly changes too (this concept will be formally defined in section 4.3.1). Depending on the context it will be clear when we are using one or the other.

4.1 The basic elements of the m-theory

In order to provide the m-theory of causal explanations, Woodward appeals to a conception of causation associated to the idea of manipulability. This conception lies within a group of approaches developed by several philosophers such as Collingwood (1940), Gasking (1955), von Wright (1971), Menzies and Price (1993) and Pearl (2000)[2]. Some differences between Woodward's m-theory and the other manipulability approaches will become clear throughout this chapter. Let us first compare and distinguish the m-theory in general from some well-known (no-manipulability based) accounts of explanation.

According to the Deductive-Nomological model of explanation (DN) that Hempel and Oppenheim proposed in 1948 [56], explanations are deductive arguments including at least one law of nature among its premises. The conclusion in the argument is called *explanandum* and it describes the phenomenon to be explained. The premises are called *explanans* and they must be true in order to account for the phenomenon.

In the m-theory, by contrast, only *some* explanations are deductive arguments of this kind, but not all of them. As it has been pointed out in the literature, there are commonly accepted explanations that do not appeal to laws of nature (see Salmon, 1989, [86]:46). Woodward argues that these are genuinely explanatory in spite of not agreeing with the DN-model. Accordingly, he proposes to "begin with a more general notion of causal explanation, understood in terms of manipulability, and then attempt to understand explanations that appeal to explicit chains of deductive reasoning and laws of nature as one specific variety within this genus" (Woodward, 2003, [106]: 20). Thus the m-theory considers laws but also a broader spectrum of generalizations as candidates to appear in the explanans of any causal explanation. Then, in addition to Coulomb's law or 'F=ma', other generalizations like 'Contraceptive pills avoid pregnancy'; 'Monopolies that take over a formerly competitive industry will raise prices' are perfectly apt causal regularities that can explain phenomena. 'This medicine caused Andrew's recovery'; 'Bad news about candidates in the press have negative repercussions for them in electoral results', or (referring to the famous example) 'Ink spills stain carpets' may play the relevant explanatory role that in other accounts is assigned exclusively to laws[3].

[2] See Woodward, 2003, [106], sections 1.1 and 1.6

[3] The status that Woodward assigns to laws in the m-theory will be discussed in section 4.6.2 of the present chapter.

The m-theory is also different from Philip Kitcher's Unificationist account of explanation (1989, [63]) and from Wesley Salmon's Causal Mechanical Model of explanations (1984, [85]). In Kitcher's unificationist view a scientific explanation must provide a unified account for a range of different phenomena. The fewer the facts needed to derive many phenomena, the more stringent the pattern of derivation and the more unified our explanation. The paradigmatic example of this sort of explanation is Newton's theory of gravitation which is able to account for terrestrial and celestial phenomena. In Salmon's model, in turn, a causal explanation consist in tracing the causal process that leads to the explained event; where *causal process* has a very specific definition. Namely, a process is characterized as causal if a *mark* is transmitted in a continuous way. A mark is an alteration in some feature of the process at some given stage in its evolution. A mark may be, for example, a physical deformation or an amount of energy transmitted from one to another billiard ball in a collision. And we say that such a mark is transmitted in a continuous way if, once the mark is introduced, it persists throughout later stages of the process. According to Salmon's account, a causal interaction involves a spatio-temporal intersection between two causal processes which modifies the structure of both – each process comes to have features it would not have had in the absence of the interaction (see Woodward, 2010, [108]: section 4.1). According to the m-theory neither unification (as conceived by Kitcher) nor spatio-temporal continuity (as conceived by Salmon) are necessary features in a satisfactory explanation.

The essence of a causal explanation, according to the m-theory, consists in "exhibiting a pattern of counterfactual dependence between explanans and explanandum –a pattern of counterfactual dependence of the special sort associated with relationships that are potentially exploitable for purposes of manipulation and control." (Woodward, 2003, [106]:13,16). In other words, according to the manipulability account, behind each causal explanation there is a practical component or *payoff* directly related with the aim of controlling what is being explained.

Thus, a basic idea of the m-theory is that causal explanations offered in natural and social sciences, do not aim simply at satisfying our intellectual curiosity, but are often guided by the goal of finding information potentially relevant for the manipulation and control of the explained events.

It is important to note that the notions of control and manipulation are understood counterfactually. Roughly it does not matter so much what is in fact manipulated as what would have been manipulated under the right conditions. Otherwise the m-theory would not be adequate for describing

explanations of cosmological phenomena, past events, or any other event beyond actual manipulations routinely carried out in practice.

In satisfactory causal explanations, *the patterns of dependence between causes and effects (or generalizations) are invariant under a set of interventions.* This claim contains the main notions of the m-theory that will be defined and explained in subsequent sections. Just as an illustration of the general idea let us consider an explanation of the trajectories of the planets. According to the m-theory, the explanation should provide information about what would happen if we change, manipulate or *intervene* upon those trajectories. In this case an intervention could be carried out by placing a massive planet between Earth and Mars and studying how this affects their trajectories. Clearly this is a hypothetical and not an actual manipulation. If the explanation in question is good enough, the generalization G postulated by the explanation (say Newton's universal law of gravitation) should hold under an intervention. Newton's theory of the planetary motion should provide an answer to what would happen under this particular situation intervention (which leads to a situation different to the actual one). If Newton's theory provides an answer, according to the m-theory, it has shown to be explanatorily relevant.

So, according to the m-theory, the import of a given explanation relies on its capacity to provide answers to counterfactual questions or *what-if-things-had-been-different* questions (from now on *w-questions*). In our example the *w-question* that the explanation is capable of answering is: "How would the planet's trajectories be affected by the insertion of another planet between Earth and Mars". As we will see in section 4.5 the wider the range of *w-questions* answered the *deeper* the explanation is, in Woodward's sense of 'depth'.

4.2 Causes

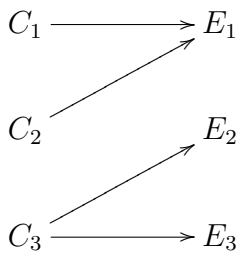
The very notion of cause in the m-theory is linked to manipulation. It is defined as follows: "C is a genuine cause of E if, given the appropriate background conditions, there is a possible manipulation of the cause C such that this is also a way of manipulating or changing the effect E" (see Woodward[106]: section 2.2, or Woodward [107]: section 1). In other words, causal relations entail some changes upon the values of E whenever the values of C are modified.[4] According to the m-theory, the manipulations carried

[4]A difference between Woodward's account and Lewis's theory of possible worlds is that counterfactuals in the m-theory are always considered with reference to the effect

out over C must be reproducible in the sense that responses to the effect E must be in some way repetitive or systematic.

A set of counterfactuals is associated with every causal relationship. When C is a deterministic cause of E, the associated counterfactuals will be of the following kind: "If C were manipulated by the intervention I, then E would experiment such and such changes". Otherwise, i.e., if C is an *indeterministic* or probabilistic cause, the associated counterfactuals will refer to $\rho(E)$, i.e. to a probability density over the outcomes of E.

Following the terminology of many other authors in the literature (among them Judea Pearl, 2000, [77]), Woodward makes use in his m-theory of directed graphs and equations to express causal relationships. Causes and effects are then represented as variables. The set comprised of those variables is denoted by $\boldsymbol{V}$. And the causal relationships are represented by arrows as shown in the directed graph 1.[5]



Directed Graph 1. Causal links between variables in $\boldsymbol{V} = \{C_1, C_2, C_3, E_1, E_2, E_3\}$.

In the m-theory equations like $E_1 = f(C_1, C_2,), E_2 = 5C_3$, etc., and the directed graphs associated with them are understood as applied to the level of the particular values of those variables. That is to say, the equation $E_n = f(C_n)$ must not be understood as the average response of the value of E_n to various average values of C_n in some population. But rather as how the value of E_n possessed by a particular individual (or object) changes in response to C_n taking the particular values $c_1, c_2, ..., c_m$ (see Woodward, 2003,

E.In Lewis' theory, by contrast, counterfactuals about what would have happened if C had not occurred in the closest possible world do not necessarily make reference to E. For more on the differences between these two theories see section 3.6 in Woodward 2003 [106].

[5]In this thesis we use the term *representation* in an ordinary way. It receives no particular philosophical interpretation.

[106]:52). An advantage of equations over directed graphs is that they enable us to represent quantitative information about the causal relationships, while directed graphs can only describe a relationship qualitatively.

Woodward classifies causes into several different kinds depending on the causal connections between the variables. It is beyond the purposes of this thesis to present and analyse them all in detail. However, we will now present the formal definitions of four kinds of causes, which are indispensable to understand the m-theory. It is worth mentioning that the different kind of causes are defined in the m-theory in terms of interventions. And the notion of intervention is in turn defined in terms of causes. This is how the m-theory is builded and cannot be avoided here. The question as to whether this circularity is vicious or not will be addressed in section 4.3.1.

4.2.1 Total Cause

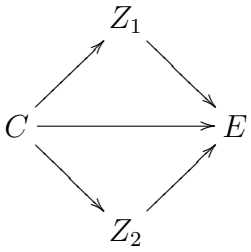
The notion of total cause is the most generic one. It captures the idea of a 'manipulable' cause in general: "C is a total cause of E if and only if there is a possible intervention I on C that will change E or $\rho(E)$."(Woodward, 2003, [106]: 51).

4.2.2 Direct Cause

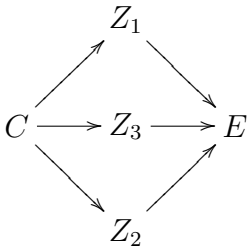
C is a direct cause of E if and only if C's influence upon E is not mediated by any other variables in the system of interest $\boldsymbol{V}$. And "a necessary and sufficient condition for C to be a direct cause of E with respect to some variable set $\boldsymbol{V}$ is that there is a possible intervention I on C that will change E or $\rho(E)$ when all other variables in $\boldsymbol{V}$ besides C and E are held fixed at some value by interventions" (Woodward, 2003, [106]: 55).

The directed graphs 2 and 3 (below) show how this necessary condition for direct causes would be violated if there were one or more intermediate variables between C and E. In both graphs C is a total cause of E because changes in the values of C would entail corresponding changes in the values of E. However C is only a direct cause in graph 2 –it fails to satisfy the conditions for a direct cause in graph 3. This is because for C to be a direct cause, when an intervention is performed on the system, the values of all the variables except C and E must be held fixed. This guarantees that the resultant change in E is not being produced by alterations in the value of any other intermediate variable. In the directed graph 2 fixing Z_1 and Z_2

does not imply that the value of E ceases to change as a response to changes in the value of C. By contrast, in the directed graph 3, fixing Z_1, Z_2 and Z_3 would block all the paths and hence would make it impossible to influence the value of E through manipulations of C.



Directed Graph 2.
C is a direct cause of E with respect to $\boldsymbol{V} = \{C, E, Z_1, Z_2\}$.



Directed Graph 3.
C is not a direct cause of E with respect to $\boldsymbol{V} = \{C, E, Z_1, Z_2, Z_3\}$.

4.2.3 Contributing Cause

According to the m-theory, if C is a contributing cause of E with respect to the variable set $\boldsymbol{V}$, then:

(CC-i)

> There is a directed path from C to E such that each link in this path is a direct causal relationship. That is, the intermediate variables along this path $Z_1, Z_2, Z_3, \ldots, Z_n$, are such that C is a direct cause of Z_1, which is a direct cause of $Z_2, \ldots, Z_{n-1}$, which is a direct cause of Z_n, which is a direct cause of E (see Woodward, 2003, [106]:57).

(CC-ii)

> And there is an intervention I on C that changes the value of E when all other variables in $\boldsymbol{V}$ that are not on this path, if any, are fixed at some value. If there is only one path from C to E, then C is a contributing cause of E as long as there is some intervention I on C that will change the value of E, for some values of the other variables in $\boldsymbol{V}$(see Woodward, 2003, [106]:59).

As the directed graph 3 shows, C and Z_3 are contributing causes of E. Besides, every contributing cause is in turn a directed cause of the intermediate variables in the path (C directly causes Z_3, and Z_3 directly causes E).

4.2.4 Actual Cause

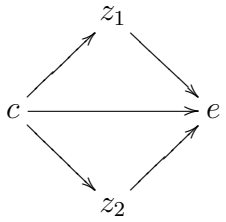
Both the notions of direct cause and contributing cause are type level causes. Woodward complements the m-theory by defining the following notion of actual cause, which applies to token level causes.

$C = c$ is an actual cause of $E = e$ if and only if two conditions are satisfied:

(AC-i) The actual value of $C = c$ and the actual value of $E = e$

(AC-ii) There is at least one route from C to E for which an intervention on C will change the value of E, given that other direct causes of Z_i of E that are not on this route have been fixed at their actual values (Woodward, 2003, [106]:77).

As the following directed graph shows, the main difference between the notion of actual cause, and the other notions of causes defined above, is that the definition of actual cause makes reference to the specific values actually taken by each variable.



Directed Graph 4.C=c is an actual cause of E=e when the actual values of Z_1 and Z_2 are z_1 and z_2.

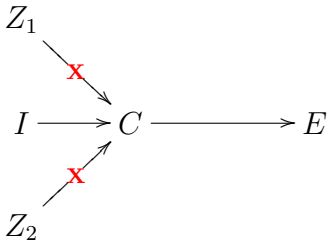
The above-presented definitions (4.2.1-4.2.4) tell us, in sum, that according to the m-theory, causal claims and hence causal explanations entail counterfactuals about what would happen under *interventions*. For that reason we turn now to assessment of the notion of intervention.

4.3 Interventions

I will discuss some general features of the notion of intervention before formally defining it. In the m-theory interventions are understood as "exogenous

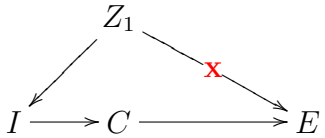
causal processes that change the value of the putative cause C in such a way that if any change occurs in the effect E, it occurs only in virtue of E's relationship to C and not in any other way" (see Woodward, 2003, [106]:47). This allows us to determine whether or not E is caused by C. As mentioned before, interventions are taken as performable *in principle* or hypothetically. They intend to capture the ideal experimental conditions that should be fulfiled to change the value of C and so study its causal link with E. The notion of intervention proposed in the m-theory aims to provide the maximum plausibility for the notions of total, contributing and actual causes. Keeping this in mind, let us consider the conditions that interventions should meet.

Firstly, we would like the intervention I to absolutely control C, in the sense that changes in C's values are not generated by the influence of any factor other than I. This means that I must "break" or "switch off", in the sense of making ineffective, the causal links between C and any other of its causes (see directed graph 5).



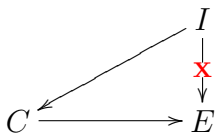
Directed Graph 5. Intervention I *switching off* alternative causes of C

Secondly, E must not be affected by the process that brings about the intervention I (see directed graph 6).



Directed Graph 6. The process Z_1 generating I must not affect E through alternative causal routes.

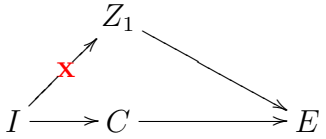
Thirdly, the intervention I must not be a direct cause of E, but can only affect it through C. In other words, all causal paths from I to E must pass though C. A direct consequence of this condition is that I and C cannot both be causes of E (see directed graph 7).



Directed Graph 7. The intervention I must not directly affect E.

Finally, the changes in E should be the result of the intervention I acting on C and nothing else. This turns interventions into relational variables. Namely, an intervention I upon C is always defined with respect to the effect variable E.

For example, we would not like the change in E to be a consequence of alterations in any other variable Z_i through an alternative causal route. That is to say, we want the intervention to act only through the path connecting C and E. One way to avoid this is to fix the values of all the variables Z_i causally connected with E except those in the path from C to E. (see directed graph 8).



Directed Graph 8. Intervention I must not affect E through an alternative route.

Taking all these conditions into account *interventions* can be defined. It must be clear that, as Woodward notes (see Woodward, 2003, [106]:114), the conditions for a manipulation of C to count as an intervention are proposed in the m-theory as *regulative ideals* rather than absolutely necessary conditions to obtain any knowledge about causal links. In other words, the failure of any of these conditions does not mean that nothing can be inferred about the causal relationship between C and E.

4.3.1 Formal definition of intervention

The notion of intervention is formally defined in the m-theory as follows (Woodward, 2010, [108]: section 5; and Woodward, 2003, [106]: 98) .

(IN)*I*'s assuming some value $I{=}z_i$ is an intervention on C with respect to E if and only if $I{=}z_i$ is an actual cause of the value taken by C, and I meets the following conditions:

(IN-i) I must be the only cause of C; i.e., the intervention must completely disrupt the causal relationship between C and its previous causes so that the value of C is set entirely by I.

(IN-ii) I should not itself be caused by any cause that affects E via a route that does not go through C.

(IN-iii) I must not directly cause E via a route that doesn't go through C.

(IN-iv)*I* leaves the values taken by any causes of E except those that are on the directed path from I to C to E (should this exist) unchanged.

It should be pointed out that the conditions (**IN i-iv**) define interventions in causal terms. However, the m-theory also defines causes in terms of interventions (see definitions of *total, direct, contributing* and *actual causes* in section 4.2). Therefore, the m-theory is not an analysis of causation that reduces causal to non-causal claims. However, Woodward argues (Woodward, 2003, [106]:22) that his theory is not viciously circular since the elucidation of whether a particular cause C causes an effect E does not presuppose information about whether there is a causal relationship between C and E but rather information about other causal relationships, in particular the one between the intervention I and the values of the variable C.

It is also worth noting two features of the definition **(IN)** that go in line with our comments in the previous section. Firstly, human agency, limitations or capacities are not mentioned at all, meaning that interventions may perfectly be natural or hypothetical. And secondly, interventions are defined not only in relation to the causal variable C but in relation to the effect variable E too. So what counts as an intervention is entirely given by the relationship that objectively obtains between the variables. These two features distinguish Woodward's m-theory from other manipulabilitist accounts as von Wright's (1971,[103]) or Menzies-Price's (1993,[71]).

4.3.2 Intervening: Some Illustrations

Now that interventions have been defined, some comments are in order regarding which variables are suitable for intervention. As mentioned before, in the m-theory interventions are understood as changes in the values of the putative causes that fulfil certain conditions. Nevertheless, there may be cases in which the very meaning of 'changing the value' of one variable is ill defined. As an example of this Woodward proposes the following claim:

> (G_L): "No physical object can travel at a velocity greater than light" (Woodward, 2003, [106]:122).

(G_L) represents a recognized scientific claim. What is not clear is whether it represents a causal claim in the m-theory terms. The set of relevant variables in this case is $\boldsymbol{V} = \{O, C\}$ where O=physical object, C=velocity of light; and the causal relationship in (G_L), if any, links O and C. The problem is that changing the values of variable O trough an intervention I makes no sense, as it seems odd to prevent an object from 'being a physical object'. Therefore, according to the m- theory, (G_L) cannot count as a *causal* claim.

In some other cases, changing the values of variables does not have a clear meaning in the original claim. Nevertheless, a simple reformulation of the claim and the right election of variables solves the difficulty. As an example of this let us analyse the following claim in terms of the m-theory:

> (G_G): "Being female causes one to be discriminated against in salary" (Woodward, 2003, [106]:115).

This claim can be expressed as the causal relationship $X \longrightarrow D$ where $\boldsymbol{V} = \{X, D\}$; D=discrimination against in salary, and there are at least two possible causes X_1 and X_2. X_1 is the worker's gender, and an intervention I on X_1 would be a change of sex. X_2 is the employer's beliefs about the worker's gender, and an intervention in this case would simply consist in changing those beliefs.

Now let us consider the range of possible interventions. Literally changing someone's gender in order to study the causal relationship in (G_G) seems impracticable. Even if worker's gender may be a cause of salary differentials, we cannot assess the causal claim containing X_1 by means of the m-theory. So the first possible cause X_1 may be tagged as inconvenient. It is much more

reasonable, for example, to consider interventions on the employer's beliefs about gender by manipulating the information in job applications. There is a well defined intervention acting on X_2 and, if indeed changes in values of X_2 are correlated with changes in values of D, then there is, according to the m-theory, a genuine causal relationship. Hence, the directed graph $X_2 \rightarrow D$ correctly describes a causal relationship in (G_G) in accordance with the m-theory.

What is remarkable about this example, is that "it matters a great deal to the interpretation and truth conditions of the causal claim which hypothetical manipulation we have in mind" (Woodward, 2003, [106]: 117). In other words, the outcome of the manipulabilist analysis of a given causal relationship may change depending on the interventions we take into account. In this example of gender discrimination it is quite clear that X_2 is preferable to X_1. Interpreting X as X_2 is convenient both because it offers more plausible interventions and because it goes along with the original intention of the claim (G_G). However, analysing more complex claims of actual scientific explanations the election between alternative interpretations X_1, X_2, X_3 or X_i and alternative interventions I_1, I_2, I_3 or I_i may not be so obvious.[6]

Yet, consider a different case. Suppose the intervention I is so strong that it affects the very relationship between the putative cause and effect. Let us imagine for example that we hang only extremely heavy weights on a string exceeding its resistance. Then we could conclude that the causal relationship expressed by Hook's law $(F = -kx)$ does not hold. This is because every time we hang a weight the string literally breaks and the law is not fulfiled. But this false conclusion is obtained from interventions in the wrong range that brought about exceptions to the causal relationship.

Conversely, it may be the case that the intervention I fails to change the values of the variable that it is supposedly affecting. As an illustration, imagine a switch on a 360° that turns on a radio when we turn it beyond 90°. The set of relevant variables is $\boldsymbol{V} = \{A, R\}$ where A= angle (0° - 360°) and R=radio (0=off, 1=on). If interventions are performed to change the value of A up to 60°, we may wrongly conclude that A is not causally related to R. Indeed, one would observe exactly the same results if the radio was broken. But this is only due to the fact that the interventions are setting the values of I within the wrong range of C's values.

[6] This will be crucial in our philosophical analysis of some particular causal explanations in chapters 5 and 6.

These third and fourth examples serve to point out that causal relationships only hold under a valid range of interventions.

The morals of all these examples can be summarized as follows. Firstly, if for a given claim "C causes E" there are no well defined interventions that would change the value of C, then according to the m-theory the claim is not genuinely causal. Secondly, how variables are chosen and how we interpret them may affect our results about the causal relationships expressed in a given claim. And finally, causal relationships are not necessarily invariant under *any* possible intervention. Quite the contrary, they are typically invariant only under some restricted set of interventions, as well as restricted range of values of such interventions.

If analysing a putative causal relationship $C \rightarrow E$ we find out that every single proposed intervention I_i disrupts the causal relationship of C and E, even after all kinds of variations in the range of interventions (that is, even hanging on the string small weights and/or turning the switch all the way from $A = 0°$ to $A = 180°$) then, according to the m-theory (see Woodward, 2003, [106]:110), the causal relationship is not genuine.

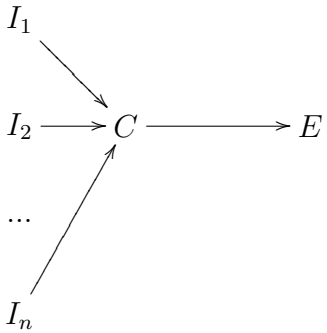
All these examples together illustrate how the notion of intervention is associated with a set of background conditions. So it is essential, in the m-theory, to spell out these conditions and the valid range of possible interventions every time that a causal relationship is postulated.

4.3.3 Possible Interventions

Let me turn now to an assessment of whether interventions are possible in some other relevant senses. It may be the case, for example, that interventions conflict with other interventions. Imagine that an intervention is physically possible, but there is no way of "switching off" an alternative cause of C (i.e., it is not possible to fulfil **IN-i**). As an illustration suppose we study the effect of the Moon over the oceanic tides. In order to inquire into the causal relationship between Moon and tides we intervene on our system by hypothetically introducing a second satellite and observing the change in the tides produced by this intervention. So far, everything fits perfectly with the m-theory. Nevertheless, the introduction of the second satellite also affects the trajectories of the Moon and the Earth. And those are in turn causal factors acting on the tides as well.

Here is another example which questions the possibility of performing interventions. According to the definition **IN**, interventions I must be themselves causes acting on the putative cause C. Thus, a particularly interesting challenge for m-theory is posted by cases in which the alleged cause C does not itself have any causes – and therefore has no possible interventions.

In other words the m-theory requires that in order for C to be causally associated with an effect E, it must also have itself a set of associated causes I_i such that $I_i \rightarrow C \rightarrow E$ (see directed graph 9). But, intuitively, this should not be relevant to the problem of whether C causes E.



Directed Graph 9.

The challenge, in terms of the directed graph 9, is whether the m-theory is capable of telling us something about the causal link between C and E when there are no such causal factors $I_1, I_2, I_3, \ldots I_n$.

In order to face both examples (the "second moon" example, and the "non-causal C" example) Woodward drives our attention to the fact that two conditions –indeed two original motivations for introducing the notion of intervention– are met in each of these examples. Firstly, "there is a basis for claims about what will happen to E under an intervention on C; i.e. we can associate some well-defined notion of change with C, and we have some grounds for saying what the effect, if any, on E would be of changing just C and nothing else. [Secondly] "there is a way of disentangling the effect on E of changing just C from the effects on E of changes in other potentially confounding variables" (see Woodward, 2003, [106]:131). Furthermore, in the second example (in which C itself has no causes), there is an advantage in comparison with many other cases. Namely, we do not even have to worry about possible causal processes of C that, in violation of requirement **IN-iii**, are also directly affecting E. In Woodward's view, all these considerations give us a purchase on whether the counterfactual "if C were to be changed by an intervention I, then E would change" is true or false (see Woodward, 2003, [106]:132).

The essence of Woodward's response to the two examples may be summarized as follows. It is a fact that for both of the examples it is physically impossible to perform interventions. This physical impossibility is not the usual impossibility related to practical limits, but a *stronger impossibility* related to the fact that every intervention I upon C would violate condition **IN-i**. Even in those cases in which interventions are physically impossible in this strong sense, Woodward believes that we can legitimately use counterfactuals to elucidate causal claims along the lines suggested in the manipulability theory [m-theory]. (See Woodward, 2003, [106]:132).

Taking into account all the examples discussed in this chapter we may conclude that there are so far only a few rare cases in which m-theory declares that the putative causal claim is not genuine. In one of these cases the causal link between C and E is broken by every single intervention on C, independently of the range of values taken by those interventions. In another, *changing the values of the variable* C has no meaning, even reformulating the claim or proposing several interpretations of the causal variables involved in the directed graph, as it happens in the "physical object" example.

4.4 Invariance

The notions of causation, explanation and invariance are closely intertwined in the m-theory. According to this theory, a generalization G –i.e. a relationship expressing a putative causal connection between two variables– counts as causal or explanatory if and only if it is *invariant* under some appropriate set interventions (see Woodward 2003, [106]:15 and 239). Like the notion of intervention, the notion of invariance is modal in the sense that it tells us whether a putative causal relationship would remain stable if, *perhaps contrary to fact*, certain changes or interventions were to occur.

Accordingly, defining the set of relevant variables $\boldsymbol{V}$ and the relations among them provides in Woodward's view no more than an *incomplete* description of the content of a given causal claim. In order to complete the description it is also necessary to specify the background conditions and the range of interventions under which the postulated relations are preserved. Analogously, one cannot assert that a directed graph correctly represents a system of causal relationships unless one shows that those causal relationships are *invariant* under intervention.

The following example may serve to introduce the notion of invariance. Suppose that the generalization G figures in a given explanation and G correctly describes the pattern of correlation between a cause and its effect. For example, let G be Hook's law $F = -kx$ (where $\boldsymbol{V} = \{F, x\}$ and the force F is the cause of the elongation x). In the m-theory a *testing intervention* is defined as an intervention I that changes the value of F (from f to f^*) in such a way that $f \neq f^*$ and hence $G(f) = x \neq G(f^*) = x^*$ (Woodward, 2003, [106]:250) –remember that we are assuming that G is the correct description of the causal connection between F and x. So, in our example, the *testing intervention* may consist in hanging different weights exerting forces of different magnitudes upon the string. Under these circumstances we can say that G is invariant under the intervention I if and only if G correctly predicts the change in the value of the elongation x (from x to x^*) given the change in the value of F brought about by the intervention I. In other words, according to the m-theory Hook's law is invariant if and only if it correctly predicts the elongations corresponding to the different forces experimentally applied upon the spring. So, the general definition of invariance may be stated as follows.

4.4.1 Definition of invariant generalization

A generalization G relating changes in the cause C (from c to c^*) to changes in the effect E (from e to e^*) is invariant under a testing intervention I if and only if it correctly describes what the new value of E, e^*, would be under this change; that is, if and only if it remains true that $G(c^*) = e^*$ (see Woodward, 2003, [106] section 6.2).

It is worth emphasising that the notion of invariance so defined is *relative* to interventions. And, given that interventions are so essential to define whether a given generalization G is invariant or not, some remarks are in order regarding what kind of interventions we can work with.

We should not identify invariance and interventions in general. For, on the one hand, G may turn out to be *invariant* under a manipulation that is not an intervention (i.e. a manipulation that fails to meet at least one of the conditions **INi-iv**). On the other hand, we may find the converse case in which G turns out to be *non-invariant* under an intervention. Let us consider these two cases in turn.

If a given generalization G is invariant under a non-intervention, according to the m-theory, this fact must simply be considered irrelevant as far as G's explanatory relevance is concerned. In the m-theory, only manipulations corresponding to *interventions* are considered relevant to *test* the explanatory import of generalizations that figure in explanations. In this sense, among all possible manipulations that may change the value of C, only those meeting the conditions **INi-iv**, namely, only interventions, play a privileged role.

In the other case, i.e., if G turns out to be *non-invariant* under a given intervention I, we should not conclude from this fact that G is not invariant and not causal at all. G may "break down" under *this particular* intervention I but may, nevertheless, have some explanatory import if it is invariant under other interventions. As mentioned in previous examples (section 4.3.2), in order to consider a generalization G explanatorily relevant it is not necessary for G to be invariant under *all* possible interventions, but only under a set of *some* interventions.

4.4.2 Degrees of invariance

Probably the most important consequence of Woodward's relativization of the notion is that invariance becomes a matter of degree.

We may associate a generalization G with the set of interventions under which it remains invariant. Some generalizations are invariant under one single intervention (let us call the set containing that intervention S_1), others are invariant under a few interventions (call the set of those interventions S_2), and yet some others are invariant under a set containing numerous and diverse interventions (call the set of those interventions S_n).[7]

Let us imagine that a given generalization G_1 is invariant under a set of the interventions S_1, and another generalization G_2 is invariant under a set of interventions S_2. If S_2 has more elements than S_1, then one may be tempted to conclude that G_2 is *more invariant* than G_1. However, this conclusion is not straightforwardly true. If, for example, S_1 is a proper subset of S_2 (i.e, if $S_1 \subset S_2$) then it is true that G_2 is more invariant than G_1. But it may also happen, in a different case, that although G_1 and G_2 are supposed to explain the same phenomenon, S_1 and S_2 have no common elements. For example, if G_1 is Newton's law of Gravitation and G_2 is Einstein's relativistic expression for the planetary orbits, the interventions in the corresponding sets S_1 and S_2 may be all different. Therefore we cannot decide whether G_2 is more invariant than G_1 by simply appealing to the number of elements (interventions) in S_1 and S_2.

In addition, one must be aware that operations mathematically valid for sets may lead to the wrong conclusions regarding invariance. For instance, the invariance of generalizations is not an additive property, in the sense that the conjunction of the two generalizations ($G_1 \wedge G_2$) is not necessarily invariant under the union of the sets of interventions under which each set remains invariant ($S_1 \cup S_2$). Similarly, we can not extract logical conclusions from the relations between sets of interventions. For example, from the fact that S_1 is a proper subset of S_2 ($S_1 \subset S_2$) we cannot infer that G_2 is logically derivable from G_1.

The m-theory tells us that, once we are aware of all these considerations, we can legitimately speak of some generalizations as more invariant than others in the sense that they are invariant under *a larger or more important* set of interventions (see Woodward, 2003, [106]:257, my emphasis). And in order to make clear whether one set of interventions is *larger or more important*

[7] The sub-index does not correspond to the number of elements in each set.

than another, one needs to know in detail the background conditions, what each of the interventions consist in, and the range of values under which each intervention is performed.

4.5 Explanatory Depth

4.5.1 The notion of explanatory depth

According to the m-theory, generalizations that are invariant under a larger and more important set of interventions can be used to provide better explanations. Hence the degree of invariance is directly related to the notion of *explanatory depth*. This notion is introduced in the m-theory using the following example (formulated by Haavelmo, 1944, [47]: 27-28). Suppose that two different explanations are offered about how to increase and decrease the speed of an automobile. The simplest explanation relates the speed to the distance of the gas pedal from the bottom of the car. And a second and more elaborated explanation details the whole inner mechanism of the car, tell us how the motor and the carburetor work and so on.

Both explanations are successful in terms of providing sufficient information to operate the vehicle. Besides, both explanations are valid under the criteria of the m-theory because they identify patterns of counterfactual dependence that enable to control the speed of the automobile. In the first case, the postulated relation between speed and the gas pedal remains invariant under some interventions, for example, under changes in the values of the pedal inclination. However, this relation postulated by the simplest explanation fails if the car runs out of petrol, or if any element inside the car does not work properly. The second explanation, in contrast, is invariant under interventions on any part of the mechanism. And thus it answers a larger number of w-questions (questions about what would happen under circumstances different to the actual ones). Among other things, it explains why the car does not accelerate if the gas tank is empty. We then say that, in m-theory terms, the second explanation is *deeper* than the first one.

The notion of *explanatory depth* is not properly defined in the m-theory in the sense that the theory does not provide us with a set of necessary and sufficient conditions for this notion. Nevertheless one may identify three criteria that the m-theory takes into account in order to assess explanatory depth. I turn now to specify those three criteria.

4.5.2 Three criteria of explanatory depth

Criterion 1. A causal explanation is *explanatorily deep* if the generalization figuring in the explanation is invariant under a *wide range of interventions.* (see Woodward, 2003, [106]: 311).

Let me illustrate criterion 1 in terms of a variation of the car's example. Suppose we define the variables that figure in the first explanation as $\boldsymbol{V}=\{P, A\}$; where P=pressing the pedal, and A= acceleration. Let us say that these are two-valued variables. That is, if the pedal is pressed then P=1, and if the pedal is not pressed P=0. Similarly, A=1 if the car accelerates and A=0 if the velocity of the car is constant. And suppose we have an alternative explanation (let us call it the third explanation) that also appeals to the causal relation between P and A but, instead of defining P and A as two-valued variables, it defines them as multi-valued variables. In the third explanation P can take any real value in the possible range of the pedal's inclination, and A can take any real value between $1m/s^2$ and $1000m/s^2$. Criterion 1 would then tell us that the third explanation appealing to multi-valued variables is *deeper* than the explanation appealing to two-valued variables. The reason is that the third explanation shows "how any one of a great number of changes in their explanans variables will lead to one of many possible changes in their explanandum variables. In other words, [it] gives us information about a much more detailed and fine-grained quantitative pattern of counterfactual dependence than the "binary" pattern".(Woodward, 2003, [106]:206). This means that the third explanation allows us to perform more interventions upon P for interventions are changes in the values of P, and there are many more ways to intervene upon the multi-valued variable P than upon the two-valuated P.[8]

Criterion 2. A causal explanation is *explanatorily deep* if the generalization figuring in the explanation is invariant, not only under a wide range of interventions (criterion 1), but also under a wide *variety of different kinds of interventions* (see Woodward, 2003, [106]: 211, 215).

Criterion 2 may be illustrated with the original example of the car. In fact, Woodward (2003, [106]:259-260) uses this example to illustrate that the second and elaborated explanation is *deeper* than the simple one because the generalizations figuring in the elaborated explanation are invariant under a wider variety of different interventions. In addition to being able to inter-

[8]For more details see Woodward's example of retrolental flbroplasia in[106]:120 and 206.

vene upon the pedal in many ways, the elaborated explanation allows us to perform many new interventions (upon the motor, the carburetor, the petrol tank, etc.) that were not considered in the first and simple explanation. Therefore, according to the m-theory, it is a *deeper* causal explanation of the car's acceleration.

This invariance under a variety of kinds of interventions is reflected, in turn, in the diverse and complex counterfactuals associated to that explanation. According to the m-theory in order "to elucidate certain kinds of causal claims, including claims about direct causal relationships and singular causal claims, one must appeal to counterfactuals with complex antecedents –counterfactuals that describe what will happen under combinations of manipulations or interventions, rather than under single manipulations" (see Woodward, 2003, [106]: 21).[9]

Criterion 3. A causal explanation is *explanatorily deep* if it is able to answer a wide range of counterfactual questions about the conditions under which the explananda would have been different (see Woodward, 2003, [106]:191). In other words, *the deeper the explanation, the wider the range of of what-if-things-had-been-different questions it answers* (see Woodward[106]:311).

Again our original example of the car illustrates how the second explanation is deeper than the fist one. The second explanation is able to answer what would happen if the carburetor breaks, if the petrol does not flow from the tank to the motor and so on. Whereas the first explanation is able to answer one single w-question, namely, what would happen if the pedal is not pressed.

In the m-theory these three criteria seem to come along together. For example, Woodward comments: "A deeper explanation for the behaviour of the car would need to appeal to [generalizations and] engineering principles that are invariant under a much wider range of changes and interventions. *Not coincidentally* such a deeper explanation could be used to answer a much wider range of w-questions".(Woodward, 2003, [106]:260, my emphasis). I have stressed the expression 'Not coincidentally' because it let us see that, according to the m-theory, a causal explanation that meets the first two criteria will also meet criterion 3. The following passage also suggest that the three criteria of explanatory depth come along together: "Some generaliza-

[9]This makes the m-theory distinct form other manipulability approaches that only consider one single manipulation for every given counterfactual.

tions are not invariant under any (testing) interventions at all and hence are nonexplanatory. Other generalizations are invariant under some testing interventions (and answer some *w-questions*) and hence are above the threshold of explanatoriness, although they are less invariant and answer a narrower range of *w-questions* than others and hence are less explanatory (Woodward, 2003, [106]:369).

A consequence of a given explanation meeting criteria 1, 2 and 3, according to the m-theory, is that the explanation will be relevant to the manipulation and control of the explained event. And this is precisely what the second explanation of the car's acceleration achieves. And, from the manipulability theory, this is what any causal explanation should aim for.

Now that we have made clear what the best scenario is (namely, fulfiling the three criteria), let us turn to a not-so-good case. Let us see what the minimum requirement is for an explanation to count as a causal explanation in the m-theory. From the previous sections we can already infer what this requirement consists in. The requirement is that, if G is the generalization figuring in the explanation that relates the event with its alleged cause, G must be invariant under *some* intervention. As we already know, this implies that the intervention fulfils the conditions **INi-iv** (otherwise it would not count as an intervention); that the intervention is performed within the valid range of invariance of G.[10]; and that G meets the conditions of invariance (in 4.4.1). Woodward integrates all these ideas and expresses them more formally in the following minimal condition for successful causal explanation.

4.5.3 Minimal condition for successful causal explanation

Suppose that there is an explanandum consisting in the statement that some variable E takes the particular value e. Then an explanans for that explanandum will consist of (a) a generalization G relating changes in the values of a variable C and changes in E, and (b) a statement (of initial or boundary conditions) that the variable C takes the particular value c. The necessary and sufficient conditions for the explanans to be (minimally) [causally]explanatory with respect to the explanandum are that:
(i) Both the explanans and the explanandum are true or approximately so;
(ii) According to G, E takes the value e under an intervention I in which

[10]In the sense explained in section 4.3.2 with the radio example.

C takes the value c;

(iii) There is some intervention [some *testing intervention*] that changes the value of C from c to $c*$, where $c_0 \neq c*$ with G correctly describing the value that E would assume under this intervention, where $e \neq e*$ (see Woodward, 2003, [106]:203).

This minimal condition is simply telling us that in order to provide a causal explanation one must present a generalization G (relating the putative cause C to the effect E) and show that there is *at least one* intervention I under which G is invariant. It would be much better if we show that G is invariant under many interventions. But showing only one is sufficient for our explanation to be considered, in the m-theory, as minimally successful. If a generalization is not invariant under any intervention, then it will fail to qualify as invariant or explanatory in the m-theory. In that case both the generalization G and the putative explanation associated to G fall below what Woodward calls 'the threshold of explanatoriness' (see Woodward, 2003, [106]:368).

Suppose that we attempt to explain a physical event (for example, the occurrence of spin-echo signals) but we are not content with our explanation only fulfiling the minimal conditions (4.5.3). What should we do if we want our explanation to do even better or to be the best? What makes one causal explanation better than another? One answer may be that 'better' means 'deeper' in the sense described in the last section 4.5.2. It is desirable that our postulated generalization G remains invariant under more than one intervention, or even better, under a wide range of interventions that vary in intensity (in agreement with criterion 1) and kind (in agreement with criterion 2). Under those circumstances, according to the m-theory, we will be able to answer a wider range of counterfactual questions also called w-questions (in agreement with criterion 3).

In sum, the higher the degree of invariance, the better control and manipulation and thus the deeper explanation is provided. This constitutes a tool that the m-theory provides for *comparing* the explanatory relevance of different explanations.[11]

[11]This tool constitutes yet another reason to choose the m-theory as the convenient philosophical framework for assessing the explanations of the SE experiments. I thank my internal examiner Federica Russo for driving my attention to this point. For recent alayses of the advantages of applying Woodward's m-theory to other cases in social and natural sciences see Russo (2009, [84]) and Suárez & San Pedro (2011,[93]).

4.5.4 Change-relating generalizations versus subsuming generalizations

Generalizations relating changes in the values of a given cause C to changes in the value of a given effect E (call them *change-relating* generalizations), play a prominent explanatory role in the m-theory compared with subsuming generalizations of the kind: "All A's have the property B". Woodward argues that this is not a mere preference, but it is rather the result we obtain when we apply his criteria of explanatory depth. It turns out that generalizations like "All A's are B's" are usually invariant under a reduced set of interventions.

This point can be illustrated by means of the following example. Suppose that somebody asks why the raven in the window is black. The first answer provided is "Because all ravens are black", and a second answer is "Because there is a gene that contains information about the color of the birds' feathers and particularly in ravens it corresponds to black color. Should this gene be modified by genetic engineering, the color of a raven could be different".

The second explanation not only provides a relation between two variables (genetic information and color of brids' feathers), but it additionally provides us with a variable suitable for intervention. It offers information about the causes of the explained event and specifies a mechanism (in this case, genetic manipulation) for changing the value of the variable "color". The significant point is that it explains both why the raven is black and the circumstances under which a raven could be "non-black".

Changing the gene that defines the color of a raven by genetic modification is an intervention under which the first explanation fails but the second one still holds. Therefore the second explanation is *deeper* than the first one in the terms of the m-theory.

This explanatory irrelevance of generalizations like "All A's are B's" may come as a shock to those who consider that many recognized laws of nature can be expressed in that form. More precisely what is shocking is the following. First Woodward tell us that laws are to be conceived in his m-theory as highly invariant generalizations, that is, as generalizations that remain invariant under a wide set of interventions. If we conceive laws as subsuming statements of the kind "All A's are B's", following Woodward we would believe that those statements are highly invariant too. However, Woodward argues that normally subsuming statements of that kind are lowly invariant. Woodward saves the m-theory of this apparent contradiction by rejecting the

subsuming conception of lawfulness. In the next (and last) section of this chapter (4.6) the conception of lawfulness defended in the m-theory will be clarified.

4.6 Solving old problems

One advantage of the m-theory is that it offers a solution to some old problems of Hempel's Deductive-Nomological Model. On the one hand, the DN-model was criticized for being to permissive in the sense of including explanations that nobody would accept; at the same time, the model was also indicted for excluding widely accepted explanations that simply do not appeal to laws, implicitly or explicitly.[12] Woodward presumes that his m-theory solves both problems, and the solutions will be separately explained in the following two sections. The debate about solving DN-model's weaknesses has been very prolific in the last decades and it goes beyond the objective of this thesis to enter its details. Each solution will be simply sketched stressing only the points that will be useful for our purposes in further chapters. (particularly in chapter 6).

4.6.1 Excluding explanatorily irrelevant factors

The m-theory correctly rejects some unsatisfactory explanations that the DN-model was not able to rule out. The famous example about Mr. Jones' pregnancy status serves as an illustration. In that example it is alleged that the reason why Mr. Jones is not pregnant is that he has been taking contraceptive pills. This bogus explanation can be expressed in terms of the m-theory as follows:

$\boldsymbol{V} = \{CP, MP\}$ where CP=ingestion of contraceptive pills and MP=male pregnancy; and $G_P : CP \rightarrow MP$ (In other words, the putative generalization is "Contraceptive pills causally prevent male pregnancy").

In accordance with the m-theory, if the generalization G_P represents a genuine causal relationship, then changes in the values of the putative cause CP should be accompanied by changes in the putative effect MP. However, as we know, changing wether a male takes birth pills does not generate any

[12] For a detailed discussion about the DN-model see Salmon, 1989, [86]. Some of the counterexamples to the DN-model presented in Salmon's book will be discussed here

change in the value of MP (male pregnancy), which is always null $MP = 0$. This means that there is *not even one* intervention under which the specific generalization G_P is invariant. And therefore, according to the m-theory, one cannot appeal to G_P to provide a causal explanation. In other words, this explanation about Mr. Jones' pregnancy is not a good explanation because does not provide information about how changes in the level of the cause may lead to changes in the value of the effect (in this case male pregnancy MP). This is to be contrasted with the second explanation in the raven's example (in section 4.5.4).

4.6.2 Dissolving the dichotomy law versus accident

The m-theory rejects the dichotomy according to which generalizations must be laws or otherwise they are merely accidental (see Woodward, 2003, [106]: 257). Instead of using this common 'law vs accident' dichotomy the m-theory suggests to talk about generalizations with different *degrees of invariance*. As a consequence, claims usually considered as laws of science in other approaches, in the m-theory are treated as highly invariant generalizations. Their high level of invariance is often reflected in the stability that the functional relations implicated in the generalization show when changing the values of the variables they relate. Just to mention an example, the universal law of gravitation remains invariant under a wide range of changes, both in the distances between the bodies and in the size and mass of the bodies implicated in the gravitational attraction.

This treatment of laws of science as strongly invariant generalizations, so the argument goes, enables us to dissolve the classic dichotomy of law versus accident within the m-theory. Therefore, the relative notion of invariance defined in the m-theory allows conceiving intermediate possibilities between laws and accidents (see Woodward [106]: 240). And this, in turn, opens new possibilities regarding the functions that generalizations play in the explanations. Accordingly, in the m-theory the only feature that ultimately serves to decide whether a given generalization G is causal and explanatory relevant is its degree invariance.

All this is compatible with Woodward's rejection of the traditional criteria of lawfulness. Among the criteria that he rejects it is worth mentioning the conception of laws of science as "exceptionless generalizations"; laws of science as "universal conditional statements"; and laws of science as "statements that make no reference to particular objects". Among the traditional criteria, the

only one that Woodward considers to be appropriate conceives laws of science as "supporters of counterfactuals"[13].

The continuum relative to invariance under interventions is a distinctive element of the m-theory. Other approaches appeal to notions like *robustness* (M. Redhead, 1987, [81]) or *stability* (S. Mitchell, 1997, [73]) which do not accommodate degrees of invariance. In Woodward's view, those approaches still face the dilemma of tracing a demarcation criterion between lawful and accidental generalizations; a dilemma that the m-theory resolves satisfactorily.

4.6.3 Does the m-theory account for every single successful explanation?

This section would be incomplete if we did not mention that Woodward's m-theory could also be criticized for not including explanations that are widely accepted in actual scientific practices.

For example, explanations about the dimensionality of space-time do not exhibit patterns susceptible of testing either by real nor hypothetical interventions. However, they seem to play an explanatorily relevant role. Also the claim "All bodies travel at velocities lower than light" seems explanatory but the m-theory classifies it as a generalization over which interventions are *ill defined* (see the first example presented in 4.3.2). Hence, this claim does not even provide a minimally acceptable explanation in the m-theory terms.

Similarly, many claims used in taxonomy, geology or mathematics, like "Reptiles have no long extremities"; "All volcanic rocks are igneous and intrusive" or "Natural numbers are closed under division" fail to reach an explanatory role in terms of providing counterfactual patterns invariant under interventions.

The way out of this criticism is simple for the m-theory: the claims we have just mentioned might be explanatory in some sense to be defined, but they are not *causally* explanatory.

[13]For more details see Woodward, 2003, [106]: 240-242, 268, where Woodward argues that whether a generalization is invariant is independent of whether it satisfies many of the standard criteria of lawfulness.

Chapter 5

Manipulability Explanations of the Spin-Echo Experiments

In this chapter the explanations of the irreversible decay of the signal in the spin-echo experiments reviewed in chapter 2 will be expressed in terms of the manipulability theory. First, I will briefly recall the conditions for expressing a causal explanation in terms of Woodward's m-theory. Subsequently, I will identify the set of relevant variables $\boldsymbol{V}$ and the patterns of dependence connecting the variables in $\boldsymbol{V}$ for each of the explanations of the spin-echo experiments. This will finally lead us to propose a directed graph[1] associated to each explanation. The aim of this chapter is simply descriptive. Once the different explanations have been cast in these terms it will be possible to proceed to a comparative analysis and a critical evaluation(which I have developed in chapter 6).

5.1 Elements to express explanations in the m-theory

Offering an explanation in m-theory terms requires us to identify the causal variables $(C_1, C_2, C_3, ..., C_n)$ and the effect variable (E) associated with the explanation. The set containing all these causal relata is denoted as $\boldsymbol{V}$. It is not a trivial matter at all to determine which variables belong to that set. For example, sometimes it is unclear whether a variable should figure as a contributing cause (thus belonging in $\boldsymbol{V}$) or as a background condition (thus staying out of $\boldsymbol{V}$).

[1] Similar to directed graph 1 in section 4.2, where the arrows stand for causal links.

Suppose we want to represent Hook's law ($F = -kx$; where F= force; x= enlongation; and k= Hook's constant). There is a certain critical value x_c of the spring elongation at which the spring breaks. This means that the law (or, more appropriately in Woodward's terminology, the generalization) only holds for a certain interval of values of x. Let us call this interval 'domain of validity' and denote it as $D = [0, x_c]$. When expressing the causal dependence of Hook's law, we can choose between two different options. We can either specify D as a background condition and then postulate the following causal relation: $F \longrightarrow x$. Or, we can alternatively incorporate D as a condition in the antecedents of the causal relationship as follows: $X_D \longrightarrow F \longrightarrow x$; where X_D means that $x \in D$.[2]

There are also other important considerations to take into account when choosing the variables in $\boldsymbol{V}$. For example, if we aim to interpret explanations in the most favourable way, we should consider whether it makes sense to "change the value" of the variables that we consider as causes, i.e., we should consider if it make sense to talk about interventions upon those causal variables $C_1, C_2, C_3, ..., C_n$.

As explained in section 4.3, interventions are changes in the value of the cause C that may allow us to determine whether a corresponding change is produced on the value of the effect E. Woodward's claim is that performing interventions one can sometimes obtain some knowledge about the causal relationship between C and E. And the conditions **INi-iv** (specified in section 4.3.1) must be taken into account in order to guarantee that the manipulations performed upon C count as interventions. Let me recall the necessary and sufficient conditions for intervention as follows:

(i) The change on the value of C must be caused *only* by the intervention I and not by the influence of any other cause of C.

(ii) Causes of I must not be total causes of E.

(iii) I must not directly cause E via a route that doesn't go through C.

(iv) The values of all other variables C_i in alternative paths form I to E must be held fixed.[3]

[2]Please note that '$\longrightarrow$' stands for the causal relation, not material implication

[3]It seems that a precondition for condition **(iv)** to hold is that there are no statistical correlations between I and any variable C_i, $C_i \neq C$, that causes E.

From now on we will call "*manipulation*" to *any change* in the value of C. Following the m-theory I will consider that, among all the possible manipulations of C, only those fulfiling the conditions **INi-iv** count as "*interventions*". And "*testing interventions*" are interventions that we actually use to asses whether or not a given causal relationship is robust (see definition of testing intervention on p.69).

It is worth noting that there might be manipulations that change the value of C, but do not count as interventions because they fail to meet at least one of the above-mentioned conditions. Invariance under those manipulations that do not qualify as interventions is not considered relevant for evaluating the depth of a causal explanation. Only invariance under interventions is considered relevant in the m-theory. And, in order to claim that a directed graph is a good representation of a causal explanation, the relationship between the effect E and its putative causes C_i must be invariant under at least one intervention.

Another point to consider when a causal explanation is provided in the m-theory terms is that the very meaning of "performing changes on the variables" must be clear. Otherwise (as the first example in section 4.3.2 shows), the notion of intervention on C, and hence the causal relationship itself, will be ill-defined.

It may also be the case that we do have a causal generalization upon which well-defined changes can be performed, but there is no unique way of understanding the causes of those changes (as in the second example in section 4.3.2). In those cases we should prefer the interpretation that makes better sense of the variables involved in the causal relationship and the very meaning of changing their values.

To sum up, in order to offer a causal explanation in terms of the m-theory, we must carefully choose the set of relevant variables $\boldsymbol{V}$; then identify the causal relationships between them within a directed graph; and finally specify the background conditions under which some *interventions* are at least hypothetically possible.

In the following sections I will employ the terms of the m-theory to express the classical interventionist explanation (section 5.2) and the quantum-based explanations (sections 5.3 and 5.4) of the irreversible decay of the echo-signal in the spin-echo experiments.

5.2 The Classical Interventionist Explanation of the SE Experiments

Let us identify the causal relata postulated by classical interventionism in the explanation of the spin-echo experiments. According to the explanation offered by classical interventionists (developed in section 2.3) during the SE experiments the spin's system transmits magnetic energy to the environment. As a consequence of this energetic transmission the frequencies of some spins are perturbed. The nuclear spins that have been perturbed by their interaction with the environment do not contribute to the echo-signal along with the rest of the spins. This explains the decay of the intensity of the echo signal. At some point in the system's evolution, more precisely, after an interval of time known as *relaxation time*, the system is unable to generate the echo-signal. In other words, the initial state of the system of spins, in which they where all aligned, is no longer recoverable. Therefore, the system of spins has reached a state of true equilibrium. So the whole process taking place during an SE experiment is an *irreversible* process.

In terms of the Olympic race analogy presented in section 2.2, the effect of the transference of magnetic energy from the spins to the environment corresponds to runners getting *fatigued*. As they become tired, they run slower during the second part of the race. Hence they take a longer time to run back to the start line. The key issue is that every runner gets fatigued at its *own personal rate*. If they were equally fatigued, they would eventually reach the start line with some delay. The point is that each runner has *its own personal delay* and this is the reason why the original alignment becomes definitively unrecoverable.

A possible way of defining the set of relevant variables $\boldsymbol{V}$ is then:
$\boldsymbol{V} = \{EP, DS, H\}$; where EP= environmental perturbations; DS= delayed spins; and H= height of the echo-signal.

$$EP \longrightarrow DS \longrightarrow H$$

Directed Graph 11. Causal links according to the classical interventionist explanation.

The causal connections relating the variables in $\boldsymbol{V}$ could then be fairly expressed by the directed graph 10 (below) which may be expressed as fol-

lows: The environmental perturbations EP increase the number of delayed ('fatigued') spins DS. This causes a decay in the height of the echo-signal H. The decay process ends when the echo-signal disappears and thus the value of H becomes constantly H=0. This means that the initial state of the system (in which $H = H_{max}$) is no longer recoverable. Classical interventionists label this state as "true equilibrium".[4]

Once the directed graph is proposed we must specify the values that each of the variables can take. Perhaps the simplest way of defining EP is as a two-valued variable whose value is EP=1 if environmental perturbations occur and EP=0 if they do not occur. However, we may also define EP as a multi-valued variable that corresponds to the "degree of environmental perturbation". Let us take this option and define the interval of values within which the "degree of environmental perturbation" may vary. The upper bound of the interval corresponds to a situation in which the environment is energetically saturated and hence cannot absorb more energy from the spins. Let us call EP_{max} the value that the variable EP takes in that situation, and let us define it as an *open* upper bound (we cannot include EP_{max} in the interval because precisely then the perturbations will cease, but the values of EP inmediately lower than EP_{max} correspond to the highest degrees of environmental perturbation). The lower bound, in turn, corresponds to the situation in which there are no environmental perturbations at all (EP=0) because the system is completely isolated from its environment. Since directed graph 11 represents the classical interventionist view, and total isolation is impossible according to this view, we will define the lower bound as an open bound too. So the interval of values of EP is]0, EP_{max}[.

The variable DS represents the number of nuclear spins that are delayed with respect to the others. In other words, we say that a spin is delayed if it has exported its energy of precession to the environment in such a way that the spin is not able to go back to phase with the rest of the spins, and thus it is not able to contribute to the echo-signal. Using Hahn's terms we would say that spins in DS have lost their "phase-memory" (Hahn, 1950, [49]:581). The variable DS represents the number of nuclear spins in the sample that possess that property. Hence if the system has n spins, the range of values DS can take is [0,n].

[4]This sums up the causal structure underlying Hahn's 1950,[49]; 1953,[49]; 1984,[51] and Ridderbos & Redhead's 1998, [82] ideas.

Lastly, the minimum value of the height of the echo-signal is H=0; and the maximum value it can take corresponds to the height of the signal that is generated at the beginning of the experiment. Let us call that height H_{max}.The interval of possible values of the variable H is thus [0, H_{max}].

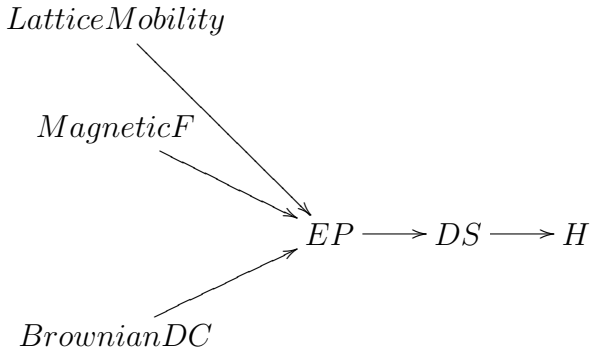
The putative causal relationships are now defined. The next step is to verify if there is some way (at least one) to change the value of the putative causal variable, in this case EP. According to classical interventionists, the environmental perturbations EP are manifested by the transformation of the spins' energy into other forms of energy (Ridderbos & Redhead, 1998, [82]:1257). Let us look for the factors that may affect this transference of energy from the spins to the environment. Then we will be able to establish some manipulations (processes that change the value of EP) that, should they fulfil conditions **INi-iv**, will count as putative interventions upon the variable EP.

Although the specific physical factors affecting EP are not listed in the classical interventionists' explanations of the spin-echo experiments, they do make reference to Hahn's papers (1950, 1953) from which I have extracted the relevant factors that may cause the environmental perturbations EP. According to Hahn's description of the SE experiments, the transference of the spins' magnetic energy of precession to the environment is brought about by the combination of three processes: The *spin-lattice thermal exchange*, the *fluctuations of the local magnetic fields*, and the *Brownian motion* of the glycerine molecules (see Hahn, 1950, [49]:581 and Hahn, 1953, [50]:7). I will consider them as three putative interventions of the causal variable EP.

The "spin-lattice thermal interaction" is the process in which the energy of the spins is exported to the lattice structure in which the nuclei are held. The lattice is normally defined as the three-dimensional arrangement of molecules in the sample. It is also appropriate to conceive it as a heat reservoir. The reason is that some of that energy absorbed from the spins is transformed into thermal energy, and the rest forms a magnetic field known as the lattice field. Whether this effect is more or less intense depends on the mobility of the lattice. In medicine, for instance, the spin-echo technique is applied to different types of tissues (brain, heart, etc. see Hashemi, 2004 [52]) instead of a sample of glycerine. The mobility in each of those lattices is different and it has been observed (see Hashemi, 2004 [52]:57) that this affects how much energy the spins can export to the lattice. Roughly, the higher the mobility of the lattice, the more likely the transmission of energy. I will define the putative intervention variable associated with this effect as '*LatticeMobility*'.

The variable associated with the second putative intervention upon EP may be defined as '*MagneticF*' meaning 'magnetic fluctuations'; and the variable associated with the third and last putative intervention upon EP may be defined as '*BrownianDC*', meaning 'Brownian Diffusion Coefficient'.

The directed graph 12 (below) illustrates these three putative interventions upon the variable EP.



Directed Graph 12. Putative interventions upon variable EP (environmental perturbations).

It is worth mentioning that during an SE experiment there is also a exchange of magnetic energy between nuclear spins. Hahn calls this effect the "mutual spin-spin flipping" (Hahn, 1950, [49]:581). Despite the fact that this magnetic effect also (partially) determines the energy of an individual spin, I do not consider it appropriate to include the spin-spin interaction as a fourth possible way of intervention upon EP. The reason is that classical interventionists talk about "the exchange between spin energy and *other forms* of energy" (Ridderbos & Redhead, 1998, [82]:1252, my emphasis). Moreover, they explain the signal decay appealing to the energy '*exported*' from the spin-system to the external environment (Ridderbos & Redhead [82]:1257). In other words, the spin's energy is dissipated into a larger system comprising the environment. The classical perspective is then considering the spin-spin interaction as a magnetic effect taking place inside the system itself, and not as a factor contributing to the system-environment energy dissipation.

The description of the classical interventionist explanation would not be complete without a specification of the background conditions under which the causal links hold. These conditions include the experimental settings

which consist of 1) a sample of a great number of spins ($\approx 10^{23}$), normally obtained by placing a glycerol solution in a strong magnetic field (0.177 Tesla); 2) an apparatus that generates the magnetic field; and 3) the radio-frequency pulse generator (see Fig.12).[5]

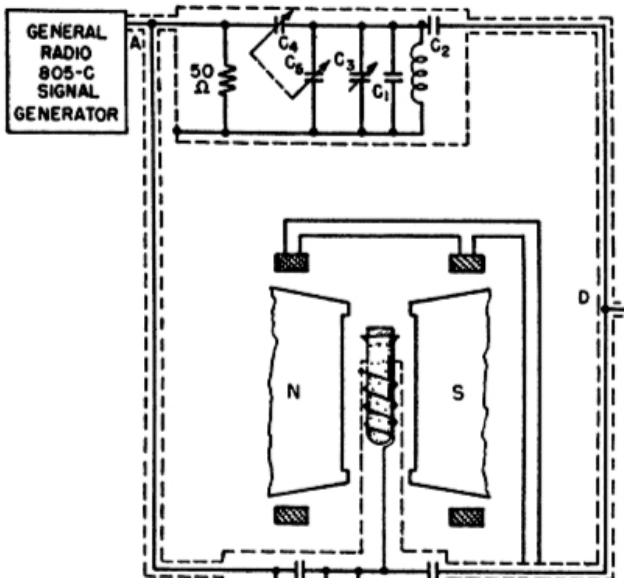


Fig.12. Experimental device.[6]

The experimental setting is treated classically. More precisely, the spin dynamics are described by a theory based on the dynamics of a classical gyroscope known as Larmor's theory of precession (see appendix D).

[5]Detailed descriptions of the experimental setting can be found at Martin, 2006, [69] and Duarte, 2008, [29].

[6]Illustration is taken form Bloembergen *et al*, 1948, [10].

5.2.1 General graph for classical interventionism

In this section an alternative and more general directed graph is proposed for classical interventionist explanations. Instead of focusing on the spins' frequencies, let us pay attention to the dissipation of correlational information. From this point of view, any thermodynamic irreversible process can be characterized as an evolution towards a state such that the correlational information about the macroscopic initial order of the initial state is completely dissipated.

A set of relevant variables $\boldsymbol{V}$ appropriate for expressing this general interventionist idea is $\boldsymbol{V} = \{EP, CI, TEQS\}$ where EP = Environmental perturbations; CI = Correlational information; $TEQS$ = True equilibrium state. And the variables in $\boldsymbol{V}$ are causally related as shown in the directed graph 10 below.

$$EP \longrightarrow CI \longrightarrow TEQS$$

Directed Graph 10. General directed graph for classical interventionism.

In this case the variable EP is defined as before. And the variable CI may be associated to the percentage of correlational information that is still contained in the system. Thus the interval of possible values of CI is [0,100]. To define the possible values of the effect-variable in the directed graph, $TEQS$, we must choose between two different options. The fist option is to associate $TEQS$ to the probability that the system reaches a state of true equilibrium. In that case the interval of possible values of the variable is [0,1]. This option may be tricky as it is still a matter of controversy how probabilities are to be interpreted in statistical mechanics. The second option is to define $TEQS$ as a two-valued variable whose value is $TEQS$=1 if the system under study reaches a state of true equilibrium, and $TEQS$=0 otherwise. If the system is in a state of quasi-equilibrium the value of the variable is also $TEQS$=0.

This directed graph embodies the very essence of classical interventionist explanations of irreversible processes. It is adequate to express the explanation of spin-echo experiments put forward by Ridderbos and Redhead in (1998,[82]); the explanation of gases evolutions based on molecular interaction due to Blatt, (1959,[9]), and the explanation based on the interaction

between the particles and the infinite parts of the reservoir in which the gas is contained (Bergmann and Lebowitz, (1955,[7])).

Particularly, the spin-echo experiments can be explained using this new directed graph as follows: At the beginning of the experiment the system contains all the correlational information (CI=100) and the perturbations from the external environment can be practically disregarded ($EP \approx 0$). However, during the experiment the spin system evolves (alternatively generating and vanishing the echo-signal) and the environmental perturbations EP begin causing a lost of correlational information (CI <100). After several r-f pulses, the spins lose their phase memory due to the energetic exchange with the environment up to the point that $CI = 0$. The true equilibrium state is then reached by the system of spins. This means that the system's evolution has been thermodynamically normal, i.e, TDB=1.

5.3 The GRW-based Explanation of the SE Experiments

As mentioned before[7] Albert's explanation of irreversibility is based upon the GRW collapse interpretation of quantum mechanics. The GRW-dynamics are genuinely stochastic, thus David Albert's GRW-based approach is indeterministic. This means that the causal relations represented in the directed graphs will now link changes in the values of the causes with changes in the probabilities of the values of the effect (see p.57). A microstate is no longer represented by a point but by a Gaussian in position. The probability of a system being in a certain microstate is replaced by the probability that such a Gaussian is centred on a given point after a real collapse.

From the GRW-based perspective, several GRW-collapses (i.e., *real* collapses) take place during the approach to equilibrium of any thermodynamical system. The essence of the explanation consists in the fact that every GRW-collapse directly affects the position of the particles and drives the system to a state with a Gaussian centred in a *thermodynamically normal state*.[8]

Recall that Albert assumes that the "thermodynamic probabilities are associated to quantum probabilities" (Albert's dynamical hypothesis, see Albert, 2000, [4]: 151-152.) According to the GRW-based explanation, if this dynamical hypothesis holds, the thermodynamic behaviour of the system is guaranteed by the GRW-dynamics, independently of whether or not the initial state of the system is thermodynamically normal. As shown in Fig.13(below), although the system may fall into an abnormal state during its evolution (including the initial state), the GRW collapses will typically drive the system back to a thermodynamically normal state. As a result, there is a high probability that the total evolution of the system is thermodynamically normal. In other words, the system evolves in accordance with the second law of thermodynamics.

[7] The main argument of this approach is in section 2.4.1.

[8] We say that a state is *thermodynamically normal* if it belongs to a trajectory in the phase space (an evolution) that follows the laws of thermodynamics. Otherwise, the state is *thermodynamically abnormal* (see Fig.11 on p.39).

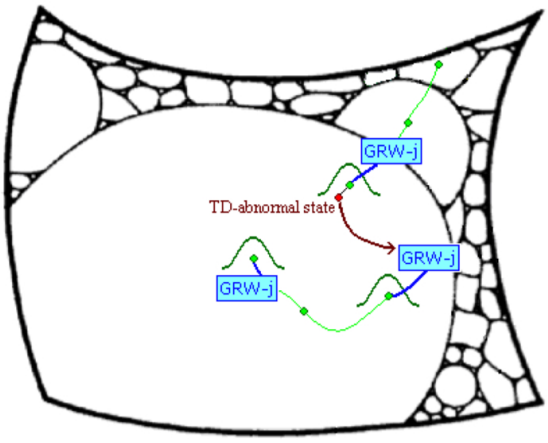


Fig.13. Every GRW-collapse drives the system into a state whose Gaussian is centred in a thermodynamically normal state.

Fig.13 represents the thermodynamic trajectory in the usual phase space. Although GRW collapses take place on the configuration quantum space, they *causally affect* the position of the particles. Therefore, they have been included in the figure.

It is worth mentioning that, in the case of the spin-echo experiments, the direct effect of GRW-collapses over the position is not enough to cause the echo signal decay. When GRW-collapses take place in position the wave function of the system remains almost unaltered. The decay is actually obtained due to an indirect effect. Given the position-spin coupling, slight changes in position generated by the GRW-collapses produce, in turn, stronger changes over the spins. It is this perturbation over the spins that really pushes the system away from the vicinity of the initial state, finally avoiding the return to it (see Hemmo & Shenker, 2003, [54]:section 3.2).

Taking all these elements into account, the simplest and more adequate way to formulate Albert's explanation of irreversibility in the spin-echo experiments is the following. Let the set of relevant variables be $\boldsymbol{V} = \{GRWC, PS, TDN\}$; where $GRWC$=GRW-collapses; PS=perturbed spins; and TDN=thermodynamic normality.

Let us interpret $GRWC$ as "the rate of GRW-collapses per second" and let $GRWC_{max}$ be the maximum number of collapses that may occur in one second. The interval of possible values of the variable $GRWC$ is thus [0, GRW_{max}]. The interval of values of the variable PS (perturbed spins) for a

system of n spins is [0,n]. And finally we may define the range of possible values of the variable TDN in terms of the evolution of the system. That is, if the evolution is thermodynamically normal TDN=1 and if the evolution is thermodynamically abnormal, then TDN=0 (see Fig.11 on p.39). I believe this is the definition of TDN's values that best captures the essence of Albert's explanation.[9]. The directed graph 13 (below) represents the causal relationship connecting these variables.

$$GRWC \longrightarrow PS \longrightarrow TDN$$

Directed Graph 13. GRW-based explanation of the thermodynamic behaviour.

Alternatively, we may use the variable H (height of the echo-signal) instead of the variable TDN. In that case, the set of relevant variables is defined as $\boldsymbol{V} = \{GRWC, PS, H\}$; where $GRWC$=GRW-collapses, PS=number of perturbed spins, and H=height of the echo-signal. And the intervals of possible values of $GRWC$, PS and H are [0, GRW_{max}], [0,n], and [0, H_{max}] respectively. The directed graph 14 represents the causal relationship connecting these variables.

$$GRWC \longrightarrow PS \longrightarrow H$$

Directed Graph 14. GRW-based explanation of the spin-echo experiments.

Since we are focusing now upon the spin-echo experiments in the next chapter I will use directed graph 14 (rather than directed graph 13) to assess the explanatory import of the GRW-based approach.

Let us now define some putative interventions upon the variable $GRWC$. In order to intervene upon this causal variable we would like to find a manipulation or mechanism that *controls* its values. However, finding such a

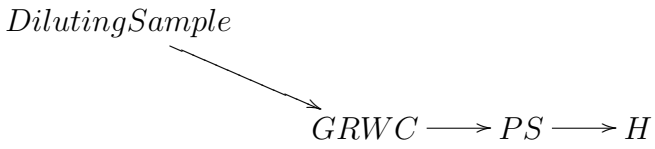
[9]We have expressed the values of TDN in Boltzmannian terms because that is the framework in which Albert works. However, we so wished, we could also define the values of TDN in terms the Gibbsian framework: if the system approaches the microcanonical distribution TDN=1; otherwise TDN=0.

mechanism is particularly tricky in this case. The quantum collapses are genuinely stochastic and we cannot generate them or stop them at our will or convenience.

In my view, we face one of the problematic cases mentioned in the description of the m-theory (in section 4.3.3). I am referring specifically to the case in which interventions on the cause are not possible, not only because they are difficult to carry out in practice, but rather because the variable upon which we want to perform the interventions has no possible causes. The good news are that in this kind of cases it is not necessary for an intervention I to switch off the causal link between the putative cause C and its multiple causes $I_1, I_2, \ldots, I_n$ (see directed graph 9 on p.67) because they do not even exist. The bad news are that, since interventions are causes themselves in accordance with condition **IN-i**, it follows from the fact that C has no cause that there can be no interventions on C that can serve to test the causal claim.

The way out suggested by Woodward in cases like this is to assess counterfactual claims about what would happen under interventions on C, even if these interventions are not in fact possible (see Woodward, 2003, [106]:13, and section 4.3.3 of this thesis).

Fortunately in our particular case, despite the fact that we cannot provoke the specific occurrence of a collapse, there is a way of increase and decrease the amount of GRW-collapses and this is the variable that we have adopted as our main object. The glycerine sample can be diluted reducing the number of spins whose states are collapsed by the GRW-collapses (see directed graph 15).



Directed Graph 15. Possible intervention upon the variable *GRW*

If Albert's explanation is right, this intervention should decrease the rate of GRW-collapses per second. As a consequence, a lower number of spins

would be perturbed and the echo-signal would take longer to disappear. That is, it would take a longer time for the variable H to take the value $H = 0$ and remain indefinitely at that value. In other words, the system of spins would take longer to reach the thermodynamic equilibrium.

Let us finally list the background conditions and background theory. Again the background conditions are given by the experimental setting illustrated in Fig.12 (on p.88) In this case the background theory is given by the GRW collapse interpretation of quantum mechanics, in particular the GRW *stochastic* dynamics in which the Schrödinger equation is replaced by a stochastic equation; and Albert's dynamical hypothesis.

It is worth noting that Albert's approach also adds the past hypothesis (low entropy initial condition defined in section 1.4), which is normally assumed to define the past-to-future direction. It is important to remark, though, that the purpose of the past hypothesis in this case is only to make the past fit with our records, but it plays no role in making the future fit with the second law of thermodynamics.[10]

[10]I thank Meir Hemmo for drawing my attention to this point.

5.4 Decoherence-based Explanation of the Spin-Echo Experiments

The decoherence-based explanation of the spin-echo experiments (introduced in section 2.4.2) is easier to express in terms of the m-theory if we divide it in three parts. The first part corresponds to the explanation of why the electromagnetic signal fades away in the early stage of the experiment during the interval of time $[0, \tau]$ (see Fig.7-9 on p.28). The second part corresponds to the explanation of why the signal re-appears at $t = 2\tau$ (return from Fig.9 to Fig.7 on p.28). And the third part explains why the maximum value of the signal gradually decreases and eventually disappears (see Fig.10 on p.29)[11].

In the early stage of the spin-echo experiments the electromagnetic signal decreases until, at time $t = \tau$, it completely vanishes. During this interval of time, the perturbations due to decoherence are ignored. In other words, according to the decoherence-based approach, the Schrödinger unitary and deterministic dynamics give us a good description of the system's behaviour during this stage.

The fact that the spins get out of phase (bringing about the first fall in the intensity of the electromagnetic signal) is explained by the action of two factors: the spin-spin interaction that randomizes the spins' states; and the in-homogeneities of the magnetic field that make the spins precess at different rates. As the system has not been perturbed during the first part of the experiment, the return to the initial state is still possible (in other words, "the runners are not yet tired"). This is precisely the return that takes place when the spins get back in phase producing the first echo-signal. The only factor responsible for this re-coordination of the spins, according to Hemmo and Shenker, is the radio-frequency pulse applied by the experimenter.

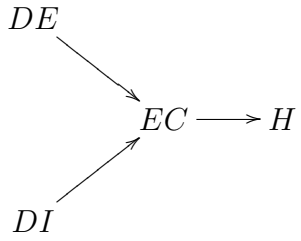
Finally, in the third part we observe the decay of the echo-signal. The perturbations that were ignored in the first two stages of the experiment come into play in the long-term evolution of the spins' system. In fact, all the explanations presented in this chapter assume that the perturbations upon the spins' system constitute the very cause of the decay of the echo signal.

[11] Classical interventionists would refer to the fist part as "the approach to quasi-equilibrium" and to the third part as "the approach to true equilibrium". However the authors of the decoherence-based approach, Hemmo and Shenker, do not adopt this terminology as they do not commit themselves to the distinction between the two kinds of equilibrium.

While the classical interventionist approach appeals to *external* perturbations, the GRW-based approach appeals to *internal* quantum perturbations. The decoherence-based approach combines both kinds of perturbations, associated with the environmental decoherence and with the spin-spin decoherence respectively.

In the next chapter I will argue that this combination of external and internal causal factors may represent an advantage for the decoherence-based approach. But first we need to express the third and last part of the decoherence-based explanation in terms of the m-theory. In order to do so we will study how both the environmental decoherence and the spin-spin decoherence work, and how the no-collapse interpretations of quantum mechanics frame the third and last part of the decoherence-based explanation.

The set of relevant variables can be defined as $\boldsymbol{V}=\{DE, DI, EC, H\}$, where DE=rate of external decoherence; DI=rate of internal decoherence; EC=effective collapses, H=height of the echo-signal. And the causal relationships are given by the directed graph 16.



Directed Graph 16. Decoherence-based explanation of the long term decay and disappearance of the echo-signal.

The variables DI and DE may be interpreted as the rate of decoherence. Let [0, DI_{max}] and [0, DE_{max}] be the respective intervals of possible variables for these variables. Let us denote as EC_{max} the maximum number of collapses that may occur in one second. The rate of effective collapses EC may take any value of the interval [0,EC_{max}]. And the range of possible values of the variable H is again [0, H_{max}].

In order to understand the decoherence-based explanation it will be useful to keep in mind that this explanation is similar to Albert's in many respects. In particular, the assumptions regarding the distribution of normal and abnormal states in the phase space remain the same. And the role played by the GRW-collapses will now be accomplished by the external and the internal decoherence together with the stochastic dynamics (assumed in no-collapse interpretations of quantum mechanics).[12]

When Hemmo and Shenker appeal to environmental decoherence to explain the decay of the echo-signal (Hemmo & Shenker, 2005, [55]:641), they are employing the process described by the standard models of decoherence[13], in which "position" is usually considered the pointer basis. This means that –in analogy with the GRW-based explanation– the spins will be affected *indirectly* though the coupling between "spin" and "position". The internal or spin-spin decoherence, by contrast, exerts a *direct* influence upon the spin states. "The system" is a particular spin and "the environment" is comprised by the rest of the spins in the sample. In this case the pointer basis is the "spin".

As a result of these external and internal decoherences the system jumps from one Schrödinger trajectory to another (see Hemmo & Shenker, 2003, [54]:338). This transition from a effective state to another effective state (*effective collapse*) may be considered to be stochastic.[14] Both external and internal decoherences ensure that those effective collapses will lead the system into a *normal quantum state*[15].

The connection between quantum probabilities and thermodynamic probabilities is in this case produced in two steps. Firstly, it is postulated that the stochastic dynamical laws yield the quantum mechanical probabilities given

[12]For instance, in modal interpretations, the dynamics are genuinely stochastic. Hemmo and Shenker refer to Bacciagaluppi and Dickson (1999, [5]) for a detailed discussion on this.

[13]The environmental decoherence, in general, is the quantum process due to the interaction between the system and its environment that leads to the suppression of interference. In other words, it is the process that turns a coherent quantum state into a diagonal state in a well-defined basis known as pointer basis. For more details see Zurek and Paz, 1994,[109]; Zurek and Paz, 2002, [76]; and Zurek, Paz et al, 2003, [110]. A recent review of decoherence is presented in Castagnino et al, 2010, [24].

[14]The collapses are called 'effective' in contrast with the 'real' collapses. I am following Hemmo and Shenker on this, for more details see[54]: 338, 349-351.

[15]Namely, a state whose quantum evolution begins and ends in well-defined states, in the sense that they collapse onto Gaussians in both position and momentum.

by the Born rule. And secondly, it is postulated that the quantum mechanical probabilities reproduce the quantitative predictions of classical statistical mechanics[16]. As Fig.14 (below) illustrates, if both postulates hold, the evolution of the system will be formed by segments of different trajectories; and the transition from one trajectory to the next is due to an effective collapse.

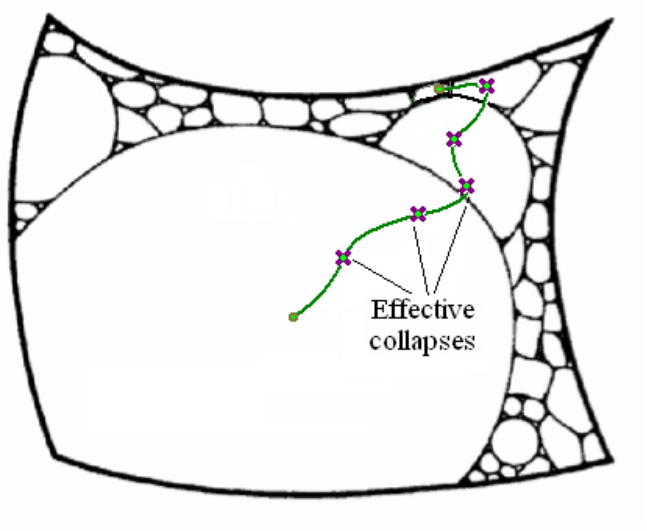


Fig.14. Evolution of thermodynamic systems according to the decoherence-based approach.

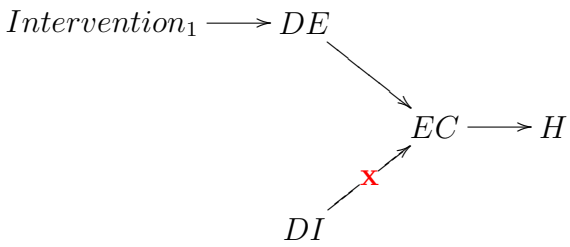
It is worth noting that, since the trajectory of the system is perturbed by the effective collapses, the evolution is *independent* of its past thermodynamic history. Precisely for this reason, it is claimed that the system will evolve in a thermodynamically normal way, regardless of whether its initial state is thermodynamically normal or not.

In order to analyze decoherence-based interventionist explanation in terms of the m-theory we need to identify possible interventions on the putative causes. There are at least two possible interventions that control the values of the causal variables DI and DE: one is diluting the glycerine sample, the other is fixing the initial quantum state in such a way that decoherence does not take place during the entire experiment. If we are able to set the experiment in such an initial quantum state, decoherence would not affect

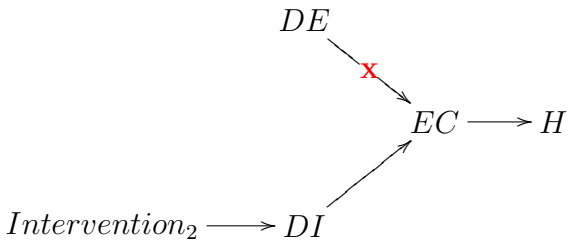
[16]This assumption is analogous to Albert's dynamical hypothesis.

the spins' system and relaxation times should be manifestly longer. This "no-decoherent states" are theoretically possible, but they have not yet been experimentally prepared.

As it happened in the classical case, in order to fulfil condition **IN-iv** it is important to consider whether it is possible to hold fixed the value of the variable DE (external decoherence) when interventions on the variable DI (internal decoherence) are performed, and vice-versa (see directed graphs 17 and 18).



Directed Graph 17. Intervention on DE while holding DI fixed.



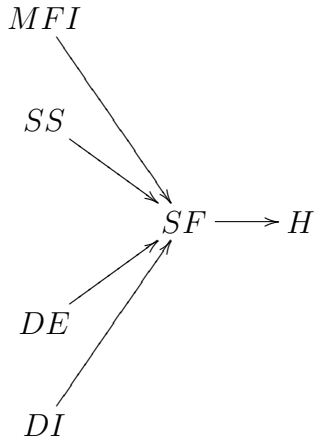
Directed Graph 18. Intervention on DI while holding DE fixed.

The background conditions are given by the experimental setting. And there are a number of important theoretical assumptions in the background theory that play a role in this explanation, but it is not necessary to list them fully here – they appear in the relevant section of appendix C.

5.4.1 Final remark about the decoherence-based explanation.

A remarkable feature about the decoherence-based approach is that spin-spin interaction plays a double role because it is considered to be a causal factor in two different stages of the spin-echo experiments. During the third part it pushes the system to thermodynamic equilibrium trough internal decoherence (represented by the causal variable DI), but also during the first part it contributes to causing the defocus of the signal trough randomization of the states.

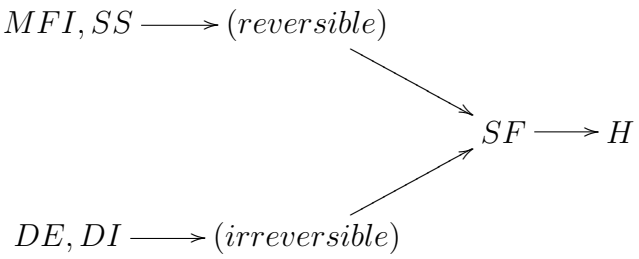
Let us define this early spin-spin interaction as the variable SS and let us try to propose a new set of relevant variables ***V'*** including SS. As we are including a causal factor of the first part of the experiment, we should also include the other, i.e., the magnetic in-homogeneities of the field (call them MFI). Instead of effective collapses (whose role is irrelevant in the first two stages of the experiment), let us focus on the alteration of spin frequencies. Then a possible ***V'*** would be ***V'***=$\{SS, MFI, DE, DI, SF, H\}$; where SS= spin-spin interaction, MFI= magnetic field inhomogeneities, DE= external decoherence, DI= internal decoherence, SF= spins' frequencies, H= height of the echo-signal. The causal relations connecting the variables in ***V'*** would generate the directed graph 19, in which the intermediate variable is not only affected by the two kinds of decoherences but also by the magnetic field in-homogeneities and the spin-spin classical exchange of magnetic energy.



Directed Graph 19. Alternative directed graph for the decoherence-based explanation of the echo-signal decay.

A consequence of proposing such a directed graph associated to the set ***V'*** is that the interventions upon the variable DI would not satisfy the m-theory conditions (**IN i-iv**). This is because there is no possible intervention that is able to bring about changes in the values of of the spin-spin internal decoherence DI, without affecting the values of SS as well.

However, the directed graph 19 is inappropriate in the sense that it mixes two kinds of factors that affect the spins' frequencies. The influence of MFI and SS upon the spins' frequencies SF is *reversible*. And the proof is that, during the second stage of the experiment, the original state is still recoverable. Decoherences DE and DI, on the contrary, exert an *irreversible* influence upon the spins, "making them tired" in the race analogy. Taking this consideration into account would lead us to a more appropriate directed graph (directed graph 20) where we have marked when the causal influence is reversible and when it is irreversible.



Directed Graph 20.

As the branch containing SS and MFI is irrelevant for the explanation of the thermodynamic behaviour of the system, we can ignore it. Therefore, from now on we will take directed graph 16 (on p.97) as the adequate expression of the decoherence-based explanation of the echo-signal decay.

Chapter 6

Manipulability test of Explanatory Depth

In the previous chapter we described the causal patterns postulated by three different explanations of the spin-echo experiments. The aim of the present chapter is to examine how *deep* those explanations are in the Woodwardian sense of *depth* (defined in section 4.5.1). The chapter is organized as follows. In section 7.1 we analyse whether the causal patterns remain invariant under some interventions. In section 7.2 we compare the capacity of the explanations for providing answers to the so-called *what-if-things-have-been-different-questions*. In other words, we compare the explanations in terms of their ability to account for counterfactual situations. Finally, in section 7.3, the results of the analysis developed in this chapter are summarized and discussed.

The analysis here presented constantly refers to experiments that have been actually performed during the last decades. This provides an interesting complement for evaluating the empirical adequacy of the explanations while we discover how *deep* they are.

6.1 Invariance under testing interventions

In this section we will first propose some actual or hypothetical manipulations to control the values of the cause(s) postulated by each explanation. Afterwards, we will verify which of them fulfil Woodward's conditions **INi-iv** (defined in section 4.3.1.). This will provide us with a set of *interventions* for testing the explanations under study. We will then proceed to analyse if

the causal links postulated by each explanation remain invariant when such *testing interventions* are applied[1].

6.1.1 Testing the classical interventionist explanation

Our description of the classical interventionist explanation in m-theory terms (illustrated in directed graph 12) assumes that the interaction between the environmental perturbations exerted upon the system of spins (denoted with the variable EP= "environmental perturbations") are brought about by the thermal interactions between the spins and the lattice, the magnetic fluctuations and the Brownian motion. In chapter 6, those factors were simply mentioned to provide a complete description of the classical explanation. We examine in greater detail how exactly those factors increase or decrease the rate of environmental perturbation. Expressed in the m-theory terms, we turn to study now how these factors can be used for manipulating the value of the variable EP.

Let us begin with the Brownian diffusion coeficient, which is frequently recognized in scientific papers as a relevant factor affecting the behaviour of the spin-echo system[2]. As Einstein pointed out in one of his *annus mirabilis* papers, the Brownian motion depends both on the *temperature* and *viscosity* of the substance in which the particles are immersed (see Einstein, 1905, [38] or Shaxby, 1914, [88]:544). Thus in the spin-echo experiments the value of the Brownian diffusion coefficient may be manipulated, for example, by diluting the glycerine in water; or by warming the sample up.

An experiment comparing the spin-echo signal for different dilutions of water and glycerine has been recently performed (Martin & Hughes, 2006, [69]). The results of this experiment show that the relaxation time of the system actually depends on the water-glycerine proportion in the sample. More precisely, the higher the glycerine content, the shorter the relaxation time. These results are consistent with the following results obtained by

[1] A "testing" intervention is no more than an intervention used to test how robust is a given causal relationship, see p.83.

[2] Widom, for example, begins his article "Fractal Brownian motion and nuclear spin echoes" as follows: "It has long been known that the microscopic Brownian diffusion coefficient D of a 'particle' in a fluid can be measured via the amplitude of the magnetic echo signal from nuclear spins subject to an appropriate sequence of magnetic field pulses" (Widom, 1995, [105]:1243)

Bloembergen, Purcell and Pound [10] in a similar experiment (see Fig.15 below).

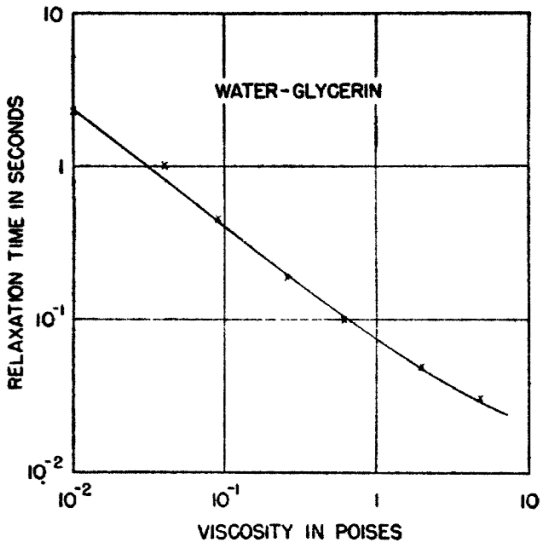


Fig.15. Relaxation time for protons in water-glycerine solutions.

Classical interventionism would account for this behaviour by arguing that the Brownian motion increases when the concentration of glycerine is high. So the perturbations are intensified making the echo-signal disappear (i.e. bringing the system to a state of true equilibrium) in a shorter relaxation time.

Suppose now that, instead of manipulating the viscosity by mixing the glycerine with water, we fix the water-glycerine proportion and vary the temperature. This situation would affect the whole experiment because, as we will see in a moment, the evolution of the system is temperature-dependent. Nevertheless, it is worth noting that warming up the sample would increase the value of the Brownian diffusion coefficient, and therefore, shorten the system's relaxation time. At the same time, however, it would make the viscosity importantly decrease, and higher viscosity leads to longer relaxation times.

In scientific practice (at the MIT Junior lab for example)[3], in order to study variations in relaxation time, experimenters use a table that indicates the viscosity values for different water-glycerine solutions at several tem-

[3]See [72].

peratures. The longest relaxation times had been achieved in experiments developed at very low temperatures.[4]

In any case, the predictions provided by classical interventionism are compatible with the experimental fact that variations in relaxation time have been seen to depend on the specific Brownian motion rate associated to the actual viscosity and temperature of the sample. This is the kind of manipulation we are looking for.

Let us now consider manipulations through variations of the magnetic fluctuations. Perhaps the rate of magnetic fluctuations can be raised during the experiment by introducing some paramagnetic impurities in the sample (Blumberg, 1959, [11]) or by moving some extra magnets around the experimental device. This would generate magnetic fields in certain areas of the sample, additional to the static magnetic field generated by the original device. In the only experiment I have been able to find performing this kind of manipulation (Stejskal, 1964, [92]) a technique is employed for generating magnetic gradients with a couple of coils wound on tapered forms. The effect on the system is then measured for several different gradients. The output of this experiment shows that the echo-signal is directly affected by the magnetic gradients in such a way that greater magnetic gradients correspond to echo-signals with weaker intensities.

An alternative would be to prevent (rather than generate) magnetic fluctuations and in-homogeneities. Some experiments have already been performed in order to achieve this. Bloembergen *et al*, for example, describe how water samples can be used to detect the field in-homogeneities and find, in that way, the more homogeneous area in the field (see Bloembergen et al, 1984, [10]:684). Also, and more recently, a computer model has been developed to simulate different in-homogeneous magnetic fields and study how much the behaviour of the spin's system is affected by these in-homogeneities (Nyenhusi, 1994, [74]). In order to reduce costs and improve the medical application of the spin-echo technique, the computer model aims to find the range at which the measurements of relaxation times are still reliable without using too expensive magnets. For our purposes, it is enough to realize that there are some experimental manipulations that effectively test the putative influence of the internal magnetic fields upon the spin-echo system.

[4]Bloembergen *et al*, for example, describe spin-echo experiments developed with glycerine super-cooled down to temperatures far below its freezing point (see Bloembergen et al, 1984, [10], section XI).

It is also worth mentioning that immersing the sample in a *perfectly homogeneous* magnetic field is also a hypothetically conceivable manipulation; therefore, it may count (if it fulfils conditions **INi-iv**) as an intervention for testing our explanation under study.

Again, classical interventionists would be able to account for the behaviour of the system under variations (increases or decreases) of the amount of magnetic field fluctuations. More precisely, they would predict longer relaxation times for the perfectly homogeneous field, and shorter relaxation times for fields with more in-homogeneities and magnetic fluctuations.

Let us finally imagine how the thermal interactions between the spins and the lattice can be modified. As Hahn explained already as back as 1950 [49], the spin-lattice thermal interaction brings about transferences in the spins of some particles (from spin up to spin down). When this occurs the magnetic energy of the precessing spin is transferred to a molecule of glycerine in the form of kinetic energy. After many spin transferences the system experiences a "cooling process" characterized in practice by the so-called spin-lattice relaxation time.

These random thermal interactions between the spins and the lattice are present during the whole experiment. They produce the de-phasing of the spins in the short term (driving the system to the state that classical interventionists called quasi-equilibrium), and the decay of the eco intensity H in the long term. According to the classical interventionist explanation, this is the dominant process affecting the spin frequencies. From their point of view, in the absence of the spin-lattice thermal interaction, the height of the echo-signal H would simply never decrease (see Ridderbos & Redhead [82]:1252-1253).

The relevant question for us is whether we can use this spin-lattice interaction to control the values of the putative causal variable EP (environmental perturbations). Intervening to stop the environmental perturbations would imply isolating the spins from the molecules to which they belong. Is this even conceivable?

From the classical interventionist perspective, all the systems in the universe are open, except the universe as a whole. But this is a contingent *matter of fact*, and not a necessary *matter of law*. Recall that, according

to the m-theory it is not necessary to actually perform the intervention in order to consider it valid. When technological limitations are involved, we invoke hypothetical interventions to analyse the causal relationships. The only requirement is that such intervention be physically possible (see Woodward, 2003, [106]: section 3.5). If we were studying some other system, a box full of gas for example, "absolutely isolating the system from its environment" would imply blocking the interaction between the gas molecules and the walls of the box. And this, perhaps, would count as a possible hypothetical manipulation. However, in the spin-echo experiments, isolating the system implies a change in the very capacity of the spins to transfer their energy to the molecules in the sample. Can we genuinely modify this?

I can hardly imagine such a manipulation, unless we replace the glycerine with some other substance (or other tissues as a brain or a heart tissue) whose molecules produce different splits between spin up and spin down. Changes in the molecular structure of the sample change the capacity of spins to flip from one state to the other; and hence the spin-lattice thermal interaction is also affected. In other words, the lattice mobility depends on the molecular structure of the sample. But, even if we manage to reduce the spin-lattice interaction, that interaction will never be *completely blocked*. So let us assume that there is no way of isolating the spins from the lattice and continue with our analysis.

Summing up, we have so far proposed four manipulations for controlling the values of the variable EP, which is the putative cause of irreversibility according to classical interventionism.

-Reducing the viscosity of the sample (M_1).

-Changing the temperature of the sample as (M_2).

-Introducing magnetic in-homogeneities (M_3); and

-Producing a perfectly homogeneous field (M_4) (hypothetical manipulation)

Let us now consider which among them fulfil the conditions **INi-iv** as defined in section 4.3.1. If these manipulations fulfil the conditions **INi-iv** they will count as *testing interventions*. If the postulated causal relations by classical interventionism (in directed graph 11) turn out be invariant under such *testing interventions*, we will be able to determine if the classical interventionist explanation meets criteria 1 and 2 of *explanatory depth* (see p.73).

In order to count as a *testing intervention* the manipulation M_1 should first fulfil the condition **IN-i**, i.e. M_1 should switch off the effect of any alternative cause(s) of the intervened variable. Directed graph 5 illustrates what a valid intervention must achieve in order to fulfil condition **IN-i**. Thus, in order to fulfil the condition in this particular case, the manipulation M_1 (diluting the sample) should switch off the effect of the magnetic fluctuations upon the spin's frequencies and, additionally, block the spin-lattice energetic exchange.

It has been found in practice, however, that both "magnetic fluctuations" and "spin-lattice thermal interactions" are still acting when the manipulation M_1 is carried out. Moreover, it is impossible to conceive any manipulation that controls the Brownian motion, and, at the same time, neutralises the exchange of magnetic energy between spins. Therefore, M_1 does not fulfil **IN-i**. And thus, despite the fact that M_1 meets other conditions on interventions, it cannot be considered as a *testing intervention.*

The second manipulation M_2 also fails to meet one condition. The evolution of the spin-system is temperature-dependent. So, cooling or warming the sample up, violates the condition **IN-ii**, according to which the intervention cannot be a direct cause of the explained event. In other words, the intervention must affect the effect only trough the path containing the postulated cause (directed graph 7 on p.62, illustrates condition **IN-ii**).

The third and fourth manipulations M_3 and M_4 fail to meet condition **IN-i**. The reason is analogous to the reason why M_1 does not fulfil the condition. But in this case the violation of **IN-i** is due to the impossibility of "switching off" the effect of the Brownian motion upon the spins frequencies, and the spin-lattice thermal interaction, while the magnetic manipulations are performed.

Our conclusion in this section could be the following: All the processes taking place during the spin-echo experiments to which classical interventionists assign a causal role (namely, magnetic fluctuations, Brownian motion and thermal interactions) are closely related to each other. The causal structure seems to be complex and no intervention is a clean single direct cause upon the putative cause. As a consequence, all the manipulations (M_1 to M_4) violate at least one of the conditions **INi-iv**. So we are left with no interventions to test the putative causal relationships postulated by classical interventionism.

But before arriving at this devastating conclusion, imagine the following scenario. Suppose we knew nothing about the spin-echo experiments. Someone then shows us the spin-echo experiment and informs us that Brownian motion is the only cause responsible for the echo-signal decay. This person offers us a fictitious explanation according to which the set of relevant variables is $\boldsymbol{V}=\{BDC, CI, H\}$; where BDC=Brownian diffusion coefficient, CI= percentage of correlational information contained in the system, H=height of the echo-signal. The following graph (directed graph 21) describes the causal links between the relevant variables; and illustrates how the manipulation M_1 (diluting the sample) serves as a testing intervention upon BDC.

$$M_1 \longrightarrow BDC \longrightarrow CI \longrightarrow H$$

Directed Graph 21. Application of M_1 (diluting the glycerine sample) to the cause postulated by the fictitious explanation.

Note that the fictitious explanation is false by our lights. For the sake of the argument (to be developed in section 7.3) we have supposed that the fictitious explanation *ignores on purpose* two causal factors, namely, the magnetic fluctuations and the lattice mobility. Nevertheless, in this case supposing the fictitious causal structure to be true, the manipulation M_1 surprisingly fulfils every single condition **INi-iv** and hence counts as a testing intervention.

Condition **IN-i** is fulfiled because there are no alternative causes of the intervened variable (BDC) whose influence needs to be blocked. In other words, as there are no alternative causes of BDC correlated with M_1 nothing prevents this manipulation from fulfiling the condition **IN-i**.

Condition **IN-ii** is fulfiled too because the manipulation M_1, i.e. diluting the sample, is not an action capable *by itself* of driving the system toward equilibrium. Translating this into the m-theory terms, the manipulation M_1 fulfils **IN-ii** because M_1 is not a direct cause of the effect (in this case, H). The directed graph 7 illustrates the m-theory forbids for interventions to be direct causes of the effect.

As shown in the directed graph 8, according to condition **IN-iii** an intervention I must be causally connected with the effect E through alternative paths. In other words, the intervention I must affect the effect E only through the path containing the putative cause C. And finally, according to

condition **IN-iv**, the values of variables Z_i in alternative paths must remain fixed while the intervention is carried out. In the fictitious explanation (illustrated in directed graph 21) both **IN-iii** and **IN-iv** are fulfiled because there are no alternative paths connecting the manipulation M_1 and the effect H.

M_1 meets all the conditions **INi-iv** and is therefore an intervention. And the postulated causal links illustrated in directed graph 21 remain invariant under such an intervention. Shall we prefer this simpler explanation to the complex one? Obviously we cannot since the fictitious explanation represents a fake experiment that is far simpler than the actual spin-echo experiment we face in reality. But since interventions are mere *tests* of causal relations, we would not like to rule out the causal relationships postulated by classical interventionists just because we do not have the appropriate interventions. I will leave this question open now, and further discuss it in the last section of this chapter.

6.1.2 Testing the GRW-explanation

Let us now reconsider David Albert's explanation (as expressed in section 5.3). Albert postulates the quantum GRW-collapses as the only cause of the echo-signal decay. Let us now consider some manipulations that change the rate of the collapses. The GRW-explanation is presumably independent of the environment perturbations. So all interventions related with the environment will also become irrelevant in this case. Besides, as mentioned at the end of section 5.3, we cannot stop or generate the collapses freely at will. They are assumed to be absolutely random and there is no way of controlling them by means of a previous cause. It seems then that, as the directed graph 15 illustrates, the only way of changing the rate of the GRW-collapses is by diluting the sample. Let us see how this presumed intervention works.

According to the GRW theory, there is a direct dependence between the frequency of the collapses and the number of particles in a given space: "the spontaneous localization mechanism [GRW-collapses] is enhanced by increasing the number of particles which are in far apart spatial regions" (see Ghirardi et al, 2007, [46]: sec 5). It then folllows that, increasing the distance

between the molecules of glycerine, by adding some water to the sample, should reduce the number of collapses.

The experimental results obtained with different concentrations of water and glycerine are plotted in Fig.15 (on p.105). And the GRW-based explanation is able to account for them because it entails that the rate of GRW-collapses (real collapses) decreases when the glycerine is diluted in water slowing down the system's approach to equilibrium (see Ghirardi et al, 2007, [46]: sec 5).

Unfortunately in this case, the manipulation consisting on diluting the sample (defined above as M_1) does not fulfil the condition **IN-i**. According with this condition, the manipulation M_1 should first of all be a cause of the variable $GRWC$.[5] However, the GRW-collapses are not properly caused by the dilution of the sample. This manipulation simply decreases the rate of GRW-collapses.

Following Woodward's suggestion *only for this kind of violations of condition **IN-i***, we can nonetheless consider M_1 as a testing manipulation, focusing on the fact that it constitutes a valid way of controlling the frequency of the collapses, and ignoring the fact that M_1 is not a proper cause of the variable $GRWC$.

The causal relationships postulated by the GRW-based explanation (illustrated in the directed graph 14, on p.93) remain invariant under this unique testing intervention. Therefore, the outcome of our analysis applying the m-theory to this case is that the GRW-based explanation is "minimally explanatory" in the sense that it meets the "minimal condition for successful causal explanations" defined in section 4.5.3.

[5]Woodward's methodology may seem circular in appealing to *causes* in order to test *causal* relationships. However, as we mention in section 4.3.1, Woodward argues that this is not "vicious cirularity" in the sense that no causal information is required about the relata under study.

6.1.3 Testing the decoherence-explanation

Our third and last explanation under study, the decoherence-based explanation, is illustrated in directed graph 16 (on p.97). In order to analyze it in terms of the m-theory, let us propose some manipulations designed to change the values of the internal and the external decoherence (represented by variables DI and DE respectively).

The manipulation M_1, consisting in diluting the glycerine in water, is again a good candidate. If we add water to the sample, the distance between the glycerine molecules increases and each spin gets away from the others. Hence, if the decoherence approach is correct, the decoherence rate decreases, thus slowing down the system's approach to equilibrium.

The experimental results (plotted in Fig.15 on p.105) are once again compatible with the explanation. It is remarkable, though, that the delay in the relaxation time predicted by the decoherence approach would be more significant than the delay predicted by the GRW approach (see Hemmo & Shenker, 2005, [55]:643). The reason is that from the decoherence perspective separating the spins produces a double effect: it decreases both spin-spin decoherence and decoherence between the spins and the environment.

Is there any other method for reducing decoherence? And furthermore, is it possible to completely avoid it? Answering these questions has become one of the main goals of researchers in quantum information (Uhring, 2009, [97]). Reducing decoherence enables us to keep quantum information for longer times. And completely avoiding decoherence would be idyllic for building the so-called quantum computer. For these reasons, and independently of its philosophical relevance for the foundations of statistical mechanics, suppressing decoherence in the spin-echo experiments has recently become a research field of great interest. As a result, during the last few years, several different sequences of radio-frequency pulses have been put forward seeking to reduce decoherence as much as possible or for as long as possible (see for example Viola, 1998, [99]; Khodjasteh, 2005, [62]; Capellaro, 2006, [20]; Uhring, 2007, [96] and 2009, [97]).

Similar multi-pulse techniques had been proposed much earlier (1950's) by Carr and Purcell [21] and soon after improved by Meiboom and Gill [70]. However, these techniques were interpreted in previous decades as ways of improving the resolution and precision of the experiments. And they were focused on getting better information about the chemical composition of a

given sample. In fact, the so-called Carr-Purcell-Meiboon-Gill sequence of pulses has been widely applied in medicine for producing images of body structures in order to detect brain tumours, osteoporosis, heart diseases, etc. The novelty of more recent work (for example Uhring, 2007, [96] and 2009 [97]) is that the improved technique for "dynamic decoupling" is explicitly proposed as a way of suppressing decoherence.

In terms of the runner's metaphor (explained in section 2.2), the essence of the multi-pulse techniques consists in "not letting the racers run too far". Imagine that all the runners are placed in the starting line. The racers begin to run at different velocities but *before they lose the alignment* (perhaps just an instant after the race begins) we stop them and ask them to run back. Again we allow them to run only a small distance. And soon multiple races are conducted. As a consequence, the original distribution remains almost unchanged.

This situation has been experimentally developed with spins obtaining successful results. As illustrated in Fig.16 and Fig.17, the improved sequences of r-f pulses bring about less diffused spin-echo patterns in which the spins' system reaches the equilibrium in a longer relaxation time. In other words, the decay of the signal is softer, the diffusion is largely circumvented and the spins are able to keep the phase-memory for much longer (2 seconds instead of 0.2 seconds in these figures).

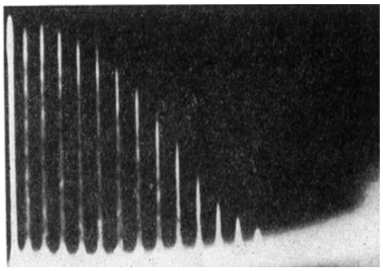


Fig.16. Original experiment.

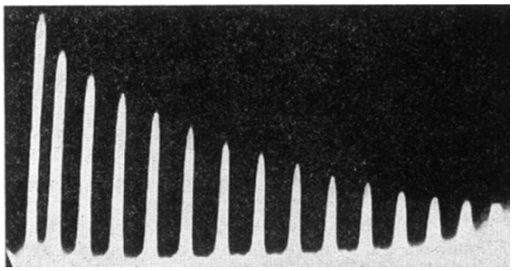


Fig.17. Modified experiment.[6]

In the latest models (Uhring, 2007, [96] and 2009 [97]).) the optimized sequence of pulses is said to suppress decoherence even more efficiently than the so far known sequences of equidistant pulses. Two additional advantages have been supposedly achieved. First, the results are less sensitive to thermal effects. And, second, the number of pulses required to get a certain prolongation of the relaxation time can be much smaller (see Uhring, 2009,

[6]Images taken from Carr & Purcell, 1954, [21].

[97]:100504-4).

As mentioned above, researchers in quantum information aim to find the way to avoid decoherence for longer times. In fact it would be ideal for them to completely avoid it. For our analysis in terms of the m-thoery it would be also very convenient to find a manipulation such that decoherence does not take place during the whole experiment. In such a situation, according to Hemmo and Shenker's decoherence-based explanation, the system should never approach equilibrium. One way of doing so is setting up the system in a "no-decoherent quantum state". This kind of state has not been prepared and, as at present, it is not considered an experimental possibility yet. However, it remains a theoreticaly conceivable quantum state, and thus counts as a manipulation in our analysis.

In sum, we have proposed the following prima facie possible manipulations for testing the decoherence-based explanation:

- Reducing the viscosity of the sample (already labelled as M_1);

- Reducing decoherence through multi-pulse sequences (M_5);

-The "no-decoherent" quantum state (M_6) (hypothetical manipulation).

Let us now run through the list and check if they fulfil the m-theory conditions for intervention.

According to **INi-iv** the values of all the variables C_i causally connected with the effect E must be fixed when an intervention is performed. This means that in order to fulfil **INi-iv** every time we vary the value of the external decoherence DE the value of the internal decoherence DI should remain fixed. This is however impossible. Due to the coupling between the spin and the position, the internal and the external decoherence, and thus the variables DI and DE associated to them, are not independent from each other. This means that it is impossible to block one kind of decoherence in order to effectively intervene in the other one.

Thus, as a consequence of the correlation between DI and DE, all the above suggested manipulations for testing this approach face a problem meeting the condition **IN-iv**.[7] It is notable, though, that this difficulty can be

[7]The intervention I must be performed in such a way that the other variables C_i in alternative paths, causally connected to the effect E, remain fixed.

overcome by replacing the variables DI and DE by a single variable DH including both of kinds of decoherence. In that case, the set of relevant variables would be $\boldsymbol{V}' = \{DH, H\}$; where DH=degree of decoherence and H=height of the echo-signal. And the causal links would be represented by the directed graph 22.

$$DH \longrightarrow H$$

Directed Graph 22. New and simpler graph for decoherence approach.

Representing the decoherence-based approach by means of this new and simpler directed graph, the manipulations M_1, M_5 and M_6 fulfil all the conditions **INi-iv**. Let us see this in detail.

The manipulations no longer violate condition **IN-iv** because now there are no variables connected with H through alternative paths whose value needs to be fixed. In other words, there is no correlation between DH and any other putative causes of H.

Conditions **IN-i** and **IN-iii** are fulfiled, because there are neither alternative causes of DH whose influence needs to be blocked (condition **IN-i**) nor alternative paths through which the cause DH may be affecting the effect H when the manipulations are performed (condition **IN-iii**).

And condition **IN-ii** is fulfiled if we accept that diluting the sample (M_1), changing the sequences of pulses (M_5) or setting the system in a non-decoherent quantum state (M_6) are not direct causes of H, i.e, the condition is fulfiled if we accept that the manipulations only affect H trough their influence upon decoherence.

Once manipulations M_1, M_5 and M_6 have met the four conditions they count as *testing interventions* under which the causal relationship postulated in the directed graph 22 remains invariant.

The conclusion of this section is the following. The decoherence based approach provides a causal explanation of the spin-echo experiments (expressed in the directed graph 16) but we cannot test this explanation by means of any of the manipulations presently available. Internal and external decoherences are so intricately related that every attempt to manipulate one of them will necessarily affect the other. However, in analogy with the case of the new interventionist classical explanation (ilustrated in the directed graph 20)

we found out that by simplifying and idealizing the underlying causal structure, we come up with theoretical manipulations that fulfil the conditions on interventions laid down by the m-theory.

6.2 Answering what-if-questions

The m-theory provides us with three criteria to assess the *explanatory depth* of causal relationships (see section 4.5.2, p.73-74). The first two criteria require the causal relationship to be invariant under a wide and diverse range of interventions. Our analysis has been so far focused on whether the explanations of the spin-echo experiments meet these two criteria. Now we turn to assessment of whether the explanations fulfil criterion 3. In other words, we analyze if the explanations are able to answer *what-if-things-had-been-different-questions* (w-questions).

All the above proposed manipulations can be associated with a corresponding counterfactual question (or w-question). For example, M_1 is associated to the w-question: what would happen to the height of the consecutive echo-signals if we dilute the sample? M_5 is associated to the w-question: what would happen to the height of the echo-signals if the initial quantum state is no-decoherent? And so on. This means that, in the previous sections, we already obtained much valuable information about whether or not the different explanations of the spin-echo experiments are capable of responding to several w-questions.

In addition to these w-questions associated to our proposed manipulations, there is a w-question that deserves special attention. The question I am referring to is related to the explanation of how Mr. Jones did not get pregnant because he took contraceptive pills (for more details see section 4.6.1). The moral of this example is the following. In order to avoid false explanations, it is always convenient to examine whether a given explanation under study is able to tell us what would happen if the postulated causes were absent.

This question (let us call it *"the full absence w-question"*) has been relevant in the philosophical debate about statistical mechanical interventionism. More precisely, interventionism has been criticized for not been able to show that the thermodynamic behaviour would disappear in absence of the interventionist causal mechanism. For example, Bricmont rejects interventionism with the following claim: "imagine a system being more and more isolated.

Is irreversibility going to disappear at some point? I cannot think of any example where this could be argued." (Bricmont, 1996, [13]:147 quoted in Shenker [89]:7).

Certainly, classical interventionists have no proof that the system will stop behaving in a thermodynamically normal way in the absence of environmental perturbations. Nevertheless, as long as the hypothetical perfect isolation is not against the laws of physics, *the full absence w-question* is genuine and, classical interventionists do provide an answer. Namely, they predict that the system will not approach equilibrium (see Ridderbos & Redhead [82]:1252-1253).

Classical interventionists could argue, for instance, that as soon as there is matter, there are perturbations. In Blatt's model, for example, perturbations arrive from intermolecular interaction and this interaction only disappears when there are no molecules, or when they are so few that they do not constitute a thermodynamic system anymore. So interventionists could argue that the reason why Bricmont cannot imagine a system without perturbations is that such a system does not exist in actual fact.

It is worth emphasizing that the decoherence approach would provide a different answer to this question. Let us imagine, as Bricmont suggests, a system becoming isolated. Hemmo and Shenker could argue that the system continues to behave thermodynamically due to the fact that internal decoherence (DI) is still acting within the system itself. So, even though the relaxation time would be significantly longer, the system would still approach equilibrium.[8]

For this reason *the full absence w-question* directed to the decoherence approach should be rather formulated as follows: What would happen if both internal and external decoherences are simultaneously blocked? The hypothetical manipulation M_6 (setting the sample in a no-decoherent initial quantum state) would lead precisely to this kind of situation. And the defenders of the decoherence approach would expect the system of spins to behave in a thermodynamically abnormal way under such circumstances.

In sum, the explanations here analyzed successfully meet criterion 3 be-

[8] As mentioned in section 3.1.1, this feature allows the decoherence approach to account for the behaviour of the universe as a whole in a more appropriate manner than classical interventionism.

cause they provide answers to several w-questions. In particular, it seems that both classical and quantum interventionist approaches do provide answers to *the full absence w-question*. It is worth mentioning though, that among all the manipulations proposed in this chapter, only those that are *entirely hypothetical* would be able to prevent the thermodynamic behaviour. And certainly none of them would be capable of reversing it (in the sense of getting the initial intensities of the echo-signal after the decay). This may seem fairly suspicious to the detractors of interventionism.

The GRW approach does not fare any better. For the dissolving sample (via manipulation M_1) actually helps to decrease the GRW-collapses frequency. However, none of the proposed interventions so far is able to completely prevent the collapses.

Huw Price has also rejected interventionism, and his argument is related to *the full absence w-question*:

> "To say that some asymmetric mechanism causes entropy to increase is to say that in the absence of that mechanism, entropy would not increase. Yet Boltzmann claims to have shown that for most possible initial microstates, entropy would increase anyway, without any such asymmetric mechanism. So friends of such mechanisms [namely, interventionists] need to say that Boltzmann is wrong –that the universe (probably) starts in *a microstate such that without the mechanism, entropy would not increase*. It's hard to see what could justify such a claim." (Price, 2004, [80]:228, my emphasis)

I believe that the "no decoherent initial quantum state" described above (manipulation M_6) corresponds to the kind of microstate that Price is asking for. Namely, a initial state such that the putative causal mechanism (decoherence) is absent and hence avoids the system's approach to equilibrium.

6.3 Summary of Results

The following table (Fig.18) summarizes the results that we have obtained in analyzing explanations of the spin-echo experiments by means of the m-theory.

	Classical	Classical Simplified	GRW	Decoherence	Decoherence Simplified
M_1: Diluting	W-answer IN-i failed	★ W-answer Invariance	W-answer Minimal	W-answer IN-iv failed	★ W-answer Invariance
M_2: Warming up	IN-ii failed		Irrelevant		
M_3: Inhomo-Field	W-answer IN-i failed	Irrelevant			
M_4: Homog-Field	W-answer IN-i failed	Irrelevant			
M_5: Multi-pulses	W-answer IN-i failed	★ W-answer Invariance	W-answer Minimal	W-answer IN-iv failed	★ W-answer Invariance
M_6: No-decohere	Irrelevant			W-answer IN-iv failed	★ W-answer Invariance

Fig.18. Results Table.

The star symbol ⋆ stands for "being invariant under the manipulation in turn".
And "Minimal" stands for meeting the minimal condition for successful explanation.

In the row corresponding to manipulation M_1 the term "w-answer" is written in every column. This indicates that all the approaches analyzed in this thesis account for the counterfactual situation in which the glycerine sample is diluted in water. Hence, they offer an answer to the w-question associated with manipulation M_1. More precisely, all the approaches predict that the echo signal decays in a longer relaxation time if the experiments are performed with diluted samples. This prediction is in accordance with the experimental results (plotted in Fig. 15 on p.105).

If besides this *qualitative* prediction some *quantitative* predictions were provided, a crucial experiment could be performed to investigate which among all the explanations describes the behaviour of the spins' system more adequately. As the putative causes are supposed to affect the system in different ways, each approach should predict different relaxation times; thus comparing them experimentally will be possible. Unfortunately, none of the approaches here analyzed has provided data about the exact values of the relaxation times that should be expected for different dilutions.[9]

Despite the fact that the explanations provide satisfactory qualitative "w-answers", in fulfilment of criterion 3 for explanatory depth, the explanations fail to met criteria 1 and 2. The reason is that it was not possible to consider M_1 as an intervention, neither for the classical explanation nor for the quantum-based ones: in every case at least one condition among **INi-iv** (but a different one in each case) was violated.

We discovered that this difficulty can be overcome, in both the classical and the quantum cases, by replacing the correlated causes for a single cause[10]. Once we do so, M_1 fulfils all the conditions **INi-iv** and hence counts as a testing intervention. Furthermore, the causal patterns postulated by the explanation under study turn to be invariant under this testing intervention. This desirable result (highlighted in the table with a star symbol $\star$) is what any explanation would wish to obtain after being evaluated with the m-theory.

Regarding the second manipulation M_2 (increasing the temperature of the sample) we found out that it is not an adequate intervention for analyzing

[9]Hemmo and Shenker commented, in a conversation, that formally deriving this kind of specific values of relaxation times is difficult given the current theories of decoherence. However, they think that making such computations will become possible in the future.

[10]For example, in the decoherence case, if external and internal decoherences (DE and DI respectively) are replaced by a single variable DH such that $DH = DE + DI$.

the explanations of the spin-echo experiments. The reason is that changes in the temperature affect the whole evolution of the system. And this means, according to the m-theory, that M_2 violates condition **IN-ii**. More precisely, the condition is violated because the manipulation is a direct cause of the decay of the echo-signal, which is the event that we want to explain (see directed graph 7 on p.62 which illustrates condition **IN-ii**).

In fact, as Martin [69] and Hughes [58] comment, when the system is manipulated (for example, by diluting the glycerine sample in water), care must be taken that all the measurements of relaxation times are carried out inside a laboratory with exactly the same temperature, and checking that the initial temperature of the sample is always the same. Ignoring the fact that the results are temperature sensitive may lead to inconclusive data.

The third and fourth manipulations are only relevant to the classical case. Both violate condition **IN-i** for not being able to block Brownian motion and thermal interactions while intervening upon the spins' system. If manipulations M_3 and M_4 had fulfiled **IN-i**, they would have counted as interventions. And if the causal relationships postulated by the classical approach would have remained invariant under those interventions, the classical interventionist explanation of the spin-echo experiments would had met criteria 1 and 2 of explanatory depth.

Nevertheless, since the classical approach provides "w-answers" to the counterfactual situations posted by these magnetic manipulations (M_3 and M_4), it successfully meets criterion 3 of explanatory depth.

Let us turn now to the fifth manipulation. Both classical and quantum approaches are able to provide "w-answers" to the counterfactual situation obtained by applying multi-pulsed sequences in the spin-echo experiments (manipulation M_5). As mentioned before, the decoherence rate is said to decrease under such manipulation. We have not mentioned, though, that the classical approach could also provide an account of the experimental results obtained through manipulation M_5 (illustrated in Fig.17 on p.114). It could be argued, for example, that shortening the time between the pulses also reduces the time during which the diffusion factors are acting act upon the system. The effects of the Brownian motion, the magnetic fluctuation and the thermal interactions are restrained, and the system's approach to equilibrium is slower and softer.

The fifth manipulation is probably the most interesting one, together with M_1, in the sense that they both are applicable to all explanations. Unfortunately, due to the lack of qualitative predictions, M_5 does not genuinely count as an intervention either except in a purely qualitative sense, where only the hypothesis of an indeterminate causal relation between EP and H remains invariant. If the corresponding quantitative predictions were provided, M_5 would be useful for comparing the empirical adequacy of the explanations.

Additionally, M_5 is related to the following philosophically relevant question: Shall we consider the pulses as a part of the system, as a part of the environment, or as a simple background condition?

All throughout our analysis, we have assumed that the system consists in a collection of nuclei with their spins. The radio frequency pulses have been included in the background conditions, as a part of the experimental device (see Fig. 12 on p.88), and we have taken those pulses as a way of intervening upon the system of spins.

However, if we define a new system of study as S'= $spins{+}magnetic field{+}rf pulses$, the system seems to evolve *without the help of any environmental perturbation*. We could even define a system S" including the electricity consumed on generating the rf pulses, or moreover a S"' including the laboratory, which is warmer after the experiments. A typical argument against interventionism consists in pointing at this regress in defining the relevant system, which only ends with a system consisting of the universe as a whole (see 3.1.1; Shenker, 2001, [89]:16; or Frigg, 2007, [41]:161).

The last manipulation M_6 (preparing the system in a no-decoherent initial state) turned out to be relevant for analyzing the simplest version of the decoherence-based explanation. It is worth mentioning though, that M_6 is not applicable to the GRW-based explanation because the GRW-collapses are supposed to drive the system to equilibrium *independently* of the quantum initial state. And, in the classical case, M_6 is not even relevant –because it is conceived in a framework where decoherence is not taken into account at all.

In terms of the m-theory this is an advantage of the simplified decoherence explanation over the rest of the explanations because it provides us with an additional way (namely, manipulation M_6) of controlling its proposed cause DH. Consequently, it is capable of answering a *what-if-things-have-been-different*-question about which the other explanations have nothing to say.

6.3.1 Dilemma regarding the criteria of explanatory depth

Finally, I would like to focus on a dilemma that follows from the analysis developed in this chapter. The analyzed explanations answer a good number of w-questions, and yet it seems that, it is not easy to find *testing interventions* to prove their explanatory depth. This means that the explanations successfully meet criterion 3 required by the m-theory to qualify as explanatorily deep, but fail to meet criteria 1 and 2 (see p.73-74).

All the explanations that we have analyzed in this thesis have valuable qualities from the manipulability perspective. For example, both classical and quantum interventionism offer mechanisms to control and study the spin-echo system; and this is precisely the kind of explanation the m-theory is looking for (see the raven's example in section 4.5.4). Besides, as shown in the table of results, all the explanations predict what would happen under several counterfactual variations (or manipulations) of the spin-echo experiments (see the term 'W-answer' in Fig.18 on p.120). Hence, they provide answers to several *w-questions*. Intuitively, the m-theory should then evaluate them as reasonably *deep* explanations of the spin-echo experiments in fulfilment of criterion 3.

Nevertheless, we have found several difficulties in proposing manipulations that control (either actually of hypothetically) the values of the postulated causes and, at the same time, fulfil the conditions for interventions **INi-vi**. Despite the fact that many manipulations have already been performed in the best laboratories of the world, they turn out not to comply with the conditions of the m-theory. Even M_1 and M_5, which have been proposed as mechanisms to contrast the empirical adequacy of the different explanations, they are not considered interventions within the m-theory, at least not for the interventionist explanations as we have expressed them in chapter 6.

We have discovered, though, that these violations of the conditions **INi-iv** can be defeated by simplifying and idealizing the explanations, i.e. by replacing correlated causes with a single cause. Actually, the versions of the explanations so simplified obtained the best results after being evaluated

with the m-theory. The reason is that, in addition to meet criterion 3 for explanatory depth, they also meet criteria 1 and 2.

Why is this? Why are these simplified explanations *deeper* than the original ones? My conjecture is that the original interventionist explanations postulate causes that are problematic in the light of the manipulability theory. Classical interventionism postulates causes whose possible interventions are correlated with each other. And quantum-based interventionism postulates causes that are correlated with each other.

Chapter 7

Conclusions

One could have at least four different reactions to the results in the previous chapter and the discussion so far. They differ in the attitude they take regarding the defining conditions on interventions **INi-iv** (on p.63). They cannot be precisely summarized as philosophical claims, and are best understood as attitudes, or stances towards these conditions, and concomitantly, towards the criteria of explanatory depth proposed in the manipulability theory. I call them the strict, flexible, simplifying, and critical attitudes.

7.1 The strict attitude

The first reaction would adopt a very strict attitude towards the conditions **INi-iv**. Only those manipulations fulfiling the whole set of conditions will be considered to be interventions. And only those explanations providing causal relationships which are invariant under such interventions will be considered appropriate.

An equally strict attitude would be adopted towards the criteria for explanatory depth. If there is not even one intervention under which a putative causal relationship holds, that relationship does not fulfil the minimal condition for successful causal explanation (see p.75). In other words, the relationship falls below "the threshold of explanatoriness" (see p.76). This means, in turn, that such a relationship is unable to meet criterion 1 and criterion 2 for explanatory depth (on p.73) for those criteria require invariance under *a wide set* of interventions (see p.73). According to this strict attitude, the larger the range of invariance, the deeper the explanation is. If there is no invariance, there is no explanatory depth. Full stop.

Adopting this attitude one would conclude that the explanations of the irreversible behaviour of the spin-echo system, as expressed in chapter 6 and given the results of chapter 7, are not even minimally explanatory [1]. This conclusion would be welcomed by critics of interventionism. They could argue, for example, that interventionist explanations are unable to show adequate interventions of the spin-echo system since interventionism approaches the irreversibility problem the wrong way. And, as the case of the multi-pulse sequences shows, interventionists still need to solve the difficulty of defining the system and the environment unambiguously.

7.2 The flexible attitude

Alternatively, one may adopt a more permissive attitude about the conditions **INi-iv**. Something like the following could then be argued: If we dismiss any manipulation failing to meet a single condition (like in the previous section) we are not acting in harmony with Woodward's suggestion of considering the conditions **INi-iv** only as *regulative ideals* (see Woodward, 2003, [106]:114)

It should be recall that if one tries to represent the GRW-based approach to irreversibility in terms of the m-theory, one faces the problem that the putative cause –namely the GRW-collapses– can not be manipulated. The reason is that GRW-collapses occur in a properly stochastic way. As a consequence, any putative intervention I fails to meet condition **IN-i** (according to which I must be a direct cause of the putative cause C). One may simply conclude that there is no way of testing this specific causal relationship. Or one may follow Woodward's suggestion (see p.68) and solve the difficulty by identifying the appropriate causal relata and interventions. In this particular case, we decided to define the variable $GRWC$ (on p.93) as representing "*the rate* of GRW-collapses". Changing the viscosity of the sample directly causes changes in *the rate* of GRW-collapses. Therefore, the variable $GRWC$ can be manipulated, for example, by diluting the glycerine sample (M_1). And, thus, the problem of finding interventions for this case is solved.

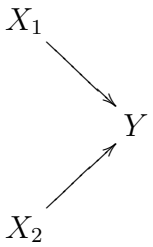
There are three aspects of this case that I want to stress. First, if a

[1] See the Minimal Condition for Successful Causal Explanation on p.75.

manipulation fails to meet any one condition, this does not mean that such manipulation is completely useless to test a specific causal relationship under study. Second, from the fact that a manipulation fails to meet a condition it neither follows that the examined causal relationship is false nor that the explanation appealing to that relationship is not reliable. And third, if one believes that a causal relationship is genuine, but it turns out that all the possible manipulations fail to meet one of the conditions **INi-iv**, then one should try to redefine the causal relata and propose new manipulations to do justice to the causal relationship.

Someone adopting the strict attitude would still agree with all these considerations. The novelty is that, in accordance with the flexible attitude, this process of redefining the variables and the manipulations, *could also imply re-interpreting or relaxing the conditions* **INi-iv**. Let me illustrate this by means of the following example (taken from Woodward, 2003, [106]:323).

Suppose that a biologist studies the growth of a plant and finds a causal relationship between the height of the plant and the amount of water and fertilizer that the plant receives. Let X_1=the amount of water, X_2 the amount of fertilizer and Y=the height of a plant. According to the biologist the functional relation $Y = aX_1 + bX_2$ and the directed graph 23 (below) represent the causal relation between these variables.



Directed Graph 23. Causal relationships between the amount of water X_1, the amount of fertilizer X_2 and the height of the plant Y.

And suppose that the only way of adding fertilizer to the plant implies adding some water too, because most of the available fertilizers are diluted in water; and even if we get some solid fertilizer and place it inside the flowerpot, the plant does not absorb it unless we add some water. This means that

manipulations of the variable X_2 unavoidably yield changes in the value of X_1 in violation of requirement **IN-iv**.[2]Therefore, there are no possible testing interventions upon the variable X_2.

The biologist, however, insists that the fertilizer is a contributing cause of the plant's growth and argues that, even though it is not possible to intervene upon X_2, the causal relation between X_2 and Y is genuine because it is invariant under a range of "controlled changes" in both X_1 and X_2. By measuring the amounts of water and fertilizer added to the plant, it is possible to know both the change in the variable X_1 (call it ΔX_1) and the change in the variable X_2 (call it ΔX_2). And the biologist argues that the functional relation is invariant under manipulation because the total change in the variable Y is exactly what the functional relation says it is, namely $\Delta Y = a\Delta X_1 + b\Delta X_2$ (see Woodward, 2003, [106]:324).

Following the m-theory we must specify the background conditions and the intervals of values for the variables in our example. Let us say that the amount of water X_1 varies within an interval of [1,3] liters and the amount of fertilizer varies within the interval [0.5,1] kg. per month. And suppose that, under these background conditions, the functional relation is invariant under changes in the variables X_1 and X_2. Someone adopting the flexible attitude would say that an explanation of the plant's growth appealing to this functional relationship fulfils criterion 1 of explanatory depth –or a flexible version of criterion 1 that admits invariance under the changes explained above, even though they are not interventions *stricto sensu*.

If, by an analogous argument, we flexibly applied conditions **IN-i** and **IN-iv** to both the classical and the quantum interventionist explanations of the spin-echo experiments, they would be evaluated as possessing some explanatory depth. The causal relationships postulated by interventionists are as if the variables X_1 and X_2 in the plant's example were intertwined in such a way that changing the value of X_1 leads to changes in the value of X_2 *and vice-versa*. In the classical interventionist explanation this is due to the fact that putative interventions are correlated with each other in violation of requirement **INi**[3] And in the decoherence-based approach this is due to

[2]According to **INiv** an intervention I upon one cause C should be performed keeping fixed the values of all the alternative causes of the effect E, should they exist.

[3]The requirement **IN-i** is not only violated when the manipulation is not a proper cause of the variable C, but also when the manipulation is performed in such a way that alternative causes of C are not switched off (see directed graph 5 on p.61).

the fact that the two putative causes are correlated to each other in a causal structure that is precisely as the causal structure of the plant's explanation (see directed graph 23).

This means that, in order to produce an argument analogous to the biologist's argument, interventionists should provide us with quantitative predictions and functional relations of the kind $H{=}f(EP)$ and $H{=}f(DI, DE)$. They should also provide the specific intervals of values that the putative causes may take (in other words, we need the numerical values of BDC_{max}, DE_{max}, DI_{max}, etc.). This would enable us to assess if the functional relations correctly describe the values of the height of the echo (H) under different changes in the putative causes (although those changes are not interventions strictly speaking). And, concomitantly, this would enable us to show to what extent the explanations meet criteria 1 and 2 for explanatory depth. Unfortunately, having reached this point, interventionists have not provided those functional relations. We only count on the qualitative directed graphs in order to assess the explanations. Therefore we can only show that the interventionist explanations are able to answer counterfactual questions meeting criteria 3 of explanatory depth. Nevertheless, the flexible attitude described in this section would allow us to argue that interventionists have offered a genuine causal account of the spin-echo experiments.

As we have just seen, according to the flexible attitude, in very particular cases some manipulations may be useful to assess the import of causal relationships *despite the fact that those manipulations are not interventions in the strict sense*. The disadvantage of adopting this flexible attitude is the ambiguity in deciding when the conditions for intervention **(INi-iv)** must be respected and when they can be relaxed or re-interpreted. Some judgments will need to be made. For example, it seems that we could be permissive with the magnetic manipulations M_3 and M_4 which are not able to switch off the influence of alternative causes. By contrast, it seems that the manipulation of the spin-echo experiments that consists on raising the sample's temperature (M_2) should be dismissed straightaway due to the fact that it is a direct cause of the whole system's evolution. In other words, it seems that not all violations of the conditions *INi-iv* are equally serious.

As the following table shows (see Fig.19 below), a flexible attitude towards both the classical and the quantum-based explanations of the spin-echo experiments would restore some *explanatory depth*.

	Classical	Classical Simplified	GRW	Decoherence	Decoherence Simplified
M_1 Diluting	⋆ W-answer (IN-i)'	⋆ W-answer Invariance	W-answer Minimal	⋆ W-answer (IN-iv)'	⋆ W-answer Invariance
M_2 Dismissed					
M_3 Inhomo-Field	⋆ W-answer (IN-i)'	Irrelevant			
M_4 Homog-Field	⋆ W-answer (IN-i)'	Irrelevant			
M_5 Multi-pulses	⋆ W-answer (IN-i)'	⋆ W-answer Invariance	W-answer Minimal	⋆ W-answer (IN-i)'	⋆ W-answer Invariance
M_6 No-decohere	Irrelevant			⋆ W-answer (IN-iv)'	⋆ W-answer Invariance

Fig.19. Flexible Results. The symbol ⋆ stands for invariance and (IN)' stands for the flexible version of the condition in turn

7.3 The simplifying attitude

A third attitude would insist on the use of the simplified versions of the interventionist explanations. The reason being that, according to the results obtained in chapter 7 (see Fig.18 on p.120), the simplified decoherence-based explanation postulates causal relationships that turn out to be invariant under a wide set of interventions –understood strictly as those that satisfy the conditions **INi-iv**.

This means, in turn, that the decoherence-based explanation of the spin-echo experiments meets *all* the criteria of explanatory depth. It fulfils criterion 1 because it appeals to a causal relationship that remains invariant under interventions for different values of the putative cause DH (decoherence rate). It fulfils criterion 2 because it provides several different ways of intervening on that putative cause: changing the viscosity of the sample, changing the frequency of the radio-frequency pulses, and setting the system in a no-decoherent initial state. And finally, it fulfils criterion 3 because it is able to explain what would happen if things had been different. Therefore, the decoherence approach in its simplified version offers the *deepest* explanation of the irreversible behaviour of the spin-echo system among all the explanations analyzed in this thesis.

Adopting this attitude one may argue that the fact that the simplified versions of the explanations turned out to be *deeper* than the complex ones in this analysis, is to be interpreted as an advantage of the m-theory. Simplicity is often considered to be an epistemic value of good explanations, and the m-theory seems to classify the simplest explanations as the deepest ones.

Although this attitude could be applicable in the case of the simplification suggested for the decoherence-based explanation, it would be mistaken in the classical case. It should be recalled that the simplified classical version (in directed graph 21 on p.110) has been deliberately constructed in an idealized fashion since it ignores two relevant causal factors and keeps only a single cause. More precisely, two causes of environmental interaction have been ignored (magnetic fluctuations, and thermal interactions). It assumes that the only cause of the echo-signal decay is the Brownian motion. It is worth noting though that the ignored causes have already experimentally proved to be relevant, and both are considered as main causes of dissipation in the scientific literature from the 1950's to the 1990's (see, for instance, Bloembergen, 1984, [10], Hahn [50] and Nyenhus [74]). Yet, we find that, by simply pretending that they do not exist, we obtain what looks like a *deeper explanation.*

7.4 The critical attitude

The fourth and last option is to conclude that the analysis presented in this thesis has revealed some weak points in Woodward's m-theory.

According to the m-theory's criterion 3 of explanatory depth, both the classical and the quantum-based interventionist explanations of the echo-signal decay are *deep explanations* given the wide set of counterfactual situations they account for. However if we apply the m-theory to analyze these explanations, they rather turn out to be quite superficial or not even minimally explanatory. Both explanations fail to meet criteria 1 and 2 for explanatory depth because they are not invariant under intervention. Why are explanations capable of answering to a wide set of *w-questions*, and hence good, in terms of one criterion, so lowly rated in terms of the other two criteria of explanatory depth? Would it not be desirable that the three criteria worked in harmony?

Similarly, even though according to the m-theory we would expect the original classical explanation to be *deeper* than the simplified and false one (because it answers more w-questions), our results do not allow us to choose the original explanation as the deeper one. Let me explain this with an example.

The original classical interventionist explanation of the echo-signal decay (as expressed in the directed graph 12 on p.87) is more detailed and takes into account more control factors than the simplified one (expressed in the directed graph 21 on p.110). In this sense, comparing these two classical explanations is analogous to comparing the two explanations of the car's acceleration. The first and simple explanation says that the acceleration of the car depends on the angle of the gas pedal. The second and complex explanation appeals to the whole internal mechanical system of the car, giving details about how the motor burns the petrol, the carburetor mixes liquid fuel with air, how the gas pedal pushed by the conductor's foot is connected with all the mechanical parts, etc.[4] According to the m-theory, the second explanation of the car's acceleration is *deeper* than the first one, because it allows us to control and manipulate the explained event in several different ways. And, as far as I can see, this is exactly what the original and com-

[4]This example was taken from Haavelmo, 1944, [47]:24, and quoted by Woodward in [106]: 258-259.

plex classical interventionist explanation, unlike the simplified explanation, is doing.

So why, then, did our analysis not arrive at the conclusion that there is a wide range of interventions (criterion 1), a diverse set of interventions (criterion 2), and *at the same time*, a wide set of w-questions answered (criterion 3) by the original classical interventionist explanation? This agreement between the three criteria of explanatory depth occurs in the car's example, but once more, it does not occur in the results of our analysis. Quite the contrary: the simpler explanations, which deliberately ignore factors that have been already recognized relevant by the scientific practice, turn out to be *deeper* according to the m-theory than the detailed ones. In this thesis we have found that, in fulfilment of criteria 3, the original classical interventionist explanation enables to answer a wider set of w-questions than the simpler version of it. The original explanation, for example, accounts for what would happen under magnetic interventions. A situation about which the new and simpler explanation has nothing to say. Indeed, the simpler explanation is only able to answer w-questions about counterfactual situations produced by variations of the Brownian motion rates. Nevertheless, the fictitious simple explanation turns out to be deeper than the complex classical explanation in the analysis. And this fact, from a manipulabilitst perspective, is at least counterintuitive because (as the car's example illustrates), an explanation that accounts for a wider set of counterfactual questions, should be preferable and deeper.

From the analysis developed in this thesis we have concluded that the reason why the original interventionist explanations fail to meet conditions **INi** or **INiv** and concomitantly criteria 1 and 2 of explanatory depth, is the following: In order to explain a phenomenon, interventionist explanations postulate two (or more) causes that are correlated with each other. And if the causes are correlated we cannot intervene upon them in the way that the m-theory considers valid.

This conclusion provides a possible answer to the questions raised in the previous paragraphs. Woodward could argue, for example, that no matter how many w-answers an explanation provides, never mind how simple or how complex it is *if the postulated causes are correlated, it will be impossible to provide testing interventions for the explanation in question, and hence it will be impossible to meet cirteria 1 and 2 of explanatory depth.*

The question is if from this position we shall conclude that explanations

that postulate correlated causes are not even minimally explanatory; or if we shall rather conclude that it is simply impossible to test those explanations by means of the m-theory.

The m-theory provides us with a definition of 'deep' explanation, but it does not define 'shallow' explanation. Is an explanation 'shallow' if it violates all the criteria for explanatory depth? Or just one, and if so which one? More precisely, are shallow explanations those that are unable to provide interventions under which the postulated causes remain invariant? Or are shallow explanations those whose causal relationships are *not invariant* under many interventions? Let us analyze the two possibilities in turn.

(A) Suppose that shallow explanations are those that are not able to provide even one intervention under which the causal relationship they postulate remains invariant. Then, we can say that the m-theory has been useful to analyze the interventionist explanations; and that far from being *deep* explanations, they turned out to be not even minimally explanatory.

(B) Now suppose, instead, that shallow explanations are those whose causal relationships are shown to be non-invariant under interventions.[5] Then, we are forced to accept that the m-theory has not adequately analyzed interventionist explanations in general. For we have not been able to propose any *testing intervention* that could serve for evaluating whether the postulated causal relationships were invariant or not. As every explanation with correlated causes will be in the same situation, we can conclude that the m-theory is not useful to evaluate a specific set of causal explanations, namely, the set of causal explanations with correlated causes.

My conjecture is that, either choosing (A) or (B), rejecting explanations that postulate correlated causes leads to some kind of skepticism. Suppose we understand shallow explanations in the sense (A) and an explanation has already been evaluated as *deep* because it is invariant under several interventions. This explanation is always at risk of becoming unsatisfactory if a new causal factor, correlated with the originally proposed cause, is discovered. Even if the new factor is genuine and provides more information for controlling the explained event, it will automatically render invalid all the *testing interventions* that were previously considered appropriate. Therefore, the explanation will fail to meet criteria 1 and 2 of explanatory depth, and hence will lose its depth. A skeptic would thus claim that we should always shed

[5] As happens, for example, with Mr. Jones' explanation. Details in section 4.6.1.

doubts on the *depth* of the explanation because we never know whether there is an *unknown* correlated cause.

Similarly if we understand shallow explanations in the sense (B) there is always some probability that an unknown causal factor is correlated with the causes postulated by a given explanation. But we said that, according to (B), the m-theory is not useful to test explanations with correlated causes. The skeptic would claim that we never know if the explanation to be analyzed belongs to the set of explanations that can be assessed by means of the m-theory.

In other words, diagnostics of explanations obtained through the application of the m-theory are as lucky as scientific theories. They may be amended or abandoned. This would be acceptable if the arrival of new causal information would produce stronger and deeper explanations. However, in diagnostics obtained via the m-theory, the new causal information far from strengthening the explanations, seems to render vulnerable explanations that were already considered to be explanatorily deep.

Last, but no least, the analysis developed in this thesis shows that the *depth* of an explanation, more precisely, whether a manipulation counts as an *intervention*, depends on how the causal relationship invoked by the explanation is represented in the directed graph. And, as we saw when we proposed the directed graphs for the putative causal relationships postulated by the statistical mechanical interventionist (in chapter 5), and later considering the rf pulses as possible manipulations (in section 6.3), it is not always easy to decide whether a factor should be considered as a direct cause, as a contributing cause, as a possible intervention, or as a background condition. This decision leads to different directed graphs, and thus to different diagnostics according to the m-theory.

In sum, the critical attitude, which is my own attitude, points at the following four weaknesses of the m-theory. 1) The criteria of explanatory depth do not work harmonically. 2) The m-theory does not specify how "shallow explanations" are to be understood. 3) The rejection of explanations that postulate correlated causes leads to skepticism about the diagnostics obtained by the m-theory. And, 4) the diagnostics obtained by the application of the m-theory are dependent on the election of the causal relata and the directed graphs, which are not always univocally identifiable in the putative causal explanations.

I would finally like to stress that this is a critical *but constructive* attitude. It is never suggested the m-theory has not served us to analyze the explanations of the spin-echo experiments. On the contrary, the m-theory has helped us to visualize the advantages and disadvantages of those explanations from a new perspective, and has provided us with some specific proposals to improve the interventionist explanations of irreversibility. The critical attitude towards the m-theory is constructive in the sense that it is requesting clarification of specific aspects of the m-theory. It would be illuminating if Woodward (or another manipulationist) could tell us how exactly a "shallow explanation" is to be understood in the m-theory.

Appendix A

Liouville's equation and Liouville's theorem

The Liouville equation describes the time evolution of phase space distribution function. Consider a dynamical system with canonical coordinates q_i and conjugate momenta p_i, where $i = 1,...,n$. Then the phase space distribution $\rho(p, q)$ determines the probability that the system will be found in the phase space volume $d^n q\ d^n p$. The Liouville equation governs the evolution of $\rho(p, q; t)$ in time t:

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \sum_{i=1}^{n} \left(\frac{\partial \rho}{\partial q^i} \dot{q}^i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right) = 0$$

Where time derivatives are denoted by dots, and are evaluated according to Hamilton's equations for the system. This equation demonstrates 'the conservation of density in phase space' (which was Gibbs's name for the theorem).

Liouville's theorem states that the distribution function ρ is constant along any trajectory in phase space, i.e. that:

$$\frac{d\rho}{dt} = 0$$

The theorem is often restated in terms of the Poisson bracket as:

$$\frac{d\rho}{dt} = -\{\rho, H\}$$

Where H is the Hamiltonian governing the system's dynamics.

Appendix B

Ergodic Theory

There are several definitions of full ergodicity. In this appendix three of them are presented. The ergodic problem and the difference between the "ergodic hypothesis" and the "ergodic theorem" will also be explained.

Intuitively, a system is ergodic if a representative point crosses the entire region in Γ that is available to the system. Let us suppose, for example, that we have a container with gas and we leave it to evolve freely. It is reasonable to believe that, after enough time, the system will pass through all its possible microstates, i.e., it will cross over all the regions of the phase space compatible with its macro constraints given its initial condition. If time tends to infinity, we can even affirm that the system will eventually pass through any compatible state. This is the essence of Boltzmann's idea[1] later on denominated **ergodic hypothesis** [2].

Also, a quasi-ergodic hypothesis was posited attempting to offer an adequate account of the relevant features of statistical mechanical systems. This new ergodic-like hypothesis asserted that "a trajectory, started at any point would, in the fullness of time, come *arbitrarily close* to every point in the allowed phase space" (Sklar, 1993, [91]:161, original emphasis). But it was not completely satisfactory and, as a result, ergodic theory lost some recognition for several years.

[1] Published in 1884 in a Boltzmann's paper where the term "ergodic" appeared for the first time.

[2] Reconstructions of the role of ergodic hypothesis in Statistical Mechanics can be founded in von Plato, 1994, [102] and Brush, 1976, [15]; To find information about the original meaning and the etymology of the term 'ergodic' see Gallavotti, 1999, [42].

The ergodic theory recovered attention when John von Neumann and G. Birkhoff obtained new mathematical results. Von Neumann investigated which conditions of the dynamical structure of the system were sufficient to prove the identity of infinite time average and phase average for any function of the microscopic state of a system. The aim was to find such condition without using the Ergodic or the quasi-ergodic hypotheses. According to von Neumann's results, the condition to be satisfied by the dynamical system is that the state of the system not be "entrapped" in some sub-region of the phase space, i.e., that the phase space is not decomposable (indecomposable).

Definition of decomposability[3] : a system is decomposable iff there exist two (or more) regions A and B of non-zero measure such that $A \cap B = \emptyset$ and $A \cup B = \Gamma$, which are invariant under the dynamics of the system: $\phi_t(A) = A$ and $\phi_t(B) = B$ for all t.

This means that, if the initial microstate of the system is in the non-zero region A, the representative point will remain inside the sub-region A during the whole evolution of the system.

A mathematical theorem that strengths von Neumann's results was later proposed by Birkhoff. **The Birkhoff theorem** (sometimes called **The Ergodic Theorem**) states: "Let a system be started in some micro-state **a**. Let R be any region of micro-states possible for the system given the system's constraints. Let R have a non-zero size in the phase space. Then, when the system is ergodic it will be the case that, except possibly for a set of initial micro-states of size zero, the trajectory form the initial microstate **a** will eventually pass through the region" (Sklar, 1993, [91]:167).

The condition that "given any set of positive measure in the phase space, the trajectories from all but perhaps a set of measure zero of phase points intersect that set" is equivalent to idecomposability. Therefore, after von Neumann's and Birkhoff's contributions, (metric) idecomposability is necessary and sufficient to assure that both the phase average of any function of the dynamical variables and the infinite-time average of the same function exist and, that they are equal for *almost every* phase trajectory (except for a set of trajectories whose initial condition belongs to a measure-zero set)[4].

[3] Also referred sometimes as metric decomposability or metric transitive

[4] This theorem may also be regarded in topological terms. That is to say, as the topological properties that the phase space must fulfil under the transformations represented by Hamiltonians. For more details see Quay (1978, p.50-60)

Definition of (full) ergodicity: The dynamical system (Γ, ϕ_t, μ) is ergodic by definition if any of the following equivalent conditions are satisfied:

1. It is indecomposable

2. For any measurable set $A \subseteq \Gamma$ such that $\mu(A) \neq 0$ and for almost every[5] $\in \Gamma$, the flow ϕ_t intersects A at some time t, *i.e.*it holds that $\{\phi_t(x)\} \cap A \neq 0$

3. Given an integrable function f, the left hand side ("time average") of the following equality is equal to the right hand side ("space average") for almost every[6] $x \in \Gamma$

$$f^*(x) =< f(x) >$$

that is to say

$$\lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{t_0+T} f(\phi(x))dt = \int_\Gamma f(x)d\mu(x)$$

4. For almost every $x \in \Gamma$, the limit of the relative time that the flow $\phi_t(x)$ spends in a measurable set $A \subseteq \Gamma$ is proportional to $\mu(A)$[7]

Definition of absolute continuity: The invariant measure μ' is absolutely continuous with respect to μ iff for any measurable region $A \subseteq \Gamma$ it is the case that $\mu(A) = 0$ then $\mu'(A) = 0$

Uniqueness Theorem: assume that (Γ, ϕ_t, μ) is ergodic and μ is normalized. Let there be another measure μ' on Γ which is normalised, invariant under ϕ_t, and absolutely continuous with respect to μ then $\mu = \mu$'.

Definition of (full) Mixing property: We say that the system (Γ, ϕ_t, μ) is mixing if and only if for any measurable sets $A, B \in \Gamma$ it holds that

[5] almost every $x \in \Gamma$, except for a set of measure 0.

[6] Except a set of measure zero.

[7] Another way to define ergodicity is known as Koopanism. Koopman, who realized a study of ergodic systems using functional analysis. This approach focuses on the stochastic properties of dynamical systems in terms of the unitary operator. By studying the spectral properties of said linear operator, it becomes possible to classify the ergodic properties of the flow ϕ_t. (Taken from Rédei, M. "Koopanism", unpublished manuscript written for a course in Pittsburgh, 1994-1995)

$$\lim_{|t|\to\infty} \phi_t(A \cap B) = \mu(A)\mu(B)$$

Intuitively, if the phase space is a surface with fluorescent properties and we illuminate a region of it, in such a way that only that region becomes green and the rest remains white, then "mixin" will be fulfiled if, as a result of the system evolution, the green region initially concentrated is spread over the whole phase space when time tends to infinite[8]

Implication Theorem: Every dynamical system that is full mixing is also full ergodic, but not vice versa.

Convergence theorem: Let (Γ, ϕ_t, μ) be a dynamical system and let μ be a measure on that is absolutely continuous with respect to the normalized measure μ. Define $\rho_t(A) = \rho(\phi_t(A))$ for all measurable $A \subseteq \Gamma$. Let $f(x)$ be a bounded measurable function on. If the system is mixing, then $\rho_t \to \mu$ as $t \to \infty$ in the sense that:

$$\lim_{t\to\infty} \int f(x) d\rho_t = \int f(x) d\mu$$

A sharp critical analysis of the ergodic program has been developed by John Earman and Miklos Rédei in reference (1996, [34]).

[8]For another intuitive explanation of "mixing" see Frigg, 2007, [41]:119

Appendix C

Graphs in chapter 6

A) Classical interventionist explanation of the SE experiments

Set of relevant variables

$\boldsymbol{V}= \{EP, DS, H\}$; where

EP= environmental perturbations;

DS= delayed spins;

H= height of the echo-signal

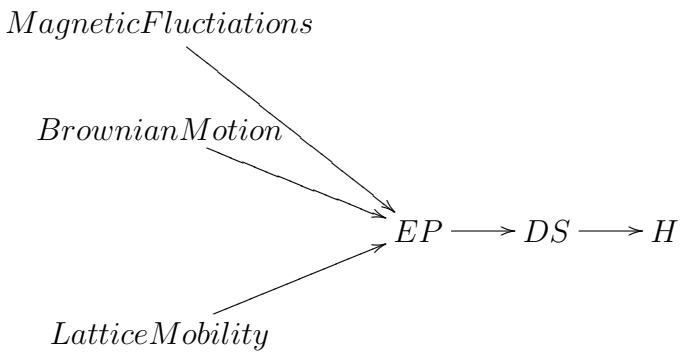
Possible interventions

I_1=Magnetic field fluctuations

I_2= Brownian motion (Brownian diffusion coefficient);

I_3= Lattice Mobility.

Graph



Directed Graph 12.

Background conditions

- Experimental settings: spin's sample, magnet and rf pulse generator;

Background theory

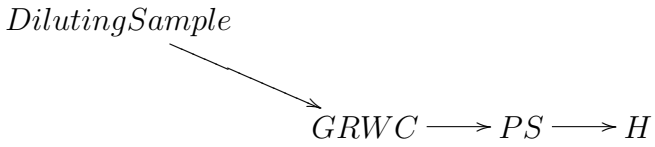
- Classical Mechanics; Larmor's theory of precession.

B) GRW-based Explanation of the spin-echo experiments

Set of relevant variables
$\boldsymbol{V}$=$\{GRW, PS, H\}$; where
$GRWC$= rate of GRW collapses per second;
PS = perturbed spins;
H= height of the echo-signal.

Possible interventions
I_4=Diluting the glycerine sample.

Graph



Directed Graph 15. Directed Graph for GRW-based approach.

Background conditions
- Experimental settings: spin's sample, magnet and rf pulse generator.

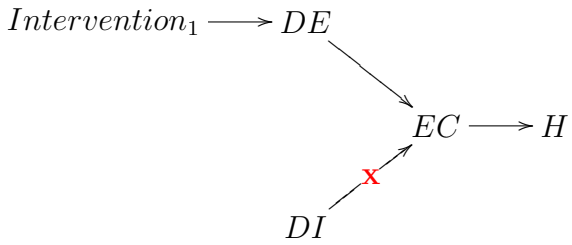
Background theory
- GRW collapse interpretation of quantum mechanics,
in particular the GRW stochastic dynamics;
- The dynamical hypothesis
-Assumption *(a)*: Among the quantum mechanically normal evolutions the set of the thermodynamic abnormal evolutions has measure zero.
-Assumption *(b)*: The thermodynamically abnormal states are uniformly distributed, in every microscopic neighborhood, among the thermodynamically normal ones.

C) Decoherence-based explanation of the spin-echo experiments

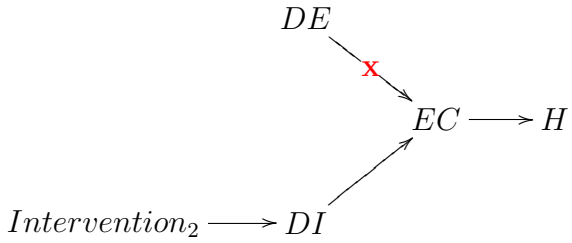
Set of relevant variables
$\boldsymbol{V}$=$\{DE, DI, EC, TDB\}$, where
DE=external decoherence;
DI=internal decoherence;
EC=effective collapses; and
H=height of the echo-signal.

Possible interventions
- Diluting sample
- Setting the initial quantum state in such a way that decoherence does not take place during the entire experiment.

Graphs



Directed Graph 17. First intervention for the decoherence-based explanation.



Directed Graph 18. Second intervention for the decoherence-based explanation.

Background conditions

- Spin- echo experiment settings

Background theory

-Non-collapse interpretation of Quantum Mechanics (see for example, Dieks and Vermaas (eds.), 1998, [28]).

- Schrödinger Quantum Dynamics (governing the system alone)

-Stochastic dynamics (governing the system plus the environment), for instance, the dynamics for modal interpretations proposed by Bacciagaluppi and Dickson, 1999, [5].

- It is assumed that these interpretations satisfactorily solve the measurement problem. As a consequence, effective collapses end up in quantum mechanically normal states, namely, states whose evolution begins and ends in states given by Gaussians in position and momentum.

- The initial quantum state must be such that decoherence is guaranteed for long enough to cover the whole evolution of the system.

-Although there is dependence of the initial quantum state, the evolution is still independent of whether the initial state is thermodynamically normal or not.

-Assumption *(a)*: Among the quantum mechanically normal evolutions the set of the thermodynamic abnormal evolutions has measure zero.

-Assumption *(b)*: The thermodynamically abnormal states are uniformly distributed, in every microscopic neighborhood, among the thermodynamically normal ones.

-Assumption *(c)*: The stochastic dynamical laws produce the quantum mechanical probabilities given by the Born rule.

-Assumption *(d)*: The quantum mechanical probabilities reproduce the quantitative predictions of classical statistical mechanics.

- In this case the past-to-future direction is entirely determined by the past hypothesis .

Appendix D

Larmor Precession

When a magnetic moment is placed in a magnetic field it will tend to align with the field. Classically, a magnetic moment can be visualized as a current loop and the influence toward alignment can be described as the torque on the current loop exerted by the magnetic field. The idea of the magnetic moment as a current loop can be extended to describe the magnetic moments of orbital electrons, electron spins and nuclear spins. In each case the magnetic moment is associated with the angular momentum, and a torque can be identified which tends to align the magnetic moment with the magnetic field. In the nuclear case, the angular momentum involved is the intrinsic angular momentum **I** associated with the nuclear spin.

When you have a magnetic moment directed at some finite angle with respect to the magnetic field direction, the field will exert a torque on the magnetic moment. This causes it to precess about the magnetic field direction (see Fig. 20). This is analogous to the precession of a spinning top around the gravity field. The torque can be expressed as the rate of change of the nuclear spin angular momentum **I** and equated to the expression for the magnetic torque on the magnetic moment; which when put in derivative form gives a precession angular velocity w.

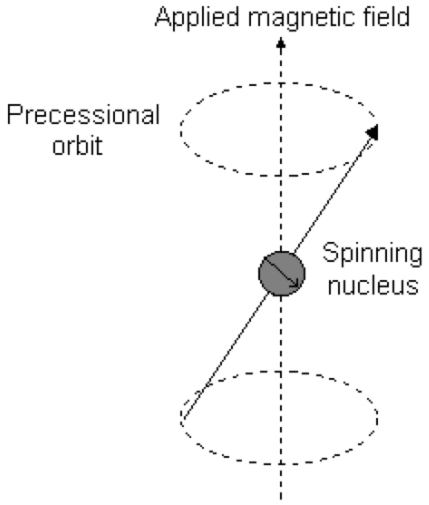


Fig.20. Larmor precession.

Sources:
http://hyperphysics.phy-astr.gsu.edu/hbase/magnetic/larmor.html
http:// teaching.shu.ac.uk/hwb/chemistry/tutorials/molspec/nmr1.htm

Bibliography

[1] Adkins, Clement John (1987). *An introduction to thermal physics*. New York: Cambridge University Press.

[2] Ainsworth, Peter (2005). 'The Spin-Echo experiment and Statistical Mechanics'. *Foundations of Physics Letters*, 18(7), pp.621-635.

[3] Albert, David (1994). 'The Foundations of Quantum Mechanics and the Approach to Thermodynamic Equilibrium', *British Journal for the Philosophy of Science*, 45, pp.669-677.

[4] Albert, David (2000). *Time and Chance*. Massachusetts: Harvard University Press.

[5] Bacciagaluppi, Guido and Michael Dickson (1999). 'Dynamics for Modal Interpretations'. *Foundations of Physics*, 29(8), pp.1165-1201.

[6] Bennet, Deborah (2011). 'Defining Randomness' In Prasanta S. Bandyopadhyay and Malcom R. Forster (eds.) *Phylosophy of Statistics* (pp. 633-638) Oxford: Elsevier.

[7] Bergmann, Peter G. and Joel L. Lebowitz (1955). 'New Approach to Nonequilibrium Processes'. *Physical Review*, 99(2), pp.578-587.

[8] Berkovitz, Joseph. Roman Frigg and Fred Kronz (2006) 'The ergodic hierarchy, randomness and Hamiltonian chaos', *Studies In History and Philosophy of Science*, Part B, 37 (4), pp. 661-691.

[9] Blatt, J.M. (1959). 'An Alternative Approach to the Ergodic Problem'. *Progress of Theoretical Physics*, 22(6), pp.745-756.

[10] Bloembergen, N., E.M. Purcell, and R. V. Pound (1984). 'Relaxation Effects in Nuclear Magnetic Resonance Absorption'. *Physical Review*, 73(7), pp.679-712.

[11] Blumberg, W.E. (1960). 'Nuclear Spin-Lattice Relaxation Caused by Paramagnetic Impurities'. *Physical Review*, 119(1), pp.79-84.

[12] Boltzmann, Ludwig ([1895] 1964). *Lectures on Gas Theory*. Reprint. Translated by Steven G. Brush. Berkeley: University of California Press.

[13] Bricmont, Jean (1997). 'Science in Chaos or Chaos in Science?'. In Gross, P.R., Levitt, N. and Lewis, M.W. (eds.), *The Flight from Science and Reason* (pp.131-176). Worcester: New York Academy of Sciences Press.

[14] Brown, Harvey. Wayne Myrvold and Jos Uffink (2009). 'Boltzmann's H-theorem, its discontents, and the birth of statistical mechanics'. *Studies in History and Philosophy of Modern Physics* 40, pp.174-191.

[15] Brush, Stephen (1976). *The Kind of Motion We Call Heat*. Amsterdam: North Holland.

[16] Brush, Stephen, Nancy S. Hall (ed.) (2003). *The Kinetic Theory of Gases: an anthology of classic papers with historical commentary* London: Imperial College Press.

[17] Callender, Craig (1999). 'Reducing Thermodynamics to Statistical Mechan- ics: The Case of Entropy'. *Journal of Philosophy*, 96, pp.348-373.

[18] Callender, Craig (2004). 'There is no Puzzle about the Low-Entropy Past'. In Christopher Hitchcock (ed.), *Contemporary debates in Philosophy of Science* (pp.240-255). Oxford: Blackwell publishing.

[19] Callender, Craig (2010). 'The Past Hypothesis Meets Gravity' In Gerhard Ernst and Andreas Hüttemann (eds.) *Time, Chance, and Reduction: Philosophical Aspects of Statistical Mechanics* (pp.34-58). New York: Cambridge Universtiy Press.

[20] Cappellaro, P., J.S. Hodges, T.F. Havel and D.G. Cory (2006). 'Concatenated Control Sequences Based on Optimized Dynamic Decoupling'. *Journal of Chemical Physics*, 125, pp.044514-1 to 9.

[21] Carr H.Y., and E.M. Purcell (1954). 'Effects of diffusion on free precession in nuclear magnetic resonance experiments' *Physical Review*, 94(3), pp.630-638.

[22] Castagnino, Mario, Olimpia Lombardi and Luis Lara (2003). 'The Global Arrow of Time as a Geometrical Property of the Universe', *Foundations of Physics*, 33(6), pp.877-912.

[23] Castagnino, Mario, Manuel Gadella and Olimpia Lombardi (2005). 'Time-reversal invariance and irreversibility in time-asymmetric quantum mechanics' (Preprint). *PhilSci-Archive* on-line, file 2595.

[24] Castagnino, Mario, Sebastian Fortin and Olimpia Lombardi (2010). 'Is the Decoherence of a System the Result of its Interaction with the Environment?' *Modern Physics Letters*, A25, pp.1431-1439.

[25] Clausius, Rudolf. (1864a). *Abhandlungungen uber die mechanische Warmetheorie*, Vol.1, Braunschweig: F. Vieweg.

[26] Coffa, Alberto (1974). 'Randomness and Knowledge'. In Robert S. Cohen and Marx W. Wartofsky (eds.), *Proceedings of the 1972 Biennial Meeting of the Philosophy of Science*. Dordrecht: Redel Publishing Company.

[27] Collingwood, Robin. (1940). *An Essay on Metaphysics*. Oxford: Clarendon Press.

[28] Dieks, D. and P.E. Vermaas (eds.) (1998). *The Modal Interpretation of Quantum Mechanics*, Dordrecht: Kluwer Academic publishers.

[29] Duarte, Javier and Sara Campbell (2008). 'Nuclear Magnetic Resonance and the Measurement of Spin-Spin Relaxation Times'. Laboratory report, Massachusetts Institute of Technology.

[30] Davies, Paul (1974). *The Physics of Time Asymmetry*. Berkeley: University of California Press.

[31] Davies, Paul. (1983). 'Inflation and Time Asymmetry in the Universe'. *Nature*, 301, pp.398-400.

[32] Eagle, Antony (2011). 'Chance versus Randomness', *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition).

[33] Earman, John (1974). 'An attempt to give a little of direction to the problem of the direction of time'. *Philosophy of Science*, 41(1), pp.15-47.

[34] Earman, John and Miklos Rédei (1996). 'Why Ergodic Theory Does Not Explain the Success of Equilibrium Statistical Mechanics', *British Journal for the Philosophy of Science* 47, pp.63-78.

[35] Earman, John (2002). 'What time reversal invariance is and why it matters', *International Studies in the Philosophy of Science*, 16(3), pp.245-264.

[36] Earman, John (2006). 'Past hypothesis: Not even false'. *Studies in History and Philosophy of Science*. 37, pp.399-430.

[37] Ehrenfest, Paul. and Ehrenfest Afanassjewa, Tatiana (1912). *The conceptual foundations of the statistical approach in mechanics*. Ithaca New York: Cornell University Press, 1959.

[38] Einstein, Albert. (1905).'On the Theory of Brownian Motion', *Ann. d. Physics*, 19, p. 371Ð381.

[39] Fink, Johannes Karl (2009). *Physical Chemistry in Depth*, Springer.

[40] Fitzpatrick, Richard (2009). 'Thermodynamics and Statistical Mechanics'. Unpublished manuscript written for course purposes in the Institute for Fusion Studies, Austin. Available in http://farside.ph.utexas.edu/teaching/sm1/lectures/.

[41] Frigg, Roman (2007). 'A Field Guide to Recent Work on the Foundations of Thermodynamics and Statistical Mechanics'. In Dean Rickles (ed.). *The Ashgate Companion to the New Philosophy of Physics* (pp.99-196). London: Ashgate.

[42] Gallavotti, Giovanni (1999). *Statistical mechanics: a short treatise*, New York: Springer.

[43] Gasking, D. (1955). 'Causation and Recipes'. *Mind*, 64, pp.479-87.

[44] Gibbs, Josiah William (1902). *Elementary Principles in Statistical Mechanics*, edited by BiblioBazaar in 2008.

[45] Ghirardi, G., Rimini, A. and Weber, T. (1986). 'Unified dynamics for microscopic and macroscopic systems'. *Physical Review*, 34, pp.470-479.

[46] Ghirardi, Giancarlo (2007). Entrance 'Collapse Theories' in the *Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu.

[47] Haavelmo, Trygve (1944). 'The Probability Approach in Econometrics'. *Econometrica* 12 (Supplement), The Econometric Society Press, pp.1-115.

[48] Hagar, Amit (2005). 'Discussion: The Foundations of Statistical Mechanics: Questions and Answers'. *Philosophy of Science,* 72(3), pp.468-478.

[49] Hahn, Erwin L. (1950).'Nuclear Induction Due to Free Larmor Precession'. *Physical Review* 77(2), pp.297-298.

[50] Hahn, Erwin L. (1953). 'Free Nuclear Induction'. *Physics Today*, 6(4), pp.4-9.

[51] Hahn, Erwin L. (1984).'Atomic Memory'. *Scientific American*, 25I(6), pp.50-57.

[52] Hashemi, Ray, William G. Bradley, and Christopher J. Lisanti (2003) *MRI: The Basics*. Lippincott: Williams and Wilkins Publishers.

[53] Hemmo, Meir and Orly Shenker (2001). 'Can We Explain Thermodynamics By Quantum Decoherence?'. *Studies in History and Philosophy of Modern Physics*, 32 (4), pp. 555-568.

[54] Hemmo, Meir and Orly Shenker (2003). 'Quantum Decoherence and the Approach to Equilibrium'. *Philosophy of Science*, 70 (2), pp. 330-358.

[55] Hemmo, Meir and Orly Shenker (2005). 'Quantum decoherence and the approach to equilibrium (II)'. *Studies in History and Philosophy of Modern Physics*, 36, pp. 626-648.

[56] Hempel, C. and P. Oppenheim (1948). 'Studies in the Logic of Explanation', *Philosophy of Science* 15, pp.135-175.

[57] Horwich, P. (1987). *Asymmetries in Time*. Cambridge: MIT Press.

[58] Hughes, Paul (2005). 'Spin Echo Nuclear Magnetic Resonance'. Laboratory report, Dept. of Physics and Astronomy, The University of Manchester. Spin_Echo_NMR_lab.pdf

[59] Jaynes, E.T.(1984). 'The Evolution of Carnot's Principle'. Opening talk at the EMBO Workshop on Maximum Entropy Methods, Orsay, France, April 24-28, 1984. Repreinted in Ericksen and Smith (1988) 1, pp. 267-82.

[60] Jesudason, Christopher G. (2003). 'Some Consequences of an Analysis of the Kelvin Clausius Entropy Formulation Based on Traditional Axiomatics'. *Entropy*, 5, pp. 252-270.

[61] Joos, E. and Zeh, H.D. (1985). 'The Emergence of Classical Properties Through Interaction with the Environment'. *Zeitschrift fur Physik*, B 59, pp.223-243.

[62] Khodjasteh, K and D.A. Lidar (2005). 'Fault Tolerant Quantum Dynamical Decoupling'. *Physical Review Letters*, 95, pp.180501-1 to 4.

[63] Kitcher, Philip (1989). 'Explanatory Unification and the Causal Structure of the World'. In *Scientific Explanation* P. Kitcher and W. Salmon (eds.), pp.410-505. Minneapolis: University of Minnesota Press.

[64] Lebowitz, J.L. and Peter G. Bergmann, Peter G (1959). 'Irreversible gibbsian ensembles'. *Annals of Physics* , 1(1), pp.1-23.

[65] Labarca, M. and Lombardi, O. (2005). 'Bolzmannian and Gibbsian approaches to the Irreversibility problem' [Los Enfoques de Boltzmann y Gibbs frente al problema de la Irreversibilidad] *CRÍTICA*, *Revista Hispanoamericana de Filosofía*, 37(111), pp.39-81.

[66] Landsberg, P.T. (1984). 'Can entropy and 'order' increase together?'. *Physics Letters A*, 102(4), pp.171-3.

[67] Lavis, D.A. (2003). 'The Spin-Echo System Reconsidered'. *Foundations of physics*, 34 (4), pp.669-688.

[68] Lombardi, Olimpia (2003). 'The problem of Ergodicity in Statistical Mechanics'. *CRÍTICA*,*Revista Hispanoamericana de Filosofía*, 35(103), pp. 3-41.

[69] Martin, Daniel and Paul Hughes (2006). 'Spin Echo NMR'. Laboratory report, School of Physics and Astronomy, The University of Manchester.

[70] Meiboom, S. and D. Gill (1958). 'Modified Spin-Echo Method for Measuring Nuclear Relaxation Times'. *Review of Scientific Instruments* 29(8), pp.688-691.

[71] Menzies, P., and H. Price (1993). 'Causation as a Secondary Quality'. *British Journal for the Philosophy of Science*, 44, pp.187-203.

[72] MIT Department of Physics (2008). 'Pulsed Nuclear Magnetic Resonance: Spin Echoes', 8.13-14 Experimental Physics I and II "Junior Lab". http://ocw.mit.edu.

[73] Mitchell, Sandra. (1997). 'Pragmatic Laws'. *Philosophy of Science*, 64 (supplement), University of Chicago Press, pp.468-479.

[74] Nyenhuis, J. A. and O. P. Yee (1994).'Simulation of nuclear magnetic resonance spin echoes using the Bloch equation: Influence of magnetic field inhomogeneities'. *Journal of Applied Physics*, 76(10), pp.6909-6911.

[75] Page, D. (1983). 'Inflation does not explain time asymmetry'. *Nature*, 304, pp.39-41.

[76] Paz, Juan Pablo and W. H. Zurek (2002). 'Environment-induced decoherence and the transition from quantum to classical'. In D. Heiss (ed.) *Fundamentals of Quantum Information: Quantum Computation, Communication, Decoherence and All That* (pp. 77-148) Berlin: Springer.

[77] Pearl, Judea (2000). *Causality*. New York: Cambridge University Press.

[78] Penrose, Roger (1989a). *The emperor's new mind*. London: Vintage books, Oxford University Press.

[79] Price, Huw (1996). "Cosmology, Time's Arrow, and That Old Double Standard" in Savitt, S. (ed.), *Time's Arrows Today*, Cambridge University Press, pp.66-94.

[80] Price, Huw (2004). 'On the origins of the Arrow of Time: Why there is still a Puzzle about the Low Entropy Past'. In Christopher Hitchcock (ed.) *Contemporary debates in Philosophy of Science* (pp.219-239) Oxford: Blackwell publishing.

[81] Redhead, Michael. (1987). *Incompleteness, Non-locality and Realism*. Oxford: Oxford University Press.

[82] Ridderbos. T.M, and M. L. G. Redhead (1998). 'The Spin-Echo Experiments and the Second Law of Thermodynamics '. *Foundations of Physics*, 28(8), pp.1237-1270.

[83] Reichenbach, H. (1956). *The Direction of Time*, Berkeley: UCLA Press.

[84] Russo, Federica (2009). *Causality and Causal Modelling in the Social Sciences. Measuring Variations*, New York: Springer.

[85] Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.

[86] Salmon, W. and Philip Kitcher (eds) (1989). *Scientific Explanation.* Minnesota: University of Minnesota Press.

[87] Serrin, James (1979). 'Conceptual Analysis of the Classical Second Laws of Thermodynamics'. *Archive for Rational Mechanics and Analysis*, 70(9), Verlag: Springer.

[88] Shaxby, John H. and E. Emrys-Roberts (1914). 'Studies in Brownian Movement'.*Proceedings of the Royal Society of London.* Series A, 89 (614), pp. 544-554.

[89] Shenker, Orly (2001). 'Interventionism in statistical mechanics: Some philosophical remarks'. *Philosophy of Science Archive* of the University of Pittsburgh. http://philsci-archive.pitt.edu/151/.

[90] Sklar, Lawrance (1974). 'Thermodynamics, Statistical Mechanics and the Complexity of Reductions'. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pp.15-32. University of Chicago Press.

[91] Sklar, Lawrance (1993). *Physics and Chance: Philosophical issues in the Foundations of Statistical Mechanics*, Cambridge University Press.

[92] Stejskal E. O. and J. E. Tanner (1965). 'Spin Diffusion Measurements: Spin Echoes in the Presence of a Time? Dependent Field Gradient'. *Journal of Chemical Physics*, 42, pp.288-293.

[93] Suárez, Mauricio and Iñaki San Pedro (2011). "Causal Markov, Robustness and the Quantum Correlations" in M. Suárez (ed.), *Probabilities, Causes and Propensities in Physics* (pp. 173-193). Berlin and New York: Springer.

[94] Uffink, Jos (2004). 'Bluff your way trough the Second Law of Thermodynamics'.*Studies in History and Philosophy of Modern Physics*, 32(3), pp.305-394.

[95] Uffink, Josh (2006). 'Compendium of the Foundations of Classical Statistical Physics'. In Jeremy Buterfield and John Earman (eds.) *Handbook of Philosophy of Physics.* (pp.923-1074.) Part B, Amsterdam: NH Elsevier.

[96] Uhring, Gotz S. (2007). 'Keeping a Quantum Bit Alive by Optimized pi-Pulse Sequences'. *Physical Review Letters*- 98, pp.100504-1to 4.

[97] Uhring, Gotz S. (2009). 'Concatenated Control Sequences Based on Optimized Dynamic Decoupling' *Physical Review Letters* 102, pp.120502-1 to 4.

[98] van Lith, Janneke (2001). *Stir in stillness: a study in the foundations of equilibrium statistical mechanics.*, University of Utrecht.

[99] Viola, Lorenza, and Seth Lloyd (1998). 'Dynamical suppression of decoherence in two state quantum systems', *Physical Review A*, 58(4), pp.2733-2743

[100] von Plato, Jan. (1988). 'Ergodic Theory and the Foundations of Probability'. In B. Skyrms and W. L. Harper (eds.), *Causation, Chance and Credence* Vol. 1, (pp.257-277), Dordrecht: Kluwer.

[101] von Plato, Jan. (1991). 'Boltzmann's ergodic hypothesis'. *Archive for History of Exact Sciences*, 42, pp.71-89.

[102] von Plato, Jan. (1994). *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective*, Cambridge University Press.

[103] von Wright, G. (1971). *Explanation and Understanding.* Ithaca, New York: Cornell University Press.

[104] Winsberg, Eric (2004). 'Can Conditioning on the "Past Hypothesis" Militate Against the Reversibility Objections?' *Philosophy of Science* 71, pp. 489-504.

[105] Widom A. and H. J. Chen (1995). 'Fractal Brownian motion and nuclear spin echoes', *Journal of Physics.* Series A, 28, p.1243-1247.

[106] Woodward, James (2003). *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press. Oxford Studies in the Philosophy of Science.

[107] Woodward, James (2008). Entrance 'Causation and Manipulability'. *Stanford Encyclopaedia of Philosophy.*

[108] Woodward, James (2010). "Scientific Explanation", *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.).

[109] Zurek, Wojciech Hubert and Juan Pablo Paz (1994). 'Decoherence, chaos, and the second law' *Physics Review Letters* 72, pp.2508-2511.

[110] Zurek, W., F. M. Cucchietti, D. A. Dalvit, and J.P. Paz (2003).'Decoherence and the Loschmidt echo' *Phys. Rev. Lett.* 91, 210403.