The Experimenters' Regress: A Conceptual and Historical Evaluation

Romina Zuppone

UNIVERSITAT DE BARCELONA

The Experimenters' Regress: A Conceptual and Historical Evaluation

Romina Zuppone

2018

The Experimenters' Regress: A Conceptual and Historical Evaluation

Romina Zuppone

Faculty of Philosophy

Phd Programme: Ciència Cognitiva i Llenguatge (CCiL)

Supervisor and Tutor: José Antonio Díez Calzada

Barcelona 2018

A mi madre,

quien me animó a cruzar el océano

Acknowledgements

This thesis has been done with the support of a scholarship received from the research network *PERSP*, *Philosophy of Perspectival Thoughts and Facts*, funded by the Consolider-Ingenio 2010 Scheme (CSD2009-00056) of the Spanish Ministry of Science and Innovation. Previous drafts of the chapters have been presented for discussion in several conferences and seminars. I would like to thank Maria Buedo for her illuminating comments of chapter two; Stefan Reining, David Rey, and Ljubomir Stevanovic for feedback and discussion of chapter three and four; Carl Hoefer for feedback on an early version of chapter five; to Javier Suárez, Santiago Ginnobili, Maria Scarpatti, Nahuel Snajderhaus and James Fraser for comments on chapter six. David Teira deserves a special mention, for reading and discussing extensively this last chapter of the thesis. My gratitude to my supervisor, José Díez, for his guidance along this period. Finally, I would like to thank my family and my friends for all their support and love.

Abstract

The goal of this thesis is to show, against Harry Collins, that there are scientific and internal grounds for the resolution of controversies in experimental practice. The approach I take is twofold: On the one hand, I develop a semantics of experimental results that seeks to show the dependence of experimental results upon theoretical knowledge and how this can be used as a conceptual way out of the experimenters' regress. On the other hand, I review, discuss, and propose an alternative reading to the cases studies Collins proposes as evidence of the external resolution of the experimenters' regress.

Introduction

I still remember my puzzlement when I first read Collins' *Changing Order*. It is in that book that he presents at length the *Experimenters' Regress* argument. I thought that something was deeply flawed with it, yet, at the same time, it attracted me as sceptical arguments have done all my life. I devoted, then, a long time to understanding how it worked, how it could be better motivated, and how it could be overcome in such a way that doesn't place the scientific enterprise at mercy of external and non-epistemic factors. This thesis is the result of that enterprise.

In chapter one, I begin by introducing the reader to some preliminary and very general considerations regarding the nature of experiments in the natural and biomedical sciences and how they have been understood in the light of the contemporary philosophy of science. In chapter two I introduce the standard ways in which to test experimental results and how to deal with experimental controversies in scientific practice, presenting different kinds of reproduction modalities, paying special attention to independent testing and the triangulation of experimental results. I finally present Collins' objections to the traditional view on reproduction. In chapter three, I present Collins' Experimenters' regress as I understand it work the best. In chapter four, I propose an initial and conceptual way out from the regress, one that appeals to the theoretical components that are necessary to understanding experimental practice and the constitution of experimental results. In chapters five and six, I present and analyse two case studies that were, according to Collins, the paradigmatic cases of regress in physics and in biomedical research. I offer an alternative reading of both. The analysis of the gravitational radiation episode mainly follows Allan Franklin's historical research, while the analysis of the Vitamin C episode is original. In both cases I fail to see that there is, *strictu sensu*, the kind of problem that Collins diagnoses.

Chapter One

Experiments in the Natural and Biomedical Sciences

1. Philosophical studies of experiment

It is quite surprising that a philosophical study of experimentation had to wait until 1980 in order to emerge as a research area. A similar delay was also experienced by the concept of observation, which was rarely studied on its own, but related to the theory-ladeness thesis and the distinction between theoretical and observational terms. Can we find an explanation for this neglect? In what follows, I will try to offer one.

Let us remember that the Neo-positivist philosophy of science arose in the context of the Linguistic Turn, the meta-methodological framework within which the Logical Empiricists conducted their epistemological research. In this respect, it seems reasonable to favour and promote the analysis of scientific discourse over the analysis of scientific practice. In this sense, the Quinean *semantic ascent* can be traced, in the philosophy of science, to the study of the function and the status of observational statements instead of the analysis of observation as an epistemic phenomenon. (Cf. Bogen, 2002, p. 131). Moreover, the emphasis on the syntactic treatment of scientific theories corresponded as well to social and political goals, especially among the Vienna Circle's left wing. As Carnap, Hahn and Neurath claimed in the *Manifesto*:

> The scientific world conception is characterised not so much by theses of its own, but rather by its basic attitude, its points of view and direction of research. The goal ahead is unified science. The endeavour is to link and harmonise the achievements of individual investigators in their various fields of science. From this aim follows the emphasis on collective efforts, and also the emphasis on what can be grasped intersubjectively; from this

> springs the search for a neutral system of formulae, for a symbolism freed from the slag of historical languages; and also the search for a total system of concepts. Neatness and clarity are striven for, and dark distances and unfathomable depths rejected. In science there are no 'depths'; there is surface everywhere. (1929).

The common language *desideratum* that could help scientists to share their knowledge and that could transform social reality was at the centre of the Vienna Circle's agenda (at least during the interwar period), suspending, probably for that reason, interest in the analysis of scientific practice and the production of scientific evidence. At the same time, given the impact of the group on the constitution of philosophy of science as a contemporary discipline, its influence set the agenda for several decades. This agenda survived, without its political dimension, after Neurath's death. (Cf. Zuppone, 2012). In that respect, the Vienna Circle contributed to the development of philosophy of science with a substantive account of what a scientific theory is, of the features of its constituents parts, as well as what it means to test a theory. But they left unexplored questions regarding the production of scientific evidence.

After the fall of Logical Empiricism, Hypothetico-deductivism rose. During the Classical period of philosophy of science, experimental practice did not receive much attention, either.[1] During this period, experimentation as an object of study was deemed unproblematic and, according to philosophers' understanding, its function was confined to theory choice. (Cf. Popper, 1954, § 30, p. 89). With Historicism, the situation did not change substantially; however, given other thesis that this movement embraced, the functions of experiment were revised. Indeed, Historicist accounts of science reconsidered the testing function of experimentation, for paradigms cannot be chosen by crucial experiments. Several functions of experiments were highlighted as part and parcel of normal science. (Cf. Kuhn, 1977, Cp. 8).

1 I am following Moulines' philosophy of science periodization. (Cf. Moulines, 2011).

Let us now try to explain what factors contributed to the rise of philosophy of experiment during the eighties.[2] In order to do so, an analogy may help us. Just as the Logical Empiricists did not provide an exhaustive account of observation, or an explication of the concept, and it was only under the challenge posed by both Hypothetico-deductivism and Historicism to the possibility of a neutral empirical basis, that systematic studies of observation began to appear. Something analogous happened to experimentation. This time, research started as a reaction to some social studies of science thesis that jeopardized the objectivity of scientific observation.

Experimentation was traditionally viewed as the anchor of empirical knowledge and several philosophers realized the need for a profound understanding of it under the pressure of the relativistic conclusions coming from other disciplines such as anthropology and sociology of science.

During the 1970s and 1980s social sciences studies began to claim that the image of science we had inherited was deeply distorted and that, in particular, the acceptance of empirical evidence, of experimental results was not as rational as was previously thought. It is as a reaction to social constructivism that contemporary philosophy of science begun to seriously study experimental practice and its epistemic virtues. In fact, during the first decades, the *philosophers of experiment* were eager to answer Latour and Woolgar's *constructions of facts* (1979), and Collins' Regress, amongst others. A typical thesis of social constructivism goes as follows:

> The main conclusion from our examples of discovery is that the existence and character of a discovered object is a different animal according to the constituency of different social networks. And by social network we refer to the beliefs, knowledge, expectations, the array of arguments and resources, equipment, allies and supporters, in short, to the whole local culture, as

[2] There are previous and isolated (albeit relevant) contributions on the topic. For example, Leibniz (1677 and 1682), Goethe (1792), Herschel (1830), Bernard (1865), Mach (1905), Duhem (1914) and Dingler (1928). Bunge (1967) is an exception within the Classical period.

> much as to the identities of individual participants. Crucially, this variation undermines the standard presumption about the existence of the object prior to its discovery. The argument is not just that social networks mediate between the object and observational work done by participants. Rather, the social network constitutes the object (or lack of it). (Woolgar, 1988, p. 65).

Every philosopher of experiment agrees in rejecting Woolgar's conclusion and defending the objectivity of experimental results, even if they differ amongst themselves as to how this objectivity is achieved.

That said, we may wonder what positive thesis philosophers of experiment offer. What is their substantive contribution to understanding the production of scientific knowledge? Can we trace different trends within this movement? What are the problems they seek to understand? And, moreover, have the New Experimentalists' answers to sceptical challenges been adequate? In what follows, I will try to sketch a general panorama of the field and to present their most distinctive thesis. The structure and reconstruction of the problems I will present is the result of a personal articulation that obeys more to a logical and systematic interest rather than an historical one.

Despite being a recent sub-discipline, philosophy of experiment already has a corpus, books, compilations and fundamental papers. Moreover, we can notice the existence of two main currents: the Anglo-Saxon and analytical branch, on the one hand, and the continental branch, mostly German, on the other.

We may assert that the Anglo-Saxon tradition started with Hacking's famous book, *Representing and Intervening*. Although this book was conceived of primarily as an introduction to philosophy of science, its second part placed experimentation at the centre of our epistemological world. Experimentation, Hacking argued, is not just a valuable object of study for philosophy of science, but understanding it will allow us to give answers to classic problems such as the reality criterion for theoretical entities. (Cf. 1983, cp. 16). It is in this very book that Hacking introduced the catchy slogan, "Experiment has a life of its own" that was repeated as a mantra over decades. Allan Franklin also deserves a mention within the analytical community. Like Hacking, Franklin is also known for defending the autonomy of experimentation and for his detailed work on the epistemology of experiment. (Cf. Franklin, 1986).

The continental branch was mainly associated with the Max Planck Institute for the History of Science in Berlin. Hans-Jörg Rheinberger, author of *Towards an Epistemology of Epistemic Things,* directed the institute until 2014 and led several research groups concerned with experimentation in the natural sciences. Hans Radder should also be mentioned here, as well as his edition of *The Philosophy of Scientific Experimentation.* We should not forget to include the contributions of Michael Heildelberger and Friedrich Steinle. Both authors represent a curious anomaly within the Germanic branch, given their stance towards the autonomy of experiments: in this respect they seem to support some approach to Neo-Empirism more akin to that of the New Experimentalists.

The two traditions can be differentiated, as we shall see, by their methodological preferences and approaches, by certain assumptions and by their conclusions rather than by the problems they consider as part of their agenda. The philosophers belonging to the Anglo-Saxon school usually displayed a form of Neo-Empirism that became evident in the defence of the autonomy of experiment thesis. In contrast, continental-spirited scholars were more likely to defend rationalistic proposals, and for them experiments are strongly influenced by theoretical interpretations of data.

It is fair to say that *Representing and Intervening* was the book that paved the ground for Philosophy of Experiment studies, despite not dealing exclusively with experimentation. It was followed by Robert Ackermann's *Data, Instruments and Theory* and Allan Franklin's *The Neglect of Experiment*, both specialized works by philosophers of science. Amongst the sociologists, we can highlight Harry Collins (1985, with a revised second edition in 1992) and Peter Galison (1987). In 1988, Hacking realized the existence of a proper philosophy of experiment that compensated for years of neglect. (Cf. Hacking, 1989). One year later, Ackermann published a short paper in which he introduced the label *New*

Experimentalism to refer to a group of researchers. (Cf. Ackermann, 1989).[3] From 1990 onwards, philosophical studies of experimentation multiplied and we can mention Franklin's long list (1990, 1999, 2002, 2004, 2005, and 2013), Deborah Mayo (1996) with which she won the Lakatos Award, Galison (1997), Sabina Leonelli's *Data-Centric Biology* (2016) which also won the Lakatos Award in 2018. Also abundant are contributed volumes such as Gooding, Pinch and Schaffer (1989), Koertge (1998), Heildelberg and Steinle (1998), Radder (2003), Galavotti (2003), Hon, Schickore and Steinle (2009), González (2011) and Léna Soler (2012). In recent years, as happens with any discipline, philosophy of experiment has become more and more specialized so we can now find that research is being done into big data, Bayesian networks in pharmacological research, decision making in diagnostic medicine, etc. I will not, however, review those topics here.

When considering the problems with which general philosophy of experimentation dealt, it is possible to detect four central concerns: to offer an explication of the concept of experiment; to offer an account of the functions of experiment; to analyse the stability of experimental results and the autonomy of experiments; and finally, to understand the dynamics of the acceptability of experimental results. The latter will be the main topic of this thesis.

2. What is an experiment?

In his work A Preliminary Discourse for the Study of Natural Philosophy, John Hershel, the discoverer of Uranus, wondered about the differences between an observation and an experiment and suggested a first distinction that is useful for our purposes. He proposed:

---

[3] The name does not seem to be a great one. The reader should notice that it became obsolete quite quickly, as happened with the "New Philosophers of Science", who we now prefer to refer as "Historicists". It is also striking that it is not very clear to me who the "Old Experimentalists" were, to begin with.

> [E]xperience may be acquired in two ways: either, first, by noticing facts as they occur, without any attempt to influence the frequency of their occurrence, or to vary the circumstances under which they occur; this is observation: or, secondly, by putting in action causes and agents over which we have control, and purposely varying their combinations, and noticing what effects take place; this is experiment. To these two sources we must look as the fountains of all natural science. It is not intended, however, by thus distinguishing observation from experiment, to place them in any kind of contrast. Essentially they are much alike, and differ rather in degree than in kind; so that, perhaps, the terms passive and active observation might better express their distinction. (Herschel, [1830], 2009, pp. 76-77).

Indeed, the idea according to which observation and experiment differ only as a matter of degree is widely accepted nowadays. The same criterion appeared, later, in Claude Bernard's *Introduction a l'etude de la médecine expérimentale.* He asked:

> Where lies the distinction between the observer and the experimenter? Here it is: an observer is he who applies simple or complex research procedures to the study of phenomena that he does not modify, those that he collects, as nature offers to him. An experimenter is he who deploys simple or complex research procedures in order to modify, with any goal, natural phenomena and to make them appear in circumstances or conditions in which nature would not have presented them. (Bernard, 1865, p. 26, my translation).

Both Herschel and Bernard agree on the difference between experiments and observations as being one of degree and manipulability; experimentation requires a directed action that could be read as an attempt, as Jean Perrin used to say, to unleash nature's treasures.

This distinction persisted in contemporary thought. For example, Rudolf Carnap, argued:

> As we have seen, all empirical knowledge rests finally on observations, but these observations can be obtained in two essentially different ways. In the non-experimental way, we play a passive role. We simply look at the stars or at some flowers, note similarities and differences, and try to discover regularities that can be expressed as laws. In the experimental way, we take an active role. Instead of being onlookers, we *do* something that will produce better observational results than those we find by merely looking at nature. Instead of waiting until nature provides situations for us to observe, we try to create such situations. In brief, we make experiments. (Carnap, 1966, p. 40, emphasis in the original).

Given this continuity, it is not surprising that certain explications of the concept of observation can be useful for thinking about experiments. Let us recall the causal approach to observation that, despite being developed in depth by Brown (1987) and Kosso (1989) had Victor Lenzen as an early advocate. In his *Procedures of Empirical Science* he understood observation as the result of inferences from hypothetical causes from perceptible phenomena. He claims:

> Nonperceptible entities are also inferred to exist as the hypothetical causes of perceptible phenomena. Such inference through causality eventually becomes observation. (Lenzen, 1938, p. 304).

The author considered that every observation involved a set of hypotheses and, moreover, he argued that the concept could be applied to accommodating several scientific scenarios that incorporated instruments and that required interpretation from well confirmed and accepted theories at a given time.

Let us now present the observability criterion Harold Brown put forward in *Observation and Objectivity*. The author considered that an entity or event is observable if we can establish a lineal causal chain between a certain entity -inaccessible to our naked eyes- and an effect that is perceptible. He claimed:

> One observes an item that is not available through direct examination by examining another item that is available to the senses, and which is the result of a causal chain that involves the item under observation. (Brown, 1987, p. 51).

The linearity condition is paramount. This claims that if there are different possible causes that could produce the phenomenon, it can't be established which one is the origin. This prerequisite can be satisfied by appealing to the most reliable knowledge available to the scientist, who can reject other possible causes of the phenomenon. If this condition is not fulfilled, the observation will remain indeterminate. Let me suggest a very simple example. Let us suppose that we would like to determine the PH of a potassium permanganate solution, whose colour changes when turning from a neutral PH (=7) to an acid PH (<7). If the only cause of this change of colour is the change in PH, then we can claim that we have observed the change of the PH of the solution. However, if there are other factors that can account for this change, then we cannot make that claim.

Brown's causal understanding of observation builds into the observation process a series of theoretical hypotheses such as those presupposed in the design of the instruments and those required to interpret the data. This, in turn, implies the fallibility of our interpretations regarding the observed phenomenon.

It is along these lines that we can conceive of experimentation in a very general way. Experimentation is an interventionist practice, one in which the phenomenon under study is properly isolated in order to guarantee (as much as possible, as we shall see in the following chapters) the lineal causal chain between input and output. The causal chain analysis will be crucial when interpreting what has been observed, that is to say, in the constitution of an experimental result.

We have pointed out the existence of a continuum ranging from observation to experiment. However, we must also recognize that the concept of experiment may well have a wider extension than that which we originally attributed to it. For example, in *Image and Logic*, Peter Galison explains how the function of the experimenter can be reduced to the interpretation of data that he didn't gather, *CERN* and *LHC* being cases at point.

> It has been a long, irregular, and often broken road between a time when it was unthinkable that a physicist be anything but someone who built equipment, designed procedures, manipulated experiments, wrote up results, and analysed them theoretically to a time when it would be a matter of near-universal consent that someone could count himself (or more rarely, herself) as an experimenter while remaining in front of a computer screen a thousand miles from the instrument itself. These alterations in practice contradict the notion that there is a single, unitary concept of experiment. Experiment and experimented are bound together, their meanings necessarily change together. (Galison, 1997, p. 5).

Technical, technological and theoretical advances have changed the experimenters' role. As a result, experimentation has become a rather complex social practice. Within this practice, according to Galison, we can detect three sub-cultures: the theoreticians, the experimenters, and the engineers, who are in charge of designing and building the facilities. (Cf. Galison, p. 1997, p. 8).

In a recent book, Sabina Leonelli (2016) also elaborates on the changing status of experiments and data. Technological advances contributed to the possibility of collecting, storing and even analysing enormous amounts of data. (Cf. 2016, cp. 3). Data, in turn, become more independent from the experiment that originated it and can “travel” and serve to address several different inquiries via data mining, for example, looking for certain patterns in what can be considered a revival of inductivism. She explains:

> As data travels, it may not always be clear how they could be used, and it is often the case that data stored in databases is not retrieved again and is thus not employed to create knowledge. This is not a problem for database curators, as long as there is an expectation that the data may serve as evidence at some point in the future, and data are therefore handled in a way that makes them available to further analysis. (2016, p. 78).

There have been several attempts to offer a proper characterization of the concept of experiment. According to Wenceslao González, for example, this explication has to be organized around different axes. In his words:

> Looking at the characterization of “experiment” there are several aspects to be considered in order to present this notion in our times. They are related to central factors of science. I) Semantically, experiment originally has a sense and a reference that differs from “observation”. II) Logically, experiment is a structural ingredient of science that is different from “theory” and, in principle, it is also distinct from “model”. III) Epistemologically, experiment is related to a kind of reliable knowledge acquired through a non-immediate process. IV) Methodologically, experiment is connected to a process that should be repeatable and, thus, it is commonly associated to reproducibility and repeatability. V) Ontologically, experiment is related to the idea of otherness (i.e. something –real or not- which is used to test). VI) Axiologically, the experiments can be oriented through different values according to distinct aims. […]. VII) Ethically, there is concern about some kinds of experiments, mainly when they are related to certain human affairs (either to people as individuals or to society as a whole). (González, 2011 pp. 26-27).

Giora Hon proposes an alternative approach to the understanding of the nature of experiments. He starts by detecting four kinds of experimental errors that haunt experiments and, on that basis, he proposes its different constituents. By way of analogy with Bacon's *idols* and emphasising *the idols of the theatre*, he proposes that an experiment consists of four parts: the script, the scenario, the spectator and the moral. In his words:

> In the spirit of the metaphoric language of Bacon and following his idols of the theatre, I suggest to discern four kind of idols that beset experiment: idols of the script, the stage, the spectator, and the moral. The image of a theatrical play constitutes a convenient and useful metaphorical setting for experiment since, like a play, an experiment is the result of an activity that truly has a "show" at its center. […]. In an experiment, nature is made, if you will, to display a show on a stage conceived and designed in some script. The show is observed and registered by a human or automated spectator and, finally, interpretation is proposed with a view to providing a moral –that is the outcome of the experiment as knowledge of the physical world. (Hon, 2003, p. 190).

Unfortunately, appealing to the Baconian idols seems nothing more than a rhetorical device. In Bacon's philosophy they play a fundamental epistemological role that is absent from Hon's proposal. As presented in the *Novum Organon,* the idols of the theatre are prejudices that arise as the result of accepting false dogmas or wrong demonstration principles. Hon's analogy seems too forced. Moreover, we can wonder how profitable his proposal is. Notice that, in an experiment, the goal is to discover or to understand an aspect of the physical world. As such, the researcher, at the end of the process, will be in a very different cognitive situation. This is something that cannot happen to the director of a play, who has an epistemic advantage when compared to the experimenter. Finally, it is not clear that the last of the components that Hon distinguishes in an experiment is indeed constitutive of a play: not every script ends up with a moral, not every play has an epistemic goal, while this is essential to experimentation. Moreover, while in a play it is the spectator rather than the director who grasps and understands the content, in the experiment, this part is performed by the experimenter.

Ian Hacking also attempted to characterize what an experiment is. In his 1992 paper he indicates the elements that constitute an experiment, with a *proviso*:

> Admitting as I do that there is less in common among experiments than we imagine, I shall, nevertheless list some elements that are often discernible. Their prominence and even their presence varies from case to case and from science to science. (Hacking, 1992, p. 43).

The elements that he highlights are:

Ideas: within this category he subsumes all the intellectual components of the experiment. These include the theoretical presuppositions with which the experimenter approaches the subject of study, the goals of the experiment and the hypotheses that will be tested (if there are any). Models of experimental design, or of the instruments and materials, are also included under this label.

Materials: this category comprises the objects that are relevant to the experiment, the object under study as well as the elements used for its study.

Marks: this category includes all kinds of data and statistical treatment of data.

But it is Hans Radder who has offered, as far as I can see, the most accurate picture of what an experiment is. In his paper *Experimenting in the Natural Sciences*, he claimed:

> In an experimental process we deal with an object to be studied and with a number of apparatus. Both object and apparatus may be of various kinds. Now, the experimental process involves the *material realization* and the theoretical description or interpretation of a number of manipulations of, and their consequences for, the object and the apparatus, which have been brought into mutual interaction. The general idea is that some information about the object can be transferred to the apparatus by means of a *suitable interaction*. That is, the interaction should produce an (ideally complete) correlation between some property of the apparatus. From this it follows that the theoretical descriptions of object and apparatus should also "interact": they need to have at least some area of intersection. […]. A typical feature of the practice of experimentation is that neither object nor apparatus is "simply available" They have to be carefully *prepared* in agreement with the goal and plan of the experiment. (Radder, 1995. P. 58, his emphasis).

Radder introduces a distinction between the material realization and the theoretical description of an experiment. As this distinction will be quite useful in our research we will return to in chapter three.

Finally, embracing as well as transcending the proposals offered by the different authors we have considered, we could think of an experiment as a technical-nomological device, that is to say, a material realization that, when oriented towards specific questions, produces marks that are of epistemic value. These marks acquire meaning through a process that we can understand as an inferential recoil that takes place when appealing to a set of laws that allow us to propose a theoretical description. Some of these laws are assumed in the experimental design, while another set of laws will allow the output of the experiment to achieve its final meaning. We will explore and deepen this suggestion in chapter three.

3. Experiments in natural sciences

Experiments in the natural sciences can be distinguished by taking into account their structural features, their function or the kind of results they offer. In accordance with their structural properties, experiments can be either material or non-material. Non-material experiments, in turn, can be mental or virtual. The form of realization of the experiment is related to its modality. In that respect, experiments can be actual, counterfactual or hypothetical. Finally, taking into account their function, experiments can be put forward in order to decide between hypotheses (testing experiments),[4] to constitute a theory (exploratory experiments), or to propitiate technological advances. That being said, nothing precludes the possibility of an experiment fulfilling the criteria of all three categories, or that it can be assigned a category only retrospectively. In that respect, the difference is pragmatic.

Material experiments are those in which the initial conditions of the physical systems under study are ones which have been created and controlled. In these kinds of experiments, the initial conditions and physical processes actually take place. There are two paradigmatic ways in which to conduct these experiments: in field situations or under laboratory conditions. Non-material experiments, in turn, can be classified into two categories: thought experiments and virtual experiments. The first are the famous *Gedankenexperimente,* in which initial conditions and the evolution of a physical system do not take place in reality, but are rather imagined. The second category depends on computer simulations that run as if the initial conditions and physical laws were taking place in a simulator. It is assumed, moreover, that all relevant variables that may influence the real system are taken into account by the simulation. In general, virtual experiments are used in circumstances in which material experiments cannot be done.

Considering now the functions of experiments in the natural sciences, we may claim that there are three main reasons for carrying out an experiment. First,

[4] We can include crucial experiments under this label.

although any experiment can, in some way or another, enlarge our knowledge, exploratory experiments are conducted with the specific purpose of discovering more about a certain domain, so as to create or develop a scientific theory. Secondly, testing experiments may help us to decide between scientific theories, while a final subset of experiments has as its goal technological advances (and possibly epistemic ones as well).

Even if material and non-material experiments can fulfil all the three functions we have just pointed out, it does not seem to be the case that they will be able to accomplish all of these with the same degree of excellence. There are, therefore, reasons why we may prefer to carry out a material experiment rather than a non-material experiment, or *vice versa*. Virtual experiments are ideal for studying the evolution of a system over the long run, such as exploring the evolution of cosmological models given certain initial conditions. Monte Carlo simulations are a case in point.[5] In cosmology, simulations have an enormous relevance, helping us to evaluate different cosmological models and to suggest new parameters that may make it possible to adjust the model in accordance with observations. For example, dark energy, an accepted parameter in most current cosmological theories, was introduced in order to reproduce, in the simulation, the state of the universe as currently observed. Virtual experiments also seem to be useful in biomedical research. Mary Morgan (2003) presents a case study in which, in order to determine bone-structure resistance to given external forces, a human skeleton is modelled over which Newton's laws are run.

When it is either technically impossible or impossible in principle to test something, thought experiments can be quite helpful. At the beginnings of modern science and in the early developments of quantum mechanics they abounded. We may wonder about the confirmatory power of non-material experiments. It is often considered, and this is especially the case with virtual experiments, that they are heuristic devices.

[5] With respect to Monte Carlo simulations see, Galison (1997, cp. 8). For an analysis of the immateriality of experiments and the extrapolation of results to the target system, see Morgan (2003).

Finally, it is possible to differentiate experiments by taking into account the kind of results they offer us. If the investigation is related to the existence of an entity or a process, it will be an existential experiment. If, in turn, it is about the attribution of a property to an entity or process, it will be an attributive experiment. Attributive experiments can be either qualitative or quantitative.

In two papers, Friedrich Steinle (1997 and 2002) has proposed an alternative classification to the one I have suggested. Steinle considers that among actual experiments we can differentiate between two types that seek different goals: exploratory experiments and theory-guided experiments. Those subsumed under the first category are typical of the early stages of research in a given domain, or typical of a scientific revolution phase (Cf. Steinle, 2002, pp. 422-423) and they contribute to the discovery of empirical regularities. According to Steinle, these experiments do not presuppose theoretical frameworks and they contribute to the formation of scientific theories. He claims:

> Exploratory experimentation typically starts when those categories have been destabilized, i.e., been revealed as being inappropriate to deal with the effects in question. Experimentation then goes hand in hand with revising, reforming, and re-stabilizing those categories. (Steinle, 2002, pp. 422-423).

The experiments that can be subsumed under the second category are ubiquitous during periods in which theories are consolidated and a paradigm is in place. These are the kind of experiments that we expect to take place during a phase of normal science. According to Steinle:

> Theory-driven experiments are typically done with quite specific expectations of the various possible outcomes in mind. Little room is left for completely unpreconceived outcomes, the very design of the instrumental arrangement may exclude many of those. (Steinle, 1997, p. S70).

And this is how he compares both categories:

> The contrast of exploratory experimentation to the theory-driven type, as understood as the standard view, is not only visible in the different epistemic goals (search for regularities vs. test of expectations), but also in the character of the guidelines of the experimental activity. The rather unspecific guidelines of exploratory experimentation bear a methodological character, and give rise to a variety of broadly dispersed experiments. The categories and concepts by which experiments are described and ordered arise typically at the end of experimental series, as their very result. Theory-driven experiments, in contrast, have such an ordering—and much more: a formulated, though perhaps provisional theory—as a precondition from the outset, and are in all essential details determined by that theory. Not a broad variety, but a single, elaborated arrangement is typically dealt with here. A third related difference is visible in the character of the instruments and apparatus used. Instruments for exploratory work have to allow for a great range of variations, and likewise be open to a large variety of outcomes, even unexpected ones. The restrictions posed by the instrumental arrangement must not be too confining. In testing well-formulated expectations, by contrast, instruments are specifically designed for a single effect. The possibilities of variations are much restricted, and so is the openness to outcomes that are not in the range of expectation. (Steinle, 2002, p. 422).

I am not interested in discussing this at length but I would like to highlight some points of agreement and disagreement. Both of us defend the idea that experimentation in the natural sciences does not consist merely in the testing of theories. There are experiments that allow us to form concepts, to determine fixed values for properties, etc. Unlike Steinle, however, I insist that all of this is independent of the availability of a theory or groups of theories that can help us to make sense of the output obtained during the experiment. Moreover, it seems incorrect to me to claim that in normal science an unexpected result is not possible, precisely because anomalies are the *surprising souvenir* within normal science. Let us consider Newton's experiment regarding refraction and the composition of white light: Steinle claimed (Ribe and Steinle, 2002) that the experiment is theoretically-driven. However, this experiment presents every feature he attributes to exploratory experiments. What about gravitational radiation experiments? Are they exploratory or theory-driven? Despite fulfilling all the criteria of the theory-driven category, their results have provoked heated controversies.

4. Experiments in biomedical sciences

Let us now explore some of the peculiarities of biomedical research, for it is quite hard to understand them in the light of was has been said so far. In this thesis, I will be particularly interested in clinical trials conceived in order to test new potential drugs for the treatment of specific diseases. In what follows, I will offer a brief introduction to and a characterization of these experiments.

In this kind of study, the experimenter usually postulates a causal claim in order to test it. There are several theoretical and practical difficulties associated with clinical trials; I will mention just a few of them. These experiments are sensitive to several types of bias which operate mainly unconsciously and which can dramatically alter the results obtained, leading to positive results even when administrating a physiologically inactive substance. There is, in addition, a lack of unifying theories to guide the research. The commercial interests associated with the research generate conflicts of interest and what can be thought of as a *tug of war* between pharmaceutical companies, the *sponsors*, who usually fund most of such research, on the one hand, and scientific imperatives, on the other (Cf. Lexchin, 2012 and also Gøtzche, 2014). Finally, the ethical implications of this kind of research are apparent. Frequently, there is a delicate tension between the imperatives of *do no harm* and the interest and urgency in finding an effective treatment.

There are different types of bias that can alter the results of a trial. Let us follow David Teira (2013a and 2013b) in considering the various actors at play and how their different and possibly conflicting desires may influence the results of a trial. For example, a participant in a clinical trial wants the treatment to work, so *patient reporting bias* -by which the patient tricks himself into thinking that he is getting better although there is no physical improvement- will probably influence the report she offers to the physician. The researcher also wants the treatment to work, so *confirmation bias* -we unconsciously give more weight to the evidence that confirms our beliefs than to that which disproves them- and *experimenters' reporting bias* are also bound to alter the results. *Selection bias* can be another source of spurious results. In this case, the groups would suffer from baseline imbalances. Differences in the characteristics of the members of the control group and the active group within a trial can be thought of as a competing explanation that can account for the results of the experiment. Furthermore, when a trial is replicated, the new researchers may have different interests and hence, when comparing experimental conclusions, they will disagree about the quality of their respective experiments, given the differences between results they get and those they were expecting.

In fact, clinical trials were created in order to neutralize the impact of these biases and to contribute to the production of safe and efficacious treatments even in ignorance of how an active principle may function. In several cases, the efficacy of a compound is known long before the reason why the compound has such a biological activity is understood. Cases in point are Streptomycin and Valium which entered the physician's pharmacopeia via "molecular lottery" (see Teira 2013b and Teira et al. 2015). Given this fact, Teira stresses that it is the statistical power of a trial which compensates for our ignorance and allows us to accept the use of a certain treatment before we know how and why it works. In the interest of the patients (and obviously, in the interest of the pharmaceutical industry), a drug may well be accepted as efficacious even though its mechanism of action is not yet well known.

Clinical trials are prospective, interventionist, comparative and statistical studies (in contrast to, for example, *natural history studies* and *case-control studies*, which are of a retrospective nature). They are regarded as the best way of understanding whether an intervention has the postulated effect. (Cf. Friedman et. al. 2010, cp. 1). They are designed to detect a rather weak signal amongst several sources of noise. To this end, a comparison is made between two treatments (for example, the active principle against the best current treatment available or against placebo)[6] and the difference between the outcomes is evaluated under the hypothesis that there is no difference between them. In other words, in these kinds of experiments the question is one of determining how probable the difference between the results is under the assumption of the truth of the null hypothesis, as the number of trials reaches infinity. (Cf. Teira et. al. 2105, pp. 11-12).

> In a frequentist approach, the probabilities of observing a given outcome are tied to one particular experimental design: if we repeat the same experiment time and again, we will observe a distribution of outcomes that will make our initial hypothesis about this distribution more or less credible. One crucial point, in making our experiment repeatable, is to define the population of patients that we are sampling in the trial. We are trying to ground an inference about the effect of a treatment in this population from the outcome we observe in the group of patients on which we are conducting the test. The probability of observing this outcome is indeed tied to a given reference class, the population of patients defined by the eligibility criteria in the trial protocol. Outside this population, the trial does not say how the treatment will work. The probability of observing a difference between treatments provides the significance of the test. If the probability is very low,

[6] However, several philosophers of science have questioned the potential benefits of randomization on different grounds. For example, there are several interesting discussions of the moral implications of conducting RCTs with a placebo arm when there is an effective treatment available. In those situations, it does seem that subjects are being harmed without any purpose beyond economic interests. (See Ethical Conduct for Research Involving Humans, 2014). Peter Urbach (1985), John Worral (2007) and Nancy Cartwright (2011) are cases in point. Amongst the medical community, Benson and Hartz (2000) and Concato, Shah and Horowitz (2000) also argued for the superfluity of RCTs.

> the event is rare enough to deserve a reconsideration of our initial hypothesis. (There was no difference between treatments) and declare one of these treatments superior. (Teira et. al. 2015, p. 12).

The gold standard within biomedical research is the *randomized-double blinded clinical trial* (RCT) methodology. It was Ronald Fisher along with Jerzy Neyman and Egon Pearson who introduced the methodology and the design of statistical experiments around 1930. Later on, around 1940, Bradford Hill began to apply this methodology to clinical trials in medicine. It was in 1962 that the FDA adopted this way of testing treatments as a regulatory standard. (Cf. Teira, 2011).

Clinical trials, pharmacological ones, in particular, involve different stages or phases (typically four), each of which evaluates a different relevant aspect of the interaction between drug and patient. Tolerance, biological activity, pharmacodynamics, adverse effects, dosage, therapeutical benefits, are tested in different kinds of populations. While Phase I studies determine the maximum amount of a drug that can be administrated to a person before unacceptable toxicity arises, Phase II studies gather knowledge regarding the kind of biological activity, if any, the substance being tested has in human subjects, if any. Phase III trials are what usually comes to our minds when thinking about clinical trials. At this stage, the goal is to assess the effectiveness of the new intervention. Finally, phase IV evaluates the long term consequences of the treatment. (Cf. Friedman et. al. 2010, cp. 1).

The peculiarities of biomedical science invite us to reconsider the classical distinction between contexts of discovery, justification and application. While finding a certain active principle via molecular lottery may be an example of how to come up with a certain hypothesis, RCTs justify statistically the causal claim relating the active principle to certain therapeutic benefits. There is, however, another aspect to take into account, another kind of justification that is important when we are considering things from an epistemic perspective: Why is this treatment effective? It is having a good answer to this question, I will claim, what would help us to offer an internal answer to the experimenters' regress in biomedical research. In a nutshell: Statistical justification of the efficacy of a drug cannot offer an explanation as to why the treatment is efficacious. In order to have that, we need to find the mechanism of action that links the treatment to the outcome of the experiment. This distinction, in turn, would help us to break a possible occurrence of reciprocity in biomedical research. Accordingly, in chapter six, I will show the importance that theoretical and methodological considerations may have in overcoming experimental disagreement.

Chapter Two

The Reproduction of Experiments and the Value of Interexperimental Evidence

1. Introduction

Experimental results are considered to be scientific knowledge only insofar they are judged to be valid (i.e. not an artefact). Even if experimental knowledge is fallible, the validity of an experimental result can be argued for by applying different strategies. Allan Franklin, in several of his papers and books, has advocated an *epistemology of experiment*, as a *set of strategies that provides reasonable belief in experimental results.* (1989). The concept of reproduction has been on the philosopher of experiment's agenda since the 1980s. Several seminal attempts have been made since then to understand the role, the varieties and the function of reproduction modalities. Chapter eleven of Hacking's *Representing and Intervening* (1983), chapter two of Collins' *Changing Order* (1985) and Franklin and Howson's (1984) paper are cases in point. In addition, Wimsatt's *Robustness, Reliability and Overdetermination* (1981), Howson and Urbach's (1989) *Scientific Reasoning* and Franklin and Howson's (1988) *It probably is a valid experimental result*, have contributed to the discussion. After a period of impasse, the topic has attracted the attention of a large group of researchers. Léna Soler, for example, edited a volume devoted to discussing the concepts of solidity and robustness and their role in experimental practice (2012). Reasons have also be given for the differences between the confirmatory power a hypothesis can receive from repeating an experiment and from independent testing. (Cf. Hartmann and Bovens, 2003, Cp. 4).

Experimental practice often astonishes us with sophisticated and novel ways of offering evidence for the existence or the properties of a phenomenon. A case in point, usually mentioned in the realism-antirealism debate, is the different ways in which Jean Perrin determined the Avogadro number. Frequently, the fact that different experimental processes lead us to convergent experimental results is a reason to consider them to be robust, and to consider the experiments conducted, reliable.

Reproduction procedures can take several forms, each of them having a different bearing on the reliability of the original results we want to evaluate. A first and usual procedure is what can be called an *internal replication*, repeating a procedure or the data gathering with the original device, or the original experimental set-up. I distinguish this from an *external replication*, the creation of a *carbon copy* of the original experiment. In such a reproduction, the theoretical principles that govern the design, the types of devices and materials used, the relevant characteristics of subjects in, for example, clinical trials, and the theoretical knowledge relevant for the interpretation of results must remain, as much as possible, constant. While internal replications are of great importance in ruling out spurious results, external replications are of a more problematic nature. Two reasons can be given for this. In the first place, as Harry Collins has argued, the (external) replication of an experiment requires the transference of tacit knowledge between researchers. In the experimental sciences, it has been argued, tacit knowledge may play an important role in making the experiment work and it may not be recognized. Thus, when incompatible results are obtained, the replica status of the second experiment could be challenged, rather than the discordance attributed to an error in the first experiment. In the second place, it seems that a successful external replication may also replicate the errors of the original experiment.[7] Reproduction of experimental results usually appeals to varying the experiments, improving original devices so as to provide an enhancement of sensitivity. In these cases, the reproduction shares most of the theoretical presuppositions of the original experiment while varying some of the features of

[7] As we shall see in the next chapter, Harry Collins challenges the epistemic credentials of replication when it comes to disconfirming an experimental result. Here, I would like to point out that replication could also be considered problematic when it comes to confirming an experimental result. The epistemic relevance of both discordant and concordant results can be, therefore, challenged, in contrast to what Collins argues.

the original design. These kinds of reproductions aim at maximizing the chances of getting a signal if there is anything to be captured “out there”. We may call them *T-reproductions*, for the theoretical component remains invariant.

However, the most acclaimed way in which the reproduction of an experimental result was thought to increase our confidence in its solidity was by means of independent testing, also known as the triangulation between results obtained via different experiments, measurements, observations or methodologies. This chapter seeks to analyse this concept. I will present a case study that seems adequate to putting to test the different possibilities discussed. Aftewards, I will review different alternatives available in the literature. Finally, I will present Collins' reasons for denying independent testing as a resource for settling experimental disagreement.

2. The Neutrino puzzle.

In this chapter, I will make use of the well-known case of the *Neutrino Puzzle* to explore the concept of experimental reproduction by exploring the different ways of checking Davis' original results. The Neutrino Puzzle is an ideal case to work with for various reasons. First, as the Solar neutrinos episode invited the scientific community to perform different experiments, it thus illustrates a wide-range of reproduction modalities. Second, the episode has been studied by several philosophers in order to argue for various thesis. This last fact will allow me to focus in detail on the aspects that are relevant to my particular concern. Readers who are interested in other aspects of the case are referred to other sources such as Dudley Shapere’s paper (1982) or Allan Franklin’s excellent book (2004).

In 1930, Wolfgang Pauli introduced a new theoretical entity to account for the missing energy in beta decay reactions. In 1956 Reines and Cowan experimentally detected these ghostly particles for the first time. Later on, the neutrinos would be involved in what can be considered a scientific mystery: how to account for a series of *convergent* independent results that could only detect 1/3 of the solar neutrinos predicted by our best solar models? This scientific episode became known as the *neutrino puzzle*. I shall outline this puzzle in the following paragraphs.

The Sun produces energy by means of nuclear reactions, in which hydrogen and other light elements fuse and liberate enormous amounts of energy as well as neutrinos. These elusive particles, that hardly interact, can offer valuable information about the processes that take place in the core of the Sun. There are different types of reaction which take place in the Sun's core, known as *PPI*, *PPII* and *PPIII*, and these occur with different frequencies and produce neutrinos with different energy levels. For example, although the *p + p reaction* occurs most of the time, the energy of the neutrinos produced in that reaction is, however, rather low, and so it is difficult to detect them. In the following table, the different reactions, frequencies, and energy are listed:

Table 1. Neutrino-producing reactions in the Sun.

Reaction	Frequency	Energy (MeV)	Name
PP I			
$p + p \rightarrow {}^2H + e^+ + \nu_e$	99.75%	0.0 – 0.42	pp
$p + e^- + p \rightarrow {}^2H + \nu_e$	0.25%	1.44	pep
${}^2H + p \rightarrow {}^3He + \gamma$	100%	—	
${}^3He + {}^3He \rightarrow {}^4He + 2p$	85%	—	
PP II			
${}^3He + {}^4He \rightarrow {}^7Be + \gamma$	15%	—	
$e^- + {}^7Be \rightarrow {}^7Li + \nu_e$	99.99%	0.86, 0.38	^{7}Be
$p + {}^7Li \rightarrow {}^4He + {}^4He$	100%	—	
PP III			
$p + {}^7Be \rightarrow {}^8B + \gamma$	0.01%	—	
${}^8B \rightarrow {}^4He + {}^4He + e^+ + \nu_e$	100%	0–14.1	^{8}B

Figure 1. Taken from Davis, 2002, p. 63.

In 1960, Raymond Davis, following an experimental design that was suggested at an earlier stage by Bruno Pontecorvo, began to study solar neutrinos with devices arranged in different abandoned mines. The Homestake mine experiment began to gather data in 1967. This was a radiochemical experiment aimed at detecting those high energy Neutrinos produced in the ^{8}B chain by the reaction of neutrinos reaching a 100000 gallons (378500 litres approx.) of perchlorethilene, C_2CL_4:

> ^{8}B decays to ^{8}Be, which splits to make two ^{4}He nuclei, a positron, and an electron neutrino, completing the PPIII chain. It is these ^{8}B neutrinos that produce most of the solar neutrino signal I detected, but there is also some contribution from *pep* and ^{7}Be neutrinos. (2002, p. 63).

$$\text{(1)}\ B^{8} \longrightarrow {}^{4}He + {}^{4}He + e^{+} + \nu_e$$

These high energy neutrinos can react with chlorine to form radioactive argon and to emit an electron by means of inverse beta decay, as Pontecorvo suggested in his 1946 paper. Each argon molecule indicates a reacting neutrino. The reaction triggered by the neutrinos interacting is the following:

$$\text{(2)}\ {}^{37}Cl + \nu e \longrightarrow {}^{37}Ar + e^{-}$$

The energy threshold of this reaction is 0.814 MeV. Since argon does not react or interact easily, it is easy to extract and to count by means of a Geiger counter when it is decaying back to chlorine. So, the Geiger counter would indicate the amount of argon detected, which in turn indicates the number of neutrinos that have interacted with the chlorine. (Calculations were made to determine other reactions that can produce ^{37}Ar, such as, cosmic rays muons, fast neutrons and internal contamination of the perchloretilene). These are the results that Davis reported:

> Over a period of 25 years, we counted a total of 2200 ^{37}Ar atoms and obtained a solar neutrino flux of 2.56 $\pm$ 0.16 (statistical error) $\pm$ 0.16 (systematic error) SNU. (Davis, R. Nobel lecture, 2002, p. 74).[8]

The predicted rate was, however, 7.6+1.3-1.1 SNU. (Cf. Bellerive, 2003, p. 4). Internal replications of this experiment were consistent. And no external replication was performed.

> The Homestake experiment was the only measurement of the solar neutrino flux for a long time. We had to wait 23 years for the Kamiokande experiment to confirm that the solar ^{8}B neutrino flux was low. In the 1990s, two radiochemical experiments that captured neutrinos using the inverse beta-decay of ^{71}Ga, SAGE and Gallex, showed that there was a discrepancy between the measured flux of lower energy neutrinos from the *pp* reaction and that expected from the standard solar model. The gallium experiments were off by a factor of two or so. (Davis, 2002. pp. 75-76).

The Gallex and Sage experiments are technically quite similar to the Homestake mine experiment. The devices also make use of chemical reactions to detect neutrinos. However, the energy threshold for these reactions is much lower than that of the Homestake. It is 0.233 MeV. Therefore, with this experimental arrangement, neutrinos from different chains could be detected, in particular, the less energetic neutrinos to which Davis' experiment was blind. These features make the experiments qualify as *t-reproductions*. The reaction that governs both devices is the following:

[8] SNU stands for: solar neutrino unit and it is equal to the neutrino flux producing 10^{36} captures per target atom per second. One capture per second and per 10^{36} target atoms.

$$\textbf{(3)}\ \nu_e + {}^{71}Ga \rightarrow {}^{71}Ge + e^-$$

In order to detect the *p-p* and *pep* neutrinos, 50 tons of gallium were required. And methods for extracting it were designed.

The SAGE experiment, a Russian-American collaboration, used a container filled with 30 tons of liquid gallium. The electron capture decay occurs within a half-life of 11.4 days. (Franklin, 2004, p. 261). Calculating the detection from 1990 to 1992 on average, SAGE reported 73^{+18}_{-16}(stat) $^{+5}_{-7}$(sys) SNU. The predictions for this experiment were 132 SNU $\pm$ 7. (Franklin, 2004, p. 265). The Gallex experiment, a European collaboration, was held beneath a mountain at Gran Sasso, Italy. Instead of using liquid Gallium, the reactive was Gallium chloride ($GaCl_3$) in an aqueous solution. The results obtained were 83 $\pm$ 19 (stat) $\pm$ 32 (sys). The expected results were the same as in SAGE.

So far we have briefly described radiochemical experiments that operate by means of a physical reaction: inverse beta decay. While the type of physical reaction is kept constant from experiment to experiment, the means by which this reaction takes place varies. As a consequence, several features of the experiments vary as well, for example, their sensitivity and their equivocity (the kinds of events that can mimic the detection of a neutrino). It is worth noticing that all these experiments are prone, at least partially, to the same sources of error. For example, if neutrinos oscillate (as was finally shown to be the case), then the three radiochemical experiments would be equivocal, because they can only capture electronic neutrinos.

Real time experiments have also been performed and they work under different physical principles. Cases in point are the IMB (Irvine, Michigan, Brookhaven), the Sudbury Neutrino Observatory and the Kamiokande II experiments. The Kamiokande II, for example, operated with a 9.3 and later a 7.5 MeV threshold, and it could also detect ^{8}B neutrinos. This experiment, unlike the previous ones, provides a real time detection of solar neutrinos. The principle that governs this experiment is that of the Cerenkov Effect. Its main inner detector consists of a cylindrical steel tank, of 14.4 m in diameter and 13.1 m in height, and contains 50000 tons of water. The surface of the tank is covered with 11000 photomultiplier tubes which record the light produced when a neutrino interacts with an electron and accelerates over the speed of light in water. It could also determine the direction of the incoming neutrinos. Several methods were used to rule out all the effects that could mimic the detection of a solar neutrino. Their results were 2.9 ± 0.4 flux units. (10^6 neutrinos per square centimetre per second), almost half of what was expected. (Cf. Franklin, 2004, p. 279).

	Chlorine	**Sage+Gallex**	**Kamiokande**
Target material	^{37}Cl	^{71}Ga	H2O
Reaction	$^{37}Cl + \nu_e \rightarrow {}^{37}Ar + e^-$	$\boldsymbol{\nu_e + {}^{71}Ga \rightarrow {}^{71}Ge + e^-}$	ν_e **+ e-→ $\boldsymbol{\nu_e}$+ e-**
Detection Method	radiochemical	radiochemical	Cerenkov
Detection Threshold	0.814 MeV	0.234 MeV	7.0 MeV
Neutrinos Detected	^7Be and ^8B	All	^8B
Predicted Rate	9 ± 1 SNU	132 ± SNU	5.7± 0.8 flux units
Observed Rate	2.5 ± 0.2 SNU	74 ± 8 SNU	2.9 ± 0.4 flux units

Table 1. A comparison between experiments. Taken from Franklin, 2004, p. 279.

3. Some considerations on independent testing

So far, we have reviewed some of the most relevant neutrino experiments. But until now, we have not presented any case of an independent test. This is because we would like to address the following question: what is it for an experiment to be an independent test of another? This is what Jacob Stegenga has named: the *individuation problem* (Cf. 2009).

What are the conditions that have to be fulfilled for two experimental tests to count as independent? To begin with, not every kind of independent knowledge, process or methodology will be pertinent to making two experiments independent and yet, relevant to counting as a test of previous findings (Cf. Collins, 1992, p. 35). We are interested in testing an experimental result built upon knowledge that we consider to be reliable. Therefore any independent test should be guided and informed by accepted theories, well-studied physical processes and properties of the objects/subjects under study. Moreover, when we are interested in offering a principle of individuation for multimodal evidence, we are interested in capturing not only a metaphysical difference, but more importantly, a difference that may have an epistemic bearing. The kind of independence we are seeking has to prevent or minimize the possibilities of the following scenario taking place: one in which both experiments offer a coincident wrong result. According to Stegenga and Menon (2017, p. 416): "The independence condition is meant to ensure that the concordant evidence from multiple methods are due to the object of investigation rather than an error-prone feature shared by the methods. (2017, p. 416). With this in mind, let us return our attention to the experiments that we presented in section two.

Although all the experiments considered attempted to determine the rate of solar neutrinos, we can still divide them in two very different kinds: 1) the radiochemical ones, such as Homestake, Gallex and Sage, and 2) the real time ones, such as Kamiokande and some others that I have not yet considered here, such as the Sudbury neutrino observatory (to which I will refer later in more detail), and Los Alamos Liquid Scintillator Neutrino Detector. On the basis of what criterion are the experiments divided into such groups? These two groups appeal to unrelated physical processes and make use of them to interact with neutrinos via their various capacities. While in the radiochemical experiments a certain reactant absorbs a neutrino, in the real time experiments based on the Cerenkov effect, a neutrino interacts with an electron transferring to it part of its momentum, and hence, accelerating it. Does this difference suffice for independent testing? Is the reliance of the experiment relies on *unrelated properties of the object under examination*, the criterion which serves to characterize independent testing?

In what follows, we will see that this way of understanding independent testing is neither sufficient to give it special credentials when it comes to clinching experimental evidence nor necessary for independent testing. For example, it would not help us to explain why we consider RCTs and cohort or case studies as independent tests, for these rely on appealing to different methodological forms of research: experimental and observational. Given that not all evidence is experimental, we may want to make room for a broader criterion of independent testing that may account for the distinction between experimental and non-experimental yet empirical evidence. I am thinking here of differentiating evidence obtained from experimental methods or highly manipulative observations from *clearly/mainly* observational evidence. And this is not just the case in biomedical research, it is frequent in physics, too. For example, let us consider the detection of gravitational radiation that we will discuss in chapter five. In that episode, there was experimental evidence, such as that provided by Joseph Weber's antennas and that of LIGO, but there was also astronomical non-experimental evidence, such as the measurement of the change of the orbital period of the PSR B1913+16 binary pulsar that Hulse and Taylor detected.

4. What kind of independence are we looking for?

In his 1981 *Robustness, Reliability and Overdetermination*, William Wimsatt understands multimodal evidence as one of the procedures that form part of the robust analysis of a target. The use of multiple means of determination to "triangulate" the existence and character of a common phenomenon is one of many ways in which different sciences argue for the stability of the phenomena under study. As Allan Franklin will later claim, there are different tools that help the scientists to "distinguishing the real from the illusory; the reliable from the unreliable; the objective from the subjective; the object of focus from artefacts of perspective; and, in general, that which is regarded as ontologically and epistemically trustworthy and valuable, from that which is unreliable, ungeneralizable, worthless, and fleeting." (Wimsatt, 1981, p. 63). Wimsatt offers different examples of multimodal testing: the sonar and the radar as alternative modes of detection (idem, p. 72). This distinction even has a bearing on a modern discussion. For example, Lockean primary qualities -shape, figure and size- are detectable through more than one sensory modality (idem, p. 76), and hence, more robust than secondary qualities such as taste, colour and sound, which can only be detected through one sense.

According to the author, one common feature of this plethora of strategies is that they require "at least partial independence of the various processes across which invariance is shown" (idem, p. 64). However, the criteria for the independence of modalities is not explored in the paper (nor in any part of his work, as far as I can tell). Hence, one of the questions we would like to address is what kind of independence is epistemologically relevant when we want to argue in favour of the correctness of an experimental result.[9] We might require an ontological form of independence, one in which via different physical processes lead to coincident results, as in radiochemical detectors and real time detectors.

[9] In their *Robustness and Independent Evidence*, Menon and Stegenga only consider two kinds of independence: *ontic independence,* which arises when "the multiple lines of evidence depend on different materials, assumptions, or theories", and probabilistic independence. (2017, p.414).

Or we might require theoretical independence, in which the experiments rely on different theories to design their material setup, as with gravitational wave antennas and LIGO, for example. We could also argue that probabilistic independence is the most relevant form of independence in these matters, since, when this form of independence is present, detection via one method does not have any bearing on the probability of its detection via another test. We might finally consider a methodological variety of independent testing, in which the ways of gathering evidence rely on different methods, such as observing, experimenting, etc., as with detecting gravitational radiation via a Weber's antenna or via the angular momentum of a binary star. We may finally wonder what the relation between all these proposals is. Is there a common feature that makes them succeed, from an epistemological point of view, when they succeed? In what follows, we will try to answer that question.

One possible advocate of the ontological independence is Ian Hacking. In his illuminating *Representing and Intervening*, he suggested a criterion for independent testing that rested on the independence of one physical process from another:

> Two physical processes –electron transmission and fluorescent re-emission- are used to detect the bodies [the dense bodies]. These processes have virtually nothing in common between them. They are essentially unrelated chunks of physics. (Hacking, 1983, p. 201).

Or regarding properties of the experimental/observational apparatus:

> Light microscopes, trivially, all use light, but interference, polarizing, phase contrast, direct transmission, fluorescence and so forth, exploit essentially unrelated phenomenological aspects of light. If the same structure can be discerned using many of these different aspects of light waves, we cannot, seriously, suppose that the structure is an artefact of all the different physical systems. (1983, pp. 203-204).

As I anticipated, the ontological independence criterion does not seem to suffice from an epistemic perspective. That is to say, it is not sufficient to offer a robustness argument. Neither is ontological independence necessary for two pieces of evidence to count as multimodal. Let me explain why. It is not a sufficient criterion because two independent tests so characterized may yield coincident results just as a consequence of sharing a problematic assumption. In fact, this is precisely what happened with the three radiochemical experiments when compared to the results offered by Kamiokande. All four experiments yielded concordant results. Yet, these results were not taken to be representative of what was going on in the solar centre. Despite those results being concordant, no modification of the Standard Solar Model could account for them (Cf. Franklin, 2004, p. 283). Therefore, two hypothesis were proposed to explain the results. The first one was the possibility of neutrinos decaying, which was ruled out by detections of neutrinos from a more distant supernova. The second one was the possibility of neutrinos oscillating into the tauonic or the muonic form. (Franklin, idem). Indeed, all the experiments presupposed that solar neutrinos do not oscillate. This assumption would make the detectors equivocal if not every solar neutrino is an electronic neutrino, since all four experiments were sensible just to the electronic variety. The experiments were ontologically independent, and yet, they shared an assumption relevant to the design of each facility that was later proved to be false and made them equivocal as solar neutrino detectors.

It was precisely the "no-oscillation" assumption which was proven to be wrong by a fifth experiment that took place at the Sudbury Neutrino Observatory. This was the first to offer evidence of the oscillation of solar neutrinos and its results were in agreement with the predictions of the Standard Solar Model: $5.44 \pm 0.99 \times 10^6$ cm^2 s^1.[10] Let me describe it briefly. The SNO is a Canadian experimental facility. The detector consists of an underground tank with 1000 tons of Deuterium, and detects all types of neutrinos coming from the Sun via the ^8B chain. Two reactions were

[10] The Solar model predicted 5.05×10^6 cm^2 s^{-1} (Cf. Franklin, p. 312).

considered (Cf. Bellerive et. Al. 2016, p. 3). A charged current sensitive to electron neutrinos:

$$\text{(CC)}\ \ \nu e + d \rightarrow p + p + e-$$ [11]

And a neutral current, sensitive to all neutrino types:

$$\text{(NC)}\ \ \nu x + d \rightarrow n + p + \nu x$$

The second reaction that takes place in this detector is of paramount importance since it allows the detection of any solar neutrino, irrespective of its flavour. Comparing the amount of neutrinos detected via (CC) and via (NC) can offer evidence regarding the oscillation of neutrinos, for example.

> A significant deficit in the B^8 ν flux measured by the CC reaction over that measured by the NC reaction would directly demonstrate that the Sun's electron neutrinos were changing to one of the other two types, without reference to solar models. (Bellerive et. al., 2016, p. 3).

The NC reaction was detected in three different ways. First, via Cerenkov radiation from the conversion of the 6.25 MeV γ ray produced when the free neutron captured on deuterium. Second, via a cascade of gamma rays when a neutron is captured in NaCl. Third, the neutral current neutrons were also detected in ^{3}He-filled neutron counters. (Cf. Bellerive et. al., 2016, p. 3). As we have seen so far, the SNO can be a fair neutrino detector even if neutrinos do oscillate. In that respect, it differs from the rest of the experiments that we considered in the previous section. In this sense, radiochemical experiments can be ontologically of real time experiments and yet, this does not suffice for the coincident results to be epistemically robust.

---

[11] *d* stands for deuteron

What about theoretical independence? As an advocate of this form, we can think of Sylvia Culp. In her *Defending Robustness: The Bacterial Mesosome as a Test Case*, she argues that independent modes of evidence must rely on different background theories. She claims:

> When comparable data can be produced by a number of techniques and the raw data interpretations for these techniques *do not draw on the same theoretical presuppositions*, this remarkable agreement in the data (interpreted raw data) would seem to be an improbable coincidence unless the raw data interpretations have been constrained by something other than shared theoretical presuppositions. (Culp, 1995, p. 448, emphasis added).

Stegenga (Cf. 2012, p. 217-218) denies that Culp's proposal could work as a proper characterization of independent testing, for three reasons: in the first place, because theory-ladeness comes in degrees. Second, because it is difficult to know what theory ladens the data. Third, because two pieces of evidence may be laden with the same theory for the production of the data and for its interpretation and yet count as independent modes. I find the first and second objections to Culp inappropriate. The first seems to be due to an uncharitable reading of her proposal. The theories that *drive* the experiment and upon which it is designed and operates are those that must be different, for something to count, in her terms, as an independent test. In that respect, the same degree of theory-ladeness is at stake here. The second complaint is irrelevant to the purpose at hand. The difficulty of a task does not undermine its importance. The third complaint raises more problems, though. Here, Stegenga is comparing case control studies, cohort studies and RCTs. He wonders: should we consider an RCT and a case control study as independent tests? He wonders. And if so, why? Do they rely on different or on the same theories? It seems rather hard to find a principled characterization of theoretical independence and one that works in the desired way.[12]

Notice, however, that the case we have presented is one in which an uncontroversial form of theoretical independence is satisfied, and yet, the results, even if concordant, fail to be robust. We should, therefore, search for a different form of independence.

The solar neutrinos episode seems to show that one of the forms of independence described above is missing: the probabilistic one. This form has been explicated in different ways by different researchers. For example, William Wimsatt (1994, p. 197), claims that "The probability of failure of the different means of access should be independent". Regarding this explication, Jonah Schupbach suggests a modification of the condition, and proposes that "if the means in question lead us astray in adopting some hypothesis, they do so for probabilistically independent reasons. Hence, learning that one of our means of detection has mislead us has no effect on the probability that the other means of detection will mislead us. Each means of detection is or isn't reliable, independent on the others." (p. 282). This is precisely what does not hold for our first four cases. Despite concordant evidence offered by ontologically and theory-independent experiments, all of them agreed on (at least) one assumption: that neutrinos do not oscillate. SNO was the first facility that did not incorporated such an auxiliary hypothesis.

In a recent paper, Stegenga and Menon highlight that many of the several conclusions that have been drawn based on ontological independence are not justified. (Cf. 2017, p. 417). We concur with the authors; as our case study showed, coincident and theoretically independent experiments failed to provide reliable evidence. Interestingly, these tests were not probabilistically independent, which is the kind of independence Stegenga and Menon defend for constructing reliable robustness arguments. In order to characterize this form of independence, the authors appeal to a Bayesian framework. They propose that two pieces of evidence are conditional-probabilistically independent (CPI) if and only if:

[12] When it comes to RCTs and Observational studies I prefer to think of methodological rather than theoretical independence.

1) Pr (H| E1) > Pr (H)

2) Pr (H| E2) > Pr (H)

3) Pr (E1 & E2| H) = Pr (E1| H) x Pr (E2| H)

4) Pr (E1 & E2| -H) = Pr (E1| -H) x Pr (E2| -H)

They also show that if it is the case that they are independent, then their conjunction will be more confirmatory than each individual conjunct. (2017, p. 429, and the appendix for the proof). In our example, these conditions are satisfied when comparing Davis' evidence with that of SNO, for example. In our case, however, we would not talk of a robustness argument, since the evidence was, in this case, discordant. Our episode then, illustrated a case of independent testing but of no independent confirmation.[13]

5. Collins' thoughts on reproduction modalities

In his book *Changing Order*, Harry Collins challenges the traditional view of how to satisfy the scientific imperative of the repeatability of experimental results. He argues that we should differentiate two contexts of the reproduction of experimental results: attempts to confirm a result and attempts to disconfirm it. These different contexts, he understands, require different reproduction modalities, for they have different epistemic credentials. He claims:

> For an experiment to be a test of a previous result it must be neither exactly the same nor too different. Take a pair of experiments -one that

[13] One problem I cannot address here is that of the comparability of the data obtained via independent tests.

> give rise to a new result and a subsequent test- If the second experiment is too like the first then it will not add any confirmatory information. […]. Confirmatory power, then, seems to increase as the difference between a confirming experiment and the initial experiment increases. (1992, p. 34).

Independent testing cannot be recommended on every occasion, but only when trying to confirm an experimental result. It is in those cases that, triangulation can have epistemic value. In disagreement scenarios, he argues, this strategy would not be of any help for overcoming discordance. In his words:

> Another complicating factor is that, though confirming power usually increases as experiments differ more […], there are circumstances in which power increases with similarity all the way to the extreme of near identity of the second experiment with the first. These circumstances arise when the second experiment is intended to disconfirm the first. This is because if a second experiment fails to see the claimed result, differences of design between the first and second may be invoked as the cause of the failure. (1992, p. 36).

I beg to differ with Collins, but we should acknowledge that even if we are committed to an epistemic asymmetry between confirming and disconfirming an experimental claim, it is also the case that, for every confirmation, Collins recommends independent testing.[14] Given that in any controversy there will be at least two discordant results to assess, independent testing would play a relevant role in the total gathering of evidence to decide in favour of one result. An independent test conducted in order to confirm *experimental result x* that fails to

[14] Always taking into consideration the kinds of procedures that are admissible within scientific practice.

confirm it, might count as a confirmation of *experimental result y* which was in conflict with *experimental result x*, i.e.: in the case in which discordant results are contradictory and exhaustive.

Collins also has a saying when it comes to the power of replication. As we will see in detail in the next chapter, he understands that in a context of disagreement, the legitimate reproduction modality is replication. He has an argument to show that replication cannot help us to overcome a scientific controversy: The experimenters' regress. The next chapter is devoted to analysing it.

Chapter Three

Getting to Know the Experimenters' Regress

In this chapter I introduce, analyse and discuss Collins' *Experimenters' Regress* (Collins, 1992) and suggest an alternative explanation of how to break out of it. In the first section I present what I take to be the problem Harry Collins wants to highlight. After presenting my reading on Collins' challenge, I show that in the experimenters' regress two different, albeit related epistemic problems are confused. These are: (i) the *replication regress* that consists in the occurrence of an infinite regress when judging whether or not a proper replication of an experiment has been carried out, and (ii) *general reciprocity*, according to which the determination of the proper functioning of an experiment and the correctness of an experimental outcome are determined reciprocally. I claim that: (1) the *replication regress* requires the soundness of the *general reciprocity* argument, so by showing the unsoundness of the second we also show that the first is untenable. (2) Reciprocity is not problematic on its own; what is problematic is Collins' explanation of how it is overcome. (i.e. his claim that non-scientific criteria are required in order to break the circularity). After offering an overview of the different proposals available in the literature either against the regress or against its external resolution and my critical comments on each of the proposals, I suggest that there is another possible explanation of the way out from *general reciprocity*, one which is extra-experimental but intra-scientific. I will pursue and elaborate that explanation in chapter four.

1. Introduction

Despite the fact that reproduction of experiments by peers has traditionally been regarded as of the utmost importance in enabling the intersubjectivity of scientific practice, reproductions may yield discordant results and deciding which result should be favoured may not be an easy task. According to Harry Collins (1992), experimental disagreement is resolved by the action of social, political and economic factors, but not by means of epistemic and scientific, or, so to say, internal reasons. His motivation for such a claim is the alleged presence of an infinite regress at the core of the experimental activity that, according to him, cannot be stopped by scientific resources: *the experimenters' regress*. The goal of this thesis is to offer an alternative account to Collins'.

Let us begin by considering the possible situations that can arise when reproducing an experiment. Given an experiment designed to test whether *x* is the case, and two research groups,[15] there are eight possible scenarios: if *x* is the case, either both research groups got the result right, or both got the result wrong, or one of them got it right and the other got it wrong. The same possibilities also appear if *x* is not the case, of course. Table 2 illustrates the possible reproduction scenarios. Possibilities 1, 4, 5 and 8 represent instances of confirmation (insofar as Team 2 confirms Team 1's findings) and hence, of agreement between the researchers; but, as we can see, it is possible to confirm an experimental finding only to later find out, for example, that an error occurred in both experimental setups. Those situations (represented by cases 4 and 5) in which we can be justified in believing false empirical claims, are testimony to the fallibility of scientific knowledge (and, particularly, to the fallibility of experimental practice). The problem of induction is also represented in the table. Consider, for example, cases 1 and 8. Despite the fact that, on the one hand, they confirm what happens to be the case, and also confirm each other results, the legitimacy of projecting those findings is, to say the least, problematic. I will not, however, dwell on these time honoured problems here.

[15] There can be disagreement among more than two groups, of course.

Instead, I will focus on a more restricted problem. The problem I will address can be summarized by the following questions: How are situations like 2, 3, 6 and 7 resolved? Which elements help researchers to overcome disagreement or, if consensus among researchers is not reached, what helps the scientific community to decide which experiment, if any, yielded an acceptable result?

	Is it the case that x?	Team 1	Team 2
1	x is the case	x/right	x/right
2	x is the case	x/right	-x/wrong
3	x is the case	-x/wrong	x/right
4	x is the case	-x/wrong	-x/wrong
5	x is not the case	x/wrong	x/wrong
6	x is not the case	x/wrong	-x/right
7	x is not the case	-x/right	x/wrong
8	x is not the case	-x/right	-x/right

Table 2. Alternative scenarios when reproducing an experiment.[16]

Not only there are several possible scenarios when reproducing an experiment but there are various ways of reproduction. In fact, as we saw in the previous chapter, the reproduction of an experiment can be done in several ways: 1)

[16] Three remarks regarding the schema: first, I take the content displayed in the second column to be, *strictu sensu*, unknowable and unreachable. It represents an external point of view, or, as Putnam would say: a *god's eye view of reality*, then, as far as reachability concerns, there are only four cases. Second, the schema portrays a scenario in which the experimental question has a categorical answer such as *x* exists/ *x* does not exist; *x* has the property *y*/ *x* does not have the property *y*; the value of property *x* is *y*/ the value of property *x* is *z*; *x* is effective for treating *y*/ *x* is not effective for treating *y*. Third, multilateral disagreement can be reduced to this schema.

Repeating a procedure with the original device. 2) By performing a *replication*: which means creating a *carbon copy* of the original arrangement in which the experimental design and the theoretical presuppositions remain constant. 3) By performing a *T-repetition*, that is to say, developing a more sensitive version, so that the new experiment shares all the theoretical presuppositions with the original but varies some of the features of the experimental design. 4) By performing an *independent test,* which consists of devising an experiment that relies on independent presuppositions.

According to Collins (1992), these different reproduction modalities diverge with respect to their testing power. This divergence constrains, he argues, the epistemic legitimacy of the reproduction strategy chosen in a given scenario (Collins 1992, pp. 35-36).[17] Confirming an experimental result, he claims, requires independent testing; disconfirming it requires replication. Given this, let us now suppose that a research group reproduced an experiment but failed to confirm the original findings (they are either in a 2, 3, 6 or 7 type of scenario). Then, according to Collins, for the results to count as a legitimate test of the original experiment, the research team should perform a replication. Collins' strategy of argumentation consists, on the one hand, in showing that the only epistemically admissible reproduction method for judging the correctness of an experimental result is replication; on the other, he points out that differences with respect to the results obtained can be attributed to an unsuccessful replication rather than to having obtained an incorrect result. The consequence is that checking the adequacy of the replication would lead to an infinite regress: *the experimenters' regress*. Discordance could be solved by providing experimental arguments, but if the experimenters' regress is an actual phenomenon, Collins claims, experiments cannot offer us a way out from the disagreement. Since, however, it is an empirical fact that science has a way out of this regress, an explanation of how it is achieved

17 Unfortunately, Collins omits *T*-repetition from his analysis.

is required. The explanation that Collins provides relies on the role of external factors in reaching consensus and settling the controversy.[18]

My challenge is to offer an alternative and epistemic explanation of how disputes regarding discordant results may be overcome even if we grant the possibility of an infinite regress in experimentation, or, more precisely, as we shall see in due course, of a reciprocity. In order to do so, I will introduce Collins' stance on the debate, presenting and discussing the *experimenters' regress*, the main reason why he advocates for an external resolution of disagreements in experimental activity. Afterwards, I will make explicit the two problems that are conflated under the experimenters' regress and the relations between them. The first is peculiar to experimental practice, and I will call it *the replication regress*. The second one is a more general problem, and it is not specific to experimental practice, but to establishing any empirical claim whatsoever. I will call it *general reciprocity*.[19]

2. Exposing two different problems behind Collins' *Experimenters' Regress*

In his book, *Changing Order: Replication and Induction in Scientific Practice*, Collins introduced the *experimenters' regress* in the following way:

> This is a paradox which arises for those who want to use replication as a test for the truth of scientific knowledge claims. The problem is that, since experimentation is a matter of skilful practice, it can never be clear whether a second experiment has been done sufficiently well to count as a check on

[18] By external factors, Collins understands non-scientific reasons. The decision that one result is correct rather than another has to do with, according to him, the persuasive skills of the actors, their influence and renown in the scientific community, etc., but not with scientific reasons. (Collins 1992, cps. 2 and 6).

[19] To be more accurate, this reciprocity concerns not only empirical knowledge but formal knowledge as well, logical reasoning being a case in point.

> the results of a first. Some further test is needed to test the quality of the experiment, and so forth. (Collins 1992, p. 2).

In a subsequent chapter he provided what he considered to be an alternative and equivalent characterization of the regress:

> What the correct outcome is depends upon whether there are gravity waves hitting the Earth in detectable fluxes. To find this out we must build a good gravity wave detector and have a look. But we won't know if we have built a good detector until we have tried and obtained the correct outcome! But we don't know what the correct outcome is until…and so on ad infinitum. (Collins 1992, p. 84).

In a typical Agrippan/Pyrrhonic/Sextan sceptical setup we would witness the divergence of incompatible experimental results and we would sensibly recommend suspending judgement. We would do so, temporarily, if the divergence could be resolved by further evidence, or *tout court*, <u>if there were</u> no fact of the matter that could help us to make a rational choice amongst the possible options. If the experimenters' regress lies at the bottom of our experimental practices, and if we adopt an empiricist and foundationalist stance towards scientific knowledge, the latter would be the case, since there would be no possible experimental resolution of the disagreement. In such scenario, any resolution of experimental discordance would have to be explained by appealing to non-experimental reasons.

In a joint book with Trevor Pinch, Collins endorsed the thesis according to which disagreement is potentially ubiquitous:

> It is worth reiterating the chain of reasoning: A *quasi-philosophical* argument shows that *no set of experimental results* can retain their potency in the face of reinterpretation by sufficiently committed critics. Or, to put this more

> positively, experimental data can retain its potency under these circumstances only when there are no such critics. (1993, p. 176, emphasis added).

And in *Changing Order* he concludes:

> There is then, *no set of 'scientific' criteria* which can establish the validity of findings in the field. The experimenters' regress leads scientists to reach for other criteria of quality. (Collins 1992, p. 88, emphasis added).

Even if Collins does not discriminate between the characterizations I presented at the beginning of the section, I understand that each of them singles out a different problem. I will, therefore, distinguish between a *replication regress* and *general reciprocity*. Since this distinction is not present in the literature,[20] in the next section, when discussing the reception of Collins' argument I will refer to the challenge as the *experimenters' regress* without further qualification. The distinction I propose will be developed in sections 4 and 5.

3. The reception of Collins' regress

Since its first appearance, the experimenters' regress and its implications have been discussed by several philosophers of science. I will briefly mention some of the criticisms made of Collins' view and some of the answers that had been given to the experimenters' regress, focusing in particular on those of Hans Radder and Allan Franklin.

20 Recently, Uljana Feest (2016) has also drawn attention to the circularity and the regress confused in Collins' work.

Probably the first reply to Collins was that of Larry Laudan. In his “A Note on Collins’ Blend of Relativism and Empiricism” (1982) he points out the inconsistency between Collins' empiricism and his dismissive attitude towards the role that empirical evidence plays in belief formation. Laudan stresses how Collins is adopting two incompatible thesis. On the one hand, he defends an empiricist approach basing his sociological research on empirical evidence, in particular, on case studies. On the other, he promotes and recommends a strong relativism, according to which, our beliefs, are not causally connected to empirical evidence. Laudan asserts:

> Unless we believe there is some linkage between a statement and a certain state of affairs in the world then we refuse to regard the statement as evidential. But if his thesis of strong relativism were correct, it would be pointless - even self-contradictory -to cite evidence for that thesis, since strong relativism denies that there is an evidential relation between our assertions and the world. Indeed, to cite empirical evidence for any claim is to concede that strong relativism is misguided, precisely because strong relativism denies the relevance of empirical evidence. (Laudan, 1982, pp. 131-132).

Another general complaint about Collins’ sceptical challenge appears in James Robert Brown’s book, *The Rational and the Social*, which devotes a chapter to some of the challenges posed by different scholars from the social studies of science, in particular those of Collins, Latour, and Woolgar. With respect to the former, Brown highlights his ill-grounded conclusions, especially the unjustified step from the existence of tacit knowledge to the social resolution of disagreement in empirical sciences. Brown calls into question the extent to which Collins' examples are really representative of experimental practice. The difficulties in replicating the *TEA laser* -a military device- that is of central importance for Collins’ argument, can be accounted for in terms of the rivalry between developers and in terms of secrecy. He also denies that the conclusions drawn from the gravitational radiation episode were accurate. Brown highlights the fact that all the researchers agreed on the theoretical background predicting the existence of gravitational radiation. This theoretical background was relevant when deciding whether or not Weber could have detected the amount of radiation he claimed. Finally, he casts doubts on the kind of externalist conclusions that can be extracted from the fact that tacit knowledge plays a role in experimentation. (cf. 1992, p. 88).

Benoit Godin and Yves Gingrass trace back Collins' challenge to Ancient scepticism in *The Experimenters' Regress: from Skepticism to Argumentation*. In the paper, they remind us how Sextus Empiricus, in discussing whether or not there is a standard of truth, points out how establishing a standard presupposes agreement upon a previously established standard, and so *ad infinitum*. Despite this, neither the authors nor Collins elaborate on the *Sextan* recommendation of *epojé, w*hich was the primarily goal of an epistemic regression.

Pleased at being related to such a tradition while trying to preserve the originality of his "discovery", Collins surprises us by claiming:

> There are forerunners of the experimenters' regress, but these were set out before the notion of controlled experiments was invented. (2016, p. 67).

Hans Radder (1992) also recognizes that Collins' proposal is an instance of a more general *knower's regress* that is, in some respects, inevitable, or, if I may, circumvented only by our language practices, to follow the suggestions of the late Wittgenstein.[21]

So far, I have presented criticisms of the more general framework Collins puts forward. In what follows, I will present the reader with some more specific ways

[21] *Forms of life* could be helping us to block general epistemological challenges, in particular, general scepticism. However it does not seem to be the most reasonable way to overcome a local form of scepticism, for which we have, as I will show in the following chapters, other more informative resources.

of discussing the social resolution of the experimenters' regress and the alternatives offered in the literature.

Sylvia Culp (1995) agrees with Collins in that there is a *data-technique circle*, but she denies that something besides the experiment is required to break it. In her opinion:

> A scientist will fail in making objective interpretations of her raw data to the extent that her interpretations are biased by dependence on idiosyncratic presuppositions. It is obvious that her interpretation will be biased if it depends on a false theoretical presupposition; but it is not so obvious that it could be biased even if it depends only on true theoretical presuppositions. (p. 440).

Culp seems to be relating Collins' concern to the theory dependence nature of the data. I believe this is not entirely correct. What Collins reports, as I will show later, is the reciprocal determination of the adequate measurement device with the correct experimental outcome. Moreover, Culp does not take into account Collins insistence on independent testing not being a legitimate strategy in disagreement scenarios. Despite this, she argues for robust sets of data generated via independent testing:

> When comparable data can be produced by a number of techniques and the raw data interpretations for these techniques do not draw on the same theoretical presuppositions, this remarkable agreement in the data (interpreted raw data) would seem to be an improbable coincidence unless the raw data interpretations have been constrained by something other than shared theoretical presuppositions. (p. 448).

While I agree with her on the value of triangulation, I still believe that Collins deserves a reply to his objection to the value of independent testing.

In his (1992), Hans Radder, differentiated the two presentations of the regress and offered a solution to what I have labelled the *replication regress*. However, he did not mention the differences between the logical structure of the arguments (an infinite regress and a reciprocal argument) or the relations between them, something that I will explore in the following sections. We both claim that the proper working of an experiment has to be judged by means of theoretical knowledge. In this sense, the strategy that I will present could be considered in line with Radder's approach while also aiming to provide a conceptually more detailed and wider empirically illustrated answer to the problem with which Collins confronts us.

Radder's strategy involves two steps. He first offers an elucidation of the concept of reproducibility. He later shows how a phenomenon can be stabilized ("delocalized") by means of different material realizations. However, Radder's concept of replication is different from the concept that Collins is dealing with. In Radder's paper, *replication* means obtaining the same experimental outcome regardless of the material realization. In his words:

> Next, remarkably enough, it is the very procedure of replication of a result *q* by means of different experimental processes that restricts its dependence on specific, local skills. A well-known example is the replication of experiments to test Avogadro's hypothesis […]. This claim was tested in the early decades of this century by means of a large number of very different experimental replications, viz. through Brownian motion, alpha decay, X-ray diffraction, black body radiation and electrochemical processes, among others. (Radder 1992, p. 70).

But the change of meaning of this central concept requires that Radder justifies the dismissal of one of Collins' thesis, namely, the idea that different reproduction modalities have different epistemic imports and therefore cannot be used interchangeably in any testing scenario. Collins would agree with Radder about the relevance of delocalization for confirming an experimental finding, but he would still disagree with him about the proper way of settling an experimental disagreement, i.e., in disconfirmation situations. In those circumstances, to replicate in Radder's sense (which is much the same as to *independently test,* in my sense), would again be unfair to Collins, who precisely denies that this is an epistemically legitimate strategy in disconfirmation situations, as we will see in the next section.

Allan Franklin's approach is mainly naturalistic, in that he make use of case studies to show how Collins' answer to the experimenter's regress is not empirically adequate, and that the regress can be broken by reasoned argument. (Cf. 1994, p. 465).

His explanation of how the regress was overcome in the gravity wave episode shows that an internal reading is possible and the *epistemological strategies* he extracts from the scientific practice help him to argue in favour of science being a rational enterprise insofar these strategies are deployed in experimental activity. Franklin proposes and develops an *epistemology of experiment,* asserting that "there are various strategies that both provide justification for rational belief in an experimental result and are used by practising scientists". And he argues that that helps us to "distinguish between a result obtained when an apparatus measures or observes a quantity and a result that is an artefact created by the apparatus" (1986, p. 165). The strategies he extracts from experimental practice can be sorted into different groups: those that help us to distinguish a valid result from an artefact created either by the instrument or the experimental design or the statistical analysis. The strategies are the following: *intervention* and *independent confirmation;*[22]*experimental checks; calibration, reproduction of artefacts that are known in advance to be present; elimination of plausible sources of error and alternative explanations of the result (the Sherlock Holmes strategy); using the results themselves to argue for their validity* (the nomic behaviour of the phenomenon); *using an independently well-corroborated theory of the phenomena*

[22] Already present in Hacking's work *Representing and Intervening*.

to explain the results; using an apparatus based on a well-corroborated theory; using statistical arguments.

Franklin claims that these strategies are routinely used throughout experimental practice and that even if their application does not guarantee that an experimental result will be correct, it does show that scientific research is built upon rational discussion and not, as Collins insists, on social tugs-of-war. Moreover, he illustrates the application of these strategies by analysing several case-studies with rigour and detail.

His strategy blocks the sociological explanation of the way out of the experimenters' regress. While I agree with Franklin's study, I will, however, try to enable a conceptual discussion of Collins' arguments[23] that could have a more general scope. However, as I will explain in chapter five, I believe his account of the gravity wave episode is right and sufficient to show that Collins' analysis is flawed.

Ujlana Feest has recently published a paper on the role of tacit knowledge in experimental practice, elaborating on some of Collins' ideas and showing how to apply them so as to arrive at some sort of *constructive scepticism*. She distinguishes two parts in the experimenters' regress argument. In her words:

> I take the argument to consist of two parts (only one of which invokes, strictly speaking, a regress): Collins first posits a circle between judgements of the validity of a measurement device and judgement of the validity of a measurement result. […] Collins himself refers to [t]he existence of this circle as "the experimenters' regress". However, it bears stressing that a circle is not the same as a regress. The regress only enters once scientists try to justify their judgements about a given outcome or about the quality of the data. It is there where Collins appeals to tacit knowledge, arguing that "[e]xperimental ability

[23] As far as I can see, it is also immune to Giora Hon's criticism, according to which Franklin's epistemology of experiment is eclectic and *ad-hoc* (Cf. Hon 2003). In fact, the analysis of the representational content of experimental results that I will present can shed some light on how to accommodate, in a principled way, Franklin's epistemological strategies. In any case, I take my approach and Franklin´s to be complementary.

> has the character of a skill that can be acquired and developed with practice. Like a skill, it cannot be fully explicated or absolutely established". It follows that for every scientific judgement there will be an inexplicable reminder, making it impossible to reduce a scientific justification to an algorithm. This become critical when scientists disagree in their judgements, because on Collins analysis such a disagreement cannot be rationally adjudicated. (2016, p. 35).

According to Feest, Collins' argument has two implications. One of these is sceptical, according to which there is no solution to the disagreement, while the other is relativistic, namely, that when controversies end, they are resolved by means other than rational argumentation. I find Feest's analysis difficult to follow on this point. The sceptical implication cannot be that there is no solution to the disagreement, since it is obvious that disagreements do in fact come to an end; it has to be that there is no *rational solution*, or epistemically justified solution to the disagreement.

Within Feest's interpretation, tacit knowledge plays an important role that becomes evident during episodes of disagreement between researchers which make it impossible to decide between two possible scenarios: (i) A and B disagree on the hypothesis being tested or (ii) A and B are not replicas. (Cf. 2016, p. 34).

She explores the sceptical implication of the argument and offers a normative analysis of the investigative process, arguing that "the explication (and critical evaluation) of tacit material assumptions required to implement specific operational definitions, and to make inferences from the resulting data, are crucial components of experimental knowledge generation" (2016, p. 36). She contends that the entire process of experimentation is governed by rules and that scientist "draw on tacit knowledge when applying those rules".

> Once we recognize that operational definitions are rules that specify what kind of experiment to run in pursuit of a given question, we may ask whether it can be unambiguously stated what it takes to apply the rule, and

> how to determine that such a rule has in fact been applied *correctly*. It is precisely here that Collins' worries about replicability derive their force. (2016, p. 37).

In a recent paper, Slovodan Perović discusses calibration as a possible way out from the experimenters' regress. By analysing the *in-situ* calibrating procedures in the LHC, he concludes that calibration is not independent of experimental outcomes, and yet, that it does not fall into the trap of the experimenters' regress, contrary to what Franklin granted to Collins (Cf. 1994, p. 465).

Perović has a broader criterion of calibration. While its typical characterization is the idea of using a surrogate signal to test the experimental device, according to Perović, "any combination of experimental techniques that ensures the proper functioning of the apparatus based on already-known phenomena may be characterized as calibration." (2017, p. 317). In his study of the measurement of the Top Quark Mass (Mt), Perović shows that calibration is combined with measurements and the different outputs are integrated and assessed by taking into account the different values accepted by the Standard Model. He claims:

> In the case of the LHC, the measurement outcome and the calibration are complex *reconstructions*, not a one-shot production of data and unrelated parameters. […]. In fact, the calibration and measurements of desired phenomena (*Mt*), are systematically co-extensive. Thus, as we have seen in the previous section, the precision measurements of *Mt* count on the improvement of *b-tagging* efficiency, while the efficiency of b-tagging will rely on improved reconstructions of Mt. And the entire *in situ* calibration is a *subsidiary of the measurement*, not a fully independent process. Since these procedures are so closely interrelated, neither them nor their validity can be understood independently. (2017, pp. 327-328).

When drawing conclusions from this case study, Perović highlights the relevance of theoretical knowledge in big science and how it is with its aid that we can overcome the experimenters' regress Collins reports. He wonders:

> What sort of agreement on calibrating procedures do we actually have, in our case? It is a comprehensive, multilevel theoretical and technical agreement developed over the course of the experiment. The broadest agreement concerns the acceptance of the background theories, Quantum Field theory and Quantum Chromodynamics. (2017, p. 329).

Indeed I agree with Perović. In the following chapters I will elaborate on the relevance of what I will call *theoretical calibration* and show how it can be used to explain in an epistemic vein Collins' favourite case-studies.

To conclude this section, I would like to mention David Teira's solution to the experimenters' regress for biomedical research (2013a).[24] In his paper, Teira proposed a contractarian solution to the regress which, at the same time, can be thought of as a defence of the use of debiasing procedures in biomedical research, and, in particular, a reply to Worrall on the relevance of randomization (2007). Rather than a way out of the regress, what Teira highlights is the importance of research groups reaching consensus on the implementation of debiasing procedures. The agreement on applying a set of debiasing procedures neutralizes the biases that each of the interested actors may introduce into the experimental arena. Even if an experimental result is biased, he claims, it will not be the case that it is biased in favour of any of the interested parties.

After this brief review of possible solutions, I will, in the next two sections, analyse each formulation of the experimenters' regress, distinguishing their logical forms and the logical relations between them. Section seven will connect the critical with the constructive part of the thesis.

[24] I will come back to Teira's proposal in chapter six.

4. Examining *Replication Regress* (*RR)*

As I suggested in section two, I understand that what has been identified and recognized as the experimenters' regress is in fact two different problems with different scope. In what follows, I will try to separate these problems apart and to show how they are related. Interestingly, we can recognize in both presentations of the experimenters' regress different *Agrippan modes*, argument forms that, according to ancient sceptics such as Agrippa, Pyrrho and Sextus Empiricus call for the suspension of judgement. According to ancient Pyrrhonic scepticism, several phenomena invite us to *epoje*. These are the *Agrippan modes*: disagreement, infinite regress, circularity and hypotheses. (Cf. Barnes, 1990). These *modes*, of course, can work in combination for a *stronger* effect. What I will suggest is reading Collins' proposal, o re-setting the experimenters' regress in terms of the combination of three Agrippan modes: disagreement, infinite regress and reciprocity. I will claim that the first two modes presuppose reciprocity. However, I will also claim that reciprocity as a sceptical tool only works under certain conditions, and that these conditions are not met in experimental practice.

In the experimenters' regress scenario, as I understand it, disagreement between researchers and the infinite regress in replication cannot by themselves, motivate the suspension of judgement. Persistent disagreement can only be a symptom of the *replication regress* which, in turn, can only be a consequence of a deeper and possibly more severe problem: *general reciprocity*, that is to say, the reciprocal determination of experimental *goodness* and experimental result *correctness*. If *general reciprocity* holds, then every time we reach agreement this is for reasons that cannot be rational, or epistemic, or internal, hence diminishing the value of the scientific enterprise.

A *reductio ad infinitum* is a destructive type of argument, one in which it is attempted to refute a certain proposition. According to Jonathan Barnes (1990, p. 43), the dialectical strategy put forward when applying the regression mode consists in the following steps: from a hypothesis under suspicion an infinite sequence is generated, the possibility of such a sequence is denied and thus, the hypothesis is rejected. Collins' main thesis would be that no scientific criteria can establish an experimental result (1992, p. 88). How does he arrive at that conclusion? Let us dive into the dialectics of *Changing Order* to find this out.

(1) *The robustness of an experimental result requires its reproduction.*

In order to be scientifically relevant, an experimental result has to be robust. In order to find if this is the case, several procedures are available, amongst them, Collins considers the following:

a- *I-Repeating* the experiment.

b- Replicating the experiment.

c- *Independently testing* the original results. [25]

(2) Reproduction procedures are not epistemically equal.

The alternatives introduced in (1) are not equivalent with respect to the degree of confirmation or disconfirmation of the original findings.

a- *I-repetition* does not increase the degree of confirmation of the original result. (Collins 1992, p. 34).

b- *Replication* is problematic when the possible sources of error that could cause a false positive or negative are unknown. (Collins 1992, p. 35).

c- While *independent testing* could be a way to confirm a result, it is not a legitimate way of disconfirming one, since it could omit relevant elements that led to the result. (Collins 1992, p. 36).

[25] Since I am following Collins' arguments, I will omit here what I called *T-repetition*, for it is a way of reproduction that he, himself, omits.

(3) Selecting the reproduction method depends on the goal of the experiment.

If it is not the case that every testing procedure is equally apt for a testing goal, then it is necessary to decide which method to use considering the purpose of the reproduction. If the reproduction is aimed at confirming the original finding, then an independent test should be performed. If it is aimed at disconfirming the original finding, then a replication is required. (Collins 1992, p. 34).

From (1), (2) and (3) we can derive the following conclusion:

(C1) *disconfirmation of an experimental result requires replicating the original experiment*. (Collins 1992, p. 36).

(4) *Replication requires tacit knowledge's transference*.

Experimentation is an activity that requires mastering certain skills that cannot be explicated or recognized as relevant even by the experts who possesses them. (Collins 1992, pp. 73-74).

From (C1) and (4) we can derive a new conclusion:

(C2): *Disconfirmation requires the transference of tacit knowledge*.

For an experimental result to be considered as a proper disconfirmation, it has to be well performed, of course. For it to be well performed, the relevant skills should have been transferred. But this cannot be determined, Collins claims:

(5) *In most experiments, the transference of tacit knowledge cannot be assessed*.

In *Changing Order*, Collins exemplifies the impact of tacit knowledge in replication by presenting a "normal science" episode of replication: the making of a

TEA Laser.[26] After highlighting the difficulties associated with making the lasers work, He claims:

> In sum, the flow of knowledge was such that, first, it travelled only where there was personal contact with an accomplished practitioner; second, its passage was invisible, so that the scientists did not know whether they had the relevant expertise to build a laser until they tried it; and, third, it was so capricious that similar relationships between teacher and learner might or might not result in the transfer of knowledge. These characteristics of the flow of knowledge make sense if a crucial component in laser building ability is "tacit knowledge". (1992, p. 56).

In this episode, there is an unequivocal symptom of proper functioning of the device: a functional laser is supposed to, for example, vaporize objects. This is how we can determine whether tacit knowledge has been transferred or not. In the absence of such an observable criterion, how can we determine that an experiment is competently performed? Collins wonders. We can only tell if they produce the proper experimental outcome:

(6) *In the absence of an observable criterion of success, proper functioning of the device and the correct experimental result are established reciprocally.*

> Proper working of the apparatus, parts of the apparatus and the experimenter are defined by the ability to take part in producing the proper experimental outcome. Other indicators cannot be found. (1992, p. 74).

We can now extract another conclusion from (C2) and (5):

[26] One can reasonably wonder how representative of normal science the building of a weapon can be. As James Robert Brown (1989) rightly points out, we are dealing with classified military research!

(C3) *In the absence of a clear indicator of success disconfirmation cannot be assessed.*

If tacit knowledge transference is invisible then we cannot tell whether an experimental result disconfirms another or if there is a failure in tacit knowledge transference.

These considerations lead Collins to suggest an infinite regress in experimental practice. Invisibility of the possession of relevant skills, would lead us to test the quality of the research by performing new tests that would be subjected to the same problem, and so, *ad infinitum*. Notice, however, that it cannot be an infinite regress what Collins wants to highlight, since this would not be much different from the problem of induction. In order to make Collins' challenge interesting, we need to stress that both disagreement and regress are rooted and rely on the reciprocal determination of results and measurements devices. It is in cooperation, that this *argumentative modes* acquire sceptical force.[27] We are now in a position to explore how reciprocity (which appeared as premise 6 in our reconstruction) is working and what would it force us to accept and under which conditions.

5. Examining *General Reciprocity* (GR)

As I have already said, Collins presents the experimenters' regress alternatively as *RR* or as *GR,* something which leads us to think that -despite the suggestions available in the literature, (Cf. Radder (2003) and my (2014 and 2017))-, he did not differentiate between them, either with regard to their scope or to their logical form. In fact, in his (1992) he refers to the experimenters' regress as a "circle" (cf. 1992, p. 84) while in his (2016) he presents it as "a kind of logical regress" (p. 66).

[27] For an excellent study on the different modes and how they operate together see Barnes (1990).

But in an infinite regress situation, we would be checking the quality of an experiment by another one, which, in turn, would have to be checked, and so *ad infinitum*. The second presentation, despite the fact that Collins uses the Latin expression, does not lead to a regression. The knowability of the correct result depends on the correct functioning of the measuring device while the knowability of the correct functioning of the measuring device requires knowing what the correct result is. No infinite regress is generated. General reciprocity has the form of the simplest type of circular reasoning: the reciprocal.

General reciprocity appears as premise (6) of the replication regress. It is quite curious that Collins does not seem to acknowledge that its scope is far more general than the argument in which it appears. If (6) happened to be the case, Collins would be able to show that, not only replication, but most reproduction modalities are problematic.[28] Contrary to Collins, I believe that the problem of determining which result is the *correct* result is not an exclusive problem of the replication of experiments,[29] but that it would haunt any form of reproduction if no particular result is anticipated.

Under which conditions would a reciprocal mode bear sceptical force? To begin with, it must be the case that there is no other way to determine the correct result. Even if it is sometimes the case that a measuring device can offer us the value of a correct experimental result, this is not necessarily so. As I will try to show in the following chapters, theoretical resources are a complementary way of assessing experimental values. If that is so, in controversial cases, we may consider that it is the experimental result which has epistemic priority, in the sense that it is the one that can be introduced in advance and that would allow us to assess whether or not the instrument is working properly.

It is possible to reconstruct general reciprocity in different ways, and it is not very clear to me which of them is the one Collins would prefer. He could be

[28] Except for (proper) independent testing with coincident results.

[29] That he takes this to be the case becomes evident when the quotation I used to present the *RR* is considered. In it, Collins explicitly claims that the experimenters' regress is a *paradox* that appears if replication is used to test an empirical claim.

claiming that (1) well-functioning device and correct result are reciprocally defined, as he suggests in the following quotation:

> Thus, the definition of what counts as a good gravity wave detector, and the resolution to the question of whether gravity waves exists, are congruent social processes. They are the social embodiment of the experimenters' regress. (1992, p. 89).

Or he could be arguing that (2) well-functioning device and correct result can only be known to be such reciprocally (which means pretty much that they cannot be known to be such). If it is the first case, I do not think it is problematic. Even if they are defined reciprocally, the correct result can be known by other means that the measuring device. There would be no epistemic reciprocity involved. The second case, if correct, is stronger, and I believe it should be resisted. We can present general reciprocity as a two premises argument:

(1) The only way to know if *x* is a good *y-detector* is by means of getting the correct experimental outcome *z*.

I don't have much to object to premise (1). Instead, I will focus my attention in (2):

(2) The only way to know if *z* is a correct experimental outcome is by means of a good *y-detector*.

To grasp correctly the scope of this premise is of the utmost importance since it will be the target of our critique. Although this thesis is central to Collins' work, he does not provide us with any compelling reason to accept it. He claims:

> But what is the correct outcome? To find this out we must build a good detector and have a look. (Collins 1992, p. 84).

Collins should have justified (2). For it is quite problematic to assert that, for example, the only criterion to determine whether the reference of a theoretical entity exists is the outcome of an experiment. In saying so, he seems to assume that in an experiment the reference of a theoretical entity can be detected directly. I will claim that this premise rests on a mistaken understanding of what an experimental result is. For it assumes that experimental results lack a theoretical component. If we can show that (2) is false, then the reciprocal mode would have no force, since there would be a reasonable and scientific way out of the reciprocity. Moreover, blocking the reciprocity also precludes the infinite regress to take place and allows a rational overcoming of possible disagreement.

6. Collins' claim: *General Reciprocity* can only be overcome by non-scientific resources.

We have presented Collins' challenge as two arguments working together, plus a symptom: persistent disagreement. I claimed that the challenge to the rationality of scientific practice arises when it is noticed that the infinite regress and disagreement are rooted in a reciprocal scenario in which the only way to determine a proper experimental result is anchored in a well-functioning experiment.

Having set up Collins' challenge to the rationality of experimentation as a sceptical system demanding an a-rational resolution, it must be explained how it is that scientists reach consensus. That they do so is an obvious empirical fact. The explanation that Collins provides is that the way out from the experimenters' regress is achieved by means of applying non-scientific strategies:

> Some non-scientific tactics must be employed because the resources of the experiment alone are insufficient. (Collins 1992, p. 143).

As I have previously suggested, we are not forced to accept such an explanation unless we presuppose that to resolve an experimental disagreement *scientifically* is to resolve it experimentaly; this assumption plays a crucial role in Collins's proposal. The way out from reciprocity has to be explained by appealing to an extra-experimental criterion, we will grant Collins that. But for Collins, as we have seen, extra-experimental equals external, contingent, social, political, economic, etc. This would be the case if scientific criteria were exclusively experimental, which is clearly not the case.

The reciprocity Collins is appealing to is a problem which any foundationalist stance with respect to the justification of empirical knowledge has to deal with. Accordingly, the denial of (2) requires the adoption of a minimal coherentist approach. Therefore, I will claim that an experimental result is a complex entity that is not introduced in a purely experimental way, but one which possesses theoretical content. If that is the case, the reciprocity would not take place, not because extra-scientific factors are providing closure to the debate, but because there is an epistemic criterion, independent of the judgment about the proper functioning of the experimental device, that helps us to determine what the correct result is or at least to narrow the set of acceptable results given the accepted scientific knowledge available. For example, if we consider the gravity wave detection episode –which is Collins' favourite example in support of his argument for the external closure of the experimenters' regress- Weber's findings were highly improbable in the context of existing physical and cosmological knowledge and, moreover, had the results been correct, they would have been accompanied by observable effects which were not in fact observed; these, when reconsidered in the light of several negative results, made it quite reasonable to consider that Weber's experimental results were incorrect. (Cf. Levine, 2004). In other words, reciprocity was broken with the aid of theoretical considerations, as well as with the aid of independent empirical evidence.[30]

In *Changing Order*, as well as in *The Golem* series, Collins presents and analyses several case studies. While the *TEA laser* case represents the good, purely empirical experimental situation in which there is a clear way of determining the proper functioning of the apparatus, the correct experimental outcome and the acquisition of the experimental relevant skills, the rest of the cases he deals with are all subject to the experimenters' regress. As such, in the case of disagreement, they would all require a non-experimental resolution. However, Collins claims (in a personal communication) that he believes that some experimental disagreement can be solved by appealing to theoretical arguments. If this is the case, and if experimental disputes can be sometimes settled by theoretical considerations, then thesis (2) of general reciprocity is false and the replication regress does not follow. We may have to understand that Collins is claiming that the regress is not ubiquitous, in contrast to what textual evidence indicates. Furthermore, in a recent publication, he claims:

> Yet the phenomenon is a general one –it is meant to apply to all deeply held scientific controversies in any field of science and if it failed to reveal itself in some other deeply controversial experimental field, it would represent a challenge to the original claim. […] Nevertheless, the point is meant to apply not only to TEA-lasers, but to all experimentation, and it can be tested by others looking at entirely different sciences. […] To draw an analytic conclusion, all we have to do is to agree that this is something that could have happened, not that it did happen. (2016, p. 76).

And later, he insists:

[30] This possibility is compatible with provisional ἐποχή, until further and complementary evidence is available.

> The importance of scientific results and non-scientific social factors will differ from case to case but the experimenters' regress shows that experiment alone cannot force a scientist to accept a view that they are determined to resist. Whatever, the eventual collective decision about what constitutes the right result will be coextensive with the collective decision about which are the competently performed experiments. It is the outcome of such experiments that sets the criterion for a well performed experiment once closure is reached. (2016, pp. 67-68).

In any event, in order to be sufficiently interesting, the scope of the sceptical challenge should be quite wide, i.e. the thesis should not be restricted to very rare and exceptional cases. If this is so and his thesis aims to be applied to enough interesting cases, then my criticism would still hold, for it claims that in most cases such a resolution of the regress does not apply. I will deny that the external resolution of the regress applies to many of the cases in which Collins claims the opposite, such as, for example, the gravity wave detection episode discussed in chapter five and the Vitamin C as a cure for cancer that I discuss in the sixth chapter.

7. My claim: *General Reciprocity* can be overcome by scientific resources.

So far, I have explained the two senses of the experimenters' regress and the relation between these two problems. The *RR* confronted us with the problem of judging whether a replication had been performed correctly. This could not be done by experimental resources because of the reciprocal determination of the experimental results and properly conducted experiments. This in turn, explains the disagreement between researchers. According to Collins this is something which cannot be overcome with the aid of internal aids, since no epistemic device would help us to avoid general reciprocity.

Given this micro-experimental panorama, Collins invoked reasons external to science itself. I suggested, on the contrary, that it was possible to deny the truth of (2), which postulates that the determination of the proper functioning of an experimental arrangement is reciprocal with determining the correctness of an experimental result. The next step is to show that (2) is false, and hence, that there is no reciprocity involved in scientific practice. I will do so by denying that the introduction of the correct result is necessarily experimental, hence rejecting thesis (2). With this goal in mind, the following chapter will present and analyse three paradigmatic physic experiments that will serve as the *empirical basis* for studying the content of an experimental result. After this presentation, I will be in a position to show how the representational content of experimental results is acquired, and how the semantics of experimental results supports the view that in conflictive situations, an experimental result may be theoretically introduced. If that is the case, then there is a way out from *GR* that does not require non-epistemic explanations. I will then present the two cases with which Collins illustrates the experimenters' regress and present a different way out from it: chapter five dwells with the Gravitational radiation case, while chapter six discusses the Vitamin C controversy. I will also add a historical analysis of those episodes, for I also believe that Collins' narrative is quite misleading and I fail to see exactly how the external factors Collins' highlights, work.

Chapter Four

Case Studies and the Semantics of Experimental Results

In what follows I will briefly describe three kinds of experiments in physics, which differ regarding the ontological status of what they purport to detect. I will call them quantitative, qualitative and existential experiments. A characterization of each of them will precede an analysis of a paradigmatic case. This chapter presents the case studies needed for understanding how to bridge the gap between what is in fact perceived as the outcome of the *material realization* of an experiment[31] and what is claimed to be its result.

1. A quantitative experiment: Michelson and the speed of light

As examples of quantitative experiments I understand experiments such as Michelson's measurements of the speed of light and Cavendish's determination of the gravitational constant. Also determining the mass of a elementary particle requires performing a quantitative experiment.[32]

All these experiments assume the existence of a kind of entity or a process-type or an event-type, and they also assume that the entity, etc. under investigation possesses a certain quantitative property. The goal of these experiments is to determine precisely a magnitude for a quantity that is already introduced by a theory. They yield a magnitude -with an associated error- for a specific as the final result of the experiment.

31 I borrow this apt expression from Hans Radder. See for example his (1992) and (2003).

32 Even though the determination of the existence of the particle would require an existential experiment. For an analysis of the determination of the mass of the top quark under the light of the experimenters' regress see Perović (2017).

I will briefly portray here one of the first experiments with which Albert Michelson measured the velocity of light (Michelson 1880). During 1877 Michelson found a way to improve Foucault's revolving mirror device so as to provide a more accurate measurement of the velocity of light. As is well known, Kepler's optical investigations show that the intensity of a light source decreases with the square of the distance, so the longer the distance that light travels through the experimental arrangement, the less distinct the output is. But the shorter the distance travelled, the harder it is to measure the output. Michelson avoided both problems by using a spherical lens of great focal length (L), placing the revolving mirror (R) within the principal focus of L and replacing the original spherical fixed mirror with a plane one (M). In figure 2 we will find a schema of the experimental design he proposed.

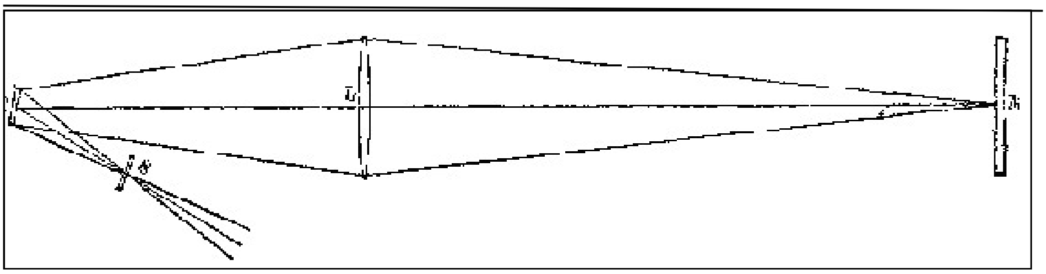

Figure 2. A schema of the experimental arrangement. Taken from Michelson (1880).

Let me now describe the principles that govern the design of the experiment. Consider a ray of light travelling from the source (S)[33] to the revolving mirror (R) through the lens (L). If R is at rest, the ray will form an image at M, and because of the law of reflection, it will return to R and finally to S. But if R were to rotate on its axis, a new light spot, deflected in the direction of rotation of R, will be formed.

[33] S is not only the source of light but also an observatory of the output of the experiment.

Figure 3 exemplifies how R's change in position can cause a second bright spot on S.

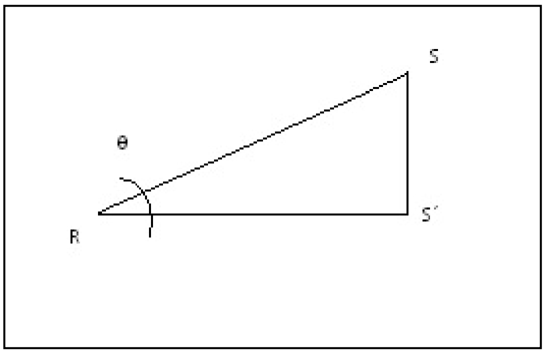

Figure 3. Θ, the deflection angle.

S'R S, (from now on, θ) is the *deflection angle*. It is subtended by the rays of light whose origin is the revolving mirror, R. θ is half the angle through which the mirror has turned since the departure of the ray of light from S to M and finally back to R. Calculating the value of θ is crucial for this experiment. For if we measure the distance between S and S' and we measure the length of the segment R-S we can then calculate the tangent of θ and its inverse function, which gives us the value of θ in radians. Once this variable is known, and together with the number of revolutions per second that R performed,[34] if velocity is the ratio between distance and time, if the distance is 2RM and time is represented by: (θ/2)/n.360,[35] then we have that:

---

[34] Michelson calculated this using a stroboscope.

[35] This equation is obtained by considering:

ν= n= cycles per second;

360 ν = angle/time

T= angle/ 360 ν

If as was said before the angle of interest is θ/2, then:

T= (θ/2)/ 360 n

$$V = \frac{\underline{2 \times 360\, n \times 2\, RM}}{\arctan(S'S/\, RS')}$$

We can then calculate the time that light took to travel the distance considered. Now that we have at least a rough idea of how the experiment works, we are able to single out different aspects of the process of producing the result. To begin with: What is the output in this experimental arrangement? Figure 4 is a magnification of the output.

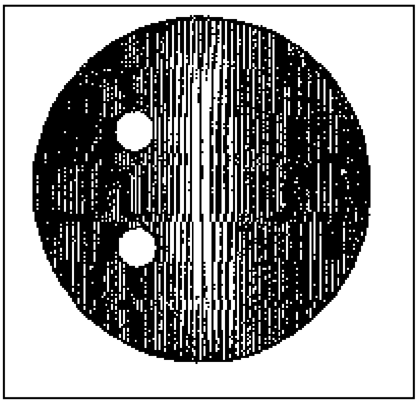

Figure 4. The output of the experiment. Taken from Michelson (1880).

It is doubtful that this image, on its own, can tell us anything regarding the nature of light and its celerity. It is only when it is interpreted as something else that it can be informative, and this requires theoretical interpretation. Michelson uses a micrometre in order to measure the distance between the bright spots. In doing so he is no longer concerned with the dots, but with a specific relationship between them: distance. In this example, the distance is the salient feature to consider because it will allow the researcher to calculate the tangent of θ. And with this, it will be possible to relate the displacement of the mirror to the time required for the formation of the second image, S’. This is a first step in the output’s acquisition of representational content, which is *laden* with two theories: a measurement theory (which establishes that between S and S’ there is a length) and a branch of geometry: trigonometry (which informs that the segment S-S’ is the tangent of θ).

Later on, every magnitude for the measured length is introduced in an equation that relates the different variables in the experiment and that enables us to calculate the velocity of light between the two mirrors of the experimental arrangement. This requires a new interpretative step, this time provided by empirical, particularly, physical theories, such as kinematics. Afterwards, the data collected is reduced by means of the application of a statistical method. This reduction implies further theoretical interpretation. In the experiment just considered, Michelson studies the different sources of error and calculates the mean of the measurements and its standard deviation. He announces the final result to be: V= 299944 $\pm$ 51 km/s (Michelson 1880, p. 141).[36]

2. A qualitative experiment: Newton and the composition of white light.

The aim of a qualitative experiment is to determine some of the properties of a previously detected entity-type. The purpose of these experiments is to gain knowledge about these entities by means of detecting the different properties that they may have. For instance, once the detection of neutrinos took place, efforts were devoted to discovering whether they oscillate or not. The experiment conducted in the Canadian observatory *SNO* is a good example. The prism experiment carried out by Newton, in order to analyse the composition of white light, also belongs to this category. The first experiment will show whether neutrinos oscillate or not; the prism experiment will show that white light is

[36] This value contemplates the correction for vacuum which was theoretically calculated. (Michelson 1880, p. 141).

composed of light of different colours.[37] The result of these experiments is the attribution of a new property to an entity or process. One might ask if the distinction between quantitative and qualitative experiments is properly justified. I consider that for epistemic reasons it is worth emphasizing the difference between measuring a property and claiming that a new property can be predicated of a system or entity. As the kind of errors associated with each kind of experiment may differ, the kind of change expected in the interpretations of the results of each experiment given theoretical change may also differ. Moreover, because not every property that a system can possess will be a gradable property, and when an absolute property is discovered it may not involve a measurement, as is the case with Newton's prism experiment, preserving the distinction is also relevant for conceptual reasons.

In what follows, I will analyse Newton's experiment on the composition of white light as a paradigmatic example of this category.

Among his several optical studies, Newton devoted himself to providing an account of the phenomenon of colour and of the nature of white light. Here I would like to briefly consider one of the experiments he presents in the letter he wrote to the Royal Society of London in 1671 according to which he demonstrated the composition of white light and the differential refraction of the simple rays that constitute it.[38] There he claims:

> Sir, to perform my late promise to you, I shall without further ceremony acquaint you, that in the beginning of year 1666, [...], I procured me a triangular glass prism, to try therewith the celebrated phenomena of colours. And in order thereto having darkened my chamber, and made a small hole in my window shuts, to let in a convenient quantity of the Sun's light, I placed my prism at his entrance, that it might be thereby refracted to the opposite

[37] By stating Newton's result in these terms, I am trying to avoid both anachronism and the appeal to concepts belonging to the wave-theory to explain his results.

[38] Newton's *experimentum crucis* generated a lot of controversy among scholars. Here I will merely offer a possible analysis of the experiment in order to show how experimental results gain their representational content.

wall. It was at first a very pleasing divertissement, to view the vivid and intense colours produced thereby; but after a while applying myself to consider them more circumspectly, I became surprised to see them in an *oblong* form; which, according to the received laws of Refraction, I expected should have been circular. (Newton 1671, pp. 3075-3076).

Notice in figure 5 the schema he offers. It should be read from right to left. It shows how a ray of sunlight passes through a prism in a minimum deviation position. The image projected in the opposite wall is oblong.

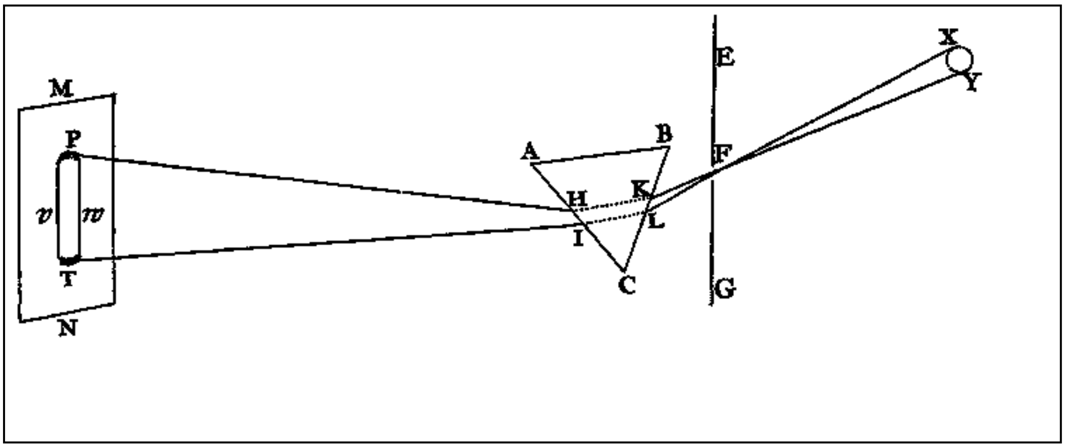

Figure 5. A representation of the first refraction. Taken from Newton (1704).

Why did Newton express surprise? Because according to the received laws of refraction, the angle of refraction of a ray of light depends only on the angle of incidence and on the variation of the refraction index of the media. Therefore, if the different rays of light pass through the same medium, the prism, and if the medium is isotropic, then they should be equally refracted, and hence, they should produce a circular image on the opposite wall. If we follow Fig. 5 from right to left, we will notice that there is a refraction that corresponds to the surprising oblong image that Newton reported in the letter I quoted.[39]

After disregarding the possibility of the oblong image being an artefact, he considered a second refraction to understand what these images suggested regarding the nature of light. By slightly rotating the prism on its horizontal axis, Newton was able to selectively project, on a second panel, regions of the spectrum formed on the first panel, and to study the behaviour of the rays when undergoing a second refraction. The conclusion that Newton reaches is that the rays that undergo the most extreme deviation during the first refraction are those which also experience the most extreme deviation during the second. These rays can be individuated by means of their colour, and, when separated, their images are circular, as was expected. Newton would show, then, that white light is a compound of rays of different colours which manifest a particular refrangibility.

Figure 6 displays a representation of the *experimentum crucis'* experimental arrangement that can be helpful to keep in mind. Again, it should be read from right to left. It represents how a ray of sunlight passes through an orifice and undergoes a prism-induced refraction. The refracted ray of light casts an image in the wall *DGE* which, as Newton states, is oblong, instead of circular. A portion of the refracted ray of light (individuated by its colour) passes through two more panels before undergoing a further refraction. Rotating the first prism allows Newton to select which part of the spectrum will undergo a second refraction. He notices that those parts of the spectrum that were most deviated during the first refraction are those that are most deviated in the second one. This filtered ray of light does indeed produce a circular image in the third panel as expected.

39 This will hold only if the prism is the position of minimum deviation. I cannot argue for this here, but I refer the interested reader to Westfall (1962) for a detailed explanation.

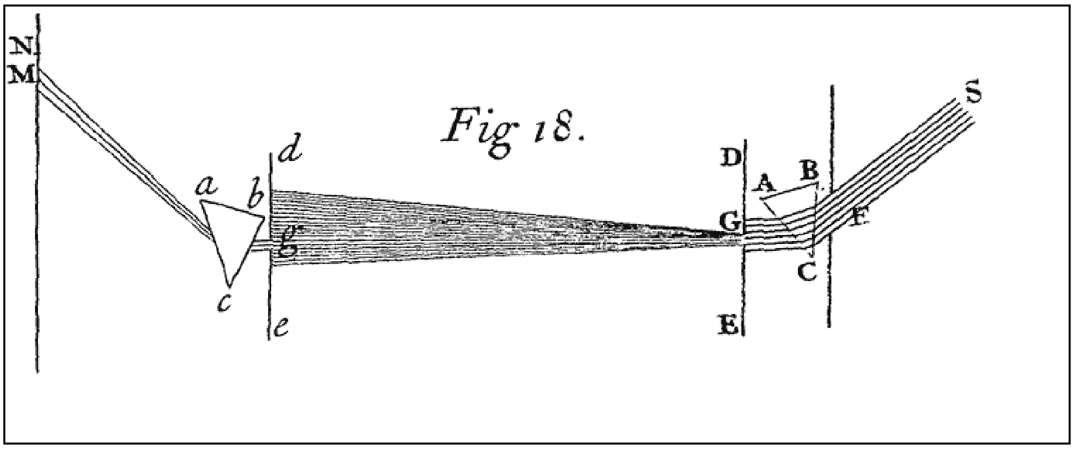

Figure. 6. An *Experimentum crucis'* schema. Taken from Newton (1704).

Let me now consider the different interpretative steps involved in this experiment. To begin with, we can find two images that will constitute the output of the experiment: the multi-coloured oblong spectrum on the first panel and the monochromatic circular image cast on the third panel. A comparison of the shape, the position and the colour of each image seems to be what is required, in this experiment, for conceptualizing the output. Again, the relevance of these features arises from the theoretical background assumed in the experiment. For instance, for conceptualizing the output, geometrical optics is presupposed. As I said before, according to Snell's laws, given that the rays of light under examination go through the same medium, they should be refracted with the same angle, if the prism is in its minimum deviation position.

In this experiment, in contrast to the previous case I presented, there isn't any statistical analysis, but an extra experiment to determine whether the oblong image could be an artefact. It consisted in making the ray undergo a second refraction through a prism in the inverse position to see if an image of the original source of light could be obtained.

Finally, in order to explain the different angle with which each ray of light refracts on the second panel, Newton introduces the property of *differential refrangibility*, a dispositional property of the simple rays that constitute white light.

3. An existential experiment: Reines and Cowan and the detection of Neutrinos

Existential experiments are conducted to find out whether an entity-type exists, or whether a process-type or an event-type takes place or does not take place. They usually involve searching for the referent of a concept introduced by a theory. Examples of this kind of experiment include Weber's attempt to detect gravity waves, predicted by general theory of relativity and Reines and Cowan's attempt to detect Neutrinos, those ghostly particles that Pauli posited in 1933 to account for the energy missing during beta decay, and hence to preserve the principle of conservation of energy. The results of this type of experiment can be stated as an affirmative or negative existential statement about a class of entities or processes. In what follows, I will introduce to the reader the detection of neutrinos after Pauli's theoretical introduction.

During the study of radioactive processes, in particular that of *beta decay*, Wolfgang Pauli acknowledged that there was a loss of energy from the nucleus that was left unaccounted and that could not be recovered by means of the measuring apparatus. Two alternative explanations were offered: the first was quite revolutionary, conservation laws were not preserved for the subatomic realm (Cf. Reines and Cowan, 1956a, p. 446). A less radical explanation consisted of attributing the missing energy to a new subatomic neutral and almost zero mass particle.

Reines and Cowan thought of a way to detect the elusive particles. Taking into account the simplest form of a beta-decay process, namely, the decay of a free neutron:

$$(1)\quad n^{o} \rightarrow p^{+} + \beta^{-} + \nu^{-}$$

They claimed:

> If the neutrino is a real particle carrying the missing energy and momentum from the site of a beta decay, then the discovery of these missing items at some other place would demonstrate its reality. Thus, if negative beta decays as in equation (1) could be associated at another location with the inverse reaction:
>
> $$(2)\ \nu^{-} + p^{+} \rightarrow \beta^{+} + n^{o}$$
>
> Which is observed to occur at the predicted rate, the case would be closed. (1956a, p. 447).

The experimenters looked for a facility that could provide an intense neutrino flux to test if the inverse reaction could take place. They made use of Hanford site facilities in a first attempt, using a detector with target protons in a hydrogen liquid scintillator (Cf. 1956b, p. 103).[40] But after some preliminary tests, they moved the experiment to the Savannah River Plant of the U.S Atomic Energy Commission in order to isolate the device from cosmic rays. The emplacement of the detector device in the vicinity of a nuclear reactor would provide a rather high flux of neutrinos they needed for the experiment to work (around 10^{12} neutrinos per second).

The operation of the device is designed upon the following physical ideas. The positron, product of the reaction between a neutrino and a proton, would interact with an electron and the energy that results of their annihilation would be detected as a pair of gamma rays emitted in opposite directions. In order to show that the specific reaction they were looking for was taking place, the researchers hoped they would also be able to detect the neutrons expected for that reaction. While the gamma rays were detected by liquid scintillators, the neutrons were detected by their interaction with cadmium 108, which absorbs a neutron and enters into an exited state, namely, cadmium 109 and emits a gamma ray. Therefore, the observational signature of this inverse beta decay reaction is the emission of a

[40] Which was a nuclear production complex devoted to the Manhattan project.

pair of gamma rays in opposite directions and of another gamma ray, 5×10^{-6} seconds after the annihilation of the electron-positron pair.

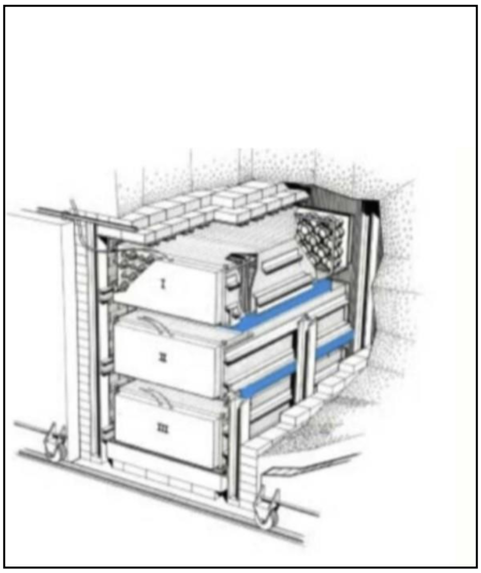

Figure 7. A sketch of the detector. Taken from Maiani (2014).

The device, as Reines and Cowan referred to it, was like a "three layered sandwich" (Cf. 1956b, p. 103), the "bread" layers corresponding to the scintillators detectors, while the "meat" layers work as targets for the neutrons, hence, they contain a water solution of cadmium chloride. It was shielded with paraffin and lead and placed underground to prevent cosmic rays and neutrons from the nuclear reactor, both potential sources of equivocity. The signals were analysed by two independent sets of equipment and the detector was periodically calibrated.

As of the results, a reactor-power-dependent-signal was detected in agreement with the cross section for the reaction. The neutrino signal rate was 2.88 ± 0.22 counts per hour and the signal-to-background ratio was about 3 to 1, and was a linear function of the protons present in the device as targets for the neutrinos. (Cf. 1956 b, p. 104).

Once again we can detect different interpretative steps in this experiment. A complex output, consisting in two perpendicular bright lines that appear in constant conjunction with another, emitted 5 x 10^{-6} seconds later. A first interpretative step, by which the output becomes *data*: the positron-electron annihilation *cum* cadmium 109 decay. Further treatment is produced offering the amount of counts of these phenomenon: 2.88 $\pm$ 0.22 counts per hour. Finally, A third interpretative step, by which these data are traced back to their context of production: an inverse beta decay is taking place: the positron is the product of the reaction between a neutrino and a proton, while the neutron is interacting with the cadmium.

4. The representational content of experimental results

Collins claimed that the determination of the correctness of an experimental result is coextensive with the determination of the proper functioning of the experimental device: thesis (2) of *GR.* He also claimed that breaking such reciprocity requires the appeal to external factors. But is this necessarily so? Can we find an alternative way of introducing the correct result that helps us to break the reciprocity without requiring an externalist explanation? I believe so. To put forward this alternative answer I will have to partially address what Marcel Weber (2012) refers to as the problem of the *representational content of data*. He claims:

> [R]eliable data are correct representations of an underlying reality, whereas so-called artifacts are incorrect representations. This characterization assumes that data have some sort of representational content; they represent an object *as* instantiating some property or properties […]. The question of what it means for data to represent their object correctly has not been much discussed in this context. (Weber 2012, his emphasis).

I will offer an approach that may help us understand how an experimental result acquires its representational content. In order to do so I will now draw some general conclusions about how the representational content of experimental results is gained. I will do so by means of proposing a *semantics of experimental results*. This proposal shares with Sabina Leonelli's view on *data*, the idea that it is not purely given *but made*,

> [T]he idea that the same set of data can act as evidence for a variety of knowledge claims, depending on how they are interpreted. (Cf. 2016, p. 71).[41]

One of the central problems regarding the content of experimental results consists in determining the process through which raw data is transformed into information about the natural world, for example, how a click in a Geiger counter is linked to the flux of solar neutrinos. If we consider the output to be the observable outcome of the experimental apparatus, it is necessary to provide it with meaning. Indeed, the final product -the outcome- of an experimental run is a directly observable event, such as the movement of a needle in a voltmeter, the sound of a Geiger counter or a line in a cloud chamber. (cf. Díez, 2002). However, *output* and *final result* seem to be quite different. Depending on the experiment, the position of the needle would allow us to claim that we have detected gravity waves, whereas the sounds emitted by the Geiger counter would, in turn, lead us to count

[41] In her *Data-Centric Biology,* Sabina Leonelli presents a brilliant proposal regarding the "travelling" of data and how it is manipulated and packaged in order to serve as evidence for different knowledge claims. Her approach emerges from studying experiments that produces large amount of outputs that can be analysed through data-mining for several purposes so that the data eventually become independent from its context of production and can be reinterpreted. I agree with Leonelli, however, given the scope and the goals of this thesis, I will only deal with interpreting experimental evidence under the light of the questions posed by a specific experiment.

solar neutrinos.[42] This can be understood as the process by which outcomes acquire representational content, representing an object as instantiating a certain property or set of properties; this is what I call a *semantics for experimental results.* Outcomes, I claim, gain their representational content by means of a sequence of interpretative steps. I will now show how this is the case by drawing some general conclusions from the experiments presented in the previous section.

Recall, for instance, Michelson's experiment. The *output* is an indirect indicator of the velocity of light insofar it is related to the laws and the theoretical assumptions that govern the experimental apparatus. *Data* are obtained when actually interpreting the relevant features of the *output* (the distance between the dots) under these laws and assumptions. Moreover, given that the measurement of a quantitative property which can be instantiated in different degrees is at stake, a suitable statistical analysis has to be applied to the *data*. The theoretical construct obtained after *data* reduction can be called an *e-result*, where the *e* stands for *experimental*. I take this to be, strictly speaking, the final contribution to the content of an experimental result that the experiment can provide. Finally, given that the attribution of a finite velocity to light is not compatible with every theory regarding its nature, the experimental knowledge obtained has to be subsumed under a theory that can accommodate the result. I call this last interpretative step *external interpretation*, and its product is what I will call a *t-result*. In this case, the *t* stands for *theoretical*. I take this to be the theoretical explanation of the *e-result*, and one of the layers of an experimental result that enables the question regarding whether or not it is a correct or an incorrect representation of the phenomena that the experiment aims at detecting.

[42] For an externalist proposal regarding the interpretation of experimental results, see Pinch (1993).

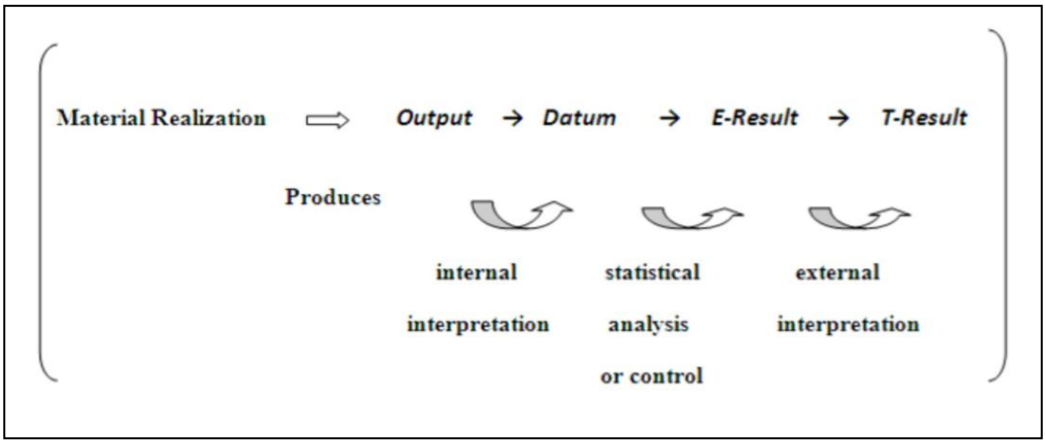

Table 3. A model for the semantics of an experimental result

More generally, the *output* produced in each material realization will undergo several interpretative steps. It will acquire part of its meaning when subsumed under a concept, whether classificatory or metrical, by means of a process that I will call *internal interpretation*. As a consequence, we will get *data* that will undergo statistical analysis or be subjected to the controls that are required for the kind of experiment in question. As a result, we will obtain what I call an *e-result*, that is the final result of the experiment, and, finally, this result will be subsumed under an explanatory theory.[43] *External interpretation* can be thought of as an explanation, in the sense that the *e-result* of an experiment can be thought of as the experiment's *explanandum* while the theories that contribute to the meaning of the *t-result* explain why we measured what we had measured or why we observed what we observed (i.e. they provide the *explanans* for the *e-result*).[44] Each of these theoretical constructs can be revised and may obviously undergo a change in

[43] In this thesis I would like to remain neutral regarding any specific account on what a scientific theory is. I take my proposal to be compatible both with syntactic and with semantic approaches. However, it has been influenced by the Structuralist conception of scientific theories developed by Balzer, Moulines and Sneed (1987) and José Díez's research on the content of scientific concepts (2002).

[44] I believe this proposal fits nicely with Giora Hon's analysis of experimental errors in terms of idols of experiment. See for example, his (2003).

interpretation and therefore in meaning, given a change in theory.[45] Table 3 is a general schema that can be abstracted from the analysis of the experiments I have offered.

While table 4 consists of an abridged presentation of the elements in each of the examples provided above.

	Velocity of Light	***Differential Refrangibility***	***Neutrinos***
Output	Image of two bright dots.	Polychrome oblong spectrum+ monochrome round images + their positions.	Two bright perpendicular lines in constant conjunction with a delayed one.
Datum	Length- Tangent of angle S'RS, relation between the variables in the experimental arrangement.	Shape variation + colour variation + same position equals same colour.	Gamma ray produced by electron-positron annihilation and emission of another gamma ray in decay of cadmium 109.
E-result	Average of the velocities measured.	Control: the oblong image is not an artefact.	Control/Analysis
T-result	The velocity of light is 299944 $\pm$ 51 km	Simple rays are differentially refrangible.	2.88 $\pm$ 0.22 counts per hour. Representing the counted neutrinos.

Table 4. The elements of the experimental results considered.

[45] This, in turn, invites us to reconsider the scope of Hacking's catchy slogan according to which experiments have a life of their own. (See my 2011).

5. The semantics of experimental results and *General Reciprocity*

What is the relevance of the semantics of an experimental result when trying to offer an account of the way out from the *GR*? Let us recall the second thesis of the general reciprocity argument, which was also a necessary premise to run the infinite regress argument, i.e. the *RR*. It stated that in order to know what the correct experimental result was, the only information available was the one that a good experimental arrangement could provide us with. However, as I showed in the previous section, an experimental result is not merely the *outcome* of an experiment, but it possesses representational content. I understand that Collins has to agree with this claim since he makes explicit reference to interpreted experimental results. (Cf. Collins, 1992, pp. 81-82). Therefore, he has to concede that for an *outcome* to correctly represent a phenomenon it has to possess representational content. Acquiring representational content requires theoretically interpreting the outcome. If that is the case, then the correctness of the representational content of an experimental result cannot be judged merely by evaluating the correct functioning of the experimental arrangement, but by taking into account the adjustment between our theoretical expectations and the available relevant accepted knowledge with the *e-result*. Hence, even if clinching an experimental result is not done exclusively by the experiment, this does not mean that it requires extra-scientific criteria, as Collins claims. It just means that experiments may require the aid of theoretical considerations, i.e: extra experimental and intra-scientific considerations, which are precisely those required for providing an *external interpretation* of the experimental result. This fits nicely with Kuhn's approach towards the reliability of experimental techniques:

> When measurement is insecure, one of the tests for reliability of existing instruments and manipulative techniques must inevitably be their ability to give results that compare favorably with existing theory. In some parts of natural science, the adequacy of experimental technique can be judged only in this way. When that occurs, one may not even speak of "insecure"

> instrumentation or technique, implying that this could be improved without recourse to an external theoretical standard. (Kuhn 1977, pp. 194-195).

Therefore, the determination of which is the correct result, especially in problematic cases, can be done by appealing to the coherence between one of the theories that would provide the external interpretation of the result, a theory that has to be compatible with current accepted knowledge.[46,47] Consequently, we would be justified in claiming that the experiment is functioning properly if it provides outcomes that are compatible with accepted and robust knowledge, in what can be considered a *theoretical calibration* of an experiment.

At this point, the interested reader may well wonder why I appeal to a *calibration of an experiment by a theory* but I do not highlight the relevance of empirical calibration when overcoming the *GR*. Let me explain why this is the case. Calibration is the use of a surrogate signal to standardize an instrument. (Franklin 1999, p. 237). In order to perform a traditional calibration, the researchers must know the signal that the experimental arrangement purports to detect and they must know how to manipulate it. But this presupposes that general reciprocity is already broken! A different kind of calibration, typical for the sort of experiments that are more prone to disagreement (those that seek to detect a hitherto unobserved phenomenon), is one in which a device is calibrated against a surrogate signal which is presumably similar in relevant respects to the signal the experiment is aimed to detect. Nevertheless, even Franklin, for whom calibration is an epistemological strategy that can help us to validate experimental results, considers that in cases such as Weber's attempts to detect gravity waves, calibration cannot be decisive. (Franklin 2002, p. 64).

---

[46] It would be interesting to determine whether (im)proper working of the experimental devices and (in)compatibility with accepted theoretical knowledge will always have the same epistemic weight. Indeed empiricist philosophers and empirically-minded scientists will tend to say that, ultimately, empirical arguments should count more than theoretical ones. However, I do not have, at least not yet, a positive proposal regarding this particular aspect and my proposal does not require that such an answer be offered.

[47] This goes in line with Perović's considerations regarding the interplay of theory and experiment at the LHC. (Perović, 2017).

I do believe, nevertheless, that there is a special kind of calibration that may play an important role in determining when two instruments are replicas: I am referring to *reciprocal calibration*, which consists in using a surrogate signal (other than the signal that the arrangements purport to detect, but assumed to be similar in the relevant respects) to check if two experimental arrangements respond in the same way to it and therefore, to infer *functional identity* between them (see my 2010). However, I do not think that a coincidence in this respect is sufficient for claiming that the experiments are working properly. What a reciprocal calibration can show is that two experiments are offering the same output given the same input, but it does not suffice for showing that the arrangements are functioning properly and are yielding correct outputs. Identity of outputs (identity to a certain extent, of course) is perfectly compatible with two experimental arrangements malfunctioning. Experimental replicability is not a sufficient condition for establishing the correctness of an experimental result, for, as I have shown, general reciprocity is a broader phenomenon than the problem of the regress in replication.

Another worry that may arise is related to the general scope of my strategy. Is it helpful to overcome the experimenters' regress by internal means in any kind of experimental situation? For example, imagine an episode in which two experiments yielded discordant *e-results* yet both compatible with accepted theories. In such a scenario it is not clear up to what extent the *t-result* may help to decide between both values in order to overcome a possible disagreement. Such a situation could arise in quantitative experiments, where determining the value of a parameter is at stake. Insofar the value with which a certain parameter instantiates cannot be inferred from a theory or group of theories, it is true that the *external interpretation* in quantitative experiments do not constraint *e-results* as much as it does in qualitative or existential experiments. It will, however, restrict the general features that *e-results* can take. In the measurement of the speed of light, for example, the most the *external interpretation* can contribute to the determination of the *e-result* concerns *surface* and *qualitative* features of the property, such as its finite character and the constancy of the value. However, since I believe that the values of certain parameters are brute facts that have to be determined empirically, I do not think that I would like to endorse a proposal that does not allow this situation to be the case. One option in such situations is to appeal not just to a theoretical calibration but also -as we claimed in chapter two- to a triangulation with indirect determinations that can contribute to the calibration of the experimental techniques. This would enable a non-experimental, yet internal introduction of the experimental result which would be compatible with the denial of thesis (2) of the reconstruction of the general reciprocity that I presented in the previous chapter, namely, independent testing.[48]

[48] There is a more general concern related to a coherentist approach when accepting an experimental result. How committed should we be to this principle? I must confess I do not have an answer to this question. What is clear to me is that coherence would only work during normal science, and would not be recommended in a revolutionary process.

Chapter Five

The Gravity Wave Case Discussed.

In this chapter, I will consider Collins' favourite example which he offers in support of his view. This is Weber's Gravity wave detection episode: the first attempt to detect gravity waves. I will begin by offering an introduction to gravity waves in the light of General Relativity. I will then explain some technicalities of the experimental procedures devised to detect them in the sixties. I will then focus on how Collins reads the episode as an illustration of the experimenters' regress overcome *via* external factors and contrast this reading with Franklin's. I will argue that Franklin's stance is better supported by historical evidence. Finally I will claim that the episode does not exemplify an instance of the replication regress but of general reciprocity, and show how the model developed in chapter four could have helped us to *epistemically* break the circle had it been needed.

1. General relativity and the quest for gravity waves.

General relativity is one of the greatest achievements of science. Its predictive power is remarkable. Not only it could accommodate all the known facts that Newtonian dynamics could explain, but was also capable of explaining currently known and unexplained phenomena, such as the advance of the perihelion of Mercury (which could be considered an anomaly within classical mechanics) and the equivalence between gravitational mass and inertial mass that was left as an unexplained coincidence in Newton's Mechanics. It also got rid of *action at distance* forces, proposing an alternative way of conceiving gravity as a geometrical property of spacetime, a result of the curvature of spacetime, where masses simply move following an inertial path.[49] If that were not enough for a theory to earn praise, it made novel predictions, that scientists took a while to put under test, such as the deflection of light near to massive bodies that change the curvature of spacetime, the phenomenon of gravitational lensing, the increase of mass due to acceleration and the increase of the life-time of a particle when it is accelerated at speeds that approximate the speed of light.[50] But among its several predictions, there is one that from 1960 onwards received more and more attention from the physics community: the existence of gravity waves. This prediction was only confirmed after almost 100 years of research when the LIGO experiment, which can be considered the first direct detection method for gravitational radiation, offered robust evidence for the existence of these elusive waves. Until 2016, there was indirect evidence in favour of their existence, arising from the behaviour of the binary pulsar PSR 1913+16, discovered by Hulse and Taylor. In a nutshell, the idea behind this indirect measurement was to calculate the loss of orbital angular momentum and to claim that the best explanation for the data obtained was that the system was radiating gravity waves. (Cf. McCulloch et al., 1980).

With respect to direct detection methods, since 1960 several research teams from different parts of the world have been designing different kinds of apparatus in order to discover if the solutions for gravity waves could be given physical meaning. Einstein himself in his (1916) and (1918) presented two solutions according to which gravity waves were produced.[51]

---

49 This could be misleading; please see the following pages for an amendment that involves taking into account the nonlinearity of the theory.

50 These two predictions came also from special relativity.

51 It is interesting to notice that whether gravity waves exists or do not exist is not only important from a purely scientific stance or because of the practical consequences of such a discovery –imagine that you could find a way of manipulating gravity waves and you could transmit messages at the speed of light!- but also from a philosophical perspective. Their existence could be thought as a good reason to favour a substantivalist approach concerning spacetime over a purely relationist conception. That leaves us wondering how faithful could Einstein have been, at the end, to his Machian project (Cf. Maudlin, 1993). In fact, there are hints towards his postulation of a new *relativistic ether*, devoid of mechanical properties, and hence, different from the electromagnetic ether that he showed to be unnecessary in the light of special relativity (Cf. Cassini and Levinas, 2009).

Let us now consider some of the peculiarities of the theory and how these have a bearing on our theoretical understanding of the waves and their detection. General Relativity is a nonlinear theory, and being nonlinear means that the variables are not independent and that the curvature of spacetime cannot be expressed as the sum of two independent components (Cf. Blair, 1991 p.6). In General Relativity, the stress energy tensor that describes the distribution of mass, energy and momentum density in the system under study is related to the curvature tensor –or, as it is usually called, to the *Einstein tensor*- that states the curvature of spacetime and which depends on the metric: *g*. As Wheeler (1962) nicely put it: "Matter tells the space how to curve, space tells matter how to move". But as Blair noted: "This hides the complexity, because in reality: matter, the motion of matter, and radiation density, including propagating waves of curvature, tell spacetime how to curve. Curvature creates curvature, and influences its propagation." (Blair, 1991, p. 6).

Consider now its field equation:

$$\textbf{(1)}\quad \mathbf{T} = \frac{c^4\,\mathbf{G}}{8\,\pi\,G}$$

The stress tensor, depends on several constants, such as the speed of light and the gravitational constant, besides a proportionality constant; but it also depends upon the curvature tensor **G**. At the same time, the curvature of spacetime, depends on how matter is distributed, that is to say, it depends on the stress tensor **T**. Therefore, you cannot express **T** without invoking **G** and vice versa. Any change in **T** or **G** will imply a change in the other tensor. The same is true for gravity waves, for they are a term that appears in the specification of the curvature tensor, but, at the same time, the field equation (1) helps us to determinate whether or not under certain circumstances there would be gravity waves. Moreover, as gravity waves can interact (like any wave-like phenomena), they may amplify or cancel each other. They can also interact with the curvature of spacetime and be deflected. All these peculiarities, as David Blair notices, make it almost impossible to test the models of the sources of gravity waves and, furthermore, to calculate precisely, given a particular source, which is the expected signal to be detected if gravity waves do exist. In his terms:

> The nonlinear complexity creates enormous mathematical difficulties. We cannot rigorously separate spacetime into the sum of a static curvature plus a time varying propagating curvature due to gravitational waves. Thus, although Einstein derived gravitational wave equations very early […] there followed a period of about 40 years during which the existence of gravitational waves was disputed. Eddington said that gravitational waves travel at the speed of thought! (1991, p. 6).

In fact, Einstein, in the aforementioned papers, introduced an idealization that could help us to overcome the nonlinear complexity: he recovered Minkowski geometry for spacetime, hence, assumed that space was plane, assumed that gravity waves were plane, and linearized the field equation, so as to express it as the sum of two terms. This could be done because, given the enormous distances between the several potential astronomical sources of gravity waves, whatever the initial amplitude the wave could have, on earth they could be considered as being plane. Again, according to Blair:

> In the case of the analysis for the detection of gravitational waves it is a different matter. For even the strongest imaginable gravitational waves from astrophysical sources we need only to consider plane, linear waves on an essentially flat space background. This is because the amplitude of waves crossing the solar system, even from relatively nearby strong sources, must always be exceedingly low, despite the fact that the energy flux may be enormous. (…) As a result, in spite of nonlinearity, we make a separation, which is not fully precise. We assume superposition, which is essential to

> our concept of gravitational waves, which is not strictly rigorous, which is bothersome to some theorists, but which need bother experimentalists not at all. (1991, p. 8).

In other terms, the curvature of spacetime is going to be considered as the sum of a background term (that will express all the relevant factors that contribute to the curvature of spacetime, except for gravity waves themselves) and a wave term, and it will be assumed that the variation of the background curvature is slow and the variation of the wave is very fast.

Furthermore, what properties does a gravity wave have? This will be very important, because it will determine how the experiment is going to be pursued and what could be measured that could provide –*ceteris paribus*- unambiguous evidence of the existence of gravity waves. As I have already said, according to General Relativity, massive bodies alter the curvature of spacetime. When a massive body changes, for instance, when a neutron star collapses, and its mass is transformed into energy, the curvature of spacetime changes dramatically and this change could be transmitted as gravitational radiation at the speed of light. This in turn, would lead to a variation in the gravitational constant, and this variation could be measured.[52] The slight variation of the constant plus the variation that a gravity wave could have until reaching the detector, plus the several sources of noise that have an impact on the detector, make the attempt to detect gravity waves an almost heroic enterprise. However, experimentalists always have a way to get things done:

> [A] gravitational wave act to distort the ring of test particles. […]. In a 1960 Physical Review article Weber (1960) showed that a mass quadrupole harmonic oscillator will be excited by gravitational waves. The simplest of such quadrupole oscillator represents a pair of particles, joined by a spring.

[52] The universal gravitation constant states the value for the force of gravitational attraction between two bodies. In a Newtonian universe, gravity is actually a force of constant quantity, but in an Einsteinian universe, gravity ceases to be considered as a force to become instead a geometrical property of spacetime. It could seem an oxymoron to use the expression *gravitational constant* to refer to a variable; it is, however an established use in physics.

> The gravitational wave will do work on the oscillator. In practice, the lumped parameter mass-spring oscillator is more easily replaced by a distributed system such as a bar or a block or sphere of material. Weber suggested that the material could be piezoelectric. In this case the wave would be observed as the piezoelectric voltage across the oscillator. Alternative the harmonic oscillator could be a metal bar, and a capacity transducer could read out the motion, or the oscillator could be the earth itself, and seismometers could read out signals at much lower frequencies. (1991, p. 43).

2. The first experiments

It was accepted that if gravitational radiation exists it would have several sources. Taking into account the properties of the sources, it could be known what kind of radiation to expect and at what frequency we could detect it. Sources could be either discrete or continuous: nova and supernova explosions, the creation of black holes, and the collisions between black holes give rise to the first kind of discrete radiation. Binary pulsars, neutron stars and the creation of spacetime itself are continuous sources of radiation. Determining the radiation source was of the utmost importance in building an antenna adequately tuned to detecting the range of frequencies that are expected and to properly analysing and modelling the data obtained. (Davies 1980, p. 96).

Joseph Weber, who was not only a physicist but also an engineer, designed one of the first gravity wave antennas. In fact, today, this kind of detector is still known as the “Weber antenna”.[53] It is not a very sophisticated apparatus. In fact, it is a compact aluminium cylinder (an aluminium bar), 1.53 m long, 0.66 m in diameter that weighs 1.4 tons. The principle behind its functioning is that it will behave as a

[53] For Weber’s papers on the detection of gravity waves see Weber (1960, 1967, 1968a, 1968b, 1969, 1970, 1972).

harmonic oscillator that will respond to frequencies close to 1660 hertz that would allow the detection of the emission of gravitational radiation from the collapse of a Supernova. (Weber 1969, p. 1320). This resonant mass antenna is linked to transducers that transform the oscillations of the aluminium bar into electrical impulses that in turn have to be amplified and recorded.

In Weber's experiment, gravity waves are to be detected by means of a Newtonian kind of apparatus: a harmonic oscillator. When a gravity wave impacts on the detector its energy is absorbed by the antenna and converted into sound waves. Since the oscillation of the bar will be transduced by means of the piezoelectric components attached to the apparatus, the outcome of the experiment will be a voltage. Despite the fact that gravity waves are supposed to carry enormous amounts of energy, the signal that they cause in an antenna is quite weak, and consequently, the antenna has to be very sensitive. But, in turn, and as a result of such sensitivity, the antenna will receive lots of extra signals that raise enormously the chances of getting a *false positive*. Hence, a threshold has to be established, and several noise reduction techniques have to be taken into account. In this experiment, the problem of noise was taken into consideration and vibrations that would stimulate the oscillator were detected independently. For instance, Weber used several independent detectors to record seismic, acoustic and electromagnetic inputs. The antenna was also kept in a vacuum chamber at a very low temperature, so as to reduce, as far as possible, thermal noise. However, this last source of noise could not be avoided completely, since the antenna could not be maintained at 0º Kelvin, which is the temperature at which molecular movement ceases. The data was also modelled by means of applying a Fourier transformation to the voltages obtained.

Only after these precautions were taken, and only after comparing the outputs of two resonant bars located at different places and taking into account as data only the coincident marks between both detectors,[54] did Weber announce his results, in a communication submitted to *Physical Review Letters* (Weber 1969). According to

[54] One of them was placed in his laboratory at Maryland University and the other was situated at the Argonne National Laboratory, in Illinois.

his records, there were seven coincident pulses per day that could not be a result of anything other than gravity waves from a frequency close to 1660 hz.

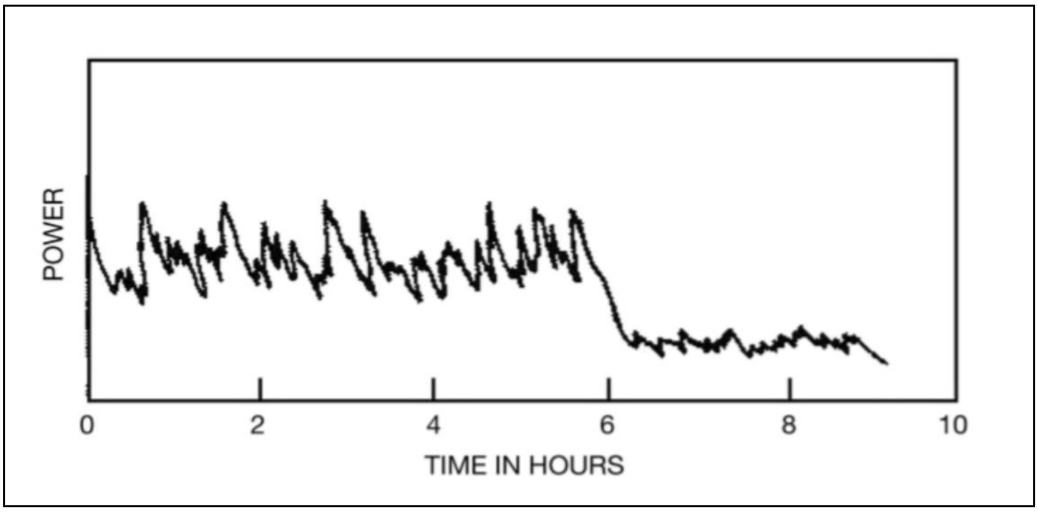

Fig. 8 A sample of the *outputs* of Weber´s experiment.

Taken from Levine (2004), p. 47.

3. Objections to Weber’s results

Weber’s results were -even though not incompatible- rather improbable in the light of the then current cosmological and astrophysical knowledge. So they were received, within the physic community with sheer scepticism. (Cf. Levine, 2004). Several research teams began to build modifications of Weber’s antenna and to share their results. They were negative. In what follows, I will concentrate on some of the reasons for being persuaded by Weber's results and those that may dissuade us from accepting them. In order to do this, I will mainly follow Allan Franklin's detailed analysis (especially his 1999 and 2002).

Weber reported the simultaneous detection of a signal by means of two distant detectors. He also reported that the signals had an anisotropic effect. Both were powerful reasons that convinced other researchers to check Weber's results, for it is improbable that artefacts manifest a regular behaviour. The evidence, however, did not last much. The coincidence between the detector placed in University of Maryland and the one situated in Argonne National Laboratory, disappeared. According to Weber, he registered about 78 coincidences between them in a determined period of time which made it very unlikely to be a result of chance. The unlikeliness of those coincidences was an invitation to regard them as genuine signals. However, as Franklin points out, the coincidence was the product of a mistake, since what Weber took to be simultaneous signals, turned out to be taking place four hours apart. (Cf. Franklin, 2002, p. 55).

Weber also claimed that the peaks presented an anisotropic effect which meant there was a sidereal correlation. In his (1970) Weber claimed that the signals displayed an anisotropic effect when read in the course of seven months of data gathering. This anisotropy would indicate that the signals were coming from a certain spatial region. If the signals detected were coming from a Supernova, we would expect them to have this anisotropic behaviour, for there would be positions of the detector in which it would be more sensitive to the radiation than when in others. According to Weber, the peaks showed a 24 hour periodicity. The anisotropy of the signals was a powerful argument against them being artefacts produced in the laboratory or by seismic perturbations. James Levine, Garwin's collaborator claimed:

> When he plotted the estimated signal strength against sidereal time for a seven-month period, he found a striking peak at a time when the antenna's most sensitive axis was directed toward the galactic centre, whereas he found no peak if he plotted the same data against local time. […]. This certainly was his strongest claim up to this time that he had detected gravitational radiation of extra-terrestrial origin. As I noted earlier, this probably was responsible for convincing many of the other experimentalists to join the search. (2004, p. 49).

However, notice that the Earth is almost transparent to gravitational radiation. Therefore, a periodicity of 12 hours is to be expected, in those moments in which the antenna is oriented normal to the source of the waves.[55] Moreover, subsequent publications ceased to mention or display it.[56]

Not only the most relevant pieces of evidence in favour of the existence of high fluxes of gravity waves disappeared, but also conflicting evidence from independent research groups appeared. Garwin and Levine from IBM; David Douglas at Bell Laboratory, in Rochester; R. Drever, J. Hough, R. Bland and G. Lessnoff at Glasgow; at Munich-Frascatti: H. Billing, P. Kafka, K. Maischberger, F. Meyer and W. Winkler all reported negative results. None of them could find signals that could be attributed to gravitational radiation. While the most powerful reasons for embracing Weber's results were being undermined, various experimental problems were detected and reported. I will refer briefly to each of them.

When a gravity wave hits an antenna, it produces a vibration that has to be amplified and processed. For that purpose, there is a system associated with the antenna that transduces the detected signals into electrical pulses. This device is known as *the electronics.* As we already know, every wave is characterized by its amplitude, its phase and its frequency. The impact of a gravity wave on a detector working as a harmonic oscillator will disturb its oscillation state, adding energy to the antenna and modifying either the amplitude, the phase or both. There are two kinds of signal processors: quadratic and linear. In this episode, Weber choose to process the signal by means of a (nonlinear) quadratic demodulator while the rest of the research groups preferred a linear one. The nonlinear demodulator, the one that Weber preferred, detects changes in the amplitude of the signal. The linear, on the contrary, can study the variation of the amplitude and the phase. Most researchers claimed that the nonlinear algorithm cannot properly separate signal from noise. According to Levine:

[55] For a detailed analysis of the sidereal correlation see Collins (2011).

[56] In his (1972), he claimed that there were some lacunae in the data that prevented him from continuing to study the anisotropy.

> A major problem with nonlinear demodulators is that they inextricably mix the signals of interest with noise and interfering signals at other frequencies. They thus degrade the signal-to-noise ratio and also distort the signals. This nonlinear mixing is a significant problem in a very lightly-damped antenna, in which thermal noise can build up the amplitude to a large value to mix with an incoming signal. This makes it impossible to measure the actual gravity-wave signal strength unless it is very large compared to the noise-induced amplitude. (2004, p. 53).

Allan Franklin considers the choice to be biased. He claims:

> Weber preferred the nonlinear algorithm. His reason for this was that it gives a more significant signal than does the linear procedure. […]. Weber remarked, "clearly these results are inconsistent with the idea that $x^2 + y^2$ [the linear algorithm] should be a better algorithm". Weber was, in fact, using the positive result to decide which was the better analysis procedure. He was tuning his analysis procedure to maximize his result. (Franklin, 2002, p. 57).

Weber justified his choice by alluding to certain features of the waves. This, in turn, led the other teams to process their data with both techniques and to compare them. The results were null in every case. How can we explain then the difference between Weber's results and the rest of the research groups? One hypothesis was the following:

> It was suggested that Weber had varied his threshold cut to maximize his signal, whereas his critics used a constant threshold. Was Weber tuning his threshold cut to create a result? This was the second reason why critics rejected Weber's result. (Franklin, 2002, p. 59).

Another complaint about Weber's experiment is the lack of calibration of the antenna. (Cf. Franklin, 1999, p. 244). According to Franklin, Weber's antenna was unable to detect calibration pulses while the rest of the teams could detect them. This is not, however, as straightforward a problem as it may seem. Sinsky, Weber's collaborator, designed a rather complex calibration system with a source of gravitational waves. It was similar to an antenna but it worked on a rotary shaft. Taking into account that the source of waves is known, it can be calculated what a device should detect. (Cf. Weber, 1960). The only shortcoming of this kind of calibration is that it cannot be performed in each data-gathering episode. The other research groups calibrated their antennas by introducing acoustic pulses in real time. Obviously this relies on the assumption that gravity waves are similar to sound waves.

The fact that the results were improbable in the light of cosmological knowledge, that those results should have been accompanied by observable events that were not observed, that the most powerful pieces of evidence were results of mistakes and that the other teams got negative results, could be considered as good internal reasons to reject Weber's results. Collins, however, disagrees. Let us now concentrate on his interpretation: why does he consider this episode as a paradigmatic case of the experimenter's regress? Why does he think that the reasons to vindicate negative results were external?

4. Discussing the episode in the light of the Experimenters' Regress

Collins was deeply involved in the detection of gravitational radiation scene when doing field work as a sociologist. His books, *The Golem: What we Should Know about Science*; *Changing Order, Replication and Induction in Scientific Practice*; *Gravity Shadow* and *Gravity's Ghost*, are testimony to a detailed approach to the research done since the sixties with the aim of detecting the elusive waves. Despite the fact of the precise case study he provides, his conclusions are deeply rooted in some of the principles of the EPOR, the *empirical programme of relativism.* The programme is committed to achieving the following goals: (1) demonstrating the interpretative flexibility of experimental data. (2) Showing the (non-rational) mechanisms by which closure is achieved. (3) Linking the closure mechanisms to social forces and political structures. (Cf. Collins 1981a, p. 7).

I do not, however, think that this episode is evidence that could favour the EPOR project. As I have explained before, Weber's results were highly contested, several errors were detected and the rest of the experiments performed gave negative results. Disagreement amongst the different research groups arose and the physics community decided in favour of the results arrived at by Richard Garwin, one of the most active researchers in the case of gravitational radiation. According to Collins, the participants were convinced by different sets of reasons either to accept Weber's results or to accept Garwin's. Although Collins insists that by 1975 all scientists has rejected Weber's results and the existence of high fluxes of gravity waves, in the interviews he mentioned different scientific reasons to justify the rejection. Indeed, Collins himself offers a detailed account of “pretty scientific” reasons to reject Weber’s findings, something which makes it quite hard to understand why he concludes from this episode the existence of an experimenters’ regress which is, furthermore, in his view, avoided by means of extra-scientific reasons. Let us try harder to follow Collins’ line of thought. The following quotation may help us to proceed:

> Obviously, the sheer weight of negative opinion was a factor, but given the tractability, as it were, of all the negative evidence, it did not *have* to add up so decisively. There was a way of assembling the evidence, noting the flaws in each grain, such that outright rejection of the high flux claim was not the necessary inference. (Collins, 1992, p. 91, his emphasis).

Apparently, Collins suggests, even if there was space to still endorse the detection of high fluxes of gravitational radiation, scientists preferred to reject it.

One major reason for that is related to the way in which Garwin wrote and presented his paper as well as the fact that he convinced another scientist to make public that there was an error in the computer programme with which Weber analysed his data.

> I have indicated how the experimenters' regress was resolved in this case. The growing amount of negative reports, all of which were indecisive in themselves, were crystalized, as it were, by Q [Richard Garwin]. Henceforward, only experiments yielding negative results were included in the envelope of serious contributions to the debate. After Q made his contribution to the transformation in socially acceptable opinion there simply were no high fluxes of gravity waves. Henceforward, all experiments that produced positive results, such as Weber's, must, by that very fact, be flawed. Owing a gravity wave detector was now much more like owing a TEA- laser. (Collins, 1982, pp. 95-96).

Let us try to reframe Collins ideas in the light of the analysis developed in chapter three. In a context of discordant results and disagreement among researchers, and where the reciprocal determination of “good apparatus” and “correct experimental result” holds, the reproduction of experiments cannot help us to make an informed and epistemic decision. The closure of the debate, thus, has to be in virtue of non-epistemic elements. Collins suggests Garwin’s eloquence and drive in criticizing Weber’s results.

I still find it hard to see, when we are considering the episode as a whole and taking into account the several errors that were discovered, as well as the absence of concomitant effects, how Collins can insist that the disagreement was resolved by appealing to extra-scientific elements. However, as we have seen in chapter three, Collins pointed out the ubiquity of the regress and its modal import by saying: *all we have to do is to agree that this is something that could have happened, not that it did happen.* (2016, p. 76). Following that line of thought I will explore in the remainder of this chapter, how we could avoid the experimenters' regress, when understood as a general reciprocity scenario, by means of the semantics of experimental results developed earlier in this thesis.

5. Applying the semantics of experimental results to overcome a possible case of General Reciprocity

In the experiment we have considered so far, *output* and final result are different and we needed a theory to help us understand what the readings indicate with respect to the detection or not of gravity waves. This reading has representational content only if we can trace it back to its generation, and relate the theoretical import of the piezoelectric crystals, and how they are to behave under the resonance of the antenna. The output will gain interpretation by subsuming it under a metrical concept, that is to say, by determining which kind of the information that the output carries is relevant for the purposes of the experiment. I will call this process the *internal interpretation* of the experiment and what we obtain from interpreting the *output* will be considered a *datum*. In this process we have to recall that if a gravity wave impacts on the detector, it will take the antenna away from equilibrium, straining it, so inducing an alteration in the relative positions of the parts of the bar. In trying to return to equilibrium, the bar will do work, and a restoring force will be produced. This force will be transduced by the piezoelectric crystals into voltage, and this voltage will be registered. The piezoelectric crystals are stimulated when waves make the resonant bar oscillate and this is transformed into voltage. The fluctuations shown in the chart, then, represent waves making the oscillator strain and this strain is proportional to the voltage. Hence, there is a relation that allows us to correlate the variation of voltage to the impact of waves on the oscillator, and the difference in voltage implies a difference in the amplitude of the wave detected. So, instead of just having the mere output of the experiment, we are faced now with a variation of amplitude over time. Within this amplitude variation, and if gravity waves do exist, we would expect to register several peaks, that could not be explained by anything else but gravity waves. In this experiment, however, things get even a *bit* more complicated, since there are several different stimuli that can induce the variation of the voltage, and a decision has to be made regarding the selection of the peaks in the chart that would count as a gravity wave going through the antenna.

In existential experiments, in contrast to measurements, internal interpretation may presuppose the theory that introduces the concept whose reference is being searched for. Let me explain why this is so. In this experiment, the output is continuous, and it consists in registering, as I have already explained, the variation of voltage over time. However, producing the data requires us to apply the Fourier transformation to convert the signal into its constituents, so as to relate this to the Riemman's tensor (Cf. Weber 1960 and also 1967). We will, moreover, have to use techniques of noise reduction, and for that it is conceptually required to establish a difference between other kinds of signals and the signal under investigation. Hence, again, we are assuming GR –and other theories as well- to perform that individuation. And, in doing so, we are appealing to the tools that GR provided us with.[57] After the interpretation process, the data selected as possible evidence for the existence of gravity waves will have to be statistically treated so as to arrive at what can, in a proper sense, be called an experimental result. The claim that the research team had detected gravity waves will stand if the results have a considerable statistical deviation and if no errors were committed during the production of the outcomes. Finally, a decision has to be made: the excess has to be related again to GR and the claim made that what was measured in fact is understandable within what is predicted by the theory, which I called *T-result*, which in this case is an excess of 7 coincident peaks per day. (Cf. Weber 1969). But, is this excess acceptable in the light of current cosmological knowledge and of General Relativity? In other words, can we provide a *T-result* for Weber's results? As we mentioned earlier in this chapter, to make the *E-result* cohere with background theories, we would have expected observational effects that were simply absent.[58] Since this result cannot be accommodated with accepted

---

[57] There are several methodological strictures in this process; of course there are criteria to be satisfied when cutting data out from statistical analysis, in this experiment, as I said a threshold was established and only the data over that threshold is to be considered.

[58] Precisely the kind of events that have been registered on August 17th related to the detection of a binary neutron star inspiral. See B. P. Abott et. Al. (2017),

knowledge, we may well claim that it is not a correct *E-result*, breaking thus the reciprocity via theoretical and background knowledge, and, therefore, making Collins' way out, unnecessary.

Chapter Six

The Vitamin C Episode Discussed

In this chapter I will analyse the studies carried out by Linus Pauling and Ewan Cameron to test the efficacy of ascorbic acid in supportive cancer treatment. The positive results their trials yielded were contested by another research group, Dr. Charles Moertel's, whose trials reached negative results. According to Harry Collins and Trevor Pinch (2005), the dispute illustrates a clear case of the experimenters' regress. As such, and according to the sociologists' explanation of how experimental disputes reach an end, the regress should have been stopped by external factors. It seems then, that given my purpose, this episode is an ideal case study to consider. I will show, as in the former episode, that the case does not represent an instance of a replication regress, but rather possibly of general reciprocity. However, I will insist on the fact that this is not pernicious in itself since there are ways to break the regress that are internal to scientific practice. I will begin by presenting in detail Cameron and Pauling's understanding of cancer and how it guided their research. After that, I will describe the experiments, both those of Cameron and of Moertel. The latter performed several RCTs to test Cameron's results. Finally, I will present Collins' reading of the controversy and my own way out that will take into account Teira's contractarian approach (2013a).

1. The Orthomolecular approach towards the treatment of cancer

During the seventies, Linus Pauling, and Ewan Cameron started a research project into medical treatments based on *orthomolecular medicine*. In their paper, *Ascorbic Acid and the Glycosaminoglycans*, they define orthomolecular medicine as "the preservation of good health and the treatment of disease by varying the concentrations in the human body of substances that are normally present in the body and are required for health." (1973, p. 181). In *The Orthomolecular Treatment of Cancer* (1974), the researchers challenge the traditional and *eradicative* approach to cancer treatment, according to which cancer is a "foreign invader" that "ought to be destroyed". That way of understanding the disease, say the authors, cannot accommodate certain biological data. In particular, the fact that although neoplastic cells were found in the peripheral blood of patients who underwent surgery, a significant amount of those patients "were alive, well, and apparently 'cured' many years later" (1974, p. 275). They claim:

> More recently it has been reported that malignant cells can be detected in the circulation of about 50% of patients undergoing resection of colonic and rectal cancers. In this study the finding did not even appear to have any prognostic significance, in that many of these patients were alive and well 5 years later; yet according to the eradicative concept of cancer treatment, all these patients should have died from tumor dissemination. (Cameron and Pauling, 1974, p. 275).

The researchers wondered how it is possible that some individuals are cured off cancer even though there is evidence of malignant cells present in their blood, the researchers wondered. The eradicative concept of illness, they say, cannot provide an adequate explanation of this phenomenon. They claim:

> The therapeutic policy of trying to eradicate every cancer cell seems to be not only illogical in principle but also usually impossible of achievement in practice. No matter what form of eradicative treatment is employed, the almost insuperable problems of selectivity require, in theory at last, that "cure" be governed by the amount of normal tissue that can be simultaneously sacrificed and still permit survival of the patient. […]. It is this doubt that gives rise to some optimism that the total treatment of cancer can be improved. (1974, p. 274).

Pauling and Cameron considered that an alternative explanation of this data could be given and that it would promote a new research avenue. According to the scientists, the positive evolution of those patients could be accounted for by taking into account the relevance of *host defence mechanisms*, that is the ability of the host to prevent the dissemination of malignant cells (and, as we will see, ascorbic has an important role in this mechanism) when dealing with the disease. A difference in the host defence mechanisms of the different patients could explain their differences in dealing with the disease and it could also explain the rare phenomenon of *spontaneous regression*, defined as "the partial or complete disappearance of a malignant tumor in the absence of all treatment, or in the presence of therapy which is considered inadequate to exert a significant influence on neoplastic disease." (Cf. Everson, Warren and Cole, 1958, p. 366). They assert:

> With respect to established cancer, the efficiency of the host resistance factor decides the final outcome. It is responsible for the great individual variations in the course of the illness, and is the determining factor in the ability or failure to respond to conventional methods of treatment. It is the only possible explanation for those patients who survive 10, 20 or even 30 years after resection of a primary tumor, and then demonstrate malignant reactivity in some hitherto unsuspected dormant metastasis. Low host resistance offers an explanation for the failures of treatment in apparently favorable situations. (Cameron and Pauling, 1974, 275).

Cameron and Pauling's research relies on a different understanding of the disease. According to them, "cancer cells are not foreign invaders, but are native cells of the host that happens to be behaving in an abnormal fashion." (1974, p. 274). This new approach leads to a series of possible treatments that could improve the life of people suffering from such a disease. Let us now explore in detail how Cameron and Pauling understood cancer, how they thought the living organism could react to the proliferation of malignant cells and the ideas behind ascorbic acid being a potential resource for supportive care.

2. The theoretical background and some motivations for the trial

As they claim in their paper, *The Orthomolecular Treatment of Cancer. The Role of Ascorbic Cancer in Host Resistance*:

> Natural resistance to neoplastic disease governs the incidence of clinical cancer and influences the individual's response to standard methods of treatment. Supportive measures, specifically designed to enhance the natural resistance of the host, should diminish cancer incidence, morbidity and mortality. […]. *It is contended that the availability of ascorbic acid plays an important role in many of these physiological processes*. (1974, p. 273, emphasis added).

The authors provide references to several studies that would support the hypothesis of the relevance of host defence mechanisms and how these mechanisms determine the final outcome of the illness. They write:

> Some of these observations even suggest that cancer may be much commoner than is generally realized and challenge our usual assumption that, unless treated, it is always a relentlessly progressive fatal illness. The very much higher incidence of cancer, particularly of the prostate, thyroid, and pancreas, found at routine autopsy than of the actual clinical incidence (about 40 to 1) would seem to indicate that such tumors have been effectively controlled in life by the majority of individuals. (1974, p. 275).

In accordance with this evidence, they propose, as I mentioned in the previous section, the relevance of host resistance difference to accounting for differences in prognosis and, taking into account the orthomolecular approach, they suggest that one of the ways in which the host-resistance factor can be improved is by administering patients with high doses of ascorbic acid. What the clinical trial they performed tried to test is whether this treatment improves the life quality of cancer patients by improving the host-resistance factor:[59]

> The objective of orthomolecular treatment is to enhance natural resistance to cancer to maximum efficiency in every patient. If we are to enhance natural resistance effectively, we must use, by definition, physiological means. Our general aim must be to create a total biochemical environment as unfavorable as possible for the progressive growth of neoplastic cells and as favorable as possible for the health of the normal tissues. *It will be noted that this therapeutic approach is quite different from the standard objective. It is not designed to kill cancer cells, instead to "restrain" them,* and ideally to bring about their reversion to a normal behavior pattern, subject again to all the usual physiological mechanisms controlling proliferation and invasiveness. (1974, p. 275-276, emphasis added).

The researchers argue that ascorbic acid may be a supplement worth testing for different reasons. Let us examine them in more detail now.

(i) Because it may contribute to improving the host resistance mechanisms. The researchers highlighted the following evidence:

Non-specific effects experienced by test subjects when taking ascorbic acid: Patients participating in trials involving the consumption of vitamin c reported that the supplementation *promotes a sense of well-being and good health*. (1974, p.

[59] However, as we shall soon see, they also think there may be good reasons to consider that vitamin C can also take part in mechanisms that directly interfere with the growth of malignant cells.

276). According to the authors, even this non-specific effect would be rather important for terminal patients from a palliative perspective.

Immunological/ physiological effects of ascorbic acid that may have a positive resistance-effect in cancer patients: Some of the effects that can contribute towards the supportive care of a cancer patient are the following: 1) the possible role of ascorbic acid saturation in lymphocytes suggested by the poor immunocompetence shown by pre-scorbutic guinea pigs. 2) As a part of the physiological response to *stressful* situations, there is a rapid depletion of the ascorbic acid found in the adrenal and pituitary glands. According to the researchers "disseminating cancer may be regarded as a *stressful* situation, and it might be that any beneficial effect of ascorbic acid in cancer could be attributed to such a non-specific protective mechanism. It is well established however that subtle variations in hormonal status and in particular variations involving the adreno-pituitary axis can profoundly influence the resistance of the host to malignant invasive growth, and it is known that ascorbic acid is involved in steroid metabolism" (1974, p. 277).

(ii) Because it may interfere directly with the metabolism of neoplastic cells.

Specific physiological functions of ascorbic acid that may have an impact in tumoural growth: 1) Ascorbic acid is required for maintaining the integrity of the ground substance, a component of the connective tissue. It is also required for collagen fibrillogenesis. Therefore, the researchers claim, ascorbic acid is required for sustaining an adequate *stromal reaction,* the growth of connective tissue as a result of a neoplasia, which purports to encapsulate the malignant process. 2) There might be a connection between a mechanism for malignant invasiveness and a hypothetical counteracting mechanism that depends on the availability of ascorbic acid. According to Cameron and Pauling:

> It has been contended that the continuous release of hyaluronidase from the neoplastic cells is an important factor in the mechanism of malignant invasiveness and tumor cell nutrition and may even be responsible for sustaining the momentum of neoplastic cellular proliferation. The action

> of the hyaluronidase is to hydrolyze the ground-substance glycosaminoglicans, and this enzyme-substrate reaction is controlled by a modified breakdown product of the substrate known as serum physiological inhibitor (PHI). There are grounds for the belief that PHI owes its powerful and highly specific inhibitory activity to the substitution of ascorbic acid for the less reactive glucuronic acid unit in a glycosaminoglycan residue, or in other words that the prime biological function of ascorbic acid is dependent upon its incorporation into an oligosaccharide hyaluronidase inhibitor complex, which exerts a restraining influence on all forms of cellular proliferation. (1974, pp. 277-278).

The authors take as the starting point for their understanding of the physiology of cancer the need to address not only the behaviour of the cell itself but also of its environment and, more fundamentally, the relationship between both. Each cell of a tissue is embedded in and surrounded by the *ground substance*, "a complex gel, containing water, electrolytes, metabolites, dissolved gases, trace elements, vitamins, enzymes, carbohydrates, fats and proteins. The solution is rendered highly viscous by an abundance of certain long-chain acid mucopolysaccharide polymers, the glycosaminoglycans and the related proteoglycans, reinforced at the microscopic level by a three-dimensional network of collagen fibrils." (1973, p. 182). According to the researchers:

> There is evidence that the interface between a cell membrane and the immediate extracellular environment is the crucial factor in the whole proliferative process. Variations in the composition of the extracellular environment exert a profound influence on cell behavior, and in turn the cells possess a powerful means of modifying their immediate environment. This interdependence is involved in all forms of cell proliferation and is particularly important in cancer. *A proliferating cell and its immediate environment constitute a balanced system in which*

> *each component influences the other.* (Cameron and Pauling, 1973, p. 182, emphasis added).

And they highlight:

> Recognition of this relationship and an understanding of the means of controlling it could lead to rational methods of treating cancer and other cell-proliferative diseases. Until recently, cancer has tended to concentrate almost exclusively upon the cell, and to ignore the other half of the proliferation equation." (Cameron and Pauling, 1973, p. 182).

Amongst the different glycosaminoglycans present in the ground substance, there is one, hyaluronic acid, which the researchers consider of special interest. They claim: "An important property of the intercellular substance is its very high viscosity and cohesiveness. This property is dependent upon the chemical integrity of the large molecules. The viscosity can be reduced and the structural integrity destroyed by the depolymerizing action of certain related enzymes […], known by the generic name of hyaluronidase." (1973, p. 183). The idea is that hyaluronic acid, also known as hyaluronan, plays a crucial role in maintaining the viscosity of the medium and, therefore, in keeping cell proliferation controlled.[60] Given this, let me quote the hypothetical mechanism they are postulating. They claim:

> All cells in the body are embedded in a highly viscous environment of ground substance that physically restrains their inherent tendency of proliferate; ***proliferation is initiated by release of hyaluronidase from the cells, which catalyzes the hydrolysis of the glycosaminoglycans in***

---

[60] Cameron introduced the hypothesis of the relevance of hyaluronic acid in his *Hyaluronidase and Cancer*. Nowadays, its importance has been recognized. See, for example, Stern, R. (Ed.) (2009).

> ***the immediate environment and allows the cells freedom to divide and to migrate within the limits of the alteration; proliferation continues as long as hyaluronidase is being released, and stops when the production of hyaluronidase stops or when the hyaluronidase is inhibited***, and the environment is allowed to revert to its normal restraining state. (1973, p. 183, emphasis added).

The degradation products can be measured (*serum polysaccharide* concentration). In pathological processes such as inflammation, tissue repair and cancer, histochemical techniques can be used to show the depolymerization of the ground substance and the increase in the concentration of the serum polysaccharide and the serum PHI (physiological hyaluronidase inhibitor). (Cf. 1973, p. 184).

According to the authors, cancerous cells may be nothing more than cells that have the ability of producing hyaluronidase continuously. Given this, the researchers propose two therapeutic possibilities. The first involves strengthening the ground substance, and the second, inhibiting or neutralizing the cellular production of hyaluronidase. What interests us now is the way in which ascorbic acid might inhibit the synthesis of hyaluronidase. They claim:

> If ascorbic acid were required for the synthesis of PHI, a deficiency of ascorbic acid would cause the serum PHI concentration to decrease toward zero. In the absence of such control of hyaluronidase by PHI, background cellular proliferation and release of hyaluronidase would produce a steady and progressive enzymatic depolymerization of the ground substance, with disruption and disintegration of the collagen fibrils, intraepithelial cements, basement membranes, perivascular sheaths, and all the other organized cohesive structures of the tissues, producing in time the generalized pathological state of scurvy. *These generalized changes, tissue disruption, ulceration, and hemorrhage, are*

> *identical to the local changes that occur in the vicinity of invading neoplastic cells*. (1973, p. 187, emphasis added).

Given the hypothetical mechanism developed, the therapeutic proposal consists in providing the patients with ascorbic acid in order to synthesize PHI so that this can control excessive cellular proliferation.

(iii) Because it actually plays a role against viral and bacterial activity, which, in turn, has a concomitant beneficial effect in cancer patients.

Links between viral activity, ascorbic acid and cancer: 1) Several tumours are known to develop at the site of herpetic infections which suggests that the virus itself may be involved in the evolution of the carcinoma. (Cf. 1974, p. 279). Given the available evidence of ascorbic acid as an antiviral agent, and the link between certain viruses and cancer, it may well be the case that ascorbic acid plays a preventive role. 2) Bacterial infections and cancer: According to the researchers, "clinicians tend to forget that almost all tumours are ulcerated, and therefore subject to secondary bacterial invasion, and that the debilitating effects of such superimposed bacterial infection must play an important part in the total cancer illness". (1974, p. 289). Ascorbic acid plays a role in the phagocytic activity of leucocytes and, therefore, improves host resistance and provides protection against secondary diseases that are superimposed on cancer.

(iv) Because there is data available that can be understood if ascorbic acid is involved in any of the above scenarios.

Measurements of levels of ascorbic acid in different situations and experimental subjects. First, it is important to notice that human beings, unlike several other mammals, have lost the ability to synthesize ascorbic acid.[61] 1) There is an increased requirement for ascorbic acid in cancer. Different studies have shown

---

[61] From an evolutionary perspective, this is a rather interesting phenomenon, but one that I cannot explore here.

that patients who suffer from cancer tend to be depleted of that substance. 2) Animals that, unlike humans, have an endogenous source of ascorbic acid, increase its production when being given methylcholanthrene, a powerful carcinogen. Also, "tumor-bearing rats increase their production to amounts extrapolated on a body weight basis equivalent to around 16 g per day for a 70 kg human." (1974, p. 278).

For the above-mentioned reasons, the researchers emphasize the importance of carrying out research at three levels: (i) To test ascorbic acid as a prophylactic method; (ii) to test it as part of a supportive treatment; (iii) to test it as a palliative treatment. The scope of the research and its objectives will be rather important in understanding the scientific disagreement that arouse when reproductions of their trial were available. Notice that the authors claim neither that ascorbic acid may cure cancer nor that this is the hypothesis being tested in the trial. Witness how explicit they are regarding the limits of the therapy:

> Because of the complexity of the intercellular ground substance and its responsiveness to external influences, many of the innumerable "cancer treatments" that have been hopefully advocated year after year might have some element of truth behind them. It is also true, however, that no form of cancer treatment based on the antineoplastic effect of modification of ground substance *can ever be more than palliative, because to render the ground substance totally resistant to hyaluronidase would create a situation incompatible with life itself*. (Cameron and Pauling, 1973, p. 186, emphasis added).

The next section will be devoted to presenting the clinical trials Cameron executed and their results.

3. The trial(s)

So far I have presented the reasons why the authors consider it is worthwhile testing the impact of ascorbic acid on cancer patients. Now it is time to present the clinical trial(s) they implemented and the conclusions they reached. The first results reported are found in Cameron and Campbell's paper: *The orthomolecular treatment of cancer II. Clinical trial of high dose ascorbic acid supplements in advanced human cancer* (1974). The results presented in that study are more of a qualitative nature, given that there is no control available, and there are no figures offering a comparison between survival times with and without the treatment. Instead, there is a detailed report describing the evolution of each of the patients and some conclusions that suggest a significant contrast between what doctors considered expectable in the absence of treatment and what they observed in the ascorbate-treated patients. They claim:

> In the absence of untreated and exactly matched controls for comparison, we have no statistical information to claim that the administration of ascorbic acid alone produced a significant increase in survival times in the terminal patients studied. However, it is our opinion that most clinicians familiar with the practical realities of terminal cancer, perusing table 1, would be inclined to agree that many of these patients survived much longer than reasonable clinical expectation. (1974, p. 310-311).

After this initial trial, the researchers broadened the number of experimental subjects and implemented a natural history control. The protocol, according to the researchers, is invariant, and they incorporate into the statistics data belonging to previous patients. For example, the second *trial* considers the progress of 100 cases: 50 new cases and 50 that "were obtained by random selection from the alphabetical index of ascorbate treated patients in the Vale of Leven District General Hospital, where the treatment of some terminal cancer patients with ascorbate has been under clinical trial since November 1971" (Cameron and Pauling, 1976, p. 3685). On this occasion, the survival time after the determination of untreatability of 1000 matched controls was analysed and compared with that of the treated patients.[62] Finally, the third report provides a new analysis by contrasting patients' survival times against a selection of 1000 new control matches. (Cf. 1978b, p. 4538). Here I offer a table in which the different results can be compared.

	Cameron and Campbell (1974)	**Cameron and Pauling (1976)**	**Cameron and Pauling (1978)**
Patients	50	100	100
Historical Control	None	1000	1000
Average survival time after date of untreatability	Not available	Ascorbate: 210 days Controls: 50	Ascorbate: 293 days Controls: 38
Ratio of average survival time after date of untreatability	Not available.	4.16	7.7

Table 5. A comparison among trials

62 The date of untreatability is determined by the following (and conventional) standards: inoperability at laparotomy, abandonment of any definitive form of anti-cancer treatment, the date of admission for terminal nursing care. Survival time is calculated from that date until the date of death. (Cf. Cameron and Pauling, 1978, p. 4538).

The research was conducted by Ewan Cameron and Allan Campbell. It took place at the Vale of Leven hospital in Scotland. It started in November 1971. The patients selected for the test were suffering from advanced cancer. Most of the patients considered for the statistics had ascorbate as their sole treatment.

The design of the second and third trial is peculiar in some respects, since they deviate strongly from a typical randomized clinical trial. However, the deviation from the gold standard seems to be justified on epistemic and ethical grounds. In their 1974 paper, Cameron and Campbell claim:

> We have made no attempt to conduct a double blind clinical trial for two reasons. Because of all the variables involved in the progress of human cancer, it would be quite impossible for us to obtain anything like "exactly matched pairs" for comparison within our own clinical practice. […] Moreover, as our clinical experience increased, we felt it to be ethically wrong to withhold ascorbic acid in otherwise hopeless situations, merely for the sake of obtaining observations of dubious significance for statistical comparison. (1974, p. 287).

To begin with, in order to have a control group (when available) an historical comparison is carried out. This means that there is no separate group of patients who are not given any treatment. Despite this fact, the researchers used medical chart data from previous patients whose disease matched the disease of each of the patients in the active arm of the trial. The controls were selected by senior members of the Medical Record Staff at the Vale of Leven Hospital, to avoid selection bias. An *ad-hoc* hired researcher examined each medical record and determined, for each patient, the date of untreatability.

> The control group was obtained by a random search of the case record index of similar patients treated by the same clinicians in the Vale of Leven Hospital over the last 10 years. For each treated patient, 10 controls were

> found for the same sex, within 5 years of the same age, and who had suffered from cancer of the same primary organ and histological tumor type. These 1000 cancer patients comprise the control group. The detailed case records of these 1000 were then analysed quite independently by Dr. Frances Meuli […], who established their presentation date of "untreatability" by such conventional standards as the establishment of the inoperability at laparotomy, the abandonment of any definite form of anti-cancer treatment, or the final date of admission for "terminal care". This presentation date of untreatability corresponds to the date when ascorbate supplementation was initiated in the treated group. Comparable survival time of the 10 matched controls could then be calculated. We accept that "the presentation of the date of untreatability" can be influenced by many factors in individual patients, but we contend that the use of 1000 controls managed by the same clinicians in the same hospital over the last 10 years provides a sound basis for this comparative study. (Cameron and Pauling, 1976, p. 3685).

The trial *is not a blind study*. The patients know they are given an experimental treatment. The physicians and nurses are also aware that the patients are being treated with ascorbate. Finally, *it is not a randomized experiment.* The patients who constitute the active group are not recruited by any random process; however, the researchers claim that the subpopulation chosen is representative:

> Even though no formal process of randomization was carried out in the selection of our two groups, we believe that they come close to representing random subpopulations of the population of terminal cancer patients in Vale of Leven Hospital. (Cameron and Pauling, 1976, p. 3685-7).

These features make the trial a *controlled observational study* where the control is based upon natural history. Given that randomization is said to prevent bias in clinical trials,[63] the scientists are quite aware that their findings could be biased due to placebo and/or anticipation effect. However, they argue:

> We are convinced that the general awareness of failure of standard treatment regimes produces a strong “reverse placebo effect” in many of these patients entering the trial. We believe that all these factors have tended to minimize placebo and anticipation effects, but we accept they cannot be entirely excluded. (1974, p. 287).[64]

Those are the peculiarities of the design. The ascorbate was administrated daily. The intake was 10 grams by intravenous infusion during the first 10 days and orally thereafter. The patients were checked periodically.

Let us now focus on the results. The first and qualitative study reports that amongst the 50 patients treated, 17 (34%), manifested no response. 10 patients (20%) showed a minimal response. 11 patients (22%) displayed a growth retardation of the tumour. 3 patients (6%) experienced cytostasis. Tumour regression occurred in 5 patients (10%) and, finally, 4 patients (8%) experienced tumour hemorrhage and necrosis. In an overview of the trials, Cameron and Pauling give the following report about this group:

> The benefits enjoyed by the majority of patients were related to relief from pain, greater wellbeing, a decrease in malignant ascites and malignant pleural effusions, relief from hematuria, some reversal of malignant

---

63 Randomization is considered to be fundamental to the prevention of unknown confounders and it is claimed to be the only justification for using a significance test, in which the hypothesis under examination is compared to the null hypothesis. However, several scholars have questioned the virtues of randomization. Peter Urbach, for example, claims: “It will be part of my case that randomization is not an essential component of good experiments and that, for the most part, it is a waste of effort and resources (1985, p. 258). See also Worrall (2007).

64 Yet, this seems rather problematic. The controls against whom the patients are compared should have experienced the reverse placebo effect without the positive placebo effect that could have been experiencing the treated patients.

> hepatomegaly and malignant jaundice, and decrease in erythrocyte sedimentation rate and in the seromucoid level, all accepted indicators of lessening malignant activity. (Cameron and Pauling, 1979, p. 119).

The lack of a control group makes it impossible to determine in a quantitative way if there is any difference in survival rates. But how probable are these results if ascorbate does not play any role in improving the host resistance factors of the patients, the researchers wondered. Considering the probability of an alternative cause for the results obtained to be very low, they declare:

> [T]he administration of ascorbic acid was able to induce tumor regression in a few patients and provoke tumor hemorrhage and necrosis in few others. (Cameron and Campbell, 1974, p. 314).

In a subsequent publication, Cameron and Pauling analysed the survival rates of 100 patients (this included 50 new patients, this time treated following the traditional protocols of the hospital, in accordance with which they were given vitamin C when they reached the date of untreatability by conventional methods) against 1000 ascorbate-untreated and matching historical controls treated in the same hospital and by the same physicians. They concluded that:

> […] The administration of ascorbic acid in amounts of about 10g/day to patients with advanced cancer leads to about a 4-fold increase in their life expectancy, in addition to an apparent improvement in the quality of life. This great increase in survival time results in part from the much larger numbers in ascorbate patients than of the controls who lived for long times. […]. Sixteen percent of the patients treated with ascorbic acid survived for more than a year, 50 times the value for the controls (0.3%). (Cameron and Pauling, 1976, p. 3688).

Here I reproduce a table with the results obtained:

3688 Medical Sciences: Cameron and Pauling *Proc. Natl. Acad. Sci. USA 73 (1976)*

Table 2. Ratios of average survival times for ascorbate patients and matched controls, with statistical significance

A	B (Days)	C (Days)	D	E (Days)	F (%)	G (%)	H	I
Bronchus (15)	136	38.5	3.53	47	47	8.7	24.5	<<0.0001
Colon (13)	282	37.0	7.61	59	54	20	7.63	<0.003
Stomach (13)	98.9	37.9	2.61	43	46	17	6.41	<0.006
Breast (11)	367	64.0	5.75	91	55	22	5.74	<0.026
Kidney (9)	333	64.0	5.21	88	67	22	8.35	<0.002
Bladder (7)	196	43.6	4.49	57	57	20	4.90	<0.028
Rectum (7)	226	55.5	4.10	71	86	33	7.57	<0.003
Ovary (6)	148	71.0	2.08	73	83	30	6.83	<0.005
Others (19)	172	56.8	3.03	67	53	27	5.28	<0.027
All (100)	209.6	50.4	4.16	65	60	25.7	55.02	<<0.0001

A, Type of cancer and, in parentheses, number of ascorbate patients. There are 10 matched controls for each ascorbic acid patient. B, Average days of survival for ascorbate patients. C, Average days of survival for controls. D, The ratio B/C. E, Average days of survival for all subjects in group. F, Percentage of ascorbate patients surviving longer than E. G, Percentage of controls surviving longer than E. H, Value of χ^2 for F and G (two-by-two calculation). I, Corresponding value of P (one-tailed).

Figure 9. A reproduction of the results reported in Cameron and Pauling (1976).

The results seem impressive. In all cases, the survival is longer for the ascorbate-treated patients than for the controls. However, and as a response to colleagues' criticisms of the possibility of selection bias, Cameron and Pauling decided to compare the treated population against a new control group. They report:

> Several experienced investigators in this field have expressed to us their doubt as to whether the ascorbate-treated patients and their controls comprised representative subpopulations of the same population and whether comparable times of untreatability had been assigned to the two groups.
>
> A new set of control patients was selected, and tests were carried out, as described in the following paragraphs, to answer the questions that had been

> raised. Our conclusion is that the results previously announced are valid, and, in fact, the increase in life expectancy of ascorbate-treated patients with terminal cancer is found to be somewhat larger than previously reported. (Cameron and Pauling, 1978b, p. 4538).

For this analysis, the researchers considered 100 ascorbate treated patients and 1000 controls (some of whom were already considered for the previous study). They claim that the cases "are representative subpopulations of the same population of untreatable patients". (1978b, p. 4538). The results seemed even more impressive: patients in the ascorbate group lived, on average, 300 days longer than the controls. 22% of the ascorbate patients lived for more than a year after the date of untreatability while only 0.4% of the controls did.

Let us consider now Moertel's reproduction of Cameron's trial.

4. The reproduction of the experiment

After the publication of Cameron and Pauling's findings, Charles Moertel and his team, based at the Mayo Clinic, designed and implemented a double blind randomized control trial in order to test the results reported. 123 patients were stratified and classified according to their primary tumour and then randomized into one of two groups: placebo or active principle. Patients were checked every two weeks and the amount and frequency of the drug taken, their symptoms and their body weight had to be reported. (Cf. Creagan et. al. 1979). The treatment was continued until death or until the patient was no longer able to take medications by mouth. (1979, p. 687). Their first trial showed no difference in survival between the placebo and the active group, both groups had a similar survival rates to those of the controls in Cameron and Pauling's trials. The Mayo researchers claim:

> Fifty-eight per cent of the patients given the placebo and sixty-eight per cent of those given vitamin C claimed some improvement in symptoms during treatment. There were no statistically significant differences in symptoms between the two treatment groups. (Creagan, E.; Moertel, C. et al; 1979, p. 688).

How can we explain the discordant results? Are the trials comparable? One relevant factor is that Scotland and the United States have different standards of treatment and protocols for cancer patients. In particular, there is an important difference that could account for the contradictory results and Cameron and Pauling complained about this difference. The experimental subjects that took part in Cameron and Campbell's first trial had ascorbic acid as their sole treatment. They were not immunocompromised patients, unlike those of the Mayo Clinic trial. In the discussion, the Mayo researchers acknowledged this fact as a possible explanation for the divergence. They assert:

> It should be noted, however, that only 9 of our 123 patients had not previously received chemotherapy or radiation therapy. It is therefore impossible to draw any conclusions about the possible effectiveness of vitamin C in previously untreated patients. In Cameron and Campbell's report of a 10 percent regression rate in 50 patients with widely disseminating cancer, none had received prior treatment and presumably were more immunocompetent than our patients. Since vitamin C may have an impact on host resistance to cancer, we recognize that earlier immunosuppressive treatment might have obscured any benefit provided by this agent. Nevertheless, the nonrandomized study that showed a fourfold enhancement of survival with vitamin C included patients who had received conventional cancer treatment. This improvement could not be substantiated by our study. (1979, p. 689).

In order to amend this relevant difference between the populations under study, Moertel's group conducted another test. On this occasion, the team recruited 100 experimental subjects who suffered from colon cancer and who were in good general conditions and who did not receive any cytotoxic therapy. Colon cancer was also the most frequent tumour in Cameron's original study and patients suffering from that sort of tumour were those who seemed to improve the most due to the vitamin C (or, at least, whose survival time were extended the most after reaching the date of untreatability).

The experimental subjects were ambulatory and they underwent an oral treatment, either with a placebo or with vitamin C. The patients were aware of the fact that they would be randomly assigned to either the active or the placebo group. Again, the researchers argued, there were no significant statistical differences between the active and placebo arms. 49% of the vitamin C group patients survived after a year while 47% of the members of the placebo group did so. However, this time, the survival rates were more similar to those of the vitamin C group in Cameron and Pauling's trials than to those of the controls. The researchers attributed the differences in results in the methodology employed and they concluded:

> It would appear that the most substantive difference between our study and that of Cameron and Pauling was that theirs was a retrospective comparison between selected study patients and historical control patients, whereas ours was prospective, randomized and double blinded. Randomization and double blinding served to protect against any possible conscious or unconscious bias on the part of the investigators as patients were selected for treatment assignment and as their therapeutic results were evaluated. There was no such protection against bias for Cameron and Pauling as they selected and then reselected the patients they decided to evaluate for their first and second reports. (Moertel et. Al., 1985, p. 140).

But, was Moertel's RCT flawless? Cameron and Pauling insisted that there were still differences between the trials that could account for the differences in the results:

(i) *Differences regarding the administration procedure*. Notice that in the original trial, the administration was intravenous during the first ten days of treatment. Even if Cameron and Campbell asserted that "with increased experience, we now tend to believe that the intravenous regime is probably unnecessary as a routine measure" (1974, p. 297), this could still be a relevant difference.

(ii) *Differences regarding the duration of the treatment*. Moertel's second trial was designed so that the "therapy was continued as long as the patient was able to take oral medication or until there was evidence of marked progression of the malignant disease" and afterwards, the medication was suppressed. (1985, p. 138). In Cameron and Pauling's trial, the administration continued indefinitely. In fact, Cameron and Campbell reported a patient with reticulum cell sarcoma that experienced regression (1974, p. 308). But after four weeks of suppressing the treatment he displayed signs of malignant reactivity. Notoriously, after recommencing the vitamin C treatment, he experienced a second regression. Therefore, the continuation of the treatment could be crucial for an improvement of the patients.

(iii) *Patients taking vitamin C on their own*: several patients in the placebo arm displayed high levels of ascorbate in urine. Notice the following:

> To further ensure compliance, 11 consecutive patients were selected during the course of our study for urinary assays of ascorbate that employed the α,α-dipyridil ultraviolet absorption method of Sullivan and Clarke. Patient selection was made at an arbitrary chosen time in our study and without knowledge of the drug assignment of individual patients. The laboratory was also blinded as to drug assignment. Patients were not told the purpose of the urine collection; they were simply asked to submit a 24-hour specimen. Five

> patients who were later determined to have been assigned vitamin C all had high urine levels of ascorbate (≥ 2 g per 24 hours). *Of the six patients assigned to placebo, five had negligible levels that were within the range of normal controls for our assay method (≤0.55 g per 24 hours). A single patient had an intermediate value between these two ranges.* This patient had poorly controlled diabetes and was taking several drugs for the control of pain and depression. Since our assay method was not completely specific for ascorbate it is possible that this patient had interfering substances in the urine that were associated with the same ultraviolet absorbance as ascorbic acid. (Moertel et al., 1985, p.139, emphasis added).

As I have already said, in this second trial, the survival times of the patients, both in the active and in the placebo group, were similar to those of Cameron and Pauling's vitamin C group. The Mayo team attributes this fact to the better health and immune conditions enjoyed by the patients in the second trial at the start of the trial. However, this fact can also be explained by assuming that both the placebo and the active group were taking vitamin C, which, of course, would invalidate the results altogether.

So far, there is a clear discordance amongst the results obtained. Moertel attributes the differences to selection bias in the original studies, while Cameron and Pauling signal several differences in Moertel's experiment that made him wrongly reject the null hypothesis. The clash between two ways of doing biomedical research is apparent.

Despite the procedural differences and the fact that there is evidence that the patients in the placebo arm of the Mayo Clinic test were taking vitamin C, which, in turn, could account for the divergent results, the scientific community dismissed Cameron and Pauling's results in favour of Moertel's. Let us try to understand how the differences can be accounted for and what kind of explanations can we offer for the fact that it was Moertel's results which were considered correct by the scientific community.

5. Discussing the episode in the light of the Experimenters' Regress

The experiments yielded conflicting results. None of the interested parties acknowledged the other's results as a refutation because both parties could find flaws in the other's experiment. Still, there was a socially accepted verdict: vitamin C does not work either as a cure for cancer (a hypothesis that, in fact, was not originally being tested) or as a palliative treatment for it (notice that Moertel claimed that the members of both the placebo and the active groups experienced improvement). Collins and Pinch take this episode to be a case of the experimenters' regress. They state:

> Readers familiar with the earlier books in the Golem series will realize that this is a classic instance of a contested experimental outcome, or what we call the "experimenters' regress". (Collins and Pinch, 2005, p. 97).

Several qualifications are in order. To begin with, describing the experimenters' regress merely as "a classic instance of a contested experimental outcome" is quite misleading. This is because the results can be contested for other reasons than the experimenters' regress, for example, because of fraud suspicion or even for religious beliefs.[65] Furthermore, and as we saw in chapter three, the controversy between investigators given discordant results arises as a consequence of general reciprocity; it is not the cause of the experimenters' regress but a *symptom* of it. The presentation also differs from the standard characterization in that more weight is given to the skills of the practitioners in making a treatment "work". I find this characterization quite implausible and careless, I must confess. They proceed:

[65] As fossil evidence for evolutionary theory is contested by religious groups.

> If vitamin C does indeed cure cancer, then Cameron has the requisite skills and Sloan-Kettering does not. If vitamin C does not cure cancer, then Sloan-Kettering are the skilled practitioners. How do we find out who has the requisite skills? The answer is we do an experiment to see whether vitamin C cures cancer, and so on. With no independent measure of skill available, the results are indecisive and we are caught in a regress. (Collins and Pinch, 2005, p. 97-98).[66]

First, it must be noted that, in clinical trials, the impact of tacit knowledge is kept at its minimum, and that is one of the reasons why the evidence-based medicine movement promotes RCTs over observational studies and over clinicians' expertise. Consider, for example, multicentre clinical trials: if the protocol is run by several physicians across the world, how can their skills be non-transferable? Or if they were, wouldn't its impact be neutralized by the different researchers working in these large projects?

Second, if this is an episode that exemplifies the experimenters' regress; would it be a case of a *replication regress* or of *general reciprocity*? Moertel's research cannot possibly qualify as a replication of Cameron and Pauling's studies, but is rather an *independent test*. Indeed, they deemed Cameron's design to be non-rigorous and subject to bias. Therefore, it cannot be a case of *replication regress*. It has to be a case of *general reciprocity*. As we showed in chapter three, according to *GR*, the determination of the correctness of an experimental result is reciprocal with the determination of the proper functioning of the experimental device. Adapting the former to fit the peculiarities of clinical trials, we may say that the determination of the correctness of an experimental result is reciprocal with the adequacy of the experimental protocol. Adapting Collins' words to the more precise framework we developed, we may say: (1) the only way to know whether the experimental protocol is adequate is getting the correct experimental result. (2)

[66] Sloan-Kettering is another American centre that run a trial on the efficacy of vitamin C without much success.

The only way to know if *x* is a correct experimental result is by means of an adequate experimental protocol.

Let us examine (1). Is it true for biomedical research? I do not think so. There are several independent ways to evaluate an experimental protocol. For example, via methodological considerations. On the basis of purely methodological considerations we can determine if a protocol can properly rule out biases. Moreover, (1) is not a sufficient condition for the adequacy of an experimental protocol, for we could get correct results with an inadequate protocol. What about (2)? Can't we find alternative ways to determine whether a treatment may be effective other than a clinical trial? I will claim that this is also the case, even though this claim will probably be resisted.

Let us consider now the explanation Collins offered for breaking the reciprocity in favour of Moertel's results:

> Are the Mayo studies definitive, then? As we have seen, this is a case of experimenters' regress. But the argument has in effect been closed in favor of orthodoxy. Experiments alone did not settled matters, but given the implausibility of Pauling and Cameron's claims within orthodox framework of cancer theory and practice, the experimental evidence offered a credible source of rebuttal. (2005, p. 109).

Is Collins putting forward any extra-scientific elements contributing to closure in this episode? No, he is not. In fact, it seems that epistemic reasons are playing a major part in determining who to believe in this controversy. I find it quite straightforward that, insofar the scientific community considers RCTs to be the gold standard for gathering evidence in biomedical research, there are general epistemic reasons to prefer Moertel's results over Cameron and Pauling's. For example, it has been argued since Fisher that RCTs are the only kind of experimental design that –via randomization- can rule out known and unknown confounders and allow the use of a statistical test. Also, that this kind of experiment can rule out more types of bias, in particular, *allocation bias*. All these features are, according to many statisticians and researchers, responsible for making RCTs internally valid. This experimental design could also help us to suspend our judgement with regards to the theoretical framework and focus exclusively on the results of the treatment.

But notice that although Moertel is conducting a gold standard protocol, in this particular episode, there are also reasons to consider Moertel's study as flawed. Remember that some patients in the placebo arm were excreting ascorbic acid metabolites and that the mean survival times of both groups approached that of Cameron and Pauling's treated patients, both facts compatible with the patients self-administering vitamin C. (Cf. Moertel et. al, 1985, p. 141). Therefore, I believe that the complexities of this episode are related to the fact that the scientific community rushed into accepting a flawed RCT over a possibly well-conducted observational experiment. Even if RCTs can rule out more sources of bias than observational experiments, it is not obvious that a faulty RCT could yield better results than a properly conducted observational research.[67]

In a nutshell, I believe that the most parsimonious explanation of how the discordance was settled in this episode is by appealing not to the impact of external factors, but rather to that of the scientific community being persuaded by the fact that Moertel tested the treatment by means of an *RCT*. These are internal reasons, albeit of a second order, since they are not empirical reasons but methodological ones.

[67] This episode calls our attention towards a related and quite interesting epistemological problem. Are RCTs, as many experts believe, methodologically superior to other ways of gathering evidence in biomedical research? Notice that according to some hierarchies of evidence, even a RCT with a high risk of bias offers better evidence than an observational study. (Cf. Weightman et. al. 2005).

6. Overcoming general reciprocity in biomedical research

Let us consider now, in a more abstract way, how general reciprocity may be avoided via internal means in biomedical research. I will explore David Teira's *contractarian approach* (2013) and I will also consider the relevance of theoretical knowledge for establishing evidence in this kind of experiments.

In his paper, *A Contractarian Solution to the Experimenters' Regress*, David Teira addresses the problem of the experimenters' regress in biomedical research, in particular, in drug testing. Teira proposes a contractarian way out of the regress, in which the involved parties agree on the application of a set of debiasing procedures, so as to guarantee that, even if their experiments are biased in some way, they will not be biased in such a way as to favour either of their respective hypothesis. Teira's paper can be read in two ways: (i) as justifying the relevance of randomization, even if this strategy cannot always avoid unbalanced distributions, or (ii) as an answer to the experimenters' regress. For the first interpretation, Teira would be arguing against Peter Urbach (1985) and John Worrall (2007), amongst others. From this perspective, he would be claiming that debiasing methods are subject to the experimenters' regress, and that even if it is not possible to judge in an objective way whether or not a debiasing procedure had worked, and even if the application of debiasing procedures can still yield a biased result, their use is still justified, since at least such procedures prevent the data from being biased such as to favour either researcher's hypothesis. If we take the second interpretation, I think his answer has to be supplemented with a second requisite in order to count as a proper solution.[68] For it would count as a solution to the regress only if the only sources of error that can lead to disagreement between researchers were biases, which is not the case.

In what follows, I will claim with Teira that agreement on the use of debiasing procedures is a necessary condition for overcoming in an internal and scientific

[68] An add-on that may not please frequentialists, though.

way the experimenters' regress.[69] I will also suggest another condition, which together with Teira's, would be conjointly sufficient to internally justify the acceptance of an experimental result in conflictive cases.

Teira argues that consensus among researchers would be easier to reach if debiasing methods were used. The contractarian approach rests on the fact that the more reasonable strategy for researchers to follow is to agree on accepting some methodological rules that establish what would count as a proper and well conducted experiment. The prerequisite of these rules is that they are impartial; they have to give each hypothesis the same probability of being correct. (Teira, 2013, pp. 714-716).

> Scientists need to decide in advance what counts as legitimate evidence in order to avoid the "temptation" of contesting someone else's discovery (e.g., for lack of data or significance) in order to maximize their own chance of making it themselves. (Teira, 2013, p. 717).

Notice that while the debiasing strategy of the contractarian approach deals with the methodology of experimental practice, it is silent regarding the evaluation of the outcomes of an experiment. However, Collins' contention is concerned precisely with how to determine the correctness of an experimental result. Despite this, Teira seems to believe that debiasing procedures generate consensus about experimental outcomes. Thus, for example, he says:

> The contractarian solution to the experimenters' regress is to implement debiasing procedures that make sure that the experimenter is impartial, even if the outcome sometimes is not. In contexts in which no objectivist alternative is available, *we do not need more than a precommitment to these procedures to make an experimental result epistemically acceptable.* (2013, p. 720, emphasis added).

[69] Nevertheless, I will not commit to randomization being necessary, just with the application of a set of debiasing procedures.

The strategy he proposes would suffice if biases were the only sources of error, but this is not the case. This is why, even granting the use of debiasing procedures as a necessary condition for the epistemic acceptability of an experimental result - and hence, as a necessary condition for an internal explanation of how an experimental controversy is solved- there will still be room for disagreement about the quality of an experiment. I take it that a proper epistemic answer to how the *experimenters' regress* is overcome is one that provides an independent but scientific criterion to determine what could count as a *correct result*, introducing, in a different fashion, what could be the correct outcome of a trial given other relevant information beside that offered by the trial itself. That is to say, we need one criterion that stands up as an alternative to premise two of general reciprocity. As such a principle, I suggest the following:

Theoretical Calibration Principle:

> (TCP) To determine the plausibility of the correctness of an experimental result, seek for compatibility with independently known mechanisms that can explain it.

Remember that one of the reasons why Cameron and Pauling started trying ascorbic acid was related to the fact that the traditional account of cancer as a foreign invader could not explain, for example, why certain people, despite having malignant cells in their bloodstream, recovered while others didn't. As I detailed in the previous sections of this chapter, they suggested the existence of a mechanism that could explain the differential recovery rates via the role of ascorbic acid for the immune system of the patients. The results they got were in accordance with the mechanism hypothesized. The truck driver case was also impressive. Finally, the role of ascorbic acid in enhancing the immune system was independently established by its use in treating scurvy. Considering all the available evidence, we have the following:

Cameron and Pauling's results	Moertel's results
Explainable under the orthomolecular understanding of cancer, which has more explanatory power.	Compatible with the traditional account, which cannot account for several recovery cases.
Mechanism of action of the active principle independently confirmed	-
Double tumour regression in a patient that is correlated with the intake and the suppression of vitamin C	-
No RCT as it was difficult to find matches in real time and because it was considered unethical to prevent terminal patients from taking the vitamin (1974, p. 287).	Negative results in RCT that could be explained by placebo patients taking vitamin C
Positive results in observational experiments	Negative results in RCT whose subjects were severely immunocompromised

Table 6. A comparison between Vale of Leuven and Mayo research

It would have been interesting to follow Cameron and Pauling's line of thought, probably with a less controversial type of trial. In fact, a systematic review which analysed 37 studies on the consequences of the intravenous administration of vitamin C in cancer patients highlights the following:

> There is limited high-quality clinical evidence on the safety and effectiveness of IVC. The existing evidence is preliminary and cannot be considered conclusive but is suggestive of a good safety profile and potentially important antitumor activity; however, more rigorous evidence is needed to conclusively demonstrate these effects. IVC may improve the quality of life and symptom severity of patients with cancer, and several cases of cancer remission have been reported. Well-designed, controlled studies of IVC therapy are needed. (Fritz et. al., 2014).

Taking into account that many statistical experiments are pragmatic experiments that seek to answer the question, “does this work?” (Cf. Sacket, 1983), it would be relevant to complement this kind of research with explanatory trials or with basic laboratory science in order to provide the mechanism by which the treatment works as a way to argue in favour of the results. That is to say, to address the question “Why does it work?” This complementary approach to biomedical research could also help us to strengthen a weak feature of RCTs, namely, their external validity.

Conclusion

In this thesis I have examined several ways out of experimental disagreement. I began by presenting the relevance of independent testing and Collins' sceptical view of this reproduction modality. I claimed that his idea of an epistemic asymmetry between confirmation and disconfirmation scenarios does not withstand the following observation: for every pair of contradictory and exhaustive results, given any subsequent independent test, it will support one of them. It is worth noting that the cases with which Collins exemplifies the experimenters' regress are precisely of this sort. Therefore, even if it is not the case that every episode of discordance can be reduced to those kinds of scenarios, it still may show that not every situation of experimenters' regress requires an external answer and Collins should offer other examples for which no other explanation could be available. Even if according to Collins, disconfirmation cannot be achieved by means of independent tests, he does agree that independent testing has a bearing on confirmation. Thus, in these situations, an independent test can confirm one of the possible results, hence causing an imbalance in the total amount of evidence gathered. It is also important to notice that -and Collins' study lacks this crucial explication- not just any kind of independence will do; theoretical independence, in particular, can lead us astray when problematic background assumptions are shared within the experiments.

As for the experimenters' regress, I claimed that, contrary to what Collins believes, it conflates two related epistemic problems that could arise in experimental practice. I argued that the experimenters' regress does not arise because of the problems that replication poses for the experimenter, but because of the possibility of a circular (reciprocal) determination of the correct result and the proper functioning of the experimental apparatus. I argued that the circularity is broken through means which are internal to scientific practice, since an experimental result is not merely the output of a material realization, but requires theory in order to be produced. If that is the case, general reciprocity can be overcome by means that are internal to scientific practice, without the need to appeal to any external explanation. This, of course, is compatible with a global circularity regarding empirical knowledge, but, I believe we have good reasons to accept that each particular experiment has a local answer.

I have appealed to several case studies in order to show different ways in which theoretical knowledge may help us to calibrate the functioning of our detectors. Were the categories of experiments proposed sufficient to cover the whole spectrum of experiments we can find in contemporary science? Probably not. This thesis left unattended the experimental knowledge produced in contemporary facilities such as the CERN, or as in some areas of biology such as genetics or plant sciences, as well as the complexities of data-intensive science. Given the goals of my research, however, I do not think it was necessary to pursue an analysis of such cases. That said, the possibility of data packaging and the delocalization of experimental evidence could also offer grounds for rejecting some of Collins' views, for example, on tacit knowledge and its transmission.

Besides offering a general analysis of the Experimenters' regress, this thesis has taken seriously Collins' case studies. I re-analized the gravitational radiation case and I agreed with Franklin in that there were several epistemic reasons that explain why Garwin's results were preferred. I also re-analysed the vitamin C discussion and I failed to see how an externalist solution to the experimenters' regress could be defended. In that respect, I believe the goals of my research have been accomplished.

There are several questions have been left unattended or that I could not address in this thesis. I hope I will be able to pursue them in due course. I will mention just a few of them. The first concerns the grounds of comparability between experimental results when they come from independent modalities and the possibility of applying the semantics of experimental results to offer an answer to this question. Another problem that I consider to be relevant but on which I couldn't dwell during my investigation is related to other ways in which social factors and culture may impact on the development of science. This topic has been explored by feminist philosophers of science and I do think it still deserves more attention, especially for the practical repercussions of this kind of research. Finally, epistemological research applied to the biomedical sciences and their practical consequences is paramount. I do believe these two lines of inquiry could promote socially relevant scientific research, contributing, in turn, to a more just society.

References

Abott, B. et. Al. (2017), “Observation of Gravitational Waves from a Binary Neutron Star Inspiral”, *Physical Review Letters*, 119: **16**, 161101-18.

Ackermann, R. (1985), *Data, Instruments and Theory: A Dialectical Approach to Understanding Science*, Princeton: Princeton University Press.

Atmanspacher, H., Maasen, S. (Eds.), (2016), *Reproducibility: Principles, Practices, and Prospects*, Hoboken, New Jersey: Wiley.

Balzer, W., Moulines, U., Sneed, J. (1987), *An Architectonic for Science*, Dordretch: Reidel.

Barnes, J., (1990), *The Toils of Scepticism*, Cambridge, Cambridge University Press.

Bellerive, J. et. al. (2016), “The Sudbury Neutrino Observatory”, *ArXiv*. https://arxiv.org/abs/1602.02469

Bernard, C. (1865), *Introduction a l’etude de la médecine expérimentale*, Paris: Flamarion.

Blair, D. (Ed.) (1991), *The Detection of Gravitational Waves*, Cambridge, Cambridge University Press.

Bogen, J. (2002), “Experiment and Observation” in: Machamer, P., Silberstein, M. *The Blackwell Guide to the Philosophy of Science*, Oxford: Blackwell Publishers. (pp. 128-148).

Boon, M. (2012), “Understanding Scientific Practices: The Role of Robustness Notions”, in: Soler, L. et. al. (2012), (Eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science, Dordretch, Heildelberg, London, New York: Springer. pp. 287-316.

Brown, H. (1987), *Observation and Objectivity*, New York: Oxford University Press.

Brown, J. R, (1989), *The Rational and the Social*, New York: Routledge.

Buchwald, J. (Ed.), (1995), *Scientific Practice: Theories and Stories of Doing Physics*, Chicago and London: The University of Chicago Press.

Bunge, M. (1967), *Scientific Research*, New York, Springer Verlag, 2 Vols.

Cameron, E., Rotman, D. (1972), "Ascorbic Acid, Cell Proliferation, and Cancer", *The Lancet*, March 4, p. 542.

Cameron, E., Pauling, L. (1973), "Ascorbic Acid and the Glycosaminoglicans: an Orthomolecular Approach to Cancer and Other Diseases", *Oncology*, **27**: 181-192.

Cameron, E., Pauling, L. (1974), "The Orthomolecular Treatment of Cancer: I. The Role of Ascorbic Acid in Host Resistance", *Chemico-Biological Interactions*, **9**: 273-283.

Cameron, E., Campbell, A. (1974), "The Orthomolecular Treatment of Cancer: II. Clinical Trial of High Dose Ascorbic Acid Supplements in Advanced Human Cancer", *Chemico-Biological Interactions*, **9**: 285-315.

Cameron, E., Pauling, L. (1976), "Supplemental Ascorbate in the Supportive Treatment of Cancer: Prolongation of Survival Times in Terminal Human Cancer", *Proceedings of the National Academy of Science of the United States of America*, **73**: 3685-3689.

Cameron, E., Pauling, L. (1978a), "Supplemental Ascorbate in the Supportive Treatment of Cancer: Re-evaluation of Prolongation of Survival Times in Terminal Human Cancer", *Proceedings of the National Academy of Science of the United States of America*, **75**: 4538-4542.

Cameron, E., Pauling, L. (1978b), "Experimental Studies Designed to Evaluate the Management of Patients with Incurable Cancer", *Proceedings of the National Academy of Science of the United States of America*, **75**: p. 6252.

Cameron, E., Pauling, L. (1979), "Ascorbate and Cancer", *Proceedings of the American Philosophical Society*, **123**: 117-123.

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada, *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*, December 2014.

Carnap, R. (1966), *Philosophical Foundations of Physics*, New York: Basic Books.

Carnap, R., Hahn, H., Neurath, O. (1929), "Wissenschaftliche Weltauffassung. Der Wierner Kreis", Vienna: Wolf.

Carnap, R., Neurath, O. Morris, Ch. (Eds) (1938), *International Enciclopedia of Unified Science*, Vol 1, Chicago: University of Chicago Press.

Cartwright, N. (1991), "Replicability, Reproducibility and Robustness: Comments on Harry Collins." *History of Political Economy* 23(1): 143- 155.

Cassini, A., Levinas, M, (2009), "El Éter relativista: un cambio conceptual inconcluso", *Crítica, Revista hispanoamericana de Filosofía*, 41: 123, 3-38.

Collins, H. (1981), "Stages in the Empirical Programme of Relativism", *Social Studies of Science*, 11: 3-10.

Collins, H. (1984), "When Do Scientists Prefer to Vary Their Experiments?" *Studies in History and Philosophy of Science* 15(2): 169-174.

Collins, H. (1992), *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press. [1985].

Collins, H. (2004), *Gravity's Shadow: The Search for Gravitational Waves*. Chicago and London: University of Chicago Press.

Collins, H. (2011), *Gravity's Ghost: Scientific Discovery in the Twenty First Century*. Chicago: University of Chicago Press.

Collins, H. (2016), "Reproducibility of Experiments: Experimenters' Regress, Statistical Uncertainty Principle, and the Replication Imperative", in Atmanspacher, H., Maasen, S. (Eds.), (2016), *Reproducibility: Principles, Practices, and Prospects*, pp. 65-82.

Collins, H., Pinch, T. (2005), *Dr. Golem, How to Think about Medicine*, Chicago: The University of Chicago Press.

Creagan, E. et. al. (1979), “Failure of High-Dose Vitamin C (Ascorbic Acid) Therapy to Benefit Patients with Advanced Cancer: a Controlled Trial”, *The New England Journal of Medicine,* **301**:13, 687-690.

Culp, S. (1995), “Objectivity in Experimental Enquiry: Breaking Data-Technique Circles.” *Philosophy of Science* 62: 430-450.

Davis, R. (2002), “A Half Century with Solar Neutrinos”, *Nobel lecture*. pp. 59-79.

Davies, P. (1980), *The Search for Gravity Waves*. Cambridge: Cambridge University Press.

Díez Calzada, J. (2002), “A Program for the Individuation for Scientific Concepts.” *Synthese* 130:13-48.

Dingler, H. (1928), *Das Experiment, sein Wesen und Seine Geschicte*. München: Reinhardt.

Duhem, P. (1914), La Théorie Physique: son objet, sa structure. Paris: M. Rivière, (1st edition 1906).

Edwards, C. (ed.), (1980), *Gravitational Radiation, Collapsed Objects and Exact Solutions*, Berlin, Springer-Verlag.

Einstein, A. (1916), “Die Grundlage der allgemeinen Relativitätstheorie”, *Annalen der Physik*, 49: 697-702.

Einstein, A. (1918), “Prinzipielles zur ellgemeinen Relativitätstheorie”, *Annalen der Physik*, 55: 241-244.

Everson, T., Cole, W. (1956), “Spontaneous Regression of Cancer: Preliminary Report”, *American Surgical Association*, 144: 366-380.

Feest, U. (2016), “The Experimenters’ Regress Reconsidered: Replication, Tacit Knowledge, and the Dynamics of Knowledge Generation”, *Studies in History and Philosophy of Science*, part A, 58: pp. 34-45.

Franklin, A. (1986), *The Neglect of Experiment*, Cambridge, Cambridge University Press.

Franklin, A. (1994), “How to Avoid the Experimenters' Regress”, *Studies in History and Philosophy of Science*, **25**: 3, pp. 463-491.

Franklin, A. (1998), “Avoiding the Experimenters' Regress”, pp. 151-165 in *A House Built on Sand: Exposing Postmodernist Myths about Science*. Koertge, N. (ed). Oxford: Oxford University Press.

Franklin, A. (1999), *Can That be Right? Essays on Experiments, Evidence and Science*. Dordretch, Boston and London: Kluwer Academic Publisher.

Franklin, A. (2002), *Selectivity and Discord: Two Problems of Experiments.* Pittsburgh: University of Pittsburgh Press.

Franklin, A. (2004), *Are There Really Neutrinos? An Evidential History*, Cambridge: WestView Press.

Franklin, A. (2005), *No Easy Answers: Science and the Pursuit of Knowledge.* Pittsburgh: University of Pittsburgh Press.

Franklin, A., Howson, C. (1984), “Why Do Scientists Prefer to Vary Their Experiments?” *Studies in History and Philosophy of Science* 15(2): 51-62.

Franklin, A., Howson, C. (1988), “It Probably is a Valid Experimental Result”, *Studies in History and Philosophy of Science,* 19-4: 419-427.

Friedman, L., Furberg, C., DeMets, D. (2010), *Fundamentals of Clinical Trials*, New York, Dordrecht, Heidelberg and London: Springer.

Fritz, H. et. al. (2014), “Intravenous Vitamin C and Cancer: A Systematic Review”, *Integrative Cancer Therapies*, 13-4: 280-300.

Galison, P. (1987), *How Experiments End*, Chicago and London, University of Chicago Press.

Galison, P. (1997), Image and Logic: A Material Culture of Microphysics, Chicago: University of Chicago Press.

Gifford, F., Gabbay, D., Thagard, P., Woods, J. (Eds.), (2011), *Handbook of the Philosophy of Science*, Oxford: Elsevier.

Godin, B., Gingrass, Y. (2002), “The Experimenter Regress’: from Skepticism to Argumentation.” *Studies in History and Philosophy of Science Part A* 33(1): 133-148.

Goethe, W. (1792), “The Experiment as a Mediator between Subject and Object” in *Scientific Studies,* Douglas Miller Editor. pp. 11.17.

Goetzche, P. (2014), “Deadly Medicines and Organized Crime”, London: Radcliffe Publishers.

González, W. (Ed.) (2011), *New Methodological Perspectives on Observation and Experimentation in Science*, A Coruña: Netbiblio.

González Moreno, C., Saborido, C., and Teira, D. (2015), “Disease-Mongering through Clinical Trials”, *Studies in History and Philosophy of Biological and Biomedical Sciences,* 51: 11-18.

Gooding, D., Pinch, T., Schaffer, S. (Eds.) (1989), *The Uses of Experiment*, Cambridge: Cambridge University Press.

Hacking, I. (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Cambridge: Cambridge University Press.

Hacking, I. (1992), “The Self-Vindication of Laboratory Science”, in Pickering, A. (Ed.), *Science as Practice and Culture*. (pp. 29-64). Chicago and London: The University of Chicago Press.

Hartmann, S., Bovens, L. (2003), *Bayesian Epistemology,* Oxford: Oxford University Press.

Heildelberger, M., Steinle, F. (Eds.) (1998), *Experimental Essays- Versuche zum Experiment,* Baden-Baden: Nomos Verlagsgesellschaft.

Herschel, J. (2009), *A Preliminary Account on the Study of Natural Philosophy,* Cambridge: Cambridge University Press. [1830].

Hon, G. (1998), “‘If this be Error’ Probing Experiment with Error” in Heildelberger, M., Steinle, F. (Eds.) (1998), *Experimental Essays- Versuche zum Experiment,* Baden-Baden: Nomos Verlagsgesellschaft. (pp. 227-248).

Hon, G. (2003), “The Idols of Experiment: Trascending the "etc. List"” Pp 174-197 in *The Philosophy of Scientific Experimentation* Edited by Hans Radder. Pittsburgh: University of Pittsburgh Press.

Hon, G., Schickore, J., Steinle, F. (2009), *Going Amiss in Experimental Research*, Boston: Springer.

Howson, C. and Urbach, P., (1989), *Scientific Reasoning: The Bayesian Approach*, Illinois: Open Court Publishing.

Jenkins, S. (2004), *How Science Works: Evaluating Evidence in Biology and Medicine*, Oxford: Oxford University Press.

Kosso, P. (1989), *Observability and Observation in Physical Science*, Dordretch-Boston-London: Kluwer Academic Publishers.

Kuhn, T. (1977), *The Essential Tension: Selected Studies in Scientific Tradition and Change.* Chicago: University of Chicago Press.

Latour, B., Woolgar, S. (1979), *Laboratory Life: The Social Construction of Scientific Facts*, Beverly Hills: Sage.

Laudan, L. (1982), “A Note on Collins' Blend of Relativism and Empiricism.” *Social Studies of Science* 12: 131-132.

Leibniz, G. (1969), *Philosophical Papers and Letters*, Chicago: Chicago University Press.

Lenzen, V. (1938), “Procedures of Empirical Science”, in Carnap, R.; Neurath, O., and Morris, Ch. (Eds.), *International Encyclopaedia of Unified Science,* Vol 1, Chicago, University of Chicago Press.

Leonelli, S. (2016), *Data Centric Biology: A Philosophical Study*, Chicago: University of Chicago Press.

Levine, J. (2004), “Early Gravity-Wave Detection Experiments, 1960-1975.” *Physics in Perspective* 6: 42-75.

Lexchin, J. (2012), "Those Who Have the Gold Make the Evidence: How the Pharmaceutical Industry Biases the Results of Clinical Trials of Medications", *Science and Engineering Ethics*, **18**: 247-261.

Machamer, P., Silberstein, M. (2002), *The Blackwell Guide to the Philosophy of Science*. Oxford: Blackwell Publishers.

Maudlin, T. (1993), "Buckets of Water and Waves of Space: Why Spacetime is probably a Substance", *Philosophy of Science*, **60**: 183-203.

McCulloch. P., Taylor, J., Fowler, L. (1980), "Gravitational Raditation and the Binary Pulsar", in: Edwards, C. (Ed.) (1980), *Gravitational Radiation, Collapsed Objects and Exact Solutions*, Berlin, Springer-Verlag.

Maiani, L. (2014), "Selected Topics in Majorana Neutrino Physics", arXiv: 1406.5503v2.

Menon, T., Stegenga, J. (2017), "Robustness and Independent Evidence", *Philosophy of Science*, 84-3: 414-435.

Michelson, A. (1880). "Experimental Determination of the Velocity of Light Made at the United States Naval Academy Annapolis." *Astronomical Papers* 1: 109-146.

Moertel, C. et. al. (1985), "High-Dose Vitamin C Versus Placebo in the Treatment of Patients with Advanced Cancer Who Have Had No Prior Chemotherapy", *The New England Journal of Medicine*, **312**:3, 137-141.

Morgan, M. (2003), "Experiment without Material Intervention: Model Experiments, Virtual Experiments and Virtually Experiments", Radder, H. (ed), (2003), *The Philosophy of Scientific Experimentation*, Pittsburgh: The University of Pittsburgh Press.

Moulines, U. (2011), *El desarrollo Moderno de la filosofía de la ciencia (1809-2000)*, México: UNAM.

Newton, Isaac. (1671), "Letter to the Royal Society." *Philosophical Transactions* 80: 3075-3087.

Newton, Isaac. (1704), *Optics*. Chicago-London: Encyclopaedia Britannica.

Ohno, S. et. al. (2009), “High-dose Vitamin C (Ascorbic Acid) Therapy in the Treatment of Patients with Advanced Cancer”, *Anticancer Research,* **29**: 809-815.

Pauling, L., Herman, Z. (1989), “Criteria for the Validity of Clinical Trials of Treatments of Cohorts of Cancer Patients Based on the Hardin Jones Principle”, *Proceedings of the National Academy of Sciences of the United States of America*, **86**: 6835-6837.

Perović, S. (2017), “Experimenters’ Regress Argument, Empiricism, and the Calibration of the large Hadron Collider”, *Synthese*, 194: 313-332.

Pickering, A. (Ed.), *Science as Practice and Culture*. (pp. 29-64). Chicago and London: The University of Chicago Press.

Pinch, T. (1993), “Testing, One, Two, Three, Testing!: Towards an Epistemology of Testing.” *Science, Technology and Values* 18: 25-41.

Popper, K. (1959), *The Logic of Scientific Discovery*, London and New York: Routledge.

Radder, H. (1992), “Experimental Reproducibility and the Experimenters' Regress.” *Philosophy of Science Association, Proceedings* 1: 63-73.

Radder, H. (1995), “Experimenting in the Natural Sciences: A Philosophical Approach”, in Buchwald, J. (Ed.), (1995), *Scientific Practice: Theories and Stories of Doing Physics*, Chicago and London: The University of Chicago Press. Pp. 56-86.

Radder, H. (2003), “Technology and Theory in Experimental Science.” Pp. 152-173 in *The Philosophy of Scientific Experimentation*. Edited by Hans Radder. Pittsburgh: The University of Pittsburgh Press.

Reines, F., Cowan, C. (1956a), “The Neutrino”, *Nature*, 178: 446-449.

Reines, F., Cowan, C. (1956b), “Detection of a Free Neutrino: a Confirmation”, Science, 124: 3212, pp. 103-104.

Rheinberger, H. (1997), *Towards a History of Epistemic Things: Synthetizing Proteins in the Test Tube*, Stanford: Stanford University Press.

Ribe, N., Steinle, F. (2002), “Exploratory Experimentation: Goethe, Land and Color Theory”, *Physics Today*, 55- **7**: 43-49.

Richards, E. (1988), “The Politics of Therapeutic Evaluation: The Vitamin C and Cancer Controversy”, *Social Studies of Science*, **18**: 653-701.

Sackett, D. (2011), “Explanatory and Pragmatic Clinical Trials: A Primer and Application to a recent Asthma Trial”, *Pol Arch Med Wewn*, 121, 7-8: 259-263.

Shapere, D. (1982), “The Concept of Observation in Science and Philosophy.” *Philosophy of Science* 49: 485-525.

Soler, L. et. al. (2012), (eds.), *Characterizing the Robustness of Science,* Boston Studies in the Philosophy of Science, Dordretch, Heildelberg, London, New York: Springer.

Steinle, F. (1997), “Entering New Fields: Exploratory Uses of Experimentation”, *Philosophy of Science*, (Proceedings), **64**: S65-S74.

Steinle, F. (2002), “Experiment in History and Philosophy of Science”, *Perspectives on Science*, 10: 408-432.

Stegenga, J. (2009), “Robustness, Discordance and Relevance”, *Philosophy of Science*, 76-5: 650-661.

Stegenga, J. (2012), “*Rerum Concordia Discors*: Robustness and Discordant Multimodal Evidence, in: Léna Soler (2012), (Ed.), *Characterizing the Robustness of Science,* Boston Studies in the Philosophy of Science 292, pp. 207-226.

Stern, R. (Ed.), (2009), *Hyaluronan in Cancer Biology*, Amsterdam, Elsevier.

Teira, D. (2011), “Frequentists Versus Bayesian Trials”, in: Gifford, Gabbay, Thagard and Woods, (Eds.) *Handbook of the Philosophy of Science,* Oxford: Elsevier. Pp. 254-296.

Teira, D. (2013a). “A Contractarian Solution to the Experimenters’ Regress.” *Philosophy of Science* 80(5): 709-720.

Teira, D. (2013b), “On the Impartiality of British Trials”, *Studies in History and Philosophy of Biomedical Sciences*, 44.3: 412-418.

Trizio, E. (2012), “Achieving Robustness to Confirm Controversial Hypotheses: A Case Study in Cell Biology”, in: Soler, L. et. al. (2012), (Eds.), *Characterizing the Robustness of Science*, *Boston Studies in the Philosophy of Science*, Dordretch, Heildelberg, London, New York: Springer. pp. 105-120.

Urbach, P. (1985), “Randomization and the Design of Experiments”, *Philosophy of Science*, **52**- 2: 256-273.

Weber, J. (1960), “Detection and Generation of Gravitational Waves.” *Physical Review* 117(1): 306-313.

Weber, J. (1967) “Gravitational Radiation.” *Physical Review Letters* 18(13): 498-501.

Weber, J. (1968a), “Gravitational Radiation from the Pulsars.” *Physical Review Letters* 21(6): 395-396.

Weber, J. (1968b), “Gravitational Wave Detector Events.” *Physical Review Letters* 20(23): 1307-1308.

Weber, J. (1969), “Evidence for Discovery of Gravitational Radiation.” *Physical Review Letters* 22(24): 1320-1324.

Weber, J. (1970), “Anisotropy and Polarization in the Gravitational-Radiation Experiments.” *Physical Review Letters* 25(3): 180-184.

Weber, J. (1972), “Computer Analyses of Gravitational Detector Coincidences.” *Nature* 240: 28-30.

Weber, M. (2012), “Experiment in Biology.” In *The Stanford Encyclopaedia of Philosophy*. Edited by Edward N. Zalta.

Weightman et. al. (2005), *Grading Evidence and Recommendations for Public Health Interventions: Developing and Piloting a Framework*, Health Development Agency, www.hda.nhs.uk

Westfall, R. (1962), “The Development of Newton´s Theory of Color.” *Isis* 53(3): 339-358.

Wheeler, J. A. (1962), *Geometrodynamics*, New York, Academic Press.

Wimsatt, W. (1981), "Robustness, Reliability and Overdetermination" in: Léna Soler et. al., Eds. (2012), *Characterizing the Robustness in Science*, pp. 61- 87, Boston Studies in Philosophy of Science.

Woolgar, S. (1988), *Science: The Very Idea*, London and New York: Tavistock Publications.

Worrall, J. (2007), "Why There's No Cause to Randomize", *British Journal for the Philosophy of Science*, **58**: 451-458.

Zuppone, R. (2010), "El argumento del regreso del experimentador y la replicación de experimentos", *Scientiae Studia* 8(2): 243-271.

Zuppone, R. (2011), "La vida propia del experimento: un análisis crítico de la autonomía de la experimentación", *Revista Latinoamericana de Filosofía* 37: 213-238.

Zuppone, R. (2012), "El Empirismo lógico en perspectiva: el olvido de Otto Neurath", *Alcances: Revista de Filosofía Contemporánea*, III.

Zuppone, R. (2014), "Overcoming the Experimenters' Regress in Biomedical Research", *Philosophical Writings* 43(1): 45-56.

Zuppone, R. (2017), "An Internal Answer to the Experimenters' Regress through the Analysis of the Semantics of Experimental Results and Their Representational Content", *Perspectives on Science*, 25-1: 95-123.