

Faculty Scholarship
Cornell Law Faculty Working Papers

Cornell Law Library

Year 2007

Science for the Environment: Need for
Reconsidering Statistical Methodologies

Elisa Vecchione
Cornell Law School, Visiting Scholar, ev63@cornell.edu

SCIENCE FOR THE ENVIRONMENT: NEED FOR RECONSIDERING STATISTICAL METHODOLOGIES

Elisa Vecchione*

INTRODUCTION

This paper is part of a larger effort to analyze the United States vs. European Union dispute on biotechnology products¹ before the World Trade Organization (WTO) Dispute Settlement Body (DSB).

Starting from May 2003, United States, Canada and Argentina requested consultations with the European Communities (hereinafter EC) regarding measures taken by the EC and its member States affecting the marketing of biotech products. According to United States, Canada and Argentina, the measures at issue were inconsistent with certain EC's obligations under the Sanitary and Phytosanitary (SPS) Agreement, the GATT 1994, the Agriculture Agreement and the Technical Barriers to Trade (TBT) Agreement. The formal proceeding initiated when the consultations ended on February 2004 with the establishment of a panel by the Director-General of the WTO.

* PhD candidate in Institutions, Economics and Law (IEL), Centre for the Comparative Analysis of Law and Economics, Economics of Law, Economics of Institutions (CLEI), University of Turin and Cornell Law School, elisa.vecchione@unito.it.

For the support and helpful comments, the author thanks in particular Silvana Dalmazzone, Douglas A. Kysar, Theodore Eisenberg, John James Barceló III, Gilles-Eric Séralini, Marco Ricolfi, Sergio A. Muro, Albert Azagra Malo, Matthias Kormaksson.

¹ In this paper, biotechnology products are also referred to as Genetically Modified Organisms (GMOs), except where differently stated. GMOs are food and feed products that have been altered using recombinant DNA technologies.

The dispute, also known as *EU-Biotech*, concerns the release of Genetically Modified Organisms (GMOs) in the European market. United States, Canada and Argentina accuse the European Union to impose a *de facto* moratorium on GMOs commercialization, due to “undue delay” in the approval procedure. In particular, the complainants claim that either the analysis of the parties’ risk assessment on GMOs was not reviewed, or the positive results of risk assessment were completely dismissed, for the approval of twenty four GMOs had never reached an end. Thus, in relation to the provisions of the SPS Agreement the European delay accounted for a non-necessary (“more trade restrictive than required”, Article 5.6) and non science-based (“maintained without sufficient scientific evidence”, Article 2.2) restrictive measure to trade. Besides, on the assumption that transgenic products are similar to their natural counterparts, the complainants accused the EC to provide a treatment less favorable than that accorded to like domestic products (Article III:4, GATT 1994) and, thus, to apply inconsistent levels of protections (Article 5.5, SPS Agreement).

This dispute is highly controversial in that it fuels the debate about the precautionary principle², which has been adopted as a fundamental part of the European Union environmental strategy. On the other side, the WTO jurisprudence seems to reveal that precautionary measures would not be allowed if not grounded on scientific evidence. Unfortunately, scientific evidence is exactly what is lacking in many environmental concerns, including biotechnologies. But when risk assessment holds that the evidence of possible environmental and health-related harms is not reached “beyond a reasonable doubt”, State-based environmental measures are by default considered disguising restrictive measures to trade, which of course are in principle prohibited within the WTO regime.

In the reading of the WTO Panel report on *EU-Biotech* dispute, recurrent notions such as “scientific knowledge”, “scientific uncertainty”, “statistical significance” seem to be abused both by the complainant (United States, Canada and Argentina) and the defendant (European Union), in that they are employed with no epistemological cognition. The parties’ argumentations result in high confusion, at least about what should be the core issue of the dispute: the legitimate protection of human health and biodiversity.

The relevance of the scientific discourse for legal trials in general, and for the *EC-Biotech* dispute in particular, originates from its crucial, if not adjudicating, role. This consists in helping determining liability of the parties in legal trials, whether this has to be linked to compensation issues or, as in the *EC-Biotech* dispute, it affects member States policies. Scientific and legal standards, as it is just suggested in this occasion and will be further treated in a second stance, can be mutually supportive, nonetheless they have to be kept separate for their genesis and heuristics are of different nature. The reason of raising this point is that the two standards seem in fact to overlap in the WTO proceeding.

² See note 3.

Thus, as a first step the aim of this paper is to apprehend the fundamental notions related to scientific knowledge in order to comprehend the stakes of the dispute and to explain their relevance within the legal trial.

1. THE PRECAUTIONARY PRINCIPLE: STATE OF THE DEBATE

Since its first public acknowledgement, the precautionary principle has raised heated debate about environmental protection and health safety issues³. The controversy between managing the risk of possible future harm and taking precautionary measures has often been filtered by the alleged complementary opposition between rational economic principles and ethical – sometimes even irrational – value judgments.

As I shall argue, such a simplified view of the debate is actually flawed. In fact, instead of sharpening the opposition between different disciplines, the precautionary principle has been creating room for new discussion and revision within the scientific, economic and legal communities, for different reasons.

From a legal point of view, there exists no unique statement of the precautionary principle⁴, nor there is consensus about its location within the sources of international law⁵. By consequence, the numerous interpretations attached to it have hampered its application and implementation, resulting in poor relevance for international disputes. For these reasons, the precautionary principle has been sometimes demoted to a “precautionary approach” in order to keep it alive.

As regard the economic discipline, the precautionary principle has raised controversy to the extent that it challenges the application of the traditional cost-benefit analysis to modern policy-making. The development of the option value literature has brought new responses to integrate uncertainty of

³ The German principle of *Vorsorgeprinzip*, or foresight, appeared in the early 1970s and became a fundamental principle of German environmental law. Its English translation to “precautionary principle” has since then flourished in international statements of policy and had been introduced in numerous international treaties. The first time in occasion of the First International Conference on Protection of the North Sea (1984), then in the Rio Declaration from the 1992 United Nations Conference on Environment and Development, the declaration stated: “Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation”.

⁴ See, e.g., D. Vanderzwaag (1999), cataloguing 14 interpretations, and P. Sandin (1999), finding 19 interpretations.

⁵ In particular, most of the debate concerns its relevance as international customary law. For those who are in support of that thesis, see e.g., T. Christoforou (2003); J. Cameron and J. Abouchar (1996). For those who oppose it, see e.g., P. Birnie and A. Boyle (2002).

future consequences and irreversibility of actions into decision-making process⁶.

Within the scientific community, the development of ever innovative technologies has progressively forced to reconsider uncertainty and, accordingly, to develop new cognitive and legal device to deal with development risks. Rather than disrupting confidence in and within scientific disciplines, uncertainty has gradually emerged as additional valuable information to integrate in scientific models. Indeed, the value of research can be improved by reconsidering scientific methodologies, such as statistical inference, and by envisaging new ways for better communicating results (Kriebel *et al.*, 2001; Lemons *et al.*, 1997; Ingeborg and Traavik, 2002).

The discussion about the precautionary principle in terms of its legal interpretation, economic rationale and usefulness to environmental policy-making runs the risk to simply add one more new normative interpretation, without offering any real guidance⁷. Indeed, the debate seems endless when reduced to a fundamental opposition between two different perceptions of managing risk in everyday life – i.e. being risk averse and take precautions or being risk-lover if not fatalist – or further, between risk as science and risk as perception⁸.

The dispute is far more complicated. That is why a major part of this chapter will be devoted to clarify the language of the precautionary principle by highlighting its relation to the scientific discourse. The aim is to detect those issues of guidance, if any, tumbling from the precautionary principle to help renovating the connection between expertise, science and public policy.

This operation must begin with the very epistemological foundations of the precautionary principle in order to capture its relation to scientific knowledge, whose relevance is all the more crucial for environmental issues characterized by fundamental uncertainty as GMOs are. The rectification process will take advantage of insights from philosophy of science and epidemiology, and finally methodological issues related to statistical inference will be reconsidered.

⁶ See *e.g.*, K.J. Arrow and A.C. Fischer (1974); C. Gollier and N. Treich (2003). Furthermore, to the extent that the option value theory has developed with tight relation to investment theory, see *e.g.*, A.K. Dixit and R.S. Pindyck (1994); C. Henry (1974).

⁷ In reality, this would be sadly consistent with the phenomenon of treaty congestions that seems to characterize particularly the international environmental discourse.

⁸ This feeling comes not only from general discussion with different people. As it shall be seen, the arguments brought by the complainants and the defendant in the WTO dispute about Genetically Modified Organisms make it clear that there is a fundamental mismatch between their interpretations of the issue, so that in some points of the panel report the very impression is one of a dispute between deaf parties.

2. SCIENCE AND POLICY – WHICH LOCATION FOR THE PRECAUTIONARY PRINCIPLE?

The relevance of the precautionary principle as guiding rule for environmental policies is still today contested through the allegation that this role is already performed by scientific knowledge. That is why much of the debate about the precautionary principle has to be analyzed in light of the relation between science and policy.

The points of opposition raised by those who contest the usefulness of the precautionary principle can be summarized as follows: current regulatory provisions are inherently precautionary⁹; the precautionary principle advocates making decisions without adequate scientific knowledge; the risk in implementing it is that technological innovation could be undermined as far as development risks associated with it would challenge the outright proof of safety of a specific product (Holm and Harris, 1999).

We are generally used to thinking about scientific knowledge as a field of neutrality and objectivity, which provides us with the tools to reach rational and shared agreements. In the same vein, we are used to conceiving policy as the field of compromise, contingency and, to some extent, of time-inconsistency¹⁰, as it were something fluid and variable.

Science has always been a synonym of certainty, stability, the expression of one common feature of all individuals, i.e. rationality, from which it has gained its power of persuasion. That is why the relation between science and policy is in our mind unidirectional, where the former provides rational grounds for decision-making.

There exist however views conveying that this kind of opposition between science and policy is contestable (Bishop, 2000), and that more precautionary policies would create opportunities for scientists to rethink the way they conduct studies and communicate results (Kriebel *et al.*, 2001; Lemons *et al.*, 1997; Ingeborg and Traavik, 2002).

The crucial role of science, biology in particular, in environmental policies is unquestionable. However, the high degree of uncertainty stemming from the inherent complexity in many environmental issues requires a special attention insofar as it poses itself incomparable challenges to the science's heuristic power. This fact induces questions about a possible reconsideration of scientific methodologies and, more provocatively, of possible contributions of the precautionary principle to that aim.

Before illustrating the reasons of skepticism about scientific, primarily statistical, methodologies, it should be briefly recalled how statistical testing is built.

⁹ This criticism can be fairly acknowledged, at least as far as the European Communities' experience of highly complex process of approval for biotech products is concerned.

¹⁰ The reference is made to the Barro-Gordon model (Barro, R. J. and Gordon, D. B., 1983), which focuses on the occurrence of bias in independent monetary policy.

2.1. FREQUENTIST STATISTICAL METHODOLOGY

At present, the dominant approach in risk assessment is given by frequentist statistical inference, even though a growing interest for Bayes methodologies is increasingly challenging the traditional system of statistical analysis (Malakoff, 1999; Carlin and Louis, 1996).

Statistical inference is operated through two distinct but related processes: estimation and hypothesis testing (Ott and Longnecker, 2001). The former consists on computing a single value – called statistic – from the sample to make a point estimation of a population parameter. From point estimation of a statistic value, such as the mean μ , the median M , the standard deviation σ , inference is made about a population value. This information is then used as the basis for hypothesis testing formulation, in such a manner that the observed changes in point estimation (e.g. the mean decrease in insomnia among people assuming a particular drug) allows the researcher to assume that there is some relation between the use of the drug and the decrease in the rate of people suffering from insomnia.

Statistical testing is made of two components, the research or alternative hypothesis and the null hypothesis. The aim is to prove an association between events by contradiction, i.e. by rejecting the null hypothesis, which is by default a no-effect hypothesis. Still, it is assumed that estimations are subject to random errors due to the size and composition of the sample. For this reason, hypothesis testing is constructed in such a way as to minimize the probability of producing either false positives (failing to detect an association when in truth it does exist) or false negatives (finding an association between two events when it does not exist). These errors are respectively identified as Type I and Type II. More precisely, Type I error is committed if we reject the null hypothesis when it is true, the probability of this being denoted by α . Type II error is instead committed if we do not reject the null hypothesis when it is false, the probability of this being denoted by β . α and β cannot be simultaneously minimized because they are inversely related, so that a choice of which error to minimize – and conversely which type of error to maximize – has to be made.

Conventionally, Type I error is the one to be minimized. α denotes the rejection region, which is the region that contains the values of the statistic that contradicts the null hypothesis or in which it is very unlikely to see the null hypothesis supported. Most importantly, the rejection region determines the significance level of the test, for it constraints the size of the interval in which the results of the repeated tests are to be fall. The lower the significance level of the test and the rejection region, the more the level of precision is imposed, the less the probability of Type I error.

2.2. BIAS AND UNCERTAINTY: CRITIQUES OF FREQUENTIST STATISTICAL METHODOLOGY

Critiques have been moved to the conventional statistical methodology, based on the type of error to minimize, the extension of time and size of the experiment, and the way uncertainty is revealed.

Concerning the first point, hypothesis testing is accused to favor less precautionary actions. Given that the null hypothesis is conventionally set as a no-effect one – such as “ x does not entail y ” – the choice of minimizing Type I error is biased by a fundamental asymmetry: the statistical test is set up to be more cautious about detecting something which in truth does not exist (Type I or α error), than about failing to find out something which in reality does exist (Type II or β error). Indeed, Type I error is set to 5%, which denotes the probability of a false positive, whereas Type II error is set to 20%, which indicates the probability of a false negative.

Translated in environmental concerns, this means to set a null hypothesis such as “chemical x does not produce effect y ”. Accordingly, the statistical test is set to be less careful about failing to detect a relation between x and y when it does exist in 20% of the cases. Otherwise said, actual statistics strictly requires a high and degree of certainty of harm before taking any preventive action¹¹, whereas a precautionary approach claims for action even though scientific certainty has not yet been achieved. In the view of some scholars (Barrett and Raffensperger, 1998; Kriebel *et al.*, 2001), the way of incorporating the precautionary approach in statistical methodologies is by minimizing Type II error, so that “when there is substantial scientific uncertainty about the risks and benefits of a proposed activity, policy decisions [are] made in a way that errs on the side of caution with respect to the environment and the health of the public” (Kriebel *et al.*, 2001: 875). Indeed, in epidemiology and biotechnology trials the language of “safety” assessment incorporates this kind of concern by requiring a sufficient statistical power or β probability (EFSA, 2006)¹².

While this solution is intuitively valid, at the same time it ignores that either probability α or β can be minimized, so that some bias will anyway persist. Furthermore, if β were the probability to be minimized, then the significance level α will rise and the threshold for rejecting the null hypothesis will be lowered (i.e. the “strictness” for accepting scientific evidence will be relaxed). Additionally, many concerns have been raised about using power analysis in environmental management studies, due to the likely smallness of sample size for which important effects may not be detected, and to the fact that growing

¹¹ This form of conservatism directly stems from the concept of opportunity-cost, where inaction is conceived to be a waste to be minimized more than an option value of being cautious (and waiting) before taking action. For references to the option-value literature, see *supra* note 6.

¹² This requirement, however, has not been followed as many drugs-related cases have revealed, one among all the Vioxx litigation.

databases may lead to a situation in which ever-smaller differences are detected (McBride, 1999). Finally, as it shall be discussed later¹³, the choice about which type of error to minimize is strictly and conceptually connected to the way the null hypothesis is set.

On the side of uncertainty about statistical results, several problems exist. To begin with, uncertainty is often confused with something that either is not valuable or lacks qualitative information. Nonetheless, it is still a fundamental character of many environmental mechanisms and surrounds the field of biotechnologies in particular. That is why its quantification in terms of limits of scientific knowledge and its communication to policy-makers is of fundamental importance.

Today, the acknowledged sources of uncertainty in statistical testing mainly converge on sampling variability through the standard *p-value* and confidence intervals. On the other hand, several additional important factors are identified, such as the choice of the statistic value according to its desirable properties and to the underlying population, and the sensitivity of the findings to the choice of the statistical model.

Further critiques have been made on the ground that assumptions like independence among variables and the existence of random errors are inconsistent with many environmental and health-related events. This inadequacy prevents considering possible systematic errors that can originate from both selection bias and the very measurement of the levels of specific factors according to the availability of detecting technologies. Finally, methodological problems concerning the size and composition of the sample and the time extension of the experiment have been specifically advanced for GMOs' safety studies, along with the difficulty if not impossibility of repeatedly replicating the experiments¹⁴.

2.2.1. THE NOTION OF STATISTICAL EVIDENCE – ASSOCIATION VS. CAUSATION

There is a second reason why the value of uncertainty is often underestimated, which stems from a fundamental confusion about the heuristic power of statistical inference. This situation puzzles the process of drawing conclusions both in case of non-significant results (i.e. those that fail

¹³ See para. 3.3.

¹⁴ The reference study is: Séralini, G-E., J. Spiroux de Vendomois and D. Cellier, "Criticism and Improvements of Strategies for the Safety Assessment of GM Plant derived Foods or Feed. An Answer to EFSA Draft Report on Animal Feeding Trials With GMOs", obtained from a personal communication with Prof. Séralini, May 9th, 2007.

to reject the null hypothesis) and in case of significant ones, where the passage from association to causation is problematic.

Hypothesis testing provides a specific kind of information, which is about association between events, but never about causal relations between them. As Sir Austin Bradford Hill affirmed in his seminal article “The Environment and Disease: Association or Causation” (1965), “[significance] tests can, and should, remind us of the effects that the play of chance can create, and they will instruct us in the likely magnitude of those effects. Beyond that they contribute nothing to the ‘proof’ of our hypothesis”.

The relation between association and causation could be better understood by taking advantage of some notions from epidemiology (Rothman and Greenland, 1998). Before claiming that association does not prove causation, the concept of cause must be better defined.

First, the concept of a one-to-one correspondence between the cause and the effect shall be cleared. Fundamentals of epidemiology instruct us that even if it is correct to define a cause of a specific event as “an antecedent event, condition, or characteristic [...] without which the disease event either would not have occurred at all or would not have occurred until some later time” (Rothman and Greenland, 1998: 8), nonetheless the concept of “causal constellation of components” informs us that numerous and less evident causes operate for producing an effect. A complete causal mechanism is therefore one in which a “sufficient cause” is composed of minimal (i.e. necessary) conditions or events, each of them being part of a multiplicative function.

The specification of the concept of causation endows us with a useful warning: in a causal chain, being each and every component necessary, the detection of possible associations is never exhaustive to account for a causal relationship. On the converse, the failure to detect an association between two events may hide a sufficient causation as soon as a third component is introduced in the chain.

In particular, biological interactions are characterized by a high degree of uncertainty since most and sometimes even all of the components of a sufficient cause are unknown (Rothman and Greenland, 1998: 12). Moreover, even though interactions of this type are known as co-actions or joint actions, this does not mean that they are simultaneous: the effect of one component may interact with the other component with a possibly large delay.

Back to the concept of significant or statistical evidence, the task of falsifying hypotheses is not aimed at giving a direct measure of, i.e. to quantify, the *probability* of the null hypothesis to be true. Rather, the task of hypothesis testing is to test the consistency, i.e. the *reliability*, of an association between experimental data and the null hypothesis (Adelman, 2004). More precisely, the association is reliable to the extent that the error rate is to be calculated by testing the likelihood of finding that particular (non-) association just by chance, either by a false positive (Type I error) or a false negative (Type II error).

Statistical significance, besides being not a causal criterion, is independent of important considerations such as the source of observational data, the degree by which the experiment conditions are controlled, the assumptions on the exposure levels (Adelman, 2004). Thus, through statistical significance alone any of this information would not be conveyed.

On the side of low level or non-significant results, i.e. failing to reject the null hypothesis, missing information stems from the fact that these results are often not reported in scientific reports (Mayo, quoted in Adelman, 2004). Nonetheless, failing to detect an association between two events does not mean that the absence of evidence is the evidence of absence¹⁵, or that there is no effect, and more than ever it does not prove a negative, which is impossible¹⁶. That is why, even though scientists proceed by *preponderance* of evidence and parsimony among theories¹⁷, it could still be possible to make inference from absence, i.e. when the null hypothesis is not rejected. This happens only when we postulate the existence of an entity or phenomenon or in case of universal laws of nature: if no evidence to the contrary ever occurs then the absence results in support of the theory. So far, however, we have learnt from frequentist statistics that our assumptions are phrased as non-existence ones. On the contrary, as it shall be discussed later¹⁸, the argument about the inference from absence is relevant to the extent that postulates or assumptions are to be questioned.

Finally, the following remark will better explain possible misinterpretations of data once the object under control, i.e. environmental and human health effects, is dismissed in drawing conclusions about the test. In fact, it is not excludable that a situation of little consistency between data and the hypothesis be due to the fact that the object of prediction is a rare event, as it is likely to be in case of long term, low risk or catastrophic environmental effects. Similarly, Sunstein's concept of "availability heuristic" (Sunstein, 2004;

¹⁵ To prevent any abuse of this idea, Mano Singham writes in his web journal (available at http://blog.case.edu/singham/2006/05/11/burden_of_proof3_the_role_of_negative_evidence), that "when you are asserting the existence of an entity, if you have not provided any evidence that they do exist, then the absence of evidence is evidence of absence".

¹⁶ The proof of the negative is a special case of the fallacy of denying the antecedent. See later para. 3.1.

¹⁷ The principle of parsimony (or *lex parsimoniae*) in scientific theories is also known as the Ockham's razor principle. According to that, among competing theories only those that are based on the fewest assumptions should be retained, and conversely those that are useless (*entia non sunt multiplicanda*, translated: entities are not to proliferate) should be eliminated, or "shaved off". As it is stated, it is by now the only method to choose between theories, for it aims at preventing congestion among theories with the risk of resulting in no scientific guidance; however, the use of negative proof in place of performing experiments is not justified, since all relevant data and all possible relevant experiments should be first exhausted.

¹⁸ See para. 3.3.

2005) shall exemplify how possible misperception occurs. If in fact we are confronted with two hypothesis, one very rare and the other very common, each of them having similar significant level of association with the same event, we will nonetheless tend to attribute the cause of the event to the most common situation so that the perception of the respective probabilities of the two events will result distorted.

3. SCIENCE FOR POLICY AND POLICY FOR SCIENCE – CONTRIBUTION FROM THE PRECAUTIONARY PRINCIPLE TO RECTIFY THE CONNUBIUM

As it has been discussed, the precautionary principle has brought about interesting critiques to the traditional, i.e. frequentist, methodology of statistics by denouncing its implicit bias toward less precautionary actions. This is a great and constructive critique, but in order to retain the very useful lessons it is crucial to understand the fundamentals of frequentist statistical analysis and the rationale of its methodology.

The analysis will proceed along two tracks starting from the same origin: the concept of falsification. The first track is indispensable, in that philosophy of science will provide critical insights to fully understand the notion of scientific knowledge by broadening the discussion to the scientific method in general. The second track, on the other hand, will restrict the analysis, to discuss the critiques that have been addressed to the traditional frequentist methodologies for environmental effects. In particular, I will evaluate statistical methodologies for their desirable conceptual methodological honesty, more precisely I will analyze the rationale behind frequentist statistical inference and its conventional way of setting the null hypothesis in order to surrender any simplistic controversy on the biotechnological debate.

3.1. FALSIFICATION: INSIGHTS FROM PHILOSOPHY OF SCIENCE

Although the critical analysis of the precautionary principle has its fulcrum on frequentist statistics, the borders of such debate extend to scientific methods in general (Barrett and Raffensperger, 1999; Jordan and O’Riordan, 1999). Philosophy of science is the discipline to refer to in that it critically instructs the discussion on the different aims and methods of science, along with its principles, practice and achievements (Salmon *et al.*, 1999).

During the twentieth century, the two dominant approaches to philosophy of science, the logical positivists (Vienna circle) and the logical empiricists (Berlin), despite their differences, focused both on the logical explanation of scientific concepts and on the relations between theory and evidence.

Through logic, science gained its unique and independent status of privileged discipline.

Nonetheless, the logic force of scientific reasoning processed both through inference and deduction encountered authoritative counter-arguments. To begin with, David Hume (1739) questioned the authoritative power of inductivism¹⁹ for it was simply based on the assumption that observed events would be eventually following the same equal pattern that had been observed ever since. According to Hume, there was no necessity behind this reasoning and, most importantly, there was no reason to assign a causal connotation to a constant sequence of events. In fact, what could be at first deemed to be a causal relation, at the end, if ever an observation to the contrary had to occur, could have turned out to be a simple coincidence. This circumstance is also better known as the “fallacy of affirming the consequent”²⁰, in which the logic failure originates from the fact that inference processes are not necessarily truth-preserving, which means that even if premises are true, conclusions can be false, and vice versa²¹ (Salmon *et al.*, 1999).

After that, scientific method was thus highly prejudiced, nonetheless the question remained: how to set a logical basis for it? This could be possible as long as the truth, what traditionally science had been devoted to, be distinguished from knowledge, or in different terms, the reason why a particular phenomenon occurs (cause and effect) had to be kept separate from the reason for *believing* that that phenomenon would ever occur.

Karl Popper (1957; 1959) thought that scientific statements could never be proved to be logically true since the only information to be inferred was about the (in-)consistency between an initial hypothesis or assumption and a specific observation. Notably, scientific knowledge would have resulted in the following fundamental asymmetry: if in fact evidence of consistency between the result and the initial assumption could not have proven the latter to be true, on the other hand inconsistency was carrying a different heuristic power in that it could²² make the initial hypothesis to be disproved. Just one

¹⁹ Inductivism is characterized by four major features (Salmon *et al.*, 1999): it is ampliative in that the conclusion has a content that goes beyond the content of its premises; it is not necessarily truth-preserving, in that there could be true premises and false conclusions; it is not erosion-proof, in that new premises can completely undermine the argument; any combination of premises and conclusion (be them true and/or false) is possible for the validity of the argumentation to rest, that is why inductive arguments come in different degrees of strength.

²⁰ The general form of the fallacy of affirming the consequent is the following (Rothman and Greenland, 1998): we have a premise which states that if H is true, then B must be true; we know through observation that B is true; therefore we conclude that H must be true.

²¹ By contrast, deductive reasoning is truth-preserving in that if premises are true, then conclusion must be true as well.

²² It is not automatic to disprove a hypothesis as soon as an observation to the contrary occurs especially if it has long been confirmed by evidence. This choice in fact depends on the scientific group conducting the analysis, which legitimately may

observation to the contrary had therefore the potential to break the supposed unequivocal deterministic chain.

The great lesson from Popper, which statistical methodology has largely learnt, is that our knowledge comes from experimental science through a process of “conjecture and refutation” and, most importantly, from ignorance. Science proceeds by elimination and is therefore tentative in its essence.

The provisional character of scientific explanations should not lead to prompt conclusions against the objective character of science and its capability of explaining how the world really works. Instead, the relevant question is whether the rejection paradigm will ultimately narrow scientific knowledge to the truth or not; in other terms, whether scientific objectivity is achievable at some end or it is rather a Kantian metaphysic²³.

The answer to this question is important in qualifying scientific uncertainty and its “waiting value”²⁴. In fact, if we suppose that scientific knowledge is progressively improving, then there is an economic rationale for postponing costly preventive measures to the point where better, if not complete and perfect, information will allow us to make the optimal choice.

In a similar vein, in epidemiology it is quite diffused the concept of the world ruled by a deterministic mechanism to be captured in its ultimate components through inference along with the hypothetico-deductive method. Room for chance considerations is thus the expression of our ignorance and it is only temporary, i.e. to be eliminated at some point in time. Nonetheless, even if temporary and incomplete, the knowledge we are endowed with at each time is still the basis of our actions and no delay will be legitimated just on the prospect of an improved knowledge. As Rothman and Greenland (1998: 22) warn, “the tentativeness of our knowledge [will] not prevent practical applications, but it should keep us skeptical and critical”. I do see in this way of defining scientific knowledge a sharp orientation to some form of precaution, by which scientific uncertainty is reconsidered for its encouragement to dialectic and so the opposition between the precautionary principle and science gradually evanishes.

deem an “alien” observation as a simple anomaly. This happens mainly in epidemiologic studies and generally in scientific disciplines where improvements occur through the criteria of preponderance of evidence.

²³ Immanuel Kant in the *Critique of the Pure Reason* (1781) believed that complete knowledge of the world was not possible, still that kind of knowledge could be interpreted as metaphysical in that, although not achievable, it provided us with the aim to which our attempts reached for in a continuous escalating process.

²⁴ Gollier and Treich (2003) marked the difference between preventive effort and precautionary effort by referring, respectively, to the static of risk management at a given time with a stable probability distribution and, on the other hand, to the dynamic of scientific progress (and scientific uncertainty) over time. “In short, while prevention aims at managing risks, precaution aims at managing the wait for better scientific information” (Gollier and Treich, 2003, *supra* note 6 at 86).

3.2. EVALUATING SCIENTIFIC KNOWLEDGE

The last statement of the previous paragraph allows us to introduce a sociological discourse about scientific knowledge. This is particularly relevant insofar as it provides serious elements for bridging the alleged opposition between science and policy and it illustrates the operational value of the precautionary principle.

In the first half of the Seventies, critiques had emerged against this sort of untouchable notion of science. Thomas Kuhn (1962) became known as one of the leaders of these sentiments and decried the unacceptable underestimation of the history of science, of its progress and cultural dimension. In his thoughts, the critical decision about which problems should be given priority and how to interpret statistical results could not be understood without a social lens. Precisely, refutation was part of a decision-making process where the consensus of the scientific community determined what to accept and what to reject²⁵. That choice in fact entailed alternative considerations about the validity of the theory being tested (the initial hypothesis) against the validity of the scientific infrastructure, intended as the prevailing science, the so-called “normal” science (Rothman and Greenland, 1998).

The claim that science is a “social construction” should terribly hurt any scholar spending his or her life searching for a rule, if not the rule, about some process, being it scientific, social or economic. Nevertheless, such a hurting feeling is the product of idolizing scientific knowledge, whose source of validity partly comes from our postulates or premises. Once acknowledged that those are hostage of our contingent interaction with the natural world, then it naturally follows that they prominently affect methodologies, aims and principles of scientific research, let alone the interpretative conclusions.

Bayesianism addresses this specific problem in that no deductive argument can ever provide the information about the truth or falsity of a scientific hypothesis unless there is certainty about the truth of the premises. This fact is the consequence of the truth-preserving characteristic of deduction, i.e. if the premises are true, then the conclusions must be true²⁶ (Samuel *et al.*, 1999).

As it has already been discussed, we construct our premises on observation, but observation is possibly undermined by random or systematic errors, so that no inference about some association between two events can ever be proved true, but only disproved (or falsified). If scientific proof is impossible, as Hume teaches us, then no premise can ever be deemed completely certain

²⁵ See *supra* note 22 (anomalies).

²⁶ This property directly comes from the complementary ones informing that, as all the content of the conclusion is at least implicitly present in the premises, deduction is non-ampliative. For sake of completeness, two more characteristics of deduction exist: it is erosion-proof, i.e. if new premises are added and the previous are not erased, then the argument remains valid; and deductive validity is all-or-nothing matter, no degree consideration of validity is admissible (Samuel *et al.*, 1999).

and consequently even no deduction reasoning can ever be certain. This is both a dilemma and a vicious cycle, for if inference carried no logical foundation, the drama is further amplified by the fact that deduction has no scientific utility (Rothman and Greenland, 1998).

Then, if it is not truth, what does deduction preserve? Deduction preserves subjective probabilities, or degrees of certainty, that we attach to premises. Once assigned, the rules of probability theory would solve the problem by producing a logical, valid, personal certainty of conclusion.

If inference is subjective because it is built upon postulates, then the scientist's state of mind becomes truly important in determining which theories and methodologies will be applied. This is a critical factor that can be linked to the uncertainty-related communication problem discussed in paragraph 2.2, for scientific results may appear to policy-makers to be more certain than they actually are, due to the missing consideration of initial assumptions.

3.2.1. THE QUALITY OF UNCERTAINTY: A NEW AMBITION FOR POST-NORMAL SCIENCE

Once acknowledged that science is not a value-free discipline, many would conclude that we are depriving its protective function of legitimating policy decisions and, mostly, its capability of providing a common ground of agreement. Different points regarding the concept of science, its tasks, and its consensus-gathering force should be made against this alarmist view.

The need to rethink scientific knowledge originates from a twofold concern, which is to denounce together its anachronistic march toward risk management and the denial of the cultural and public dimension of that same march. It is no wonder that the human capacity to address problems such as loss of biodiversity, climate change, resource depletion and other related issues, is in doubt, because "...facts are uncertain, values in dispute, stakes high, and decisions urgent" (Funtowicz and Ravetz, 1991: 138). This state given, the need for a new conceptualization of science has emerged as impellent for adjusting its ambitions to the new global challenges.

Barrett and Raffensperger (1998) make an interesting conceptualization about scientific knowledge that allows us to integrate the notion of uncertainty, which instead has generally been separately advocated as the core reason for invoking precautionary actions as opposed to science-based actions. They identify "sound" or "industrial"²⁷ science as the authoritative source of decision-making for environmental policies. Industrial science is characterized by the strict association with risk assessment and considers uncertainty as "a temporary and surmountable lack of data" (Barrett and Raffensperger, 1998: 2). Conversely, they suggest the need of an alternative conceptualization of

²⁷ The industrialization of science is characterized by capital-intensity and commissioned research.

science, now called “biosystems” science, without which it is not possible to comprehend the rationale behind the precautionary principle. If we concede that uncertainty covers a large portion of environmental science – not only due the lack of data but also for the ecosystems’ inherent complexity, non-linearity and impracticable experimental replication of environmental testing; and if we conceive that this uncertainty, rather than being a “temporary misalignment of theory and observation” (Barrett and Raffensperger, 1998: 5), is the product of indeterminacy as opposed to determinacy, then the value of waiting changes. In fact, instead of leading necessarily to more certainty, progressive environmental studies can increase the areas of disagreement, as it is to be expected from non-linear systems. This point is truly crucial in that it marks a detachment from anything it has been said so far: besides arguments that science is neither deterministic²⁸, nor it will ultimately provide complete understanding of the world, still the supposed progressive narrowing of scientific uncertainty through falsification is now challenged.

Whether or not scientific progress will ultimately achieve complete explanation of phenomena to let calculations predict the future, it is still arguable that science communalizes opinions and, by consequence, provides a protective shield of legitimacy for any action based on its prescriptions. For that reason, instead of truth, science should tend toward quality of information, which is defined as “the degree to which the recommended policy choices are robust against [...] underlying uncertainties” (Funtowicz and Ravetz, 1994: 202). In this approach, the relation between quality and uncertainty is not at all of antithesis, but rather of transparency, completeness and even “honesty of evidence”²⁹. Moreover, being uncertainty pervasive and stakes high for many environmental issues, quality, as the new leading principle of the “post-normal” science, advocates democratization of knowledge through extended peer-communities (Funtowics and Ravetz, 1991, 1994; Norgaard, 2006).

The idea of democratizing knowledge is strictly connected to a critical approach toward economics, for its one-dimensional standard of valuation, i.e. monetization³⁰, is incapable of comprehending all the uncertainties and complexities of the environmental problem. This sentiment introduces the second part of the scientific discourse in its social dimension. As Wynne (2002) tries to explain, our time is experiencing an increasing disconnection

²⁸ Cfr. association and causation at para. 2.2.1.

²⁹ In this context, the expression “honesty of evidence” has an epistemological connotation, but that will be further extended to a methodological discourse (*see* para. 3.3.).

³⁰ In this critique, Funtowics and Ravetz (1994) try to define the new paradigm of ecological economics as a “post-normal” science, in which all the complexities of environmental problems – including economic, ethical and social considerations – are to be held together in the rule of multidisciplinary. Nonetheless, to give monetary value to natural resources and even to ecosystem’s services is still advocated by some authors, e.g. Costanza *et al.* (1997), as the meaning to create a common perception of the costs of global problems such as natural resources depletion, climate change, etc.

between the public on the one hand and policy and science on the other. The importance of Wynne's article is that it draws attention to the institutional component of the scientific discourse. His analysis of environmental policies, to the extent that they are related to technological progress, goes beyond the mere legacy of uncertainty: it is not just about not having of all information to take action, it is further about denying what it is at stake, which represents the source of ambiguous decisions. Science has been building its authoritative role in policy-making upon the task of calculating and controlling risk. Hence, we have reached the stage in which any decision based on risk assessment should naturally be acceptable, legitimate. Risk appears as the *deus ex-machina* to the crucial economic problem of aggregating people's different preferences, due to its universal flavor. Once risk has been calculated and consequences have become known and controllable, no room is left for other considerations, such as the source of risk or the motivations behind the decision of (not-)incurring the risk. Thus, ambiguity of decisions originating from possible additional considerations is simply written off and attention to what is at stake is diverted.

The denial of ambiguity produces disaffection toward the "dominant scientific-institution risk culture" (Wynne, 2002: 460) among lay people because they are deprived of their dialectic space. Furthermore, the problem of this disconnection between the public on the one hand and policy and science on the other is far from being acknowledged in the form that has been expressed. Instead, it is often reduced to a fundamental opposition between risk-perception and risk-science, respectively related to lay people and the experts. In this vein, disagreement on technocratic policy-making is no more than the result of misperception of risk and, more provocatively, this status is at the origin of supporting precautionary actions. Sunstein (2005: 35) introduces one of his paragraphs by stating: "I suggest that the [precautionary] principle becomes operational if and only if those who apply it wear blinders – only, that is, if they focus on some aspects of the regulatory situation but downplay or disregard others". It is needless to say that the aspects on which Sunstein believes supporters of the precautionary principle would base their actions, are the product of risk misperception³¹. His arguments of behavioral economics carry their validity if and only if they are abstracted from "what it is at stake". And that is precisely what he does by denying ambiguity.

³¹ Sunstein (2005), by borrowing fundamental concepts from behavioral economics and cognitive psychology, identifies five sources of misperception of risk: the availability heuristic, probability neglect, loss aversion, a belief in the benevolence of nature and system neglect.

3.3. FALSIFICATION: THE SOURCE OF METHODOLOGICAL INTEGRITY

Besides scientific epistemology discourses, the falsification process shows its relevance in that it promotes what can be called methodological “honesty” and controls for the object of analysis. To begin with, the first question to ask is: “Why should we prove something by following a counterintuitive process, i.e. by falsifying it?”.

The relevance of complicating the process of statistical inference lies in the sake of providing objective analysis, or, equally, avoiding possible biases.

The initial point to retain is that before setting the null hypothesis there is an earlier stage in which we formulate a conjecture (also called research hypothesis), if not a suspect, about some possible event. This conjecture is precisely what triggers the statistical analysis and what we subsequently try to support by refuting the null hypothesis. Anyway, rejecting the null hypothesis requires a high degree of certainty so that the evidence of some effect could be confidently deemed scientific. The concept of “supporting by falsifying” is pivotal and it has to be strictly retained along with all the discussion.

For instance, we may suspect a chemical to have some negative effects on human health. If we were to set the null hypothesis in the same form of our conjecture, such as “chemical x causes harm to human health”, then our analysis would be biased from the very beginning, since we would concentrate all our efforts to prove what actually is a personal allegation. In other terms, because we tend to commit ourselves to personal beliefs, unexpected discoveries could turn out to be very disappointing, therefore to be avoided.

Discussions about how we should formulate assumptions have rarely been made, but as we have learnt from the history of philosophy of science, those are fundamental to the result of the analysis. From Bayesianism we know that *posterior* probabilities strictly depend on *prior* probabilities, but even without this crucial lesson it is possible to reach a similar conclusion by directing our mind to the concept of falsification.

3.3.1. THE RATIONALE OF STATISTICAL CONVENTIONS: ADAPTATION TO ENVIRONMENTAL CONCERNS

Refutation is the basis of scientific knowledge, whose soundness is so crucial in our society that it requires a high degree of certainty and thus compels to restrict as much as possible the probability of error. The crucial question is then “erring about what?”. By answering this question, we know what we want to control: in fact, we want to avoid producing scientific information which is not scientific at all, and the best way to preserve the basis of science, i.e. falsification, is minimizing the possibility of rejecting the null hypothesis when in fact it is true (i.e. Type I error).

Type II error is not as serious as Type I, because failing to refute the null hypothesis when it is false does not entail that we accept it as true. We just do not know. Because of this asymmetry between Type I and II errors, we already know that the probability α is more important to minimize, for sure. What is then to be discussed is its complement, the null hypothesis.

As it was explained in paragraph 2.1, it is convention in significance hypothesis testing to postulate a no-effect null hypothesis, such as “chemical x does not cause harm to human health (or to a specific environment)”. For those who are familiar with environmental concerns, this particular setting could immediately raise some conceptual problems. In fact, that hypothesis is far from expressing concern and precaution about potentially harmful effects. Statisticians would reply to this kind of perplexity by saying that the way the null hypothesis is set is the product of convention. Then, to better understand the situation let us temporarily overcome skepticism and maintain the default rule of a no-effect null hypothesis. The subsequent stage is to minimize one type of random error (and, conversely, maximize the other). Again, conventionally, the error to be minimized is Type I or α error, which occurs when the null hypothesis is rejected while it is in fact true. Translated in environmental terms, this choice would entail minimizing the possibility of detecting an environmentally dangerous effect, which in reality does not exist, and, consequently, to maximize the possibility of failing to detect the same effect when in reality it does exist.

This time, the earlier skepticism against conventional statistical methodology could not be surmounted: in fact, even supposing that the way the null hypothesis is conceived is, at least intentionally, not biased, certainly the preference for controlling Type I error is.

Why then these conventions should be maintained? One plausible reason consists in our legal and cultural apprehension that someone is presumed innocent until proven guilty. This is the same reason why even those defendants who themselves claim to be guilty – as it happened for example during the terrorism period in Italy in the Seventies – are to be given the possibility of a trial. Until no positive evidence has been delivered and a burden of persuasion has been reached, nobody can be said to be guilty. In statistical terms, this entails that the null hypothesis be a no-effect one.

Given the effect to be controlled, the choice of Type I error is due to a reasonable concern, which is convicting someone who is in fact a victim. This kind of error is abhorred in our western society, so that we can all agree that the concern about controlling Type I error is truly founded and therefore is the one to be minimized. Nevertheless, is this not precisely a kind of precautionary approach toward the weakest part, i.e. the victim? I believe it is and, differently from what is generally considered, this attitude does not directly stem from the type of error we choose to minimize, but rather from the way the null hypothesis is constructed to integrate a specific concern.

The reason of this claim is rooted on the fact that the way the null hypothesis is constructed and the decision of minimizing one type of error are part of the same rule: first do no harm. On the one side, the scientific attitude instructs us

to discourage at most the probability of committing an error, especially the error of claiming that something is certain (i.e. scientific) when in fact it is not; on the other side, the object of our concern is not shielded by the type of error we want to minimize, but, rather and before, by the assumption we make of it. We are already precautionary about a specific circumstance at the moment we construct the null-hypothesis, so that the decision about which error to minimize naturally follows. As it shall be discussed in paragraph 3.3.2, if the null-hypothesis already includes the object of our concern, then the type of error to minimize is always Type I.

Let us now move specifically to the GMOs issue. We know from above how the null hypothesis is set and which error to control for³². Following the same rationale as for not convicting a victim, we should assure that we do not conclude that GMOs are unsafe (or at least that they exceed a certain risk standard) when actually they are not, so as to avoid overregulation³³. At the same time, as we cannot simultaneously minimize Type I and Type II errors, the latter is conversely maximized in such a way that the possibility of underregulating a specific GMO, when in fact it poses risks to human health (or the environment), raises less concern.

In this case, something seems to be changed. That is precisely our concern, due to the fact that what was before identified as a potential victim raising feelings of protection, now has become a possible risk or hazard which creates totally different perception of who is at this point the weakest part to protect. This change happens because the object of the analysis itself, along with the new distributional issues it raises, has been modified.

At this time, political as well as social considerations enter the discussion insofar as they attach different values to different objects of regulation and, consequently, different types of errors. Type I and Type II errors suffer from an asymmetry of “contingent importance”, insofar as the cost of erring respectively on the side of underregulation and overregulation pivots around the contingency of the concern. It is precisely that, along with the available scientific knowledge, which triggers decisions in favor of more or less precaution. For this reason, the peculiarity of the object of analysis should never be dismissed against conventional methodological rules and our postulates should always find their proper analytical contextualization and never become obsolete alongside the path of research.

³² In the special case of GMOs, the null hypothesis setting is based on the assumption that GMO and their wild counterpart, non-GMOs, are “substantially equivalent”. The concept of substantial equivalence is the key for a comparative assessment, in which traditionally cultivated crops have gained a history of safe use so that they provide the baseline for determining any substantial difference between the GMO and its wild counterpart (EFSA, 2006).

³³ See para. 3.3.2, Shapiro (1995).

3.3.2. ISSUES OF GUIDANCE - SUBJECT, POTENTIAL VICTIM AND OBJECT: QUESTIONS OF BIAS

As it should have emerged, the potential victim of the case has been changed. Furthermore, the victims are potentially numerous, probabilities of incurring in the same risk are likely to be correlated among individuals, and finally risk is largely replaced by uncertainty and, possibly, by fundamental uncertainty³⁴. Once agreed that different issues entail different costs and concerns, it is useful to find a guiding rule for setting the null hypothesis appropriately and, accordingly, decide which error to minimize. Bearing in mind the complementarity between the two, the attentive reader should have realized that the kinds of environmental critiques moved against statistical methodologies (para. 2.2.) – along with suggestions to simply minimize Type II instead of Type I error – may hide methodological as well as interpretative failures. In fact, as α or β cannot be simultaneously minimized, either power analysis is performed while β is still maximized, or β is minimized with the inconvenience of conversely raise the rejection region denoted by α , which will enlarge the boundaries of accepting evidence as scientific when in fact it is not.

Besides these reasons, additional issues are to be developed on the side of the burden of proof.

To begin with, as discussed by Shapiro (1995), in condition of uncertainty – which is precisely the case for GMOs – the errors of overregulation and underregulation have different costs and different bearers, respectively the industry and the individuals. This said, his position is one of advocating fairness by setting the burden of uncertainty about potential risk on the industry rather than individuals, since the former is the least cost bearer. Again, we should always keep in mind the question “who is the weakest part, who is the potential victim, who should be protected”, given a particular concern. The rule for setting the burden of proof directly comes from its complementary, i.e. the identification of the potential victim, so that the onus of proof is on the non-victim part.

So far, we know the object of the study, i.e. GMOs, we know the potential victims, we know the least-cost bearer that should carry the burden of proof, but we still do not know what we have to prove: either the existence of some harm or the existence of some safety³⁵. To this matter, it is crucial to consider

³⁴ Since Knight (1921) the distinction between risk and uncertainty has become relevant to the extent that the first is characterized by an objective probability distribution, while for the second there exists no precise statistical estimates. Fundamental uncertainty, or ignorance, is distinguished from simple uncertainty by the fact that the impossibility of attaching accurate probabilities to possible scenarios derives from the impossibility of even conceiving those scenarios, because the information required does not exist at the decision time.

³⁵ Neither of the two is of course possible to prove in absolute and deterministic terms (cfr. para. 3.1.), so that by using the word “proving” the referee is always some preponderance of evidence.

the subject conducting the analysis, since, while keeping the rationale of falsifying the null hypothesis, possible bias will arise on the setting of the null hypothesis and on the choice to minimize one specific type of error. In fact, it is needless to say that a study on the same matter committed by an environmentalist organization and by an oil company would be biased in opposite directions.

For these reasons, I believe that the allocation of the burden of proof could not be reasoned without at the same time reconsidering statistical hypothesis testing. More precisely, this is not just a question of bias, but also and further a question of giving an incentive that is consistent both with the personal interests of the agent (e.g. selling GMOs) and the need to have honest scientific conclusions. Thus, a consistent incentive is one that leads the agent to canalize all his efforts to find out what he desires, that is to falsify (make *scientific* highly certain statements about) what he does not desire, i.e. the GMOs to be unsafe. This twofold concern stemming from the burden of proof is easily explained through an example: let us set the null hypothesis in the conventional manner, "GMOs have no effect on human health". Bearing in mind the WTO case of the *EU-Biotech* dispute, the parties willing to commercialize GMOs in the European market are the ones that, according to the precautionary principle, should prove that no risk to human health arises from their action. However, in the WTO Biotech dispute the complainants are of course to do their interest, which is to market GMOs, but the starting assumption that GMOs are safe would clearly bias their research.

To make the point clearer, rejecting the null hypothesis would mean that the experiments have provided high evidence that undesirable effects caused by GMOs actually exist. In order to come up with such a preponderance of evidence, a high degree of certainty – and effort – is required by setting a very low, 5% or 1%, significance level α . If, on the contrary, this preponderant evidence is not reached, then the null hypothesis cannot be rejected, neither can it be accepted as true: this means that scientific statements about the safety of the product are not allowed and the potential danger cannot be excluded. However, it is practice to draw scientific conclusions of no-effect evidence by failing to reject the null hypothesis or, conversely, by failing to detect a statistically significant association (Parkhurst, 2001).

Let us now try to build a counter-example: the null hypothesis is "the GMO has adverse effect y on human health (or on biodiversity)". The GMO producer, who has the burden of proof, will convey all his effort to reject the null hypothesis in order to build up evidence about their safety. Differently from the scenario of a no-effect null hypothesis, the GMO producer will be required a high level of certainty to prove that the GMO is preponderantly safe, whereas the benefit of doubt about their harmfulness directly coming from the non-refutation of the null hypothesis, will be left to consumers.

However, there is a practical problem of putting the hypothesis in the affirmative form. In fact, in this case the adverse effect to test has to be specified. A no-effect hypothesis, in accordance with the principle of

substantial equivalence³⁶, presumes that the ingestion of GMOs produces no change in the key nutrients or anti-nutrients of the plant. This form of hypothesis does not require specifying all kinds of expected effects, since in case some effects were detected, their specification would be made at the time they are observed. On the contrary, an effect-null hypothesis requires indicating *a priori* the specific effect to be concerned with.

Given that there is a huge portion of uncertainty about possible effects of GMOs both on biodiversity and on human health, this situation is truly problematic. How to find a compromise between the incentive effect that has been exposed and the practicability of it?

At this point, it is helpful to revise alternative statistical testing. Proposals have been advanced to introduce equivalence testing as substitute for (McBride, 1999) or additional stage to (Parkhurst, 2001) traditional hypothesis testing. Equivalence testing requires the null hypothesis to incorporate the concern that two similar products may differ in some of their characteristics. This, for example, happens to be used for drugs testing. Filtered down into the GMOs issue, this implies that instead of assuming that GMOs and non-GMOs are substantially equivalent³⁷, the two are presumed to be “bioinequivalent” (McBride, 1999: 3). This is certainly the first step for incorporating a precautionary approach. Furthermore and strictly related to the problem of specifying the type of effect to be studied, zero-valued³⁸ point estimates are replaced by intervals determining the bioequivalence region. The null hypothesis is set in such a manner as the difference between the parameters of the test GMO and reference non-GMO is lower and greater of, respectively, the lower and upper bounds of the interval. The size of the equivalence interval is decided before performing the test (or it can be extracted from an already set regulatory standard) to formalize all the set of differences which are of no practical importance.

Hence, equivalence testing is an important statistic tools in that it makes it possible to incorporate the precautionary principle while maintaining the scientific rationale of falsifying the null hypothesis and controlling for Type I error. On one side, refutation of the null hypothesis now implies that high certainty is required to conclude that GMOs are as safe as their natural counterparts, whereas failing to reject the null hypothesis now implies that the doubt³⁹ on possible danger will at most err on the side of more precaution. On the other side, given that Type I error was previously identified as the crucial one for deriving statistically significant inferences, the issue of concern is now to reject that GMOs are not as safe as their counterparts, i.e. finding that they are as safe as their counterparts, when this is not true.

³⁶ See *supra* note 32.

³⁷ See *supra* note 32.

³⁸ The point estimate is zero because it informs that the point hypothesis is a no-effect one.

³⁹ Remember that failing to reject the null hypothesis does not entail that it is to be accepted or confirmed to be true.

To conclude, the relevance of equivalence testing has emerged gradually along with a train of thoughts that took into consideration the object of concern, the potential victim and the least-cost bearer carrying the onus of the proof. The key message is that all those components have to be – and had been – balanced against the “scientific power” of statistical methodologies, in order to build a scenario of highly scientific results to be produced consistently with the interests of both the potential injurer and the potential victim. The former will in fact be required to put all his effort in producing high-evidence results that GMOs are safe instead of benefiting from doubts about their harmfulness; the latter, conversely, will benefit from precautionary science.

CONCLUSIONS

When Sir Austin Bradford Hill (1965) illustrated in his article nine aspects to consider in order to deduce *likely* causation from association, his prescriptions were welcomed as rules for cause-effect decision-making. As some scientists have pointed out, more important lessons were unfortunately missed (Phillips and Goodman, 2004).

His stress on contingency of evaluation (“the evidence is there to be judged on its merits”) and on the importance of the object at stake (“...we may surely ask what is involved in our decision”) perfectly incorporated all the relevant concerns and reasons for taking precautionary actions. The estimation of the value of waiting against the value of taking action “will depend upon circumstances”, upon the information we have at our disposal (“the whole chain [of cause-effect] may have to be unrevealed or a few links may suffice”). On the same idea, the aim of our action “almost inevitably leads us to introduce differential standards before we convict”, so that even in the presence of weak evidence precautionary actions can be undertaken and “if we are wrong in deducing causation from association no great harm will be done”⁴⁰.

The importance of environmental studies does not need to be shouted any more; nonetheless this has often resulted in big noise about environmental policy rules (Wynne *et al.*, 1996). I do believe this is the product of the difficulty of tackling the environmental issue through a multidisciplinary perspective. The debate about the precautionary principle has further sharpened this situation and a new need for clarification has become compelling. In particular, the paradox of scientific knowledge being used to

⁴⁰ In this statement, Hill was presenting the example of introducing a drug for early-morning sickness in pregnant women. As he said, the doctor can decide to restrict the use of the drug even on relatively slight evidence, the fact being that the “good lady and the pharmaceutical industry will doubtless survive”.

delegitimate environmental policies at the international level need to be solved, and for that a tentative preapprehension of science in its statistical formulation has been made to eventually determine the relation between scientific and legal standards of persuasion.

REFERENCES

- Adelman, D. (2004), "Scientific Activism and Restraint: The Interplay of Statistics, Judgment, and Procedure in Environmental Law", *Notre Dame Law Review*, 79(2) 497-584.
- Arrow, K.J. and A.C. Fischer (1974), "Environmental preservation, uncertainty and Irreversibility", *Quarterly Journal of Economics*, 88(2): 312-19.
- Barrett, K. and C. Raffensperger (1998), "Precautionary Science", Paper prepared for the Wingspread Conference on Implementing the Precautionary Principle, available at <http://www.johnsonfdn.org/conferences/precautionary/barraf.html> [last access: May 2nd 2007].
- (1999), "Precautionary Science", in C. Raffensperger and J. Tickner (eds), *Protecting Public Health and the Environment: Implementing the Precautionary Principle*, Washington, DC: Island Press.
- Barro, R. J. and D.B. Gordon (1983), "A Positive Theory of Monetary Policy in a Natural Rate Model" *Journal of Political Economy*, 91(4): 589-610.
- Birnie, P. and A. Boyle (2002), *International Law and the Environment*, Oxford: Oxford University Press.
- Bishop, W. (2000), "Risk Assessment vs. Precautionary Principle: Is It really either/or?" *Risk Policy Report*, 20: 35-38.
- Bothe, M. (2006), "Whose Environment? Concept of Commonality in International Environmental Law", in Gerd Winter (ed.), *Multilevel Governance of Global Environmental Change, Perspective from Science, Sociology and the Law*, Cambridge: Cambridge University Press.
- Brand, R.A (1996), "Sustaining the Development of International Trade and Environmental Law", *Vermont Law Review*, 21: 823-872.
- Cameron, J. and J. Abouchar (1996), "The Status of the Precautionary Principle in International Law," in D. Freestone and E. Hey (eds.), *The Precautionary Principle and International Law: The Challenge of Implementation*, The Hague: Kluwer Law International.
- Carlin, B. and T. Louis (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, New York: Chapman and Hall.
- Charnovitz, S. (1994), "Free Trade, Fair Trade, Green Trade: defogging the debate", *Cornell International Law Journal*, 27: 459.
- Christoforou, T. (2003), "The Precautionary Principle in EC Law and Science," in Joel Tickner (ed.), *Precaution, Environmental Science and Preventive Public Policy*, Washington, DC: Island Press.
- Costanza R. *et al.* (1997), "The Value of the World's Ecosystem Services and Natural Capital" *Nature*, 387: 253-260.
- Dixit, A.K. and R.S. Pindyck (1994), *Investment under Uncertainty*, Princeton University Press.
- Dixon, C.A.A (2000), "Environmental Survey of WTO Dispute Panel Resolution Panel Decisions Since 1995: "Trade at All Costs?"" *William and Mary Environmental Law and Policy Review*, 24(1): 89.

- EFSA (2006), *Guidance Document of the Scientific Panel on Genetically Modified Organisms for the Risk Assessment of Genetically Modified Plants and Derived Food and Feed*, European Food Safety Authority: Parma.
- Funtowics S.O. and J.R. Ravetz (1990), "A New Scientific Methodology for Global Environmental Issues", in Costanza, R. (ed.), *Ecological Economics. The Science and Management of Sustainability*, New York: Columbia University Press.
- (1994), "The Worth of a Songbird: Ecological Economics as a Post-Normal Science", *Ecological Economics*, 10: 197-207.
- Gollier, C. and N. Treich (2003), "Decision-Making under Scientific Uncertainty: The Economics of the Precautionary Principle", *Journal of Risk and Uncertainty*, 27(1), 77-103.
- Henry, C. (1974), "Option Values in the Economics of Irreplaceable Assets", *Review of Economic Studies Symposium*, 41: 89-104.
- Hill, B.A. (1965), "The Environment and Disease: Association or Causation?", *Proceedings of the Royal Society of Medicine*, 58: 295-300.
- Holm, S. and Harris J. (1999), "Precautionary Principle Stifles Discovery", *Nature*, 400: 398.
- Hume, D. (1888), *A Treatise of Human Nature*, Oxford: Oxford University Press; 2nd ed, 1978 (original publication, 1739).
- Ingeborg, A. and T. Traavik (2003), "Genetically Modified (Gm) Crops: Precautionary Science and Conflicts of Interests", *Journal of Agricultural and Environmental Ethics*, 16: 227-247.
- Jordan A. and T. O'Riordan (1999), "The Precautionary Principle in Contemporary Environmental Policy and Politics, in C. Raffensperger and J. Tickner (eds), *Protecting Public Health and the Environment: Implementing the Precautionary Principle*, Washington, DC: Island Press.
- Knight, H. Franck. (1921), *Risk, Uncertainty and Profit*, New York: Augustus, M. Kelley.
- Kriebel, D. et al. (2001), "The Precautionary Principle in Environmental Science", *Environmental Health Perspectives*, 109 (9): 871-876.
- Kuhn, T. (1962) *The Structure of Scientific Revolutions*, 2nd ed., Chicago: University of Chicago Press.
- Lemons, J., K. S. Shrader-Frechette, and C. Cranor (1997), "The Precautionary Principle: Scientific Uncertainty and Type-I and Type-II Errors," *Foundation of Science*, 2, 207-236.
- Malakoff, D. (1999), "Bayes Offers a 'New' Way to Make Sense of Numbers", *Science*, 286: 1460-1464.
- McBride, G.B. (1999), "Equivalence Tests Can Enhance Environmental Science Management", *Australia and New Zealand Journal of Statistics*, 41(1): 19-29.
- Norgaard, R.B. (2002), "Optimists, Pessimists, and Science", *BioScience*, 52(3): 287-292.

- (2006), “Values and Valuation in a Rapidly Changing World. Lessons from the Millennium Ecosystem Assessment”, in *Biodiversity: The Interplay Between Science, Valuation, and Policy*, AIBS Annual Meeting.
- Orebeck, P. *et al.* (2005), *The Role of Customary Law in Sustainable Development*, Cambridge: Cambridge University Press.
- Ott, R.L. and M. Longnecker (2001), *An Introduction to Statistical Methods and Data Analysis*, 5th ed., USA: Duxbury.
- Parkhurst, D.F. (2001), “Statistical Significance Test: Equivalence and Reverse Tests Should Reduce Misinterpretation”, *Bioscience*, 51(12): 1051-1057.
- Phillips and Goodman, (2004), “The Missed Lessons of Sir Austin Bradford Hill”, *Epidemiologic Perspectives and Innovations*, 1: 3.
- Popper, K.R (1957), “Probability Magic or Knowledge out of Ignorance”, *Dialectica*, 11: 354-374.
- Popper, K.R (1959), *The Logic of Scientific Discovery*, New York: Basic Books.
- Rothman, K. J. and S. Greenland (1998), *Modern Epidemiology*, 2nd ed., Philadelphia, U.S.: Lippincott Williams&Wilkins.
- Salmon, M. H. *et al.* (1999), *Introduction to the Philosophy of Science*, Indianapolis/Cambridge: Hackett Publishing Company.
- Sandin, P. (1999) “Dimensions of the Precautionary Principle”, *Human and Ecological Risk Assessment*, 5: 889-907.
- Shapiro, S. A. (1995), “Keeping the Baby and Throwing Out the Bathwater: Justice Breyer’s Critique of Regulation”, *Administrative Law Journal*, 8: 721-732.
- Sunstein, C. R. (2005), *Laws of Fear. Beyond the Precautionary Principle*, Cambridge: Cambridge University Press.
- Vanderzwaag, D. (1999), “The Precautionary Principle in Environmental Law and Policy: Elusive Rhetoric and First Embraces”, *Journal of Environmental Law and Practice*, 8: 355-375.
- Von Moltke, K. (2006), “On Clustering International Environmental Agreements”, in Gerd Winter (ed), *Multilevel Governance of Global Environmental Change, Perspective from Science, Sociology and the Law*, Cambridge: Cambridge University Press.
- Wirth, D. (2006), “The Transatlantic GMO Dispute against the European Communities: Some Preliminary Thoughts”, *Boston College Law School*, 144: 173-208
- Wynne, B. (2002), “Risk and Environment as Legitimatory Discourses of Technology: Reflexivity Inside Out?”, *Current Sociology*, 50(3): 459-477.