van Zwanenberg *et al.*
Environmental Sciences Europe (2025) 37:138
https://doi.org/10.1186/s12302-025-01176-9

Environmental Sciences Europe

REVIEW

Open Access

Asymmetric evaluations of scientific evidence indicating harm compared to evidence indicating an absence of harm in regulatory appraisals

Patrick van Zwanenberg[1,2], Erik Millstone[3*] and Alice Livingston Ortolani[3]

Abstract

This paper asks whether, when assessing the safety of regulated products, the standards of scrutiny and evaluation deployed by regulatory officials and scientific advisors differ for evidence indicating that a product might be harmful compared to evidence indicating an absence of harm. Four cases from the field of food chemical regulation are analysed for which safety appraisals were conducted by European and US regulatory institutions between the late 1980s and the 2010s. The cases concern selected areas of the possible toxicity of ethylene bisdithiocarbamate fungicides, a genetically modified variety of Bt maize, the artificial sweetener Aspartame, and the herbicide Glyphosate. We find that evidence that those products were unlikely to be harmful was routinely accepted by regulatory bodies as reliable, relevant, and sufficient to support judgements of safety, even when that evidence was incomplete, equivocal or the underlying studies were inadequate or flawed or both. By contrast, evidence indicating possible or actual hazards and risks was subjected to far more critical scrutiny to try to discern any possible grounds for discounting it, including reasons that were deemed not to be a problem when they characterised evidence indicative of a lack of harm, or when those reasons were entirely speculative or were contradicted by available evidence. We identify and characterise several different types of evaluative asymmetry and argue that all are antithetical to the effective protection of public and environmental health. Several also violate indispensable scientific requirements for making valid inferences and reaching well-founded conclusions; that is, they are scientifically defective. Their deployment misleads many policy decision makers and most of the public. Their effect is to conceal the scope for diminishing possible harm. We outline hypotheses as to why asymmetric patterns of scrutiny and evaluation appear to be a relatively widespread phenomena across different regulatory jurisdictions and time periods.

Keywords Regulatory policy-making, Scientific evidence, Asymmetric evaluations

Background

When bovine spongiform encephalopathy (BSE), commonly known as 'mad cow disease', was discovered in UK cattle herds in the mid-1980s, it was not obvious that the British government's assessment of the threat posed by the novel disease, or its response to it, would cause any more political problems than its actions in relation to a wide range of other high profile food safety and industrial risk issues during that era. There was

*Correspondence:
Erik Millstone
e.p.millstone@sussex.ac.uk
[1] Centro de Investigaciones para la Transformación, Escuela de Economía y Negocios, Universidad Nacional de San Martín, San Martin, Argentina
[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina
[3] SPRU - Science Policy Research Unit, University of Sussex, Brighton, UK

nothing conspicuously distinctive about the government's approach to BSE.

Nevertheless, that approach went on to trigger a severe science-based political crisis. In March 1996, a decade after BSE had first been identified, Ministers announced that a new, fatal brain disease in humans had been discovered and was very likely to have been caused by consumption of BSE-contaminated food. A handful of cases in young people had by then been identified, but the government's advisors made it clear that the number of future human deaths was at that stage totally unpredictable [1]. Whilst that announcement was always going to be a shock, the key reason why it triggered a political crisis was because medical and veterinary officials and Ministers had insisted repeatedly, throughout the previous decade, that BSE posed no threat to human health ([2], para. 1180).

A subsequent public inquiry identified many 'shortcomings' in the UK government's management of BSE, but it emphasised that a fundamental problem was the pervasive failure to acknowledge, articulate and respond adequately to scientific uncertainties ([2], paras. 1260–1301). When BSE was first discovered, agricultural officials, and subsequently an early advisory committee, sanctioned the optimistic hypothesis that BSE originated from, and would behave just like, a disease of sheep and goats called scrapie, which was widely assumed to be harmless to humans. That hypothesis was provisional, based on very limited, circumstantial evidence—or 'largely on guess work' as one of the government's expert advisors privately admitted to a colleague at the time ([3], para. 10.33). Yet, in early reports to Ministers and to the public, veterinary officials, in collaboration with expert advisors, represented that hypothesis if it were more robust than was in fact the case ([3], paras. 10.25–10.37 and 11.1–11.13). When, subsequently, emerging pieces of evidence progressively undermined that theory, a small circle of senior scientific advisors and officials failed explicitly to convey those facts to Ministers, the rest of government or to the public ([4], paras 4.498–4.759 and 5.287–5.294 and 5.344). The initial over-confident endorsement of the 'BSE is scrapie in cows' hypothesis, and its status for many years within most of the UK government as received wisdom, were the main reasons why the most important regulatory interventions to diminish human exposure to the BSE agent were not introduced until nearly 3 years after BSE was first discovered, and why few of those responsible for the implementation and enforcement of those regulations believed them to be necessary ([2], para. 1186).

A key characteristic of the UK government's institutional evaluation of BSE risk, albeit one under-acknowledged by the Inquiry, was its asymmetrical character. Therefore, whilst the hypothesis that BSE was 'scrapie in cows' and would behave just like scrapie was portrayed as if it were more robust than was in fact the case, alternative hypotheses were glossed over in early advisory committee reports to Ministers and subsequently ignored in official representations of BSE policy ([3], paras. 10.25–10.37, [5]) Those alternative hypotheses, namely, that BSE might not have derived from scrapie or that even if it had, it might not behave the same way once passaged into cattle, were also supported by circumstantial evidence, and were no less plausible given what (little) was known about the nature and behaviour of the infectious agent [5]. The handful of academic scientists who insisted on discussing those alternative hypotheses in public and their implications (which were that we had no idea of the potential risks to humans) were treated with hostility and contempt by ministers and officials ([6], para. 5.16). Veterinary officials often claimed that that there was no evidence or rational basis to suggest the possibility that BSE might pose a risk to humans, and that the academic critics were making scientifically spurious, politically motivated interventions ([4], paras. 4.559 and 4.522).

Evaluative asymmetries are frequently encountered

Our central argument in this paper is that, in the context of technology regulatory policy-making there was, and continues to be, nothing particularly unusual about that form of official evaluative asymmetry. Our research, and the research of others we discuss in this paper, indicate that evidence that a product is unlikely to be harmful is routinely accepted by official regulatory bodies in the UK and elsewhere as unproblematically reliable, relevant, and sufficient to support judgements of safety, even when that evidence is incomplete, equivocal or the underlying studies were inadequate or flawed or both. By contrast, evidence indicating possible or actual hazards and risks is frequently subjected to far more official critical scrutiny, to try to discern any possible grounds for discounting it, including reasons that were deemed not be a problem when they characterised evidence indicative of lack of harm, or when those reasons were entirely speculative or even when they were contradicted by available evidence.

Since this evaluative asymmetry occurs during the production of regulatory-scientific claims and advice, policy makers are often unaware that such choices have been made; or that those evaluative choices are being made for them by officials and scientific advisory bodies. We argue that this pattern of asymmetrical evaluation is sometimes scientifically defective and always antithetical to the effective protection of public and environmental health. It misleads many policy decision-makers and most of the public. It renders inconspicuous the scope for measures that could diminish possible harm and, more

generally, for exercising technological choices. It reflects and reproduces institutional biases in favour of whichever technological products industrial firms choose to commercialise.

In what follows, we illustrate our argument with four cases from the field of food chemical regulation for which safety appraisals were conducted by European and US regulatory institutions between the late 1980s and the 2010s. The cases concern selected aspects of the putative toxicity of (1) ethylene bisdithiocarbamate fungicides, (2) a variety of genetically modified Bt maize, (3) the artificial sweetener Aspartame and (4) the herbicide Glyphosate. The cases were selected in part, because we are familiar with them: the first two are based on research that we have previously undertaken on the ways in which evidence was evaluated during regulatory appraisals [7, 8], and the remaining two are based partly on work that others have undertaken on how evidence was evaluated. The section of the glyphosate case on European regulation is based on work that we undertook specifically for this paper, and has not previously been published, and therefore, the discussion of that case is longer than the other three. Our choice of case studies also reflected the fact that we wanted to include cases that occurred both before and after important changes to regulatory practice introduced in the wake of the BSE saga at the end of the 1990s and early 2000s.

Our discussions of cases 1, 2 and 4 differ in one key respect from the discussion of aspartame. The latter summarises the totality of the European Food Safety Authority's assessment of toxicological studies on aspartame in respect of a broad set of toxicological endpoints. In relation to cases 1, 2 and 4; however, the discussions focus more narrowly on selected sub-sets of toxicological endpoints. The discussions are not comprehensive, let alone exhaustive, but they may be sufficient to illustrate our key claim about pervasive evaluative asymmetries in official regulatory toxicological assessments. We are not claiming that those cases are entirely representative of the totality of regulatory topics or settings, but they do provide evidence that evaluative asymmetries are a quite widespread phenomenon across several jurisdictions and time periods. Our subsequent discussion distinguishes between different kinds of evaluative asymmetry evident from the four cases, and points to some possible reasons for those phenomena.

Value-laden judgements and regulatory-scientific appraisals

Contestation over the knowledge claims that regulatory institutions produce and rely on to inform policy decision-making, and to justify their decisions, has been an enduring feature of regulatory politics for many decades [5, 9, 10]. On controversial regulatory issues, academic scientists and non-government organisations sometimes criticise official appraisals and make competing claims about the nature of the potential threats posed by industrial products and processes. The orthodox regulatory response to such challenges is to insist that official regulatory-scientific appraisals are objective, value-free processes of assembling facts [10, 11]. It is not difficult to see why that response is expedient, because the implication is that dissenting accounts of industrial risks must be scientifically flawed or politically biased, but that is often wishful thinking.

An intellectually more honest perspective, widely accepted amongst science policy analysts, philosophers of science and sociologists of scientific knowledge, and by at least some regulatory policy-makers and expert advisors, is that entirely objective, value-free appraisals of technological and industrial risks are unattainable. On all regulatory issues, scientific officials and advisors must make or endorse a set of judgements that cannot be settled solely by reference to scientific evidence and logic [12–15]. Those judgements include choices about which categories of potential harm to address, what questions to ask of and within an appraisal, which hypotheses to formulate, what kinds of evidence to draw on, commission, or require, and how to analyse and interpret those empirical data.

One reason why evidence and logic are insufficient to settle those choices in regulatory contexts is because the kinds of evidence that are available, or practically feasible to obtain, are often subject to a range of chronic, often irreducible, uncertainties. For example, regulatory scientists are often unable to measure the actual endpoint of harm that is of regulatory interest and so must make assumptions about which kinds of secondary indicators may be useable as (measurable) surrogates, and then to choose amongst that range. They often need to choose how to extrapolate experimental data on the effects of high doses of a regulated substance on small groups of genetically homogenous rodents to estimate effects in large, diverse human populations exposed to lower doses. In addition, they often need to choose which potential exposure pathways to include within an appraisal and how best to model exposures. There are almost always competing ways in which those and many other kinds of choices can be made, but typically no unique epistemic reason why one should be preferred over another. Yet, how such choices are exercised often have very substantial implications for the ways in which possible threats are assessed and subsequently understood, with obvious consequences for the kinds of regulatory policy responses that are then considered, adopted and implemented. It is precisely because of those policy consequences that

judgements as to which choices are appropriate are value-laden [16].

In this paper our focus is concerned with that subset of value-laden choices within regulatory-scientific appraisals that are concerned with how empirical findings are evaluated. Whilst that focus circumscribes the main parameters of the following sections, there are at least three main sub-sets of often related choices that we will be highlighting. First, choices about the criteria that guide decisions about whether particular scientific studies, or types of studies, are considered reliable and relevant to a specific appraisal. Second, choices about methods and practices for interpreting and drawing inferences from empirical evidence. Third, choices about how much and what kinds of evidence are deemed necessary and/or sufficient to support scientific judgements and advisory conclusions.

Part of the academic literature that discusses the implications of recognising that evaluative (and other types of) choices within regulatory appraisals must be value-laden has focused on whether the choices made or endorsed by regulatory-scientific officials and advisors are appropriate, given legislative mandates to protect public and environmental health [17–19]. The concept of precaution is often central to those discussions. Precaution is frequently represented by policy-makers as a consideration that is relevant only after scientific appraisals have been completed, and only if the scientists report policy-relevant uncertainties [20, 21]. Those assumptions serve to diminish the meaning and scope of precaution and reduce its application to rare and marginal occasions. That perspective assumes that scientific appraisals are entirely objective activities that, were it not for any acknowledged uncertainties, would also deliver scientifically definitive and complete judgements. Scholarly understandings of precaution (including our own), by contrast, generally view it as a concept that, amongst other things, can and often should inform appraisals; risk assessments can be more or less precautionary. Explicitly adopting that interpretation of precaution contributes to making explicit the fact that a wide range of credible value-laden choices are often available when deciding which question(s) to ask, which evidence to include and which to omit, which knowledge claims to advance, and how to construct them. Consequently scholars who articulate that type of analysis advocate the use of transparent and accountable processes to choose and justify which choices are made [18, 19, 21, 22].

Other parts of the academic literature have been concerned more with identifying the political, institutional and cultural factors that influence why particular value-laden choices are made or reproduced during regulatory appraisals [23–27]. A large subset of this literature is concerned with how regulated industries routinely try to protect their commercial interests by shaping the evidence supplied to regulatory institutions, as well as the assumptions that frame the substance and conduct of regulatory appraisals, to minimise the likelihood that their products will face regulatory restrictions [24–33].

The empirical focus in this paper is not primarily on whether specific evaluative judgements are or are not appropriate to legislative mandates, or on why they have been deployed by officials and/or advisors, but rather on how and why they can and do differ when comparing evidence that ostensibly indicates a hazard or a risk as compared to evidence suggestive of the absence of such problems.

One reason why that focus is interesting and important is that it is very widely accepted, especially amongst professional academic scientists, that empirical evidence should be evaluated in ways that are consistent and uniformly sceptical, especially where the objective is to inform scientific judgements, for example, by checking carefully if mistakes might have been made. That standard is implicit in very widely held norms of professional science as required for rigour and well-grounded inferences and conclusions [34, 35]. A second reason is because issues of evaluative (a)symmetry have often been overlooked by many scholarly investigations of social and political influences on the production of regulatory knowledge claims, despite their profound significance.

A notable exception, highlighted by several authors, concerns asymmetrical efforts to identify and/or minimise potential false-positive errors (i.e., experimental data indicating that there is a significant effect or difference when, in reality, there is none), as compared to the efforts made to identify and/or minimise potential false-negative mistakes (i.e., suggesting that there is no significant effect when, in fact, there is one) [36–39]. Lemon, Shrader-Frechette and Cranor, for example, noted that the epistemic values that should guide academic research in a range of disciplines require strenuous efforts to minimise false positives to try and avoid adding mistakenly to the stock of scientific knowledge. Conversely, however, there is often relatively less concern, at least in some scientific fields, to minimise false negatives That asymmetry is apparent, for instance, in statistical conventions. Experimental studies are typically designed and interpreted in ways that require that there should be no more than a 5% chance that a reported effect is a random statistical artefact. However, researchers often tolerate anywhere between a 5% and a 20% chance that a result indicating no effect is a statistical artefact, with 20% representing the standard value, though they are sometimes much higher [17, 38, 39]. In other words, statistical

conventions in academic science tolerate four times as many false-negative errors as false-positive errors.

Lemon, Shrader-Frechette and Cranor [36] pointed out that, in relation to such statistical errors, it is often not possible to keep the odds of accepting both false positives and false negatives equally low without very large (and prohibitely expensive and impractical) sample sizes. Instead there is typically a trade off between efforts to minimise false-positive and false-negative mistakes. They argue that the extent to which efforts to diminish the chance of false positives ought be greater than or less than efforts to minimise the odds of false negatives (as well as the absolute levels of evidential proof or persuasion required in each case) are important value judgements, and ought to vary depending on context. For example, those value-laden judgements will have different implications in regulatory arenas, where false-negative errors have important consequences for human and/or environmental health.

This particular evaluative asymmetry (greater efforts to identify and/or minimise false positives than to identify and/or diminish false negatives) is quite widely acknowledged, at least amongst analysts of science and regulation, and it is one of the kinds of asymmetry that we identify in what follows; we will refer to it as an example of 'inferential asymmetry'. However, it is not the only form of evaluative asymmetry between evidence suggestive of harm and evidence suggestive of safety that we will identify; the others that we shall identify are far less widely appreciated.

The empirical discussion that follows examines examples drawn from the field of food chemical regulatory toxicology. It is, therefore, helpful, at this stage, to clarify a slightly curious feature of the linguistic conventions that toxicologists have adopted. When toxicologists talk about 'negative' studies or evidence, they mean studies that did not provide evidence of any adverse toxicological effects. On the other hand, when they refer to a study or its results as 'positive', they mean one that has provided at least *prima facie* evidence of one or more adverse toxicological effects. Using that toxicological idiomatic convention, we will illustrate several different ways in which 'positive' and 'negative' experimental evidence can be and have been evaluated asymmetrically.

Ethylene bisdithiocarbamate fungicides

Our first case concerns a British review, conducted in the late 1980s, of the potential risks posed to food consumers by a group of fungicides called the ethylene bisdithiocarbamates, or EBDCs, which had been in widespread use for several decades as agricultural fungicides, for example, to control fungal growth in harvested and stored potatoes. It was prompted by reports that US regulatory authorities were intending to cancel most uses of the EBDCs on food crops because of concerns about carcinogenic risks, particularly from a metabolite, and degradation product common to all the EBDCs called ethylene thiourea, or ETU.

UK regulatory practice in the 1980s was highly opaque and very little information was publicly available about the ways in which evidence had been selected and evaluated by British regulatory officials and advisors. The case was analysed within a PhD project, in which both published and unpublished scientific evidence about the EBDCs, and information about UK regulatory practice, was possible to obtain or infer using information publicly available in the USA, under that jurisdiction's freedom of information legislation [7, 40].

At the time of the British assessment, evidence of carcinogenicity from five long-term rodent feeding studies was available on ETU. Two mouse studies on ETU had both reported increases in liver tumours in exposed mice and three rat studies on ETU all reported increases in thyroid tumours [41–44]. Seventeen long-term feeding studies, mostly industry-commissioned assays, were also available on the individual EBDCs [45]. Of these, seven had reported increases in tumours in exposed animals, whilst the remaining ten reported no carcinogenic effects.

In a brief published appraisal, the UK's Advisory Committee on Pesticides (ACP) argued that the positive carcinogenicity evidence on ETU was not relevant to assessing human risk to consumers. It suggested that the mouse liver tumours were likely to have occurred either by chance, or due to secondary mechanisms that either would not occur in humans or not at the doses humans were typically exposed to [46]. It also argued the rat thyroid tumours were caused by a mechanism that would not occur in humans. All seven positive carcinogenicity studies on the individual EBDCs were omitted, without explanation, from the ACP's review (as were three of the negative studies). The seven negative studies on the EBDCs included in the UK's appraisal were reported without any comment, as if they had provided reliable evidence of the absence of carcinogenic effects. The ACP concluded that "…there was no evidence of a risk of cancer or other adverse effects to consumers arising from the use of ethylene bisdithiocarbamates…" ([46], p. 76). After the Ministry of Agriculture Fisheries and Food received the ACP's report, it made no changes to the regulatory status of those fungicides.

The fact that all the ostensibly positive carcinogenicity studies were subject to critical dismissive scrutiny, and the results deemed to be chance occurrences or irrelevant to human risks (or were omitted), whilst all seven nominally negative studies were reported without

comment, suggests a *prima facie* case of evaluative asymmetry. Nonetheless, it might have been the case, that the positive findings were, in fact, either very likely to be chance occurrences, or plausibly caused by mechanisms that were not relevant to humans or not at dose levels that humans were typically exposed to, whilst the negative studies were all robust, just that the ACP omitted to explain that was so. It is worthwhile, therefore, examining briefly the arguments used by the ACP, and some of the studies.

In respect of the two positive mouse feeding studies on ETU, one study had been conducted by the US National Institutes for Health in the late 1960s [41], and the other by the US National Toxicology Programme in the 1980s [44]. Both had reported statistically significant increases in liver tumours in mice exposed to ETU. One of the ACP's arguments to challenge the relevance of those findings for humans was to claim that "most" chemicals that are non-genotoxic (as the ACP argued was the case for ETU) and that cause increases in liver tumours also cause "…hepatocellular necrosis [cell death in the liver] or a sub cellular change such as peroxisome proliferation." ([46], p. 73). Both hepatocellular necrosis and peroxisome proliferation are effects that can themselves cause cancer, but peroxisome proliferation does not occur in humans, and hepatocellular necrosis occurs only after continuous and high levels of exposure to a chemical.

No evidence was provided by the ACP that ETU might have caused hepatocellular necrosis or peroxisome proliferation in exposed mice; rather the Committee just asserted that 'most' non-genotoxic chemicals that are liver carcinogens do so. Interestingly, US regulatory archives show that Rohm and Haas, one of the manufacturers of the EBDCs, had conducted several research studies in the 1980s that had been designed to provide an understanding of the mechanism of ETU-induced mouse liver carcinogenesis. One of those (unpublished) projects had investigated ETU's effects on peroxisomal proliferation but, as US regulators noted, the project had found no evidence of such an effect [47]. Furthermore, a review of one of the two mouse studies, conducted by US regulators, pointed out that "…cellular necrosis was not observed in any group…" ([48], p. II-39). In other words, the ACP had outlined speculative reasons to dismiss the relevance to humans of positive evidence of carcinogenicity in mice, but not only was there no evidence to suggest that ETU might have caused hepatocellular necrosis or peroxisome proliferation, there was actually evidence indicating that ETU did not cause those effects.

A second reason the ACP gave for discounting the relevance of the mouse liver tumours to human risk, specifically in relation to the 1969 National Institutes for Health study, was because "…these [liver] tumours occur frequently in mice and the increased incidence is unlikely to be compound related." ([46], p. 19). In all long-term animal studies, there are often spontaneous tumours that are not produced by the test agent, and these can be highly variable among control groups of the same animal species and strain in different studies. Regulatory authorities sometimes recommend that historical control data can be used to help assess whether concurrent controls constitute the typical species/strain pattern for background tumour rates and thus help in identifying possible false-positive or false-negative results [49]. The ACP did not refer to historical control data in its assessment, yet one of the two mouse strains used in the National Institutes for Health study—the $B6C3F_1$ mouse—has been used routinely in studies performed by the US National Cancer Institute, and in nearly 200 studies the incidence of spontaneous liver tumours in control $B6C3F_1$ mice ranged from 18% to 47% in males, and 2.5–8% in females [50]. The incidence rate in ETU-treated $B6C3F_1$ mice in the National Institutes for Health study was 88% for males and 100% for females [41]. Those rates significantly exceeded the tumour incidence in comparable historical control groups, particularly for the female mice, as well as in the concurrent controls, where the incidence was 21% for male and 0% for female mice [41]. The ACP's claim that since liver tumours occur frequently in mice the increased incidence was 'unlikely to be compound related' was again inconsistent with the evidence then available.

It is illuminating to contrast the readiness of the ACP to dismiss positive findings in the National Institutes for Health mouse study on the grounds that they were likely to be chance occurrences with the committee's lack of criticism of a negative rat study on Zineb, one of the parent EBDCs. That study, completed in the early 1950s, was conducted on groups of ten rats, each of which were fed different doses of Zineb for 2 years [51]. Tumours were reported in the dosed groups and in one animal in the control group, but the difference in the incidence between dosed and control groups was not statistically significant. The study was reported by the ACP without comment, as if it were evidence of the absence of any carcinogenic effects. However, with only 10 rats in each group (instead of the more typical number, in more modern rodent studies, of about 50) the study was so underpowered that with one of the ten animals in the control group developing tumours, there would have had to be a minimum of six of the ten animals in the treated groups developing tumours before a statistically significant difference could be identified, as was pointed out in a U.S. regulatory evaluation of that study, which for that reason rejected it as unreliable ([48, 52], p II-11). Moreover, even if the true cancer rate in the treated groups were

60% compared with 10% in the control group, the number of animals in the study was so low that there was only a 66% chance of detecting an effect if it existed ([52], p. II-11). The ACP's failure to criticise a negative study that was quite likely to have been a false negative, if Zineb were a carcinogen, and its readiness to criticise and dismiss a positive study (on ETU) that, by contrast, was very *unlikely* to be a false positive is revealing.

In summary, distinct kinds of evaluative asymmetry were evident in those aspects of the EBDC case discussed here. First positive evidence was critically scrutinised for potential problems, whilst it appeared that little or no efforts were made to identify potential problems with negative evidence. Second, the critiques themselves were asymmetric in that the same kinds of issues that provoked scepticism about the reliability or relevance of positive evidence were also evident in negative evidence but those were seemingly ignored. Third, the evidential basis for those critiques was also asymmetric in that criticisms of positive studies either had no evidential backing, or evidence existed to the contrary, whilst substantive evidential reasons for doubting the reliability of negative evidence were ignored.

Genetically modified maize

Our second case concerns a dispute in the late 1990s and early 2000s about a genetically modified (GM) variety of maize that had been engineered to express the insecticidal bacterial toxin *Bacillus thuringiensis* (Bt), and specifically its possible adverse effects on what are called 'non-target' insects, i.e., insects that the Bt toxin is not intended to kill. Initial testing of GM Bt maize, both in the laboratory and under field conditions, had shown no adverse effects on non-target insect species and the maize had been approved for cultivation in Europe.

In the late 1990s, a group of academic agricultural ecologists examined the effects of Bt toxins on a beneficial predatory species, the green lacewing, which is often used in the biological control of pests and is typically present in maize fields. Although the green lacewing had been routinely used in regulatory experimental tests for non-target effects, those tests usually fed the lacewing larvae with moth eggs, the surface of which were coated with the Bt toxin. Hilbeck and colleagues argued that lacewing larvae usually pierce and suck out the contents of the eggs and so would be unlikely to ingest the experimental Bt toxin from the surface of the eggs. Instead, they either administered Bt toxin directly into the gut of lacewing larvae or they fed lacewing larvae with caterpillars that had first been fed with Bt maize leaves. In both circumstances statistically significant lethal effects were observed in lacewing larvae exposed to the Bt toxin, as compared to controls [53, 54].

The new evidence was reviewed by the EU's Scientific Committee on Plants, after the Austrian and German governments cited the Hilbeck et al. studies as a reason unilaterally to suspend authorisation of the Bt maize. It was also reviewed by the UK's Advisory Committee on Releases to the Environment (ACRE) and by the US Environmental Protection Agency (EPA). All three bodies argued that the studies were incomplete and questionable, especially given their 'unrealistic' experimental conditions, and did not, therefore, constitute new evidence of harm sufficient to alter the authorisations to cultivate Bt maize, or in the EU Committee's case, to justify the Austrian and German governments' suspension [55–57].

Both the UK and US advisors outlined speculative reasons as to why the harm observed in the laboratory studies might not materialise in farmers' fields. ACRE suggested that lacewing larvae would have a wider variety of plants to eat under field conditions (and thus be exposed to less GM maize in their diet), and they might be healthier than in a laboratory setting, and so more resistant to the Bt toxin [57]. ACRE also suggested that alternative explanations for the high level of mortality observed in the laboratory study were possible, for example, because of noxious substances produced in caterpillars as a result of exposure to Bt. ACRE also argued that better support for the direct toxicity of Bt toxin to lacewings could have been provided by dose/response data [57]. The EPA argued that the lacewing larvae in the laboratory study were not given a choice of what to eat, unlike in a field setting, and suggested that the larvae had consumed a suboptimal diet consisting of sick or dying prey which may have been septicemic (and, therefore, indirectly toxic), or of limited nutritional value, or unpalatable to the lacewings [58]. It also argued that the lethal effects were in any case so slight as to suggest no significant impact in field conditions [58]. Members of the EU's Scientific Committee on Plants and the EPA highlighted other methodological weaknesses, especially a relatively high mortality rate in control groups [56, 58].

Yet, as Levidow and Wynne pointed out [57, 59], other laboratory studies which reported no harm to non-target insects were also not representative of real field situations either, yet the committees were willing to infer reassurances of safety from those studies. Furthermore, some of those negative laboratory studies warranted the same criticisms, because, for example, control insects in some of those studies had even higher mortality rates than those in the Hilbeck et al. studies, yet their relevance was not challenged by official experts on either the EU or UK committees. Hilbeck, Meierand and Trtikova [60] also pointed out that the problem Hilbeck and other colleagues identified with standard lacewing laboratory exposure studies, namely, that lacewings are unlikely to

ingest Bt toxin coated on the outside of moth eggs, did not subsequently prompt critical scrutiny by the EU's regulatory agencies of the previous studies that had reported no harm, and which had been relied upon in part to approve cultivation of Bt maize. Those authors noted that inconsistent levels of critical scrutiny were acknowledged in an interview with a former European Food Safety Authority GMO panel member who stated: "Of course, studies that describe potential negative [i.e., adverse] environmental effects of GMOs are discussed particularly intensively." ([60], p. 3).

Similar criticisms were made of the US Environmental Protection Agency's (EPA) assessment of the Hilbeck et al. studies; in this exceptional case the criticisms came from the EPA's Federal Insecticide, Fungicide and Rodenticide Act Scientific Advisory Panel, which is responsible for providing oversight of the scientific quality of EPA decision-making on pesticide-related issues. In 2001 the Advisory Panel complained that:

> *"[t[he Hilbeck data was dismissed by the agency, based on standards that were not applied to all the work reviewed by the agency, and the Hilbeck work was singled out for an excessively critical analysis. Control mortalities were not unusually high (especially compared to control mortalities in other studies reviewed favorably by the Agency), dead insects were not fed to Chrysopa … the observed effect was not small (30% increase in total immature mortality), and the Agency should have concluded that a potential hazard to Chrysopa had been identified." ([61], p. 54).*

The key point, for our purposes, is that the Scientific Advisory Panel had judged EPA´s dismissal of the adverse effects identified by Hilbeck et al. to have been unreasonable, in part because (as in the UK and EU) studies reporting no harm were not subjected to the same type and level of scrutiny and criticism, even when they presented the same kinds of potential shortcomings.

As in the EBDC case, asymmetrical levels of critical scrutiny appeared to have been directed at positive and negative data on GM Bt maize non-target adverse effects. In addition, the critiques themselves were asymmetrical, in so far as some of the same issues and limitations that provoked criticism of positive evidence (such as mortality rates in controls) were evident in studies with negative findings too, but were seemingly ignored or discounted. Furthermore, the basis for those critiques also appeared asymmetrical insofar as a series of hypothetical reasons were offered as to why the lethal effects observed in lacewing larvae might not materialise in farmers' fields (and on that basis the results discounted), and yet actual evidence that lacewing larvae may not have been exposed to Bt toxin in the negative studies failed to change advisors' judgements that the studies' findings were reliable indicators of the absence of harmful effects.

In a commentary on the Bt maize case, Wynne argued that a somewhat different, albeit closely related, form of asymmetry characterised this case. Wynne noted that appraisals of this kind are almost always faced with a combination of limited evidence and the need to make extensive inferential judgements about possible risks [57]. He argued that the inferential license to make such judgements is often attributed scientific credibility when the direction of the conclusions heads towards permissiveness and a judgement of 'safe'; however, the inferential license informally tightens when the direction of the conclusions heads towards a more precautionary judgement of 'unsafe', with more direct and comprehensive evidence of harm required to justify a non-permissive conclusion. Wynne also pointed out that this kind of 'inferential asymmetry' operated outside of any normal mechanism of accountability, with the conclusions of the appraisals portrayed as objective science.

Thus, one reading of the Bt maize case is that expert advisors were not willing to infer from what was necessarily limited evidence of harm (i.e., the Hilbeck studies) to reach a conclusion of 'unsafe' to non-target biodiversity. Instead, they implied that more direct and conclusive evidence of harm (e.g., quantitative dose/response data and/or replication of the findings in more realistic field conditions) would have been needed before a judgement of 'unsafe' to non-target insects could be adopted. Conversely, however, the evidence of no harm was deemed adequate and sufficient to infer a conclusion of 'safe' to non-target insect species, even though it too was limited as it was based in part on laboratory studies, and there were uncertainties as to whether lacewing larvae actually ingested the Bt toxin.

This inferential asymmetry might appear to be just a restatement of the kinds of asymmetries that we have identified (as regards asymmetric levels of critical scrutiny, the critiques themselves, and the evidential basis for those critiques), but it is also concerned with asymmetric judgements about whether individual studies, or a body of evidence, are *sufficient* to support conclusions of 'safe' versus 'unsafe'. The asymmetries we identified earlier concern judgements about the *reliability* and *relevance* of individual studies indicative of harm versus those indicative of an absence of harm. There is of course considerable overlap, since judgements about the reliability and relevance of individual pieces of evidence will inevitably comprise part of the argument and justification for judgements about whether that collection of evidence is sufficient to support particular conclusions about possible hazards or risks. As we argue later, the distinction

between inferential asymmetries and asymmetrical judgements about the reliability and relevance of evidence is important, because the former can be a scientifically reasonable, albeit an anti-precautionary judgement, but the latter are both anti-precautionary and scientifically flawed forms of reasoning.

Aspartame

Our third case concerns an assessment, published in December 2013, of the toxicological risks posed by the artificial sweetener aspartame by the European Food Safety Authority (EFSA), specifically its Panel on Food Additives and Nutrient Sources added to Food (ANS) [62]. EFSA's assessment included far more studies and far more details about those studies, and about how they were interpreted and evaluated by EFSA, than had ever previously been published by any regulatory authority. The section of the report that focussed specifically on the putative toxicity of aspartame, as opposed to the sections that discussed an impurity called Diketopiperazine and its metabolites, reviewed a total of 154 separate studies.

Millstone & Dawson published a detailed and critical discussion of that assessment, focused on both cataloguing and critiquing asymmetries in the ways in which the ANS panel interpreted toxicologically positive and negative studies [8]. Their critique was lengthy and comprehensive, so in this context a summary discussion of their analysis might be sufficient. Table 1 reproduces Millstone & Dawson's quantitative summary analysis of the panel's interpretations of the individual studies.

Thus, whilst the ANS panel judged 24% (19 of 81) of studies that did not indicate adverse effects to be unreliable, it deemed 100% of the 73 studies that had provided evidence of adverse effects to have been unreliable.

Kass & Lodi responded on behalf of EFSA to that *prima facie* evidence of asymmetry; they provided a table with the same structure, as shown in Table 1, but with entirely different numbers [63]. Kass and Lodi claimed that the ANS panel had treated 35% (27 of 78) of negative studies as unreliable, whilst in marked contrast to Millstone and Dawson's figures, only 43% (16 of 37) of putatively positive studies had been treated by the panel as unreliable. Whilst Millstone & Dawson's 2019 paper had provided a detailed tabulation (with 224 rows and 6 columns) itemising all the studies cited in the ANS panel's review, and its interpretations of their findings on the putative toxicity of aspartame, Kass and Lodi failed to provide any corresponding details to support their numbers. Millstone & Dawson had also provided individual study-specific reasons for all of their categorisations. Kass and Lodi provided no list, no details and no study-specific reasoning. Millstone & Dawson responded by challenging EFSA's representatives to publish EFSA's list and to provide its individual study-specific reasons for their categorisations [64]. Since when (i.e., until 2025) no response had been forthcoming from EFSA.

Table 1 ANS panel's interpretation of the reliability of studies for those that had, and had not, indicated possible harm, by number of studies

	Number of studies reviewed	Number treated as reliable	Number treated as unreliable
Studies not indicating possible harm	81	62	19
Studies indicating possible harm	73	0	73

Table 1 shows that the ANS panel accepted 62 of the 81 ostensibly reassuring studies of aspartame and treated them as unproblematically reliable. Yet, several of those studies were based on very small sample sizes or were affected by confounding factors or had used methods that did not comply with Good Laboratory Practice. For example, study E4, a sub-chronic rat feeding study, used only 5 animals per dose group; E86 used only 5 dogs per dose group; Sasaki et al. used only 4 mice per group; E97 and E101 were reported by the ANS panel not to have met Good Laboratory Practice standards, whilst E51 was reported by the ANS panel to be confounded by poor health of the animals and the gavage technique ([8], p. 14). All were nevertheless treated as reliable negative results by the ANS panel.

The Panel's treatment of ostensibly worrying studies was strikingly different; each and every one of the 73 studies indicating possible harm was discounted as unreliable. Imperfections in those studies were treated as grounds for dismissing the results, even though the panel was indifferent to similar or more serious imperfections in negative studies. For example, adverse consequences of nitrosating aspartame reported by Meier et al and Shepard et al were discounted on the grounds that the conditions for nitrosation were 'harsh' ([8], p. 15). E87, which re-examined rat brain tissue from long-term studies E33–34 and E70, reported brain tumours but these were discounted by the panel, because they were random with respect to dose and gender. E20, an 8-week rat feeding study, reported a significantly higher liver to body weight ratio for the high dose group males but the panel dismissed that result as 'not unequivocal' ([8], p. 15).

Overall, the number of different factors that the ANS panel invoked as critical benchmarks that the putatively positive studies allegedly failed to satisfy, as well of the height of the hurdles that the studies failed to reach, was

remarkable, and were far more demanding than those applied to studies with seemingly reassuring results. If the benchmarks that the ANS panel invoked as grounds for discounting putative positive evidence of adverse effects from aspartame are aggregated together, they collectively imply that nothing could count for EFSA's ANS panel as a reliable positive study unless:

1) the results derived from a long-term study conducted with a large sample of people or large groups of laboratory animals;
2) those studies followed an orthodox protocol (which were implicitly assumed to be valid), but not a more sensitive one;
3) the magnitude of evidential differences between test and control groups satisfied the conventional benchmark of statistical significance, namely, that there was less than one chance in 20 of it having been a random fluctuation;
4) the results were entirely unequivocal;
5) the findings were consistent, e.g., across genders and studies and were monotonically dose-related;
6) those studies were entirely free of any imperfections; and
7) demonstrated causality.

A particularly controversial part of the panel's report discussed a study conducted by the Bologna-based Ramazzini Foundation, which is a non-commercial research institution. In 2005 a team of Ramazzini researchers published a paper that reported the results of one of their carcinogenicity studies on aspartame [65]. One striking feature of that study was that, rather than following an orthodox protocol and using 400 rats (50 males + 50 females at 3 dose levels + control groups), they endeavoured to improve on orthodox practice using 1,800 rats. Instead of testing aspartame at three dose levels plus controls, they tested it at six dose levels plus controls. Instead of killing the rats prematurely, before they reached the ends of their caged lives, the rats were allowed to live until their 'natural' deaths so that longer term effects could be studied. Keeping the animals until they die may not be common practice, but since European Union food safety policy legislation stipulates that "Assuring that the EU has the highest standards of food safety is a key policy priority…" [66] we might have expected that EFSA's benchmark would be the protection of all consumers throughout their entire lives, rather than, for example, only until they reach retirement age.

In those, and several other ways, the Ramazzini study was more thorough, sensitive, reliable and relevant to human exposure than those conducted in accordance with conventional protocols. The authors said their study: "…demonstrated for the first time that APM [aspartame] is a multipotent…carcinogenic agent…" with dose-related tumour increases in both males and females [65]. In 2010 Ramazzini also published the results of a study showing that aspartame induced tumours in the livers and lungs of male mice [67].

EFSA's ANS panel discounted those findings, and criticised the Ramazzini studies (as had other official advisory bodies in the USA, the UK and WHO previously) for reasons that were mainly provoked by the fact that those studies had not followed orthodox protocols. For example, one reason why the findings had been discounted was because the ANS panel claimed that increases in lymphomas and leukemias in treated rats were more likely to have been caused by respiratory disease than exposure to aspartame [62, 68]. There were relatively high rates of respiratory infections in the elderly rats (though those rates did not actually differ significantly between treated and control groups), but the incidence of chronic respiratory disease tends to be higher in animals that are allowed to live until their 'natural' death rather than being killed prematurely. The Ramazzini protocols were non-standard, but their unorthodox innovations provided greater sensitivity and specificity than could be obtained from an orthodox study. Those Ramazzini protocols can reasonably have been expected to have provided better models of the risks to public health than orthodox studies. The statistical power of the Ramazzini's studies was mirrored by the severity of the invective that the official bodies invoked in their dismissals of the Ramazzini findings.

An illuminating aspect of the evaluative asymmetry that characterised the ANS panel's assessment of aspartame is not just that the hurdles that positive studies were expected to reach were more demanding than those applied to negative studies. The Ramazzini studies were discounted by the ANS panel for reasons that were mainly linked to the fact that they were intended to be a more sensitive test than orthodox protocols stipulated. In contrast, as we noted earlier, several negative studies were problematic, because they were less sensitive than orthodox protocols stipulate—for example, because they had used very small numbers of test animals, and so would have had a relatively high chance of producing false-negative results—but those reasons were ignored by the ANS panel and the findings accepted as valid.

Glyphosate

Our final case concerns assessments of the herbicide glyphosate, conducted in the mid-2010s by the US Environmental Protection Agency's Office of Pesticide Programmes (OPP) and Germany's Federal Institute for Risk Assessment, or *Bundesinstitut für Risikobewertung* (BfR), which had assessed glyphosate on behalf of

all EU Member States. Specifically, we focus on those two regulatory institutions' evaluations of evidence on glyphosate's putative genotoxicity. This comprises studies designed to test a chemical's potential to cause genetic alterations, whether mutations (alterations of DNA) and/or DNA damage [69]. Although our focus is narrowly on Glyphosate's putative genotoxicity, this section is lengthy, because numerous studies were available and debates about their interpretation have been complex, lengthy and vigorously contested.

In 2015 the regulatory status of glyphosate became acutely controversial after the World Health Organisation's International Agency for Research on Cancer (IARC) completed a review of glyphosate and classified the herbicide as 'probably carcinogenic to humans' [70]. IARC is not a regulatory institution, but it is responsible for providing an evidence base for the cancer control policies of the World Health Organisation and its members. IARC's panel had concluded that there was 'limited' evidence of cancer in humans from epidemiological studies, 'sufficient' evidence of cancer in experimental animals, and 'strong evidence' of genotoxicity both for pure glyphosate and glyphosate-based formulations [70]. In complete contrast, both the BfR, in a 2013 Renewal Assessment Report [71], and in its 'final' Addendum to that report in March 2015 [72], and in 2017 the OPP's Cancer Assessment Review Committee [73], concluded that glyphosate does not pose either a carcinogenic or a genotoxic risk to humans—as indeed had other major regulatory institutions for many years previously.

The OPP's assessment

In an illuminating analysis of IARC and the OPPs' divergent assessments of genotoxicity, Benbrook and colleagues [74, 75] provided evidence indicating that asymmetrical patterns of evaluation were one of the three reasons that accounted for the two institutions' conflicting conclusions on genotoxicity. They showed that the OPP and IARC drew on quite different data sets. The IARC only evaluated genotoxicity evidence that had been reported in the peer-reviewed academic literature [76]. The OPP evaluated proprietary genotoxicity studies provided by industry applicants in response to regulatory licensing requirements, along with a report on research published in the peer-reviewed literature that industry applicants were required to provide. That difference is important, because a striking feature of the genotoxicity data on glyphosate is that virtually all the studies commissioned by regulated firms, and submitted to regulatory authorities, reported no genotoxic effects, whilst the majority of published, peer-reviewed studies reported positive evidence of genotoxicity. Specifically, 99% (94 of 95) of the registrant-commissioned studies included in the OPP's assessment had been reported by their authors as negative [74], whilst of the 118 peer-reviewed studies included in IARC's assessment (of which a smaller and less diverse fraction were included in OPP's assessment), 70% (83 of 118) had reported positive findings [74]. Glyphosate provides yet another example in which toxicity studies sponsored by the chemical, pharmaceutical and food industries were far more likely to arrive at conclusions favourable to the compounds under assessment than those that had not been sponsored by those industries [39, 77–79].

Second, the OPP and IARC had asked and answered different questions.As part of its statutory remit, the OPP focused mainly on the risks from exposure to glyphosate in its pure or technical form (i.e., the active ingredient). Although the OPP also included studies on glyphosate formulations in its review, the agency's weight of evidence evaluation focussed on the active ingredient, with little or no weight given to evidence from tests on the formulations, which contained other chemicals, such as surfactants and adjuvants [73, 75]. Those co-formulants often have their own toxicological profiles which can affect the relevance and conclusions of pesticide risk assessments [80]. Furthermore, the OPP's risk assessment took into account both the hazard posed by glyphosate and exposure to the compound. Its conclusions on glyphosate safety were conditional on exposure to glyphosate at doses relevant to human exposures [73]. In contrast, IARC's evaluation focused only on the potential hazard of exposure to glyphosate, both in the form of the active ingredient and as commercial formulations. The questions IARC asked were, therefore, more open-ended than the OPPs, because it examined the potential for glyphosate to cause a genotoxic/carcinogenic response without reference to the dose applied, and because its evaluation was not limited to technical glyphosate but included formulated products too.

Third, although some of the available evidence on glyphosate genotoxicity was evaluated by both the IARC and the OPP, the two institutions interpretations of that evidence diverged markedly. IARC accepted as valid a wide range of published evidence indicating positive genotoxic effects, but the OPP almost never did [75]. Furthermore, whilst the OPP provided reasons for discounting almost all of the positive evidence on genotoxicity, it accepted as valid all the studies that were reported as negative.

A 2016 EPA Health Effects Division (HED) memo summarised 65 genotoxicity studies on technical glyphosate, both registrant-commissioned and published, which the OPP had included in its assessment, along with comments from HED reviewers [81]. The HED reviewers gave reasons for placing limited weight on almost all the

positive studies (19 of 21 studies) included in the memo. All such comments cast doubt on some aspect of study design, data collection and interpretation, and/or the biological significance of the results [75]. In contrast, the reviewers did not make a single comment on any of the 36 registrant-commissioned studies that were reported as negative, nor on 7 of the 8 published studies reported as negative. All were accepted as negative with no critical appraisal or comment. The OPP commented on just one of the negative, published studies, but only to point out that it was unclear from this paper why glyphosate with a purity of only 62% had been used in the study.

The implication of that analysis is that unless, by way of a remarkable set of coincidences, virtually all the studies on technical glyphosate included in the OPP's assessment and reported as positive were problematic, whilst none of the studies reported as negative had any shortcomings, the conclusion must be that the OPP's pattern of scrutiny and/or its pattern of criticism was asymmetrical.

Benbrook [74] also noted that 'dozens' of the registrant-commissioned studies reported as negative had in fact provided some evidence of positive genotoxic responses, although the authors of those studies had chosen to classify those findings as 'negative'. Those study authors had done so either on the grounds that the route of administration was not regarded as relevant to a human-health risk assessment, or because the reported result occurred at a high dose level, or because the dose was considered toxic to cells via a non-genotoxic mechanism [74]. However, if the route of administration was not relevant to a human-health risk assessment it is puzzling why the registrant had ever chosen to commission those studies in the first place. Benbrook remarked that "…the criteria and decision process regulators apply in determining whether the authors of regulatory studies are justified in dismissing a given positive result are generally unknown…" and deserve 'further research' [74]. That is indeed so, but it is worth stressing that agency reviewers did not challenge any of those authors' decisions to discount their positive findings and to classify them as negative. Neither did they ever seek to reclassify a borderline negative study as equivocal or positive.

The BfR's assessment

In this section, the focus is on asymmetries that were evident in the BfR's 2013 discussion of glyphosate's possible genotoxicity. Like the OPP, the BfR based its assessment on both unpublished registrant-commissioned studies and a report on published peer-reviewed studies supplied by the registrants. The BfR reviewed that evidence in three parts: (i) a review of registrant-commissioned studies, (ii) a review of peer-reviewed literature published up to 2000, and (iii) a review of peer-reviewed literature published after 2000 [71]. Some care needs to be taken in identifying and characterising the BfR´s evaluation of those data, because much of the text that comprises that evaluation was directly copied and pasted from dossiers submitted by Monsanto [82, 83]. The BfR explained that due to the large number of submitted toxicological studies and data, study descriptions and analyses were reproduced from the material provided by the pesticide industry; however, the BfR stressed that it had provided its own comments on each study, in italics, clearly labelled as such ([71], p. 1).

All three parts of the BfR's genotoxicity assessment provided summary tables, descriptions and conclusions for each individual study, whilst the second and third parts also contained commentaries on the reliability and relevance of each study, all of which were presumably written by Monsanto. The third part also contained a weight of evidence evaluation and 'Klimisch evaluations' of 16 of the published studies which, as Molander et al. have shown [84], provide relatively superficial assessments of a study's reliability.

The BfR own comments on the individual studies were confined to those studies described in the first two parts of its genotoxicity assessment. In the first part the BfR's italicised comments categorised the studies variously as ´acceptable´, ´supplementary´ or ´unacceptable´, along with brief remarks ([71], pp. 301–376). The second part provided a commentary on the reliability, relevance and sometimes the implications of each study ([71], pp. 376–391). For the third part the BfR provided no comments on any of the individual studies ([71], pp. 372–415). Instead, as a plagiarism analysis commissioned by the European Parliament revealed, the entire third section, including the Klimisch evaluations, was an unacknowledged word-for-word copy of Monsanto's submission [83].

We have compiled Tables S1–S3, and provide them as supplementary material; they provide lists of all the studies included in the three sections of the BfR's assessment. The text in the final columns of each table summarises the BfR's evaluation of each study. Table S1 reproduces the BfR's classification of acceptable, supplementary or unacceptable for registrant-commissioned studies. S2 provides the gist of BfR's descriptive evaluations of the reliability and relevance of each published study prior to 2000, and by way of comparison in the penultimate column, the gist of Monsanto's evaluations. The final column of Table S3 provides the gist of Monsanto's evaluations of the reliability or relevance of studies published after 2000, but presumably endorsed by the BfR, since the agency did not provide any comments of its own.

As those supplementary tables indicate, the BfR evaluated a total of 109 genotoxicity studies. They included 45 unpublished registrant-commissioned studies, of

which only one was reported as positive by its author. The remaining 64 studies were published peer-reviewed studies, of which 48 were reported by their authors as positive. We have categorised the BfR's comments, or the comments that they endorsed, on all those studies (the gist of which are included in the tables in the supplementary material) as either a judgement that the study was: (i) reliable with no limitations noted (ii) reliable with some limitations noted but the reported result endorsed, or (iii) unreliable or irrelevant with the reported result discounted.

The summary data are presented in Table 2 and they show that the BfR judged 5% (3 of 60) of the studies that did not indicate any adverse genotoxic effects to be unreliable, but found reasons to discount 98% (48 of 49) of the studies indicating positive genotoxic effects as unreliable or irrelevant. Furthermore, of the 18 negative studies (out of a total of 60) that were deemed to exhibit some flaws or limitations, the BfR concluded that 15 of those were nonetheless valid negative findings, with only 3 discounted as unreliable. Of the 48 positive studies (out of a total of 49) that were deemed to exhibit some flaws or limitations, all 48 were discounted as unreliable or irrelevant. Thus, the BfR always concluded that the positive evidence was unreliable or irrelevant, and should be discounted, whilst problems with most of the negative studies were disregarded. Moreover, in almost no instances did BfR infer that a presumption of genotoxicity should be drawn from what it concluded to be limited or flawed, but nominally positive evidence. Inferential judgements, whether for evidence reported as positive or negative, were largely made in one direction only, towards the conclusion of an absence of harm.

The three negative studies discounted by the BfR as unreliable (all registrant-commissioned studies) were judged to have been tested with dose levels that were much too low (Table S1). The forty eight positive studies discounted as unreliable or irrelevant (all peer-reviewed studies) were discounted on one or more of the following grounds: methodological deficiencies, reporting deficiencies, lack of biological significance, failure to comply with OECD guidelines, possible cytotoxicity, lack of consistent dose responses, results contradicted by other studies that had been reported as negative, test systems employed not considered to provide evidence of relevance to humans of genotoxicity, positive findings likely to be caused by a component other than glyphosate, or studies unable to differentiate exposure to glyphosate from exposure to other pesticides (Tables S1, S2).

Overall, then, unless by remarkable coincidence almost all of the positive studies (98%) were characterised by problems, of design, methodology, conduct, failure to comply with guidelines, biological significance and so on, of sufficient seriousness to discount their findings, whilst very few of the negative studies (5%) suffered similarly significant flaws, then the BfR's evaluation was based on deploying asymmetric standards of reliability and relevance.

A possible counter argument to that claim is that since almost all the positive evidence of glyphosate genotoxicity was provided by published, peer-reviewed studies, it might be that peer-reviewed studies were far more likely than registrant-commissioned studies to fail to meet BfR's standards of reliability and relevance. But that argument is weak. Sixteen of the peer-reviewed studies were reported by their authors as negative, and *all* were deemed reliable and relevant by the BfR, but of the 48 peer-reviewed studies reported as positive *none* were considered reliable and relevant by the BfR.

Table 2 BfR's interpretation of the reliability and relevance of studies for those that had, and had not, indicated possible harm, by number of studies

	Number of studies reviewed	Number treated as reliable with no limitations	Number treated as reliable with some limitations	Number discounted as unreliable or irrelevant
Studies not indicating possible harm	60	42	15	3
Studies indicating possible harm	49	1	0	48

Discussion

Across the four cases, we have identified and characterised several distinctive, albeit overlapping and interrelated, types of evaluative asymmetry. Each of the cases involved one or more of those categories. They include:

1. *Asymmetric levels of critical scrutiny.* Positive evidence was critically examined for potential flaws and/or biological significance but little or no efforts were made to identify potential problems with negative evidence.
2. *Asymmetric critiques.* Uncertainties and/or methodological flaws provoked scepticism about the reliability or relevance of positive evidence, but the same or similar kinds of uncertainties or flaws were ignored and/or discounted for negative evidence.
3. *Asymmetric bases for criticism.* Criticisms of positive studies were hypothetical, with little or no evidential backing, or were even spurious, whilst substantive

reasons for doubting the reliability or relevance of negative evidence were ignored and/or discounted.

4. *Asymmetric inferences from inconclusive evidence*. Inferences from limited or inconclusive evidence (whether nominally positive or nominally negative) to conclude 'safe' were deemed scientifically credible, but inferences from limited or inconclusive evidence to conclude 'unsafe' were considered unjustified, at least in the absence of more direct evidence of harm.

All four evaluative asymmetries are practices that are antithetical to the effective protection of public and environmental health. They involved ignoring, understating or denying uncertainties and problems associated with evidence suggestive of the absence of harm whilst emphasising, overstating or even inventing uncertainties and flaws associated with evidence suggestive of harm. One key consequence is that the scope for diminishing possible harm was repeatedly made inconspicuous [85]. Policy makers responsible for acting on the evaluations provided by officials and advisors, and much of the wider public, would not have realised that regulatory options to diminish or eliminate exposure to potentially harmful technologies were legitimately available. The fact that those asymmetries appear to have been routinely evident in mainstream official regulatory policy practices implies that the policy frameworks within which they have been deployed are also antithetical to the effective protection of public and environmental health, and that they function, in effect, to deny the availability of technological choices.

There is, however, a very important contrast between the first three of those four types of evaluative asymmetry and the remaining one. The first three types of asymmetries are evaluative practices that are scientifically flawed, and so policy regimes that deploy those asymmetrical practices produce poor quality science. They are scientifically flawed, because they fail to abide by widely accepted and important scientific norms. Those norms include what Robert Merton referred to as 'organised scepticism' [34]. Merton identified a normative benchmark that is, or at any rate should be, characteristic of science, which is both a methodological and an institutional injunction. He argued that all members of the scientific community are subject to the imperative to presume that 'Nothing is sacred' and everything should be equally subjected to critical scrutiny.

Critically analysing toxicologically positive evidence but failing to do so with respect to negative evidence is not subjecting everything to uniform and consistent critical scrutiny. Identifying problems with toxicologically positive evidence (and dismissing that evidence on those grounds), when the same or similar kinds of problems are ignored or discounted with respect to negative evidence, is not doing so either. Nor is the practice of identifying entirely hypothetical problems with toxicologically positive evidence yet ignoring or discounting evident problems with negative evidence. Such practices amount to cherry-picking evidence, in ways that are contextually useful to industrial and commercial interest groups, though often uninformative. Deployment of those practices distorts and retards our collective knowledge of the potential threats posed by the products of the food and chemical industries and misleads others' understandings of the nature of those threats. It also has the effect of inhibiting inquiries that could have advanced our understanding of the factors influencing public and environmental health and pathologies.

The remaining type of asymmetry (i.e., asymmetric inferences from inconclusive evidence) is not in itself scientifically flawed, but it is an anti-precautionary judgement and, crucially, misleadingly so. The need to make inferences from incomplete or limited evidence to reach regulatory-scientific conclusions is commonplace, and asymmetric benchmarks of *evidential sufficiency* are legitimate, and maybe even indispensable, when interpreting evidence and constructing advice for policymakers. In the medical regulatory domain, for example, evidence that a vaccine or drug has serious side effects typically needs to meet a much lower evidential threshold to justify regulatory restrictions than the threshold required to establish the absence of serious adverse effects. The former can take the form of a single well-conducted study on patients taking the drug, whilst the latter usually requires extensive evidence from each phase of the drug approval process, including several independent, large clinical trials. Asymmetric judgements about the kinds or strengths of evidence that are sufficient to justify a judgement of safety versus those that are sufficient to justify judgements of harm or risk are relatively common. As in that example from medical regulation, they are often uncontroversial. Yet the drug/vaccine example is legitimate, not because it happens to be a precautionary approach to making asymmetrical inferences from evidence, but because the asymmetry is explicit and reasoned.

Insofar as asymmetric inferences from inconclusive evidence were deployed in our cases, they might have been scientifically valid if prior judgements about the reliability and relevance of individual studies had been exercised symmetrically. That is to say, if we had a relatively fair account of the limitations of both evidence indicting harm and evidence indicating an absence of harm, as the basis upon which to then make (possibly asymmetric) judgements as to the *sufficiency* of that evidence to support regulatory-scientific conclusions of

'safe' as compared to conclusions of 'harm'. They might also have been democratically valid if those inferential asymmetries were explicit and justified. Neither of these conditions characterised the reasoning in our four case studies.

Explaining evaluative asymmetries

Evidence from the cases can contribute to explaining how policy regimes operate in ways that deploy evaluative asymmetries that are antithetical to the effective protection of public and environmental health. For example, the BfR's asymmetrical evaluation of glyphosate's genotoxicity data almost entirely mirrored Monsanto´s assessment of both registrant data and pre-2000 published data and it directly reproduced Monsanto´s evaluation of post-2000 published data. However, evidence explaining why officials and advisors engaged in, or endorsed, such asymmetric evaluative practices is scarce, but a few hypotheses can be outlined.

One is that resource or time constraints, and/or a culture of mutual trust between regulators and industry, mean that, institutions responsible for regulatory appraisals may uncritically reproduce industrial firms' own, self-interested asymmetric assessments of evidence. Confronted by the need to evaluate tens of thousands of pages of regulatory-scientific evidence, and with officials likely to be overworked and under resourced, and advisors unpaid and over-committed, regulatory agencies may come to rely on, or be influenced by, industry's evaluations of the reliability, relevance and meaning of toxicological evidence.

A second hypothesis is that officials and advisors in risk assessment institutions may share, or come to share, or are chosen because they share, a broader policy presumption to favour industrial innovation and/or the *status quo* unless there is direct and strong evidence that a technological product causes unacceptable harm. Given that there are almost always chronic uncertainties as to what much of the evidence practically available to regulators really means for human and environmental health, the rationale for that presumption is to diminish the possibility that regulators place restrictions on products (and impose costs on industry and innovation) that turn out not, in fact, to pose an unacceptable threat (i.e., what is known as a 'false-positive' error). For example, in the UK, at a meeting of the Committee on Toxicity (CoT) held in March 2006, and in the context of a discussion of evidence indicating that a group of compounds might exert a carcinogenic effect in human consumers, a member of the committee said: "We [i.e., the members of CoT] have a particular responsibility to seek and to avoid false positives." ([86], p. 6). When that remark was made, none of the other committee members commented on, or contested, that remark; none suggested that avoiding false negatives was equally, or at least as, important. A similar normative stance is apparent in the comments, noted earlier in the Bt maize case, of a EFSA GMO panel member who stated that "[o]f course, studies that describe potential negative [i.e., adverse] environmental effects of GMOs are discussed particularly intensively." ([60], p. 3).

It is not difficult to appreciate why an over-riding concern to avoid erroneously concluding that a product is unacceptably harmful, will tend to mean that officials and advisors make inferential judgements from inconclusive or limited evidence towards conclusions of 'no harm' rather than 'harm'. Having made those subjective, and contestable judgements, officials and advisors might be unwilling to make them explicit. They might, therefore, prefer to categorise positive evidence that falls short of direct causal evidence of harm as unreliable or irrelevant, rather than as inconclusive or limited evidence of harm. If they had been explicit about those judgements, they then might have to explain why inconclusive or limited evidence of harm does not constitute grounds for concluding 'possibly unsafe'. The rational may be that advisors and officials recognise that an overtly anti-precautionary approach to appraisal would be politically difficult to sustain, or because they think their pronouncements would lack authority unless they were represented as flowing directly from evidence alone [87, 88], or just because they do not want to be challenged. In such circumstances, the deployment of (scientifically legitimate) inferential asymmetries about the regulatory-scientific meaning of a body of evidence (asymmetry no. 4 in our earlier list) might also slip into making (scientifically flawed) evaluative asymmetric judgements about the reliability and relevance of positive versus negative evidence (asymmetries no. 2 and 3 in our earlier list).

A third hypothesis is that officials employed in and by science-based regulatory institutions have organisational incentives to protect the reputation of the institution in which they work, by trying to give the impression that the institution had never previously made any mistakes. This produces what might usefully be referred to as 'institutional inertia'. If a science-based technology regulatory institution changes its regulatory judgements or recommendations, without new evidence that could be portrayed as legitimating those changes, then the institution would in effect be admitting that it had previously made mistakes in its judgements and advice. Such an admission would provoke questions along the lines of: which other mistakes have you made? In the glyphosate case both US and EU agencies had the opportunity to reconsider the evidence of carcinogenicity, once the documents from US civil litigation against Monsanto documented

potential scientific misconduct relevant to the evidence presented for the glyphosate assessment [89, 90]. However, both agencies maintained their position on the reliability of industry sponsored studies. Those considerations help to explain why such regulatory institutions are keen to endorse and reinforce as many of their previous judgements and decisions as possible. Therefore, the types and quantities of data required to persuade such institutions to change any of their previous judgements has often been far more than was previously sufficient to justify their earlier judgements.

Implications for policy and practice

Insofar as one or more of the hypotheses listed above might explain the evaluative asymmetries identified in this paper, it is useful to sketch out the kinds of institutional policies and practices that would help diminish their prevalence. First, in so far as resource or time constraints, or an institutional culture of trust in regulated industries, tempt regulatory officials and advisors to reproduce industrial firms preferred asymmetric assessments of toxicological evidence, the response is straightforward. Regulatory institutions need sufficient resources to perform their own independent analyses of unpublished and peer-reviewed evidence and, more generally, a culture that recognises, and can respond procedurally to, the fact that industrial actors have commercial interests in portraying evidence about their products in as favourable a way as is possible to sustain.

This problem is not novel. Over 40 years ago a US Senate oversight subcommittee took evidence from Environmental Protection Agency (EPA) staff who felt that industry submissions often contained 'questionable scientific arguments' ([91], p. 13). The subcommittee noted that

> *"[o]ne current member of the [EPA's] Toxicological Branch says that, while he thought he had encountered "every trick in the book" during his career evaluating chronic toxicity experiments and data, he has recently been amazed by new levels of "ingenuity and cleverness" employed by some pesticide registrants. He expressed admiration for the capacity of registrants to advance new arguments minimising the significance of negative experimental findings [in this context, findings indicative of harm] ..." ([91], p. 121)*

Yet, EPA scientists told the committee that they only had sufficient resources to detect and pursue a small proportion of findings that were evident in experimental data but had been ignored or dismissed in submitted reports. Indeed, the Senate subcommittee discovered that some EPA evaluations were nothing more than verbatim transcriptions of summaries submitted by the registrants, and that the EPA had accepted without question the accuracy and interpretation of the industry's assessments [91]. The subcommittee recommended that programmes be established to ensure that the EPA could realistically be expected to detect, and act as a deterrent against, what it termed 'shoddy science' ([91], p. 14).

That important recommendation was made in 1983. In the subsequent four decades, extensive evidence has accumulated showing that regulated firms, across many different industries, can and do steer the production and representation of knowledge developed for regulatory purposes in accordance with their private interests, and that they do so in numerous ways, including by seeking to minimise or discount adverse findings and to overstate flawed but negative findings [25–32, 92, 93]. A critical issue for regulatory policy concerns the extent to which, and ways in which, officials, advisors and managers accede to or challenge that kind of influence over regulation. Michaels [28–30], for example, outlines several proposals to limit and challenge industrial influence.

The second hypothesis is interesting, because, in principle at least, a series of regulatory reforms in the early 2000s, especially in European jurisdictions, should have diminished pressures on officials and scientists to conduct appraisals in ways that favour support for industrial innovation and/or the *status quo* unless there is strong, direct evidence of unacceptable harm, and they should have made it much more difficult for officials and advisors to try and disguise the subjective nature of their judgements in reaching regulatory-scientific conclusions.

Those reforms, introduced in the wake of the BSE saga, typically involved: (a) the relocation of regulatory decision-making from government departments responsible for sponsoring regulated industries to Ministries (or in the case of the European Commission, Directorates General) of Health, Environment or Consumer Protection; (b) the institutional separation of organisations responsible for scientific appraisal and advice from the departments responsible for regulatory decision-making; (c) the imposition of much greater levels of transparency and openness on appraisal bodies; and (d) a requirement that officials and advisory scientists make all uncertainties and assumptions underlying their appraisals explicit, in a form comprehensible to decision makers [94].

Those reforms were intended, in part, to limit unwarranted political influence on scientific appraisals and/or its representations by policy-makers and the regulated industries, in the wake of carefully documented evidence that this is exactly what happened during the BSE saga, both in the UK and at the European Commission [2, 95], and to restore public confidence in food safety regulation after the debacle of BSE.

The fact that those reforms do not appear to have diminished the kinds of asymmetric evaluations of evidence we have illustrated in our post-2000 cases suggests at least three possibilities. One is that the reforms have only been partially implemented (for example, it is very rare that appraisals do make explicit all uncertainties and assumptions despite guidance in many institutions that they are obliged to do so); a second is that the reforms were not sufficient to ensure that appraisal bodies function in ways that are sufficiently independent from political and industrial influence; and a third is that the problem may partly lie elsewhere, for example, in regulatory appraisal cultures that have long embodied permissive, non-precautionary approaches to analysis. All three possibilities are simultaneously plausible, and useful responses might include full implementation of regulatory guidelines on being accountably explicit about uncertainties and assumptions; and procedural recognition that scientific regulatory appraisals are irredeemably partly constituted with normative judgments, and therefore, those need to be explicit, and ideally chosen by policy-makers, for example, in compliance with commitments agreed at meetings of the Codex Alimentarius Commission, and set out in the Codex Procedural Manual [16, 96].

Conclusions

Our purpose in this paper has been to identify whether the standards of scrutiny and evaluation deployed by regulatory officials and advisors differ for evidence indicating that a product might be harmful and evidence indicating an absence of harm. Across four case studies, we identified different forms of evaluative asymmetry, and outlined hypotheses as to why regulatory institutions, in different jurisdictions and at different times, have deployed asymmetric forms of evaluation. Our analysis and findings are important for several reasons.

One of these, as we have already argued, is that as a consequence of deploying one or more of the evaluative asymmetries we have highlighted, opportunities for diminishing possible harm from the use of the products under appraisal are concealed. Our analysis suggests that if officials and advisors had adopted consistent standards of scrutiny and evaluation across positive and negative data then a conclusion of unambiguous safety would probably not have been reached regarding appraisals of the carcinogenicity of the Ethylene bisdithiocarbamates, non-target harm from a variety of Bt maize, the toxicity of Aspartame, or the genotoxicity of Glyphosate. How, exactly, regulatory understandings of the hazards or risk posed by those four products might otherwise have looked are unclear, but they would have been far less likely to have been deemed unproblematically safe. They would, therefore, have made explicit to policy makers responsible for acting on those understandings that regulatory options to diminish or eliminate exposure to those products, or seek alternatives, were legitimately available.

Another reason is that the types of evaluative asymmetry we have identified are not only anti-precautionary, some are also scientifically defective. They are scientifically defective, because they violate indispensable scientific requirements for making valid inferences and reaching well-founded conclusions. This is important but also revealing, because it highlights some of the ways in which groups of scientists recruited to serve policy-makers can violate some of the basic norms of professional science.

A final reason why our analysis is important is that questions about the existence, extent and explanation for evaluative asymmetries in regulatory institutional structures and procedures have only rarely been explored in the social science literatures concerned with understanding social, political and cultural influences on the production of regulatory-scientific knowledge claims. Evaluative asymmetries have often been identified in research into how the chemical, pharmaceutical, tobacco and food industries themselves produce and interpret evidence [24, 27, 28], but far less attention has been directed at the behaviour of regulatory institutions. Partly that has been because many of the institutions operate in ways that are at least partly opaque. Furthermore, scrutinising the behaviour of ostensibly science-based policy-making processes require social scientific research skills and an adequate comprehension of the underlying scientific issues, a combination that is not possessed by most social scientists. In practice, much of the research has been accomplished by social-scientifically sophisticated natural scientists and by environmental and public health NGOs, who have uncovered many of the inconsistent ways in which positive and negative evidence have been evaluated within regulatory institutions [60, 97–99].

Our analysis has focused on evaluative asymmetries, but whilst we highlight this form of asymmetry, because we judge it to be important, we are not suggesting that it is the only important kind of asymmetry. As we noted earlier, numerous other kinds value-laden choices must be deployed or accepted within appraisals, for instance when deciding on the scope of assessments, the questions to ask, and the design of empirical studies. Such judgments can often be made in ways that structure in asymmetries in the treatment of nominally positive versus nominally negative evidence within appraisals [19, 36, 38]. Furthermore, wider asymmetries in power and policy-making will also affect knowledge production. For example, Cranor [37] noted that actors with a strong interest in continuing to sell and use

potentially harmful products tend to be well-organised and resourced, whereas those who might be harmed by those products tend to much more diffuse, less organised, and often unaware of the threats that they face. Those kinds of asymmetries in the political constituencies with a stake in appraisals and regulation mean that, amongst other things, there are few incentives for industrial corporations, and often governments, to produce knowledge about environmental and health threats. We recommend that research should evolve to include a broad portfolio of studies of wider sets of asymmetries in and impinging on technological risk regulatory institutional structures and procedures. This paper is intended as a partial contribution to that larger project.

Abbreviations

ACP	Advisory Committee on Pesticides
ACRE	Advisory Committee on Releases to the Environment
ANS	Scientific Panel on Food Additives and Nutrient Sources added to Food
BfR	Bundesinstitut für Risikobewertung
BSE	Bovine spongiform encephalopathy
Bt	*Bacillus thuringiensis*
CoT	Committee on Toxicity
EBDCs	Ethylene bisdithiocarbamates
EFSA	European Food Safety Authority
EPA	Environmental Protection Agency
ETU	Ethylene thiourea
GM	Genetically modified
HED	Health Effects Division
IARC	International Agency for Research on Cancer
OPP	Office of Pesticide Programmes

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12302-025-01176-9.

Supplementary material 1: Table S1. BfR's Evaluations of Registrant-Commissioned Genotoxicity Studies of Glyphosate. Table S2. BfR's Evaluations of Studies of Glyphosate Published before 2000. Table S3. BfR's Evaluations of Studies of Glyphosate Published after 2000.

Acknowledgements
The authors are grateful to Prof Andy Stirling for reading and commenting on a draft of this paper.

Author contributions
PvZ and EM: conceptualisation and original draft. ALO: glyphosate case study investigation and data analysis. All 3 authors discussed and developed the arguments, revised the draft, and read and approved the final draft.

Funding
The authors received no specific funding for this work.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 31 March 2025 Accepted: 5 July 2025
Published online: 04 August 2025

References

1. Bowcott O, Wintour P, Elliot C (1996) The crisis ministers ignored. The Guardian 22 March.
2. Lord Phillips of Worth Matravers, Bridgeman J, Ferguson-Smith M (2000) The BSE inquiry: report, Volume 1, findings and conclusions. London: The Stationery Office
3. Lord Phillips of Worth Matravers, Bridgeman J, Ferguson-Smith M (2000) The BSE inquiry: Report, Volume 4, The Southwood Working Party 1988–1989. London: The Stationery Office
4. Lord Phillips of Worth Matravers, Bridgeman J, Ferguson-Smith M (2000) The BSE inquiry: report, Volume 6, Human Health 1989–1996. London: The Stationery Office
5. van Zwanenberg P, Millstone E (2005) BSE: risk, science and governance. Oxford University Press, Oxford
6. Lord Phillips of Worth Matravers, Bridgeman J, Ferguson-Smith M (2000) The BSE inquiry: report, Volume 11, Scientists After Southwood. London: The Stationery Office
7. van Zwanenberg P, Millstone E (2000) Beyond skeptical relativism: evaluating the social constructions of expert risk assessments. Sci Technol Hum Val 25(3):259–282
8. Millstone E, Dawson E (2019) EFSA's toxicological assessment of aspartame: was it even-handedly trying to identify possible unreliable positives and unreliable negatives? Arch Pub Health 77:34
9. Nelkin D (1984) Controversy: politics of technical decisions. SAGE Publications, Beverly Hills
10. Jasanoff S (1990) The fifth branch: science advisers as policymakers. Harvard University Press, Cambridge
11. Url B (2018) Don't attack science agencies for political gain. Nature 553:381
12. US National Research Council (1983) Risk assessment in the federal government. Managing the process. National Academy Press, Washington, DC
13. Jasanoff S, Wynne B (1998) Science and decision-making. In: Rayner S, Malone E (eds) Human choice and climate change. Vol 1 the societal framework. Batelle Press, Washington DC, pp 1–87
14. Felt U, Wynne B (2007) eds Taking European knowledge society seriously. Report of the expert group on science and governance to the science, economy and society directorate, Directorate-General for Research. Brussels: European Commission
15. Douglas HE (2000) Inductive risk and values in science. Philos Sci 67:559–579
16. Elliot KC (2019) Managing value-laden judgements in regulatory science and risk assessment. EFSA J 17(S1):e170709
17. Cranor CF (1990) Some moral issues in risk assessment. Ethics 101(1):123–143
18. European Environment Agency (2001) Late lessons from early warnings: the precautionary principle 1896–2000. Office for Official Publications of the European Communities, Luxemburg
19. Stirling A (2007) Science, precaution and risk assessment: towards more measured and constructive policy debate. EMBO Rep 8:309–315
20. Commission of the European Communities (2000) Communication from the Commission on the Precautionary Principle, COM (2000)1 final, 2 February 2000, Brussels
21. Wynne B (1992) Uncertainty and environmental learning: reconceiving science and policy in the preventive paradigm. Glob EnvironChange 2:111–127
22. Raffensperger C, Tickner J (1999) Protecting public health and the environment: implementing the precautionary principle. Island Press, Washington, DC

23. Jasanoff S (1987) Cultural aspects of risk assessment in Britain and the United States. In: Johnson BB, Covello VT (eds) The social and cultural construction of risk: essays on risk selection and perception. D. Reidel, Dordrecht, pp 359–397
24. McGarity TO, Wagner WE (2008) Bending science: how special interests corrupt public health research. Harvard University Press, Cambridge
25. Lexchin J (2012) Those who have the gold make the evidence: how the pharmaceutical industry biases the outcomes of clinical trials of medications. Sci Eng Ethics 18:247–261
26. Holman B, Elliott KC (2018) The promise and perils of industry-funded science. Philos Compass 13(11):e12544
27. Legg T, Hatchard J, Gilmore AB (2021) The science for profit model - how and why corporations influence science and the use of science in policy and practice. PLoS ONE 16(6):e0253272
28. Michaels D, Monforten C (2005) Manufacturing uncertainty: contested science and the protection of the public's health and environment. Am J Public Health 95:S39–S48
29. Michaels D (2008) Doubt is their product: how industry's assault on science threatens your health. Oxford University Press, Oxford
30. Michaels D (2020) The triumph of doubt: dark money and the science of deception. Oxford University Press, Oxford
31. Oreskes N, Conway EM (2010) Merchants of doubt: how a handful of scientists obscured the truth on issues from tobacco smoke to global warming. Bloomsbury Publishing, London
32. Sismondo S (2007) Ghost management: how much of the medical literature is shaped behind the scenes by the pharmaceutical industry? PLoS Med 4(9):e286
33. van Zwanenberg P, Millstone E (2015) Taste and power: the flavouring industry and flavour additive regulation. Sci Cult 24(2):129–156
34. Merton RK (1942) The normative structure of science. In: Merton RK (ed) The sociology of science: theoretical and empirical investigations. University of Chicago Press, Chicago, pp 267–278
35. Resnik DB, Elliott KC (2023) Science, values and the new demarcation problem. J Gen Philos Sci 54:259–286
36. Lemons J, Shrader-Frechette K, Cranor C (1997) The precautionary principle: scientific uncertainty and type I and type II errors. Found Sci 2:207–236
37. Cranor C (1999) Asymmetric information, the precautionary principle, and burdens of proof in environmental health protections. In: Raffensperger C, Tickner J (eds) Protecting public health and the environment: implementing the precautionary principle. Island Press, Washington, DC
38. Needleman HL, Bellinger DC (1986) Type II fallacies in the study of childhood exposure to lead at low dose: a critical and quantitative review. In: Smith MA, Grant LD, Sors AI (eds) Lead exposure and child development: an international assessment. Kluwer Academic Publishers, Boston, pp 293–304
39. Hayes JP (1987) The positive approach to negative results in toxicology studies. Ecotoxicol Environ Saf 14:73–77
40. van Zwanenberg P (1996) Science, pesticide policy and public health: ethylene bisdithiocarbamate regulation in the UK and USA. PhD thesis, University of Sussex.
41. Innes JRM, Ulland BM, Valerio MG, Petrucelli L, Fishbein ER, Hart AJ et al (1969) Bioassay of pesticides and industrial chemicals for tumorigenicity in mice: a preliminary note. J Natl Cancer Inst 42:1101–1114
42. Ulland BM, Weisburger JB, Weisburger EK, Rice JM, Cypher R (1972) Thyroid cancer in rats from ethylene thiourea intake. J Natl Cancer Inst 49:583–584
43. Graham SL, Davis KJ, Hansen WH, Graham CH (1975) Effects of prolonged ethylene thiourea ingestion on the thyroid of the rat. Food Cosmet Toxicol 13:493–499
44. Chhabra RS, Eustis S, Haseman JK, Kurtz PJ, Carlton BD (1992) Comparative carcinogenicity of ethylene thiourea with or without perinatal exposure in rats and mice. Fund Appl Toxicol 18:405–417
45. U.S. Environmental Protection Agency (1979) The carcinogen assessment group's risk assessment on ethylene bisdithiocarbamates (EBDC). U.S. Environmental Protection Agency, Washington, DC
46. Advisory Committee on Pesticides (1990) Position document on consumer risk arising from the use of ethylene bisdithiocarbamates. Advisory Committee on Pesticides Evaluation no. 16. London: Ministry of Agriculture, Fisheries and Food.
47. Fenner-Crisp PA (1989) August 21 meeting with EBDC registrants on ETU toxicology studies and liver tumour formation. Memorandum to EBDC Special Review docket. Washington, DC: Office of Pesticide Programs, U.S. Environmental Protection Agency.
48. U.S. Environmental Protection Agency (1989) EBDC special review: technical support document 2/3. Report no. EPA/540/09-90/077. Washington, DC: Office of Pesticide Programs, U.S. Environmental Protection Agency.
49. International Programme on Chemical Safety (1990) Principles for the toxicological assessment of pesticide residues in food. Environmental health riterion 104. World Health Organization, Geneva
50. Doull J, Bridges BA, Kroes R, Golberg L, Munro IC, Paynter OE et al (1983) The relevance of mouse liver hepatoma to human carcinogenic risk: a report of the International Expert Advisory Committee to the Nutrition Foundation. Nutrition Foundation, Washington, DC
51. Blackwell-Smith R, Finnegan JK, Larson PS, Sahyoun PF, Dreyfuss ML, Haag HB (1953) Toxicologic studies on zinc and disodium ethylene bisdithiocarbamates. J Pharmacol Exp Ther 109:159–166
52. U.S. Environmental Protection Agency (1982) Ethylene bisdithiocarbamates decision document: Final resolution of rebuttable presumption against registration. Report no. 540/09-87-164. Washington, DC: Office of Pesticides and Toxic Substances, U.S. Environmental Protection Agency.
53. Hilbeck A, Baumgartner M, Fried PM, Bigler F (1998) Effects of transgenic Bt corn-fed prey on immature development of Chrysoperla carnea (Neuroptera: Chrysopidae). Environ Entomol 27:480–487
54. Hilbeck A, Moar W, Pusztai-Carey M, Filipini A, Bigler F (1999) Prey-mediated effects of Cry1Ab toxin and protoxin and Cry2A protoxin on the predator Chrysoperla carnea (Neuroptera: Chrysopidae). Entomol Exp Appl 91:305–316
55. Scientific Committee on Plants (2000) Opinion of the scientific committee on plants on the invocation by Germany of Article 16 of Council Directive 90/220/EEC regarding GM Bt maize 176 notified by Ciba-Geigy, (now NOVARTIS), notification C/F/94/11-03 (SCP/GMO/276Final - 9 November 2000). https://food.ec.europa.eu/document/download/dd33fa02-98fb-439f-8a14-80980e292898_en?filename=sci-com_scp_out78_gmo_en.pdf. Accessed 31 Mar 2025
56. Levidow L, Murphy J (2003) Reframing regulatory science: trans-atlantic conflicts over GM crops. Cahiers de Econ et Sociol Rurales 68:47–74
57. Wynne B (2006) GMO risk assessment under conditions of biological (and social) complexity', in Bundesministerium für Gesundheit und Frauen (BMGF), the role of precaution in GMO policy. BMGF, Berlin, pp 30–46
58. U.S. Environmental Protection Agency (2000) Response of the environmental protection agency to petition for rulemaking and collateral relief concerning the registration and use of genetically engineered plants expressing bacillus thuringiensis endotoxins, submitted by petitioners Greenpeace international, International federation of organic agriculture movements, International center for technology assessment, et al. April 19, 2000. https://www3.epa.gov/pesticides/chem_search/reg_actions/pip/greenpeace-petition.pdf. Accessed 31 Mar 2025
59. Levidow L (2003) Precautionary risk assessment of Bt maize: what uncertainties? J Invertebr Pathol 83:113–117
60. Hilbeck A, Meier M, Trtikova M (2012) Underlying reasons of the controversy over adverse effects of Bt toxins on lady beetle and lacewing larvae. Environ Sci Eur 24:9
61. Scientific Advisory Panel (2000) Sets of scientific issues being considered by the Environmental Protection Agency regarding: Bt plant-pesticides risk and benefit assessments, SAP Report No. 2000-07, FIFRA Scientific Advisory Panel Meeting, October 2000. https://archive.epa.gov/scipoly/sap/meetings/web/pdf/octoberfinal.pdf. Accessed 31 Mar 2025
62. EFSA Panel on Food Additives and Nutrient Sources added to Food (2013) Scientific opinion on the re-evaluation of aspartame (E 951) as a food additive. EFSA J 11(12):3496
63. Kass G, Lodi F (2020) Letter to the editor regarding the article 'EFSA's toxicological assessment of aspartame: was it even-handedly trying to identify possible unreliable positives and unreliable negatives? Arch Public Health 78:14
64. Millstone E, Dawson E (2020) Why did EFSA not reduce its ADI for aspartame or recommend its use should no longer be permitted? Arch Public Health 78:112
65. Soffritti M, Belpoggi F, Esposti DD, Lambertini L (2005) Aspartame induces lymphomas and leukaemias in rats. Eur Oncol 10:107–116

66. Commission of The European Communities (2000) White Paper On Food Safety, 12 January 2000, COM (99) 719 final. Brussels.
67. Soffritti M, Belpoggi F, Manservigin M, Tibaldi E, Lauriola M, Falcioni L, Bua L (2010) Aspartame administered in feed, beginning prenatally through life span, induces cancers of the liver and lung in male swiss mice. Am J Ind Med 53:1197–1206
68. EFSA (2006) Scientific opinion of the panel on food additives, flavourings, processing aids and materials in contact with food (AFC) related to a new long-term carcinogenicity study on aspartame. EFSA J 356:1–44
69. OECD (2017) Overview on genetic toxicology TGs, OECD series on testing and assessment, No. 238. OECD Publishing, Paris
70. International Agency for Resarch on Cancer (2017) Some organophosphate insecticides and herbicides. IARC Monog Eval Carc 112.
71. Bundesinstitut für Risikobewertung (2013) Renewal assessment report, 18 December 2013, glyphosate, vol 3, annex B.6.1 toxicology and metabolism. Bundesinstitut für Risikobewertung. https://corporateeurope.org/sites/default/files/attachments/glyphosate_rar_08_volume_3ca-cp_b-6_2013-12-18_san.pdf. Accessed 31 Mar 2025
72. Bundesinstitut für Risikobewertung (2015) Final addendum to the renewal assessment report: glyphosate, public version, risk assessment provided by the RMS Germany and co-rapporteur MS Slovakia for the active substance, October 2015, Bundesinstitut für Risikobewertung. https://www.efsa.europa.eu/en/consultations/call/public-consultation-active-substance-glyphosate. Accessed 31 Mar 2025
73. US Environmental Protection Agency (2017) Revised glyphosate issue paper: evaluation of carcinogenic potential. https://www.regulations.gov/document/EPA-HQ-OPP-2009-0361-0073. Accessed 31 Mar 2025
74. Benbrook C (2019) How did the US EPA and IARC reach diametrically opposed conclusions on the genotoxicity of glyphosate-based herbicides? Environ Sci Eur 31:2
75. Benbrook C, Mesnage R, Sawyer W (2023) Genotoxicity assays published since 2016 shed new light on the oncogenic potential of glyphosate-based herbicides. Agrochemicals 2(1):47–68
76. International Agency for Resarch on Cancer (2006) IARC Monographs on the Evaluation of Carcinogenic Risk to Humans: Preamble. https://monographs.iarc.who.int/wp-content/uploads/2019/05/Preamble_updated2015.pdf. Accessed 31 Mar 2025
77. Bero L (2017) Addressing bias and conflict of interest among biomedical researchers. JAMA 317(17):1723–1724
78. Fabbri A, Lai A, Grundy Q, Bero LA (2018) The influence of industry sponsorship on the research agenda: a scoping review. Am J Public Health 108(11):e9–e16
79. Rocha GM, Grisolia CK (2019) Why pesticides with mutagenic, carcinogenic and reproductive risks are registered in Brazil. Dev World Bioeth 19(3):148–154
80. Mesnage R, Benbrook C, Antoniou MN (2019) Insight into the confusion over surfactant co-formulants in glyphosate-based herbicides. Food Chem Toxicol 128:137–145
81. US Environmental Protection Agency (2016) Memo: Glyphosate. Study summaries for genotoxicity studies. https://www.regulations.gov/document/EPA-HQ-OPP-2016-0385-0098. Accessed 31 Mar 2025
82. van Zwanenberg P (2015) Chemical reactions: glyphosate and the politics of chemical safety' The Guardian 13 May. http://www.theguardian.com/science/political-science/2015/may/13/chemical-reactions-glyphosate-and-the-politics-of-chemical-safety. Accessed 31 Mar 2025
83. Weber S, Burtscher-Schaden H (2019) Detailed expert report on plagiarism and superordinated copy paste in the renewal assessment report (RAR) on glyphosate. https://bit.ly/Copy-Paste-Glyphosate. Accessed 31 Mar 2025
84. Molander L, Ågerstrand M, Beronius A, Hanberg A, Rudén C (2015) Science in risk assessment and policy (SciRAP): an online resource for evaluating and reporting in vivo (eco)toxicity studies. Hum Ecol Risk Assess 21(3):753–762
85. van Zwanenberg P (2020) The unravelling of technocratic orthodoxy? contemporary knowledge politics in technology regulation. In: Scoones I, Stirling A (eds) The politics of uncertainty: challenges of transformation. Routledge, London, pp 58–72
86. Committee on Toxicity (2006) Minutes of the meeting held on Tuesday 28 March 2006 in Conference Rooms 4 and 5, 4th Floor, Aviation House, London Minutes, TOX/MIN/2006/02. 2006; p 6, para 25. https://webarchive.nationalarchives.gov.uk/ukgwa/20130802142956/http://cot.food.gov.uk/cotmtgs/cotmeets/cot2006/cotmeeting060328/cotfinalminutes28march2006.
87. Linnerooth J (1984) The political processing of uncertainty. Acta Psychol 56:219–231
88. Wagner WE (1995) The science charade in toxic risk regulation. Columbia Law Rev 95:1613–1723
89. Glenna L, Bruce A (2021) Suborning science for profit: monsanto, glyphosate, and private science research misconduct. Res Pol 50(7):104290
90. McHenry LB (2018) The Monsanto papers: poisoning the scientific well. Int J Risk Saf Med 29(3–4):193–205
91. US Congress house committee on agriculture, subcommittee on department operations, research, and foreign agriculture (1983) Regulation of pesticides, appendix to hearings, vol. III, 98th Cong., 1st sess.
92. Wilholt T (2009) Bias and values in scientific research. Stud Hist Phil Sci Part A 40(1):92–101
93. Saltelli A, Dankel DJ, Di Fiore M, Holland N, Pigeon M (2002) Science, the endless frontier of regulatory capture. Futures 135:102860
94. European Food Safety Authority (2009) Guidance of the EFSA scientific committee on transparency in the scientific aspects of risk assessment carried out by EFSA. Part 2: general principles. in response to Question No EFSA-Q-2005-050Ba, and adopted on 7 April 2009. EFSA J 1051:1–22
95. Ortega MM (1997) Report on alleged contraventions or maladministration in the implementation of Community law in relation to BSE, without prejudice to the jurisdiction of the Community and national courts. Temporary committee of inquiry into BSE, 07 February 1997, European Parliament session document.
96. Millstone E (2009) Science, risk and governance: radical rhetorics and the realities of reform in food safety governance. Res Pol 38(4):624–636
97. Testbiotech (2012) The European Food Safety Authority: using double standards when assessing feeding studies. Testbiotech Background 30.10.2012. https://www.testbiotech.org/wp-content/uploads/2016/10/the-double-standards-of-EFSA_0.pdf. Accessed 31 Mar 2025
98. Mesnage R, Antoniou M (2018) Ignoring adjuvant toxicity falsifies the safety profile of commercial pesticides. Front Public Health 5:361
99. Meyer H, Hilbeck A (2013) Rat feeding studies with genetically modified maize – a comparative evaluation of applied methods and risk assessment standards. Environ Sci Eur 25:33

Publisher's Note