

# Science for the Environment: Examining the Allocation of the Burden of Uncertainty

Elisa Vecchione\*

*The aim of this paper is to review the basic literature on scientific uncertainty in its statistical paradigm in order to provide enlightenment on one pivotal facet of the precautionary principle, i.e. the allocation of the burden of proof to demonstrate that an activity is not harmful to the environment. The purpose is not to explain a new theory of statistical inference, but to show how regulatory policymaking that is properly informed by scientific expertise and designed to avoid one type of error, may actually make other errors more likely and thus expose the public to danger. This problem is explained in terms of the conceptual as well as operational conflicts that arise when knowledge about statistical-inferential methods is applied to policymaking. The paper argues that this issue can be resolved by first reconsidering the burden of proof as a burden of uncertainty.*

## I. Introduction

The aim of this paper is to reconsider the relationship between science and policy, using the opportunity provided by the debate over the precautionary principle.

While this undertaking may not be new in the literature on policy science, the premises of the discussion are unconventional. For instance: they hinge on legal reasoning; the perspective is focused (i.e. only one feature of the precautionary principle, namely allocation of the burden of proof, is considered relevant to the discussion); and the conclusions are

novel. Despite the conventional separation between risk assessment and risk management, it is contended that the way to integrate the precautionary principle into policymaking is by choosing which kind of uncertainty – rather than evidence – the proposer of a potentially dangerous activity should bear – rather than provide – in terms of costs.

The subject of this paper came to my mind quite unexpectedly while analyzing the World Trade Organization's (WTO) Panel Report on the *EC-Biotech* dispute between the groups of the United States, Canada and Argentina, and the European Union.<sup>1</sup> The main question was to understand how the latter could invoke the precautionary principle in defence of the safeguard measures that six European Member States had adopted in order to temporarily ban the importation of certain genetically modified organisms (GMOs).<sup>2</sup> After all, competent European scientific committees had already reviewed and approved certain risk assessments, and they concluded that these products presented no cause for concern. In actual fact, the precautionary principle applies only to situations of scientific uncertainty, but this did not seem to be the case in this instance, since the risk assessments had allegedly disclosed evidence that these products posed no risk (or no greater risk) to human and environmental health.

In the *EC-Biotech* dispute, science was considered as a proof-deliverer that, based upon its supposedly

\* Chaire de Développement Durable, Sciences Po, Paris, Email: <elisa.vecchione@sciences-po.org>. Acknowledgements: For their support and helpful comments, the author thanks in particular Silvana Dalmazzone, Douglas A. Kysar, Olivier Godard, Theodor Eisenberg, Marie-Angèle Hermitte, Claude Henry, Bruno Contini, Antoine Méssean, Gilles-Eric Séralini and Gérard Pascal.

1 WTO Panel Report, 2006, European Communities – Measures Affecting the Approval and Marketing of Biotech Products, WT/DS291/R, WT/DS292/R, WT/DS293/R, Corr.1 and Add.1, 2, 3, 4, 5, 6, 7, 8 and 9, adopted 21 November 2006.

2 Genetically Modified Organisms are products that have been altered using recombinant DNA technologies. Drawing from the European legislation on the subject (Directive 2001/18/EC on the Deliberate Release into the Environment of Genetically Modified Organisms, OJ 2001 L 106/3, and Regulation (EC) No 1829/2003 on Genetically Modified Food and Feed, OJ 2003 L 268/1) GMOs pertain to two fundamental categories: food and feed products, consisting or made from GMO, no longer considered living organisms themselves and intended for consumption; crops or plants in a living form, capable of and intended for growing.

undisputable status, may have been a determining factor in ending the legal process. In the same vein, science is regarded as the ultimate source of rational policymaking. Hence, the European measures seemed inordinately irrational, if not trade-restrictive and contrary to WTO obligations.

However, the preliminary analysis described above is soon shown to be partial: in fact, once science is specified in its epistemological paradigm, it becomes clear that it is anything but undisputed, and it is certainly not self-evident. This places the realm of science not at the opposite extreme to that of precaution, but rather as merging the two spheres together according to the degree of uncertainty that surrounds the issue at stake. As a first consequence, the type of scientific proof required to support policy actions – namely risk assessment in the WTO system – is not to be taken as indisputable evidence of rationality. Most importantly, as we discuss later, the assumptions on which scientific regulatory analysis is based are not indifferent to policymaking purposes and may actually induce unwanted policy outcomes. This paper devotes extensive attention to this specific problem, approaching the issue in terms of the conflicts created by the application of statistical-inferential methods (risk assessment) for policymaking. The rationale of shifting the burden of proof onto the producer, as prescribed by the precautionary principle, will provide the conceptual framework to elucidate these conflicts.

To fill this theoretical framework with substantial arguments, the discussion will proceed as follows: The EC's position at the WTO dispute, in justifying the safeguard measures of Member States based on the precautionary principle, will provide a bridge for the analysis, linking the legal to the policy arena. The opposition between science and precaution is first theoretically destabilised, then arguments are made regarding its untenability: the subsequent section reviews both the critiques of the non-precautionary bias of statistical hypothesis testing and the fundamental rationale of statistical inference, which lies at the origin of risk assessment. After highlighting that scientific knowledge proceeds from a process of "conjecture and refutation", and hence that falsification is an indispensable element in claiming certain information to be scientific, explanations will be given of why and how conflicts arise when regulatory policymaking is based upon conventional risk assessment. Indeed, if the type of caution underpinning scientific knowledge corresponded with the policy objectives,

then such conflicts would not arise. Furthermore, where regulations are based on the precautionary principle, the rationale for allocating the burden of proof according to the criteria of the potential injurer and the least-cost bearer should receive two additional considerations. First, the burden of proof is actually a burden of uncertainty, determining which party (consumers or producers) is the potential victim and due this status should benefit from the doubt that specific events could take place. Second, coming as a consequence of the first, the type of proof required from scientific assessment for discharging this particular burden is not neutral in the finalization of precautionary policies.

## II. Science and policy: The precautionary principle in between?

Ever since the first time it was publicly acknowledged, the precautionary principle has been provoking heated debate. The debate over managing the risk of possible future harm by taking precautionary measures has been described in many ways, such as the opposition between rational economic principles and ethical – sometimes even irrational – value judgments; or between two different perceptions of managing risk in everyday life – i.e. being risk averse and taking precautions or being a risk-lover or fatalist; or going further, between risk as science and risk as perception. Moreover, the relevance of the precautionary principle as a guiding rule for environmental policies is still contested today through the allegation that this role is already performed by scientific knowledge. This general contention assumes that the precautionary principle pertains to a realm totally opposite to that of scientific/rational principles. Indeed it is precisely upon this view that much of the debate about the precautionary principle has built up around the opposition between science and policy.<sup>3</sup>

However, the allegation that actions taken on the basis of precautionary considerations are at the opposite extreme to measures grounded in scientific

3 The points of opposition raised by those who contest the usefulness of the precautionary principle can be summarised as follows: current regulatory provisions are inherently precautionary; 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 (S. Holm and J. Harris, "Precautionary principle stifles discovery", 400 *Nature* (1999), pp. 398 *et seq.*).

knowledge reveals not only an anachronistic perception of science in its “normal” connotation,<sup>4</sup> but also, more importantly, a confusion between the two forms of science, namely innovation and regulatory science. The latter is politically driven and operates under a legal pressure of advising policymakers or instructing legal disputes;<sup>5</sup> whereas the former is dedicated to the advancement of the state of the art and is constantly submitted to extensive analysis and debate within the scientific community. Despite this distinction, it is self-evident that regulatory science has its origins in laboratory science, but it has the specific characteristic of being forced to trespass on its own limits of knowledge due to policy constraints.<sup>6</sup> The problem here is not so much that scientific information is biased, but that its ambition needs to be adjusted to the new global challenges. Funtowicz and Ravetz<sup>7</sup> advanced the concept of “post-normal” science precisely referring to the inherent complexity of scientific disciplines and to the problems that this status creates in the instruction of policymaking. If we concede that uncertainty touches upon a large proportion of environmental science – not only due to a lack of data, but also to the inherent complexity of ecosystems and the non-linearity and impracticality of the experimental replication of environmental testing; and if we conceive that this uncertainty, rather than being a “temporally misalignment of theory

and observation”<sup>8</sup> is the product of indeterminacy as opposed to determinacy, then the value of waiting changes. In fact, instead of leading necessarily to greater certainty, progressive environmental studies can even enlarge the areas of disagreement, as is to be expected from non-linear systems.<sup>9</sup>

While the incontestable value of these contributions to the reconsideration of “normal” science cannot be denied, I believe that they have nevertheless failed to stress the fact that the process of “adjusting” scientific knowledge to address contemporary challenges starts off by re-evaluating the method of its construction. In this scenario, the assumptions from which scientific regulatory analysis proceeds deserve greater appreciation and more discussion since, as will be explained later, they are not unconnected to the achievement of certain policy outcomes.

### III. The relation between risk assessment and scientific evidence

The section above presented a brief overview of the complexity of science and cast some doubt about its role as neutral arbiter between contradictory positions, whether for international legal disputes or in the policy arena where the regulatory activity of national governments requires rational instruction.

This investigation was triggered, as already mentioned in the introduction, by the analysis of the *EC-Biotech* dispute, the main contention of which pivoted around the determination of the legal status of precautionary actions with respect to science-based actions. The United States, Canada and Argentina accused the European Union of having imposed a *de facto moratorium* on GMO commercialization since 1998. Taking the opposite side, the EU invoked the precautionary principle as a reason for prolonging the approval procedures of certain GMOs under the condition of scientific uncertainty. However, there were no cases where the WTO Panel had concluded that scientific evidence was insufficient to perform an adequate risk assessment. Not only that, risk assessments were already available that had found in favour of the commercialization of certain GMOs.

As the Panel’s proceedings were keen to determine whether risk assessments had either already been performed or would be achievable in order to support the approval or rejection of GMOs, it is interesting to investigate what constitutes the basis of a risk assessment and how such assessments are intended

4 In the age of post-positivism, Thomas Kuhn became one of the leaders of a critical sentiment toward the untouchable notion of science as inherently capable of solving the world’s problems. “Normal” was the science that prevailed over competing theories, hence setting the scientific paradigm, but was not completely successful in solving problems. Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 3rd ed. (Chicago: University Of Chicago Press 1996).

5 See Sheila S. Jasanoff, “Contested Boundaries in Policy-Relevant Science”, 17 *Social Studies of Science* (1987), pp. 195 et sqq.; Sheila Jasanoff, *Designs on Nature. Science and democracy in Europe and the United States* (Princeton and Oxford: Princeton University Press 2005); Philippe Roqueplo, “Entre savoir et décision, l’expertise scientifique”, in *Sciences en questions* (Paris: INRA 1996); Pierre-Benoit Joly and Marc Barbier, “Séparation de l’évaluation et de la gestion des risques. Les leçons de la ‘guerre du bœuf’”, *Actes du Colloque “L’organisation du recours à l’expertise scientifique en situation d’incertitude”* (2002), pp. 1 et sqq.

6 Roqueplo, “Entre savoir et décision”, *supra* note 5, at p. 3.

7 Silvio O. Funtowicz and Jerome R. Ravetz, “The worth of a songbird: Ecological economics as a post-normal science”, 10 *Ecological Economics* (1994), pp. 197 et sqq.

8 Katherine Barrett and Carolyn Raffensperger, “Precautionary Science”, p. 5, presentation held at the Wingspread Conference on Implementing the Precautionary Principle, 26 January 1998.

9 See section IV.3 on “Uncertainty, The Information Paradigm And Risk Acceptability” about the difference between information and knowledge, the former being progressive for a matter of accumulation, the latter indicating closed information.

to instruct certain policy decisions, such as those on human and environmental safety.

## 1. Statistics applied to the environment

It is well understood that the assessment of risk is largely based on statistical hypothesis testing.<sup>10</sup> The type of information, or evidence, resulting from this procedure is anything but unequivocal; it is open to interpretation and evaluation and, most importantly, is biased due to the precise selection of the research questions and initial assumptions.

It is with regard to shared knowledge that statistical testing consists of two components, the null hypothesis (which is by default a no-effect hypothesis) and the alternative hypothesis. The aim is to prove the existence of an association between the experimental data and the null hypothesis by contradiction, i.e. by rejecting the null hypothesis.<sup>11</sup> Still, it is assumed that estimations are subject to random errors due to the size and composition of the sample. To give a measure of the reliability of the association, an error rate is 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). For this reason, hypothesis testing is constructed in such a way as to minimise, respectively, the probability of rejecting the null hypothesis when it is true, or that of not rejecting the null hypothesis when it is false. Their respective probabilities are denoted by  $\alpha$  and  $\beta$ , which cannot be simultaneously minimised because they are inversely related. Therefore a choice between which error should be minimised – and conversely which type of error should be maximised – has to be made.

### a. For more precautionary statistics

Criticisms are increasingly arguing against the standard statistical methodology for the type of error that is conventionally minimised.

Hypothesis testing has been accused of favouring less precautionary policy actions because the targeted error is Type I;<sup>12</sup>  $\alpha$  in this case denotes the rejection region, which is the region that contains the values of the statistic that contradict the null hypothesis or in which it is very unlikely to support the null hypothesis. Most importantly, the rejection

region determines the significance level of the test, for it places constraints on the size of the interval in which the results of the repeated tests may fall. The lower the significance level of the test and the rejection region, the greater the level of precision imposed and the lower the probability of Type I error, i.e. the probability of wrongly rejecting the null hypothesis. When this principle is coupled with a no-effect null hypothesis – such as “ $x$  does not entail  $y$ ” – it becomes evident that the decision to minimise Type I error is biased by a fundamental asymmetry: the statistical test is set up to be more cautious about the risk of detecting something which in truth does not exist (Type I or  $\alpha$  error), than about the risk of failing to discover something which in reality does exist (Type II or  $\beta$  error).

To appreciate this in environmental terms, we can consider a null hypothesis such as, for instance, “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 one does exist. In other words, the actual statistics strictly requires a high degree of certainty of harm before any preventive actions are suggested, whereas a precautionary approach calls for action even though scientific certainty has not yet been achieved.

In the view of some scholars, the way to incorporate the precautionary approach into statistical methodologies is by minimizing Type II error,<sup>13</sup> 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

10 See David R. Anderson, Kenneth P. Burnham and William L. Thompson, “Null hypothesis testing: Problems, prevalence, and an alternative”, 64 *Journal of Wildlife Management* (2000), pp. 912 et seq.; David F. Parkhurst, “Statistical Significance Tests: Equivalence and Reverse Tests Should Reduce Misinterpretation”, 51 *Bio-science* (2001), pp. 1051 et seq.

11 For the sake of clarity, it should be specified that this action is not intended either to prove causation between events, nor to give a direct measure of, i.e. to quantify, the probability that the null hypothesis is true (David E. Adelman, “Scientific activism and restraint: The interplay of statistics, judgment, and procedure in environmental law”, 79 *Notre Dame Law Review* (2004), pp. 497 et seq.).

12 This criticism is clearly expounded in Kristin Shrader-Frechette and John Lemons, “Methodological Rules for Four Classes of Scientific Uncertainty”, in John Lemons (ed.), *Scientific Uncertainty and Environmental Problem Solving* (USA: Blackwell Science 1996); cf. also David Kriebel et al., “The Precautionary Principle in Environmental Science”, 109 *Environmental Health Perspectives* (2001), pp. 871 et seq.

13 Shrader-Frechette, “Four Classes of Scientific Uncertainty”, *supra* note 12, at p. 6; Barrett and Raffensperger, “Precautionary Science” *supra* note 8, at p. 4; Kriebel et al., “The Precautionary Principle”, *supra* note 12, at p. 6.

side of caution with respect to the environment and the health of the public.”<sup>14</sup>

#### b. A personal critique: Maintaining the rationale of falsification

The criticisms explored in the previous section are significant in that they focus attention on the problem of how evidence is constructed, namely by deciding the significance level  $\alpha$ , and how this can affect the results of studies aimed at detecting possible hazards. The fact that scientific information is biased by the way in which it is constructed may initially be regarded as reasonable due to the very empirical nature of environmental science, where the source of heuristics is given by repeated experiments and so induction is the only available methodology.<sup>15</sup> How-

ever, such appraisals do not take into account several problems arising from the decision to minimise Type II error, and in particular, they do not retain the rationale for which frequentist statistics was conceived.

Regarding the first point, several concerns have been raised about using power ( $\beta$ ) analysis in environmental management studies: important effects may not be detected because sample sizes are likely to be small; equally, in order to discern chronic (instead of acute) effects from low doses of GM food in toxicological studies, sample sizes should be large enough to contain and give reliable representation of natural variability, which is practically impossible.<sup>16</sup> On a more fundamental level, we know that either probability  $\alpha$  or  $\beta$  can be minimised, but not both, unless we are able to increase the sample size. Hence, if  $\beta$  were the probability to be minimised, then the significance level  $\alpha$  would rise and the threshold for rejecting the null hypothesis would be lowered (i.e. the “strictness” for accepting scientific evidence will be relaxed). This raises the problem of maintaining the rationale of statistical analysis. Indeed, it is important not only to understand what kind of information arises from hypothesis testing (more or less precautionary), but also how this information originates. Once these points have been clarified, it soon emerges that the choice of which type of error should be minimised is strictly and conceptually connected to the way the initial hypothesis is set (see later, in the section on the so called “precautionary hypothesis”).<sup>17</sup>

The rationale of frequentist statistics is built upon a fundamental feature of science, i.e. falsification, which requires us to know how to proceed by a process of elimination and to advance with caution, which is indeed a principle that is assimilated into scientific research. This feature is clearly dismissed when claiming to minimise Type II error instead of Type I. Indeed, increasing the probability of wrongly falsifying (Type I error) implies increasing the chance to say that something is “scientific” when it is not; instead, non-falsification proves nothing, and we should always exercise caution when discriminating scientific knowledge from something else. Hence, falsification carries a different and more important heuristic value than non-falsification.

Epistemology should help us to understand this latter point better. In *A treatise of human nature* (1739) David Hume was already questioning the authoritative power of inductivism, since it was based simply on the assumption that observed events would

14 Kriebel *et al.*, “The Precautionary Principle”, at p. 875, *supra* note 12, at p. 6. Indeed, in epidemiology and biotechnology trials the language of “safety” assessment incorporates this kind of concern by requiring a sufficient statistical power or  $\beta$  probability (cf. EFSA, “Guidance document of the scientific panel on genetically modified organisms for the risk assessment of genetically modified plants and derived food and feed”, 99 *EFSA Journal* (2006), pp. 1 *et seq.*).

15 Environmental science is considered a “soft” science as opposed to other “hard” sciences such as physics and chemistry. The two categories differ in that the latter have more predictive power than the former (Jordan and Miller, 1996). This is fundamentally due to the fact that inductivism is the methodology affiliated to experimentation. 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 conclusions (be they true and/or false) is possible for the validity of the argumentation to test, which is why inductive arguments have different degrees of strength (Merrilee H. Salmon, John Earman, Clark Glymour and James Lennox, *Introduction to the Philosophy of Science*, 1<sup>st</sup> ed. (Indianapolis and Cambridge: Hackett Pub Co Inc 1999)).

16 To the same extent, methodological problems concerning the size and composition of the sample and the time extension of the experiment have been specifically advanced for GMO safety studies. On this point, see Gilles-Eric Séralini, Dominique Cellier and Joël Spiroux de Vendômois, “New analysis of a rat feeding study with a genetically modified maize reveals signs of hepatorenal toxicity”, 52 *Archives of Environmental Contamination and Toxicology* (2007), pp. 596 *et seq.*

17 For the purposes of this paper, only critiques concerning the choice of the error to minimise have been reported. However, more technical critiques exist that focus on the intrinsic simplification of the sources of uncertainty against hypothesis testing, denouncing its uninformative structural character and highlighting two main points: the acknowledged sources of uncertainty in statistical testing mainly converge on sampling variability through the standard p-value and confidence intervals; and the high chance of incurring systematic errors can originate from both selection bias and the actual measurement of the levels of specific factors according to the availability of detecting technologies. On these issues, see Anderson, Burnham and Thompson, “Null hypothesis testing”, *supra* note 10, at p. 5.

continue to follow the same pattern that had been previously observed.<sup>18</sup> According to him there was no logical 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 at first be deemed a causal relation could ultimately turn out to be a simple coincidence if ever an observation to the contrary occurred.<sup>19</sup>

Later on, the scientific method became highly prejudiced. Nevertheless, the question remained: how to set a logical basis for science? This was possible as long as the truth, to which science had traditionally always paid service, could be distinguished from knowledge or, put in other words, as long as the reason why a particular phenomenon occurs (cause and effect) could be separated from the reason for believing that such a phenomenon would ever occur.<sup>20</sup> Karl Popper believed that scientific statements could never be proven to be logically true since the only information to be inferred concerned the consistency (or lack of it) between an initial hypothesis or assumption and a specific observation.<sup>21</sup> Notably, he viewed scientific knowledge as resulting from the following fundamental asymmetry: if evidence of consistency between the result and the initial assumption cannot prove the latter to be true, inconsistency wields a different heuristic power in that it can disprove the initial hypothesis. Just one observation to the contrary therefore has the power to break the supposedly unequivocal deterministic chain.<sup>22</sup>

The great lesson from Popper, one that 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 in essence tentative.

## 2. From science, to individual choice, then to policymaking

The provisional nature of scientific explanations should not lead to hasty conclusions against the objective character of science and its capability to explain how the world works. In fact, even if temporary and incomplete, the knowledge we are endowed with still provides the basis for our actions, hence any delay to our advancement of knowledge should not be easily countenanced. As Rothman and Greenland warn, “the tentativeness of our knowledge [will] not prevent practical applications, but it should keep

us sceptical and critical.”<sup>23</sup> Those who might keep up their scepticism are not only scientists whose approach is indeed based on caution<sup>24</sup> but also, and especially, policymakers. Thus scientific uncertainty may be reconsidered as encouraging a dialogue that will gradually wear down opposition between policy and science to nothing. Moreover, the objective providing the reason for testing, such as environmental protection or human health safety, is to be considered against the usefulness of the statistical methodology. To the extent that the test may be applied to interactions as uncertain as biological ones, we cannot rule out that a situation where consistency between data and the hypothesis is small could be due to the fact that the object of prediction is a rare event, as is likely to be the case for long term, low risk or catastrophic environmental effects.

The theory of individual choice should illuminate this last point. Sunstein’s concept of the “availability heuristic” exemplifies how possible misperceptions occur: if we are confronted with two hypotheses, one very rare and the other very common, each of them having similar significant levels of association with the same event, we will still tend to attribute the cause of an event to the most common situation; hence, our instincts will be to maintain the “availability heuristic” hypothesis rather than

18 David Hume, *A treatise of human nature*, 2nd ed. (Oxford: Oxford University Press 1978).

19 These circumstances are better known as the “fallacy of affirming the consequent”, in which a 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 versa (Salmon et al., *Philosophy of Science*, supra note 15, at p. 7).

20 Salmon et al., *Philosophy of Science*, supra note 15, at p. 7.

21 Karl Popper, “Probability Magic or Knowledge out of Ignorance”, 11 *Dialectica* (1957), pp. 354 et seq.; Karl Popper, *The Logic of Scientific Discovery* (New York: Basic Books 1959).

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 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 Kenneth J. Rothman and Sander Greenland, *Modern Epidemiology* (Philadelphia, U.S.: Lippincott Williams & Wilkins 1998), at p. 8.

24 Within scientific theories, there exists an established principle of due care which is the principle of parsimony (or *lex parsimoniae*), also known as the Ockham’s razor principle. According to which, 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”.

the most rare.<sup>25</sup> This conclusion is certainly neither new nor surprising,<sup>26</sup> and least of all unreasonable: why indeed should we be inclined, based on personal experience, to regard a remote assumption as more plausible than a common one? But this conclusion is interesting in that it creates a parallelism between the general theory of individual choice and scientific research. The latter proceeds by preponderance of evidence and parsimony among theories, firstly by selecting the most plausible hypothesis and then imposing a strong burden of proof on whoever attempts to refute it.<sup>27</sup> The same is true for individuals in terms of how they form their preferences: they select only part of the information available, usually, the part that is more common, so easier to process.<sup>28</sup>

However, if we were to ask the same question, not to individuals or scientists, but to policymakers: “why should we be inclined, based on experience, to regard a remote assumption as more plausible than a common one?”, then this would raise another preliminary question: “for what purposes should the two elements (of the hypothesis testing) be similar?”, or why should they not be different? In this case, while plausibility and rationality should not be dismissed, they should nonetheless be combined with desirability questions, set intentionally by public policymakers. This remark introduces the problem areas of the following section, namely the division, and possibly the incongruity, that exists between the world of science and the world of policy.

#### IV. Examining the allocation of the burden of uncertainty

From lectures on scientific epistemology it has emerged that falsification is at the core of scientific

information, hence its preservation lies in restricting as much as possible the probability of wrongly falsifying (i.e. false positive). This means that the aim should be to avoid producing scientific information that is not in any way scientific, at all, and the best way to do so is by minimizing the possibility of rejecting the null hypothesis when it is in fact 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 simply do not know. Because of this asymmetry between Type I and II errors, it is certainly clear that the probability  $\alpha$  is the most important one to minimise. The matter for discussion then is its complement, the null hypothesis, i.e. the assumption that should be falsified.

##### 1. Precautionary error or precautionary hypothesis?

We know from conventional statistics that the initial hypothesis is one of no-effect or no-difference between two variables being compared. This creates two problems. First is that in ecology and biology it is not particularly relevant to determine whether or not any difference has been detected, but rather more important to estimate the value of this difference. In fact, although a statistical test may be significant, it does not provide any information on the biological significance of its result. Secondly, as Parkhurst claims, “significance tests make some sense in situations for which there is good reason [...] to believe a null hypothesis, and we wish to place a strong burden of proof on those who attempt to refute that null hypothesis.”<sup>29</sup> This means that the object of our concern is not merely included through the type of error we decide to minimize, but is also included from the very set-up of our hypothesis.

To elucidate this point, we shall re-examine a previously cited example while keeping in mind the conclusions just reached regarding the rationale of hypothesis testing, thus: according to convention, we postulate a no-effect null hypothesis, such as “chemical  $x$  does not cause harm to human health (or to a specific environment).”<sup>30</sup> We already know that proceeding in this manner for environmental risk analysis has raised concerns over the anti-precautionary bias of hypothesis testing, the source of this bias being the preference for controlling Type I error.<sup>31</sup> Indeed, the decision to minimise the chance of detect-

25 Cass Sunstein, *Laws of Fear. Beyond the Precautionary Principle* (Cambridge: Cambridge University Press 2005).

26 For a discussion of how individuals construct their preferences, see Daniel Kahneman and Amos Tversky, *Choices, Values, and Frames*, 1<sup>st</sup> ed. (New York: Cambridge University Press 2000).

27 See *supra* note 24, at p. 10.

28 Kahneman and Tversky, *Choices, Values, and Frames*, *supra* note 26, at p. 10.

29 Parkhurst, “Statistical Significance Tests”, *supra* note 10, at p. 1051.

30 This formulation is simplified. Generally, a certain chemical substance is compared with another one already in usage, and it is assumed that there is no difference between the two in terms of some parameters. For instance, if the parameter is the toxicity of the chemical, the null hypothesis is that the chemical being tested is no more toxic than the conventional one.

31 Cf. section III.1.a “For more precautionary statistics”.

ing an environmentally dangerous effect, which in reality does not exist, does sound anti-precautionary.

Why then should statistical conventions be maintained? One common response is that the origin of this convention is our legal and cultural apprehension that a person is presumed innocent until proven guilty. Indeed, even those defendants who themselves claim to be guilty – as happened, for instance, during a period of terrorism in Italy during the 1970s – are to be given the possibility of a trial. Until some positive evidence has been delivered and a burden of proof has been discharged, nobody can be said to be guilty. In statistical terms, this entails that the null hypothesis should be a no-effect one. Then, given the effect to be controlled, the choice of Type I error is due to a reasonable concern, which in this case is convicting someone who is in fact a victim. This kind of error is so abhorred in our society that we can all agree that the concern about controlling Type I error is well-founded and therefore should be the one to minimise. Nevertheless, does this not precisely represent a type of precautionary approach towards what is considered to be a potential victim? I believe it is. In divergence from the generally held opinion, this attitude does not stem from the type of error we choose to minimise, but rather from the way the null hypothesis is constructed to integrate a specific concern.

Both the manner in which the conventional null hypothesis is constructed and the decision to minimise one specific error type form part of the same rule: first do no harm. Translated into the language of public policy, regulations serve the public purpose. Policymakers should then instruct scientific experts on which issues to focus according to policy objectives, to enable them to evaluate available scientific information upon instructed regulatory priorities and to see whether there exists compatibility between their initial assumptions and the objectives set by the policymakers themselves.

On one side, scientific attitude teaches 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 hand, the object of our concern is not shielded by the type of error we want to minimise, but, rather and beforehand, by the assumption we make about it. This is why uncertainty is firstly managed by structuring our assumptions and priorities, giving us the cognition of our preferences: we are already precautionary about a specific circumstance at the moment we

construct the initial hypothesis, so the decision about which error to minimise (Type I) follows naturally.

## 2. Setting priorities: Who is the potential victim?

Moving specifically to the issue of GMOs, we know from conventional statistics how the null hypothesis should be set and which error should be controlled: The initial assumption is that GMOs and their wild counterparts, non-GMOs, are “substantially equivalent”, hence the former does not entail greater risk than the latter.<sup>32</sup> Following the same rationale as that for not convicting a victim, we should ensure that we do not conclude that GMOs are unsafe (or at least that they exceed a certain risk standard) when in fact they are not. This means that the possibility of the underregulation of a specific GMO, when in fact it poses risks to human health (or the environment), raises less concern.

In this specific setting, something seems to have changed; which is precisely our concern. This is due to the fact that what had previously been identified as a potential victim, thereby raising feelings of protection, has now become a possible risk or hazard, which creates a completely different perception of what public policy should be about and who at this point could be regarded as the weakest party (i.e. a potential victim) to protect. This change occurs because the object of the analysis itself, along with the new distributional issues it raises, has been modified. Nonetheless, it is important to underline that designation of the potential victim is a matter of policy arbitrariness: this paper assumes that public policy should be devoted to the public good, which, in the light of distributional considerations and according to the author’s personal choice, specifically corresponds to protecting the weakest party, i.e. the party in the weakest position to bear the cost and burden of regulatory failures.<sup>33</sup> In the situation where the potential victim is designated according to other

32 The concept of substantial equivalence is the key for a comparative assessment, in which traditionally cultivated crops have gained a history of safe use upon which they provide the baseline for determining any substantial difference between the GMO and its wild counterpart (cf. EFSA, “Scientific panel on genetically modified organisms”, *supra* note 14, at p. 7).

33 This means that in a welfare function we suppose that the coefficient of the marginal utility of the potential victims (or weakest parties) is higher than any other coefficient of other categories.

contingent policy priorities, then a different initial hypothesis should be set, while the error to control for (Type I) should remain the same for completing the design of specific regulations.

### 3. Uncertainty, the information paradigm and risk acceptability

As demonstrated, the potential victim in this particular case has been changed. Furthermore, as there are potentially many victims, the probabilities of incurring the same risk are likely to be correlated among individuals, and finally, risk is largely replaced by uncertainty and possibly also by fundamental uncertainty.

These elements represent the core reasoning behind the precautionary principle, which advocates allocating the burden of proof onto the proponent of a supposedly dangerous activity. This principle of wisdom will in any case be balanced against its effectiveness, i.e. against the question we want the potential injurer to answer: are GMOs safe or are they unsafe? Which of the two should require the highest level of scientific certainty and be proved by falsification?

From epistemology we know that nothing can be proven in absolute terms, neither the complete safety nor the complete harmfulness of a product or activity, precisely because science proceeds by disproof or falsification,<sup>34</sup> or by preponderance of evidence. Secondly, and consequently, it is a matter of linguistic and epistemic correctness to specify that the burden of proof is always complemented by a burden of uncertainty. Indeed, a residual uncertainty corresponds to any given proof: if experimental science can never produce “certain evidence” – truth – but only “preponderance of evidence” – from which we derive our knowledge – then proving something always implies the risk of an error, i.e. the risk that the knowledge we have validated does not correspond to the truth. The gap between truth and knowledge is the same as the one between causation and association: experimental science, as already described, aims to prove association between events by testing whether experimental evidence rejects or does not reject the

initial hypotheses; from that we can infer causation, which nonetheless remains a matter of choice and not a necessary conclusion. Since the chance to make an error is very high, as many scientists confirm, we can qualify this risk as the uncertainty we are not able to grasp. Therefore, there always exists a residual uncertainty which, like the burden of proof, requires to be allocated. According to distributional concerns, policymakers should indicate the identity of the potential victim of a wrong scientific conclusion and consequent regulation, and allocate the burden of the possible error, i.e. the burden of uncertainty, to the non-victim party. This does not mean that the latter (i.e. the potential injurer) should be able to prove that there exists no uncertainty, but rather that he should bear the consequences of potential errors, i.e. bear the burden of uncertainty while leaving the benefit of the doubt to the potential victim. For this to happen, the type of proof to be provided by the potential injurer must be specified with respect to its residual component, which is the type of error the potential victim should be shielded from.

If we adjust our expectation according to the only kind of knowledge we can produce, i.e. partial and uncertain knowledge, then it is clear that the management of uncertainty begins from the very shaping of our assumptions. This means accepting the fact that uncertainty is pervasive and is inherent to scientific knowledge, not merely consisting of “a temporary and surmountable lack of data”<sup>35</sup> rendering any search for the optimal regulation a senseless operation. To better understand this point, allow me to specify uncertainty with respect to the information paradigm.

It is generally acknowledged that incomplete information is one of the major sources of suboptimal choice. Following this argument, we implicitly suppose that through its accumulation the information will become progressively complete. This assumption is not under question. However, information as such has no value if it is not productive of knowledge. Knowledge is a specific form of information that has been given a meaning by consensus. If information generates neither knowledge nor understanding about a particular phenomenon because of disagreement, contrasting preferences or ambiguity, then uncertainty cannot be assumed to decrease with time and even the most complete information would be of no use for determining the optimal regulation. Furthermore, if uncertainty is inherent to scientific knowledge and if, consequently, the point in time

34 Popper, *The Logic of Scientific Discovery*, *supra* note 21, at p. 9.

35 Barrett and Raffensperger, “Precautionary Science”, *supra* note 8, at p. 7.

where knowledge is perfect will never reach a cognitive status, we will never be able to determine *ex-post* whether or not a specific public policy has erred on the side of overregulation or underregulation. This is why the optimality criteria for decision-making under uncertainty must be balanced against bounded rationality, against intertemporal knowledge and against the contingency and urgency of the issue at stake. Knowing that we will inevitably err, we have to decide which risk of error we prefer. Hence, the question of optimality is better replaced by one of acceptability or preference for risk.

The revelation of preferences about alternative scenarios is fundamental to direct regulatory efforts. Notably, when regulatory decision-making takes advantage of scientific information, it is crucial that it builds tight and clear communication bridges with scientific experts, not only to state priorities clearly and to guide research, but also to understand the value and utility of the evidence produced. Particularly, with regard to information produced from risk assessment, it must be remembered that the way this information is constructed by statistical inference affects the effectiveness of the allocation of the burden of uncertainty.<sup>36</sup>

It is clear that the selection of the triggering element of the analysis requires decisions that cannot be purely scientific.<sup>37</sup> Hence, the way the initial hypothesis is set will be in line with our assumptions adjusted to our concerns. It is important to identify clearly what possible scenarios are concerned since this can improve communication between the scientific community and the policymaker.

In order to anticipate certain criticisms, I put forward the suggestion that this practice would not lead to an invasive intrusion of policy values into scientific/objective statements, because the control of plausible concerns would ultimately be determined by scientists through scientific methodologies. Indeed, the desire for improved communication in terms of clearer policy instructions comes directly from scientific experts, who often blame decision-makers for resigning from their public responsibilities when they do not know how scientific information should be used.<sup>38</sup> Moreover, the fear that such improved communication would account for science being biased ignores the fact that science is in any case biased. The positive effect of revealing preferences would increase the chance of posing the right question<sup>39</sup> and would effectively shift the burden of uncertainty onto the potential injurer, who would be required to

prove with a high degree of certainty that GMOs are safe, instead of benefiting from the doubt regarding their harmfulness.<sup>40</sup>

As Shapiro points out, even though public concerns that trigger risk analysis are certainly to be balanced against the unavoidable conditions of resource constraint and economic feasibility, they also entail non-economic considerations.<sup>41</sup> To begin with, as discussed by the author, in a situation of uncertainty – which is precisely the case for GMOs – the errors of overregulation and underregulation have different costs and different bearers, respectively industry and individuals. That said, his position is one of advocating fairness by setting the burden of uncertainty about potential risk on the industry rather than on individuals, since the former is the least cost bearer. Again, given any particular concern we should always keep in mind questions such as “who is the weakest party?”, “who is the potential victim?” and “who should be protected?”, given a particular concern. The rule for setting the burden of uncertainty derives directly from its complementary element, i.e. the identification of the potential injurer, so that the onus of proof is on the non-victim.

36 In fact, as explained in the section below, to shift the burden of proof is not just a matter of deciding who should prove or bear the risk of uncertainty about a possibly dangerous event, but it is also a matter of how the proof (i.e. preponderance of evidence) about the occurrence of that event is constructed to err on the side of more or less precaution.

37 Vern R. Walker, “The Myth of Science as a ‘Neutral Arbiter’ for Triggering Precautions”, 26 *Boston College International and Comparative Law Review* (2003), pp. 197 *et seq.*  
Walker’s statement was made in reference to lawmakers, to define the factual predicate for taking precaution. Nonetheless, scientists face the same dilemma when selecting the relevant element for conducting their studies.

38 Interview with Gérard Pascal, Chairman of the Scientific Steering Committee of the European Union, Paris, France (April 8, 2008). Gérard Pascal remembers just one instance, during the years of the mad cow crisis when he served at the European Commission, DG SANCO, as president of the scientific committee, where he experienced policy confrontation with the responsible Commissioner.

39 As some authors have warned, as well as Type I and Type II errors, there exists another type which demands a more fundamental question: what is the problem? Hence, Type III error accounts for the risk to produce an accurate answer for the wrong question (see Kriebel *et al.*, “The Precautionary Principle”, *supra* note 12, at p. 5; H. Sanderson and K. R. Solomon, “Precautionary Limits to Environmental Science and Risk Management: Three Types of Errors”, 2 *The Journal of Transdisciplinary Environmental Studies* (2003), pp. 1 *et seq.*).

40 Cf. with section below on the burden the potential injurer is supposed to discharge.

41 Sidney A Shapiro, “Keeping the Baby and Throwing Out the Bathwater: Justice Breyer’s Critique of Regulation”, 8 *Administrative Law Journal* (1995), pp. 721 *et seq.*

#### 4. Setting priorities: Which question should the potential injurer answer?

So far, we have identified GMOs as the object of the study, we know the identity of the potential victim and we know the least-cost bearer that should carry the burden of proof. However we still do not know what we have to prove: either the existence of some harm or the existence of some degree of safety. On this matter, it is crucial to consider the subject conducting the analysis, since, while maintaining the rationale of falsifying the initial (null?) hypothesis, possible bias will arise regarding the setting of that hypothesis and on the choice to minimise one specific type of error. Indeed, it is obvious that a study on the same subject undertaken 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 cannot be contemplated without at the same time reconsidering statistical hypothesis testing. More precisely, this is not just a question of bias but also a question of providing an *incentive* that is consistent with both the personal interests of the agent (e.g. selling GMOs) and the need in public policy for honest scientific conclusions. Thus, a consistent incentive is one that leads the agent to channel all her efforts into finding out what she desires, that is to ‘falsify’ (make scientific, highly certain statements about) what it is that she does not desire, i.e. that GMOs are unsafe.

The twofold concern of bias and uncertainty 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 *EC-Biotech* dispute, the parties willing to commercialise GMOs in the European market are the ones that, according to the precautionary principle, should provide proof that no risk to human health arises from their action. However, in the Biotech dispute the complainants were of course serving their own interests, i.e. marketing GMOs, but the starting assumption that GMOs are safe<sup>42</sup>

would clearly bias their research. To make the point clearer, rejection of the null hypothesis would mean that experiments had provided strong evidence that undesirable effects caused by GMOs actually exist. In order to arrive at such a preponderance of evidence, a high degree of certainty – and effort – is required by setting a very low (5% or 1%) significance level  $\alpha$ . 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 admissible and the potential danger cannot be excluded. However, it is accepted 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.<sup>43</sup>

Let us now try to build a counter-example: the initial hypothesis is that “the GMO has an adverse effect  $y$  on human health (or on biodiversity)”. The GMO producer, who has the burden of proof, will devote every effort to rejecting the initial 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 to prove with a high level of certainty that the GMO is preponderantly safe, whereas the benefit of doubt about their harmfulness, which comes directly from the non-refutation of the initial hypothesis, will be left to consumers.

However, there is a practical problem of stating the hypothesis in the affirmative form. In fact, in this case the adverse effect under testing has to be specified. A no-effect hypothesis, in accordance with the principle of substantial equivalence,<sup>44</sup> 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 the specification of all types of expected effects, because if some effects are detected, their specification would be made at the time of observation. On the contrary, an effect-hypothesis requires the a priori indication of the specific effect of concern, i.e. what makes the difference.

Given that there is a huge amount of uncertainty about the possible effects of GMOs both on biodiversity and on human health, this situation is truly problematic. How can a compromise between the incentive effect for the GMO producer that we have just explored and the practicability thereof be found?

At this point, it is helpful to review alternative statistical testing. Proposals have been advanced to introduce equivalence testing as a substitute for<sup>45</sup> or

42 GMOs are presumed safe to the extent that they are substantially equivalent to their natural counterparts. See *supra* note 32, at p. 13.

43 Parkhurst, “Statistical Significance Tests”, *supra* note 10, at p. 5.

44 See *supra* note 42, at p. 18.

45 Graham B. McBride, “Equivalence tests can enhance environmental science and management”, 41 *Australian & New Zealand Journal of Statistics* (1999), pp. 19 *et seq.*

as an additional stage to<sup>46</sup> traditional hypothesis testing. Equivalence testing requires the initial hypothesis to incorporate the concern that two similar products may differ in some of their characteristics. This approach, for example, is used for drug testing. Filtered down into the GMO issue, it implies that instead of assuming that GMOs and non-GMOs are substantially equivalent, the two are presumed to be “bioequivalent”.<sup>47</sup> This is certainly the first step for incorporating a precautionary approach. Furthermore and strictly in relation to the problem of specifying the type of effect to be studied, zero-valued<sup>48</sup> point estimates are replaced by intervals determining the bioequivalence region. The interval hypothesis is divided into two one-sided sub-hypotheses and is set in such a manner that the difference between the parameters of the test GMO and the reference non-GMO is lower and greater than, respectively, the lower and upper bounds of the equivalence interval. The size of the equivalence interval is decided a priori, according to the contingent state of knowledge that indicates all the sets of differences which are of no practical relevance. The interval hypothesis is rejected when both sub-hypotheses are rejected, leading to the conclusion that the difference between the parameters of the test GMO and reference non-GMO falls within the equivalence region, hence that they are not relevant to the purpose of the test.

In this manner, equivalence testing makes it possible to incorporate the precautionary principle while maintaining the scientific rationale of falsifying the initial hypothesis and controlling for Type I error. On one side, refutation of the initial hypothesis now implies that high certainty is required to conclude that GMOs are as safe as their natural counterparts, whereas failing to reject the initial hypothesis now implies that the doubt<sup>49</sup> over the 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 error 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.

In conclusion, the relevance of equivalence testing has emerged gradually along with a train of thought that has taken into consideration the object of concern, the potential victim and the least-cost bearer bearing the onus of proof. The key message is that all these components have to be – and had indeed 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 devote all efforts to producing high-evidence results proving that GMOs are safe instead of benefiting from doubts about their harmfulness; the latter, conversely, will benefit from precautionary science.

## V. Conclusions

When Sir Austin Bradford Hill illustrated in his article the nine aspects to be considered in order to deduce likely causation from association,<sup>50</sup> his prescriptions were welcomed as rules for cause-effect decision-making. As some scientists have pointed out, more important lessons have unfortunately been missed.<sup>51</sup>

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. As he contended, the estimation of the value of waiting against the value of taking action “will depend upon circumstances” and upon the information we have at our disposal (“the whole chain [of cause-effect] may have to be unraveled 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.”<sup>52</sup>

46 Parkhurst, “Statistical Significance Tests”, *supra* note 10, at p. 5.

47 McBride, “Equivalence tests”, at p. 3, *supra* note 45.

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

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

50 Sir Austin Bradford Hill, “The Environment and Disease: Association or Causation?”, 58 *Proceedings Royal Society of Medicine* (1965), pp. 295 *et seq.*, available on the Internet at <<http://www.edwardtufte.com/tufte/hill>>.

51 Carl Phillips and Karen Goodman, “The missed lessons of Sir Austin Bradford Hill”, 1 *Epidemiologic Perspectives & Innovations* (2004), pp. 3 *et seq.*

52 In this statement, Hill (see *supra* note 50, at p. 21) 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.”

These claims belonged to a scientist, whose attitude towards precaution was already structured within his own profession that instructed him in the first place to do no harm. The same rule is not necessarily relevant to policymaking. Indeed, public choice theory would definitely support a different view of governmental attitude characterised by rent-seeking and policy actions that deviate from public interest. However, this paper has espoused a theoretical approach for which, since the precautionary principle is the pivot of the discussion, the only plausible assumption can be that decision-makers are civil servants working for the public good. In this regard, the approval of GMOs falls within the government's responsibility to protect consumers. Why protect consumers and not industry? Because in this specific case the former represent the "weakest party" due to asymmetric information, and on the basis of this status they should be protected by national governments.

To accomplish their public goals, governments take advantage of many types of expertise, in this case it is scientific expertise. This enables governments not only to endorse "better" decisions, but also to reinforce their legitimacy. However, this paper has tried to demonstrate that incongruence may arise between policy goals and the scientific production of knowledge. Indeed, there are cases in which distributional concerns integrated into regulatory actions may not find corresponding concerns about the way in which scientific information is produced. To explain this point, risk assessment and statistical hypothesis testing have been taken as exemplary to describe the way science proceeds, i.e. through a process of conjecture and refutation,

and the manner of interpreting the results of this process to instruct policymaking. It has emerged that conventional hypothesis testing, which is typically required before a certain GMO is approved for marketing within European boundaries, produces a kind of information that does not correspond to the rationale for which it has been required, which is to take precautionary actions to protect the public. This is due to the fact that the choice over distributional issues, such as deciding which party should carry the burden of proof, is only one part of the precautionary story. Indeed, knowing that scientific knowledge is partial, contingent and constantly submitted to revision, policy mandates should enlighten scientific experts about the type of concern that justifies their interventions. The definition of what type of evidence should be provided and hence which kind of residual uncertainty the burdened party should endorse is the missing point that renders shifting the burden of proof ineffective in terms of precautionary actions.

Acknowledging that uncertainty is a pervasive and inherent condition of scientific knowledge does not make science less useful or important for decision-making; neither does the acknowledgment that (empirical) science is biased and open to interpretation because it is based on induction. Spelling out these conditions is instead fundamental for adjusting our expectations to the heuristic power of scientific knowledge and for realizing that the first step to manage uncertainty is not to search for "facts" to make scientific knowledge "harder", but to shape our assumptions according to our preferences or priorities about what type of error we wish to avoid.