Risk Agoras: Using Dialectical Argumentation to Debate Risk

Peter McBurney and Simon Parsons
Department of Computer Science
University of Liverpool
Liverpool L69 7ZF United Kingdom
{P.J.McBurney,S.D.Parsons}@csc.liv.ac.uk

May 9, 2000

Abstract

Public policy debates about the risks of new technologies and substances often hinge on statistical tests of scientific hypotheses. Such tests have the appearance of mechanistic rules but in fact require subjective judgement and interpretation. Particular interpretations developed originally for experiments in agricultural research after WW I have become accepted as standard scientific practice across the social and biomedical sciences. Among other flaws, these standards ignore the consequences of potential errors of statistical inference, consequences which may be uneven in their distribution, impacts and durations. To begin to redress this situation, the authors propose the use of a form of Dialectical Argumentation, such as Habermas's Discourse Ethics, to frame debate over errors of statistical inference in areas of environmental and health risk. If applied, such a framework could enable wider and more effective democratic participation in risk debates and better understanding by non-experts of the issues involved. Such understanding in turn could lead to greater appreciation of the extent of subjective interpretation and judgment involved in scientific assessments of risk.

1 Introduction

Modern society is filled with risks. New technologies and substances have the potential to cause adverse and unintended effects, on people, on other living species, and on our environment, and these effects are increasingly global in scale. Because of such risks, most chemicals and many innovations require Government regulatory approvals before they can be sold to the public. These approvals typically require the conduct of one or more scientific experiments, to assess impact on users or those likely to be exposed to the innovation. Even when no regulatory approvals are required, or when the approval process does not require a program of scientific assessment, public pressure may lead to companies and Governments being forced to undertake experiments to assess impacts, as is the case currently with assessment of the possibility of a link between the use of cellular mobile telephones and brain cancers.[1]

To non-statisticians, the statistical procedures used to assess scientific hypotheses may appear to be objective and deterministic. This objectivity may be a delusion, as the adoption of a standard procedure may be masking prior subjective choices and interpretations. This is certainly the case with the standard procedures used for statistical hypothesis testing, a methodology underpinning almost every scientific assessment of the impact of new technologies and substances. This article commences therefore in Section 2 with an outline of the standard approach to hypothesis testing, and we present what we consider its major flaws in the domain of environmental risk assessment. These are well known to statisticians, where the debate over hypothesis testing methodologies throughout this century has been fierce, but less-well known to others. In Section 3 we present an approach to debate about issues of risk which aims to respond to these flaws, using a form of dialectical argumentation. Section 4 concludes with a discussion of our ongoing work which aims to implement these ideas.

The primary objective of this paper is to identify the subjective component in the standard procedure used for testing scientific hypotheses in the domain of risk assessment, and to propose an alternative, qualitative, approach. Our presentation is aimed at a general audience and we try to avoid technical details. It is important to note that we are not attempting here an overall review of general quantitative risk assessment procedures in the environmental domain, procedures which have been the subject of much recent debate and criticism.[2] Nor is our purpose a discussion of issues relating to the public communication of risk or the resolution of conflicts, although these too are important topics and our proposals could have some bearing on them. Instead, we are seeking to make explicit the full extent of subjective interpretation and judgment involved in just one crucial component of the scientific assessment of risk.

2 Statistical Inference

Informal tests of hypotheses date back almost 300 years [27]. Over the last seventy years, a standard formal methodology has developed for the conduct

[1] Such links are currently under investigation in both Britain and the US, although the phones have been sold commercially for 15 years [41].

[2] For example: "All attempts to establish risk yardsticks, such as probability estimates, threshold values and calculations of costs, founder, as far as late industrial risks are concerned, on the incommensurability of hazards and the problem of the subjective assessment of the probability of occurrence." [26], cited in [4, p.36]. An American critic, Mark Eliot Shere, concluded a detailed review of U.S. Government risk assessment procedures in the USA with the words: "This article posits that quantitative risk assessment is so unreliable that the results are meaningless." [44, p.480].

of experiments involving samples from a larger population, known as statistical hypothesis testing. Because it is generally impossible to undertake tests on all potential members of an animal or human population, experiments are undertaken on a representative sample of the population, and statistical inference is used to derive conclusions about the population from the sample results.

While there are several approaches to hypothesis testing,[3] most scientists use some form of the theory developed by Jerzy Neyman and Egon Pearson in the decade from 1928 [31, 6]. This starts with an hypothesis, the null hypothesis, which we test against an alternative. Typically, the null hypothesis is precisely specified, and asserts that the treatment being studied has no effects. The Neyman-Pearson approach constructs a value, called a test statistic, from the sample and compares this to a set of standard values. Depending on the value of the test statistic, we conclude either in favour of the null hypothesis (we "accept" it) or in favour of the alternative hypothesis (we "reject" the null hypothesis).

In this structure, there are two generic types of errors possible when inferring from sample to population. Type I error occurs when, based on the sample evidence, we wrongly reject the null hypothesis (i.e. when it is actually true in the population). Type II error occurs when we wrongly accept it (i.e. when it is actually false). The probabilities of these errors occurring are typically denoted by α and β, respectively. Under certain assumptions, and for a given sample size, it can be shown that α and β are inversely related: decreasing α can only be achieved at the expense of increasing β. We can only decrease both probabilities simultaneously by increasing the sample size. Hence, it has become customary to fix α at a certain level (say, 5% or 1%) and then use that statistical test procedure which minimises β. The tests taught in text-books and commonly used in the bio-medical and social sciences are those established by statisticians as the best (in this sense), given certain assumptions about the underlying population distribution and the process used to generate the sample. In any particular case, the level of β may be well above the 5%-level pre-selected for α.

Formal hypothesis testing was first applied in agricultural experiments for new crop varieties and treatments, experiments undertaken subsequent to the famines which followed World War I [21]. An American statistician, Abraham Wald, in applying it to manufacturing problems, realised that this approach ignored the consequences of the errors. Accordingly, he developed a decision theory which selected the "best" statistical test not only on the basis of α and β, but also informed by the losses anticipated from each type of error [50]. In deciding whether to accept or reject products coming off a manufacturing assembly line, for example, we can compare the costs of compensating customers for defective items wrongly accepted (and then sold) with the costs of wrongly rejecting non-defective items (and then having to replace or repair them). Upper bounds for direct costs in both cases can generally be estimated for manufacturing applications.

In bio-medical and environmental domains, however, it is usually difficult

[3] And statisticians still argue the relative merits of each approach. See, for example: [43].

to quantify the consequences of inference errors. A new chemical substance, for instance, wrongly found to be safe and then used, may impact millions of people. Also, the full consequences may only be evident many years after exposure, as appears to be the case with BSE-CJD [51]. How does one quantify the subsequent misery or loss of life? Conversely, the same chemical wrongly found to be unsafe, and so never sold, may cause financial losses to the company which undertook the initial research. These losses can be large, given the costs of pharmaceutical research. Moreover, not using the chemical, when it would be safe to do so, may adversely impact those who could benefit from its use; these people too may number in their millions and their (unrelieved) misery may also be great. Again, how can such consequences be quantified?

Moreover, as Talbot Page noted in 1978 [32], and as this generic example illustrates, the consequences of the errors may be asymmetric. Different people may suffer under each type of error, and, indeed, some people may even gain (e.g. competitors to the company developing a chemical wrongly refused approval). The consequences also commonly manifest asymmetries in their relative timings, durations and degrees, all of which may be hard to estimate. Even envisioning the possible consequences may be difficult. For instance, although Thalidomide was tested on animal and human subjects before being sold commercially, none of the tests involved pregnant subjects [46], presumably because no one thought to do so.

As mentioned above, the standard approach to statistical hypothesis testing in areas of public policy ignores any consideration of the consequences of inference errors. Ignoring them in this way effectively treats the consequences of each type of error as equal, which may or may not be valid or desirable. Whether desirable or not in any particular case, this is a decision for society as a whole to take, informed by the views of those potentially impacted. It should not be a decision solely for scientists to make or, worse, a decision made by default as a result of a standard testing procedure applied unthinkingly.

Apart from Page's article, we know of no public discussion of the issue of the consequences of inference errors.[4] However, other questions arising from the use of the standard Neyman-Pearson approach in the bio-medical and social sciences have been raised for some time [2, 30]. One serious flaw is the mechanistic use of the threshold values for α (typically at 5% or 1%) while ignoring the value of β. As mentioned above, formal hypothesis testing theory was developed in association with agricultural experiments following WWI. Typically, a new crop variety was developed and tested to see if it resulted in higher yields than did existing varieties. Setting the null hypothesis to be the hypothesis of equality of yields from the two varieties meant that a Type I error would have resulted in

[4]In 1980, the first author worked on the design of a statistical study to assess the impacts of exposure to certain herbicides, where he argued unsuccessfully for the consequences of inference errors to inform the study design. Other than Page, the only published discussion we know is an article which uses a maximum-expected-utility formulation of the hypothesis testing decision to argue that the process is inherently subjective in environmental risk domains [47]. Current work by philosopher Deborah Mayo on meta-rules for hypothesis testing in risk assessment is related [28].

the release of the new variety onto the market when in fact it was no better than the existing variety. A Type II error would have resulted in not releasing it when in fact it was an improvement. In these circumstances, it was generally thought preferable to wrongly forego a new improved crop variety (Type II error) than to wrongly deploy a new variety offering no improvement (Type I error). Hence the probability α was set at a specific low level (5% or 1%) and β minimised (through choice of test), rather than the other way round.

This structure is not necessarily appropriate for other domains, even without detailed articulation of the consequences of inference errors. Suppose, for instance, we were testing the carcinogenicity of some new chemical, and we set the null hypothesis to be that the chemical has no carcinogenic effects versus an alternative hypothesis that it does. Here, the Type I error would result in a conclusion of carcinogenicity when this was not in fact the case, while the Type II error would see a conclusion of no carcinogenicity when in fact this was the case. Arguably, the general consequences of a Type II error here are more serious than for a Type I error, which should lead to a test procedure which fixed the value of β at a low level and aimed to then minimise α.[5]

Moreover, ignoring these issues can lead to a high risks of erroneous conclusions being drawn from the testing process. For instance, one study reexamined 71 randomised-control medical therapy trials, where each had concluded in favour of a null hypothesis of no effect, i.e. each reported a "negative" scientific finding about the respective medical therapy under consideration [13]. This study found that, due to the small sample sizes used, no fewer than 67 of the 71 trials had a β value of greater than 10% when the alternative hypothesis was that the treatment under study was 25% more effective than the control. In other words, in these 67 studies, even if the treatment being tested was 25% more effective than the control treatment, there was a greater than 10% chance that the hypothesis testing procedure would not detect such a difference.

Because of issues such as these, many scientists and statisticians have argued for the use of p-values rather than hypothesis tests in published reports [40], and many journals now require these. With this approach, the scientific researcher does not conclude either in favour or against the null hypothesis, but instead merely reports the probability that a value of the test statistic equal to or more extreme than the one obtained from the sample would have been observed if the null hypothesis were really true. Thus, the smaller the p-value, the less likely it would be the case that the null hypothesis could generate that particular test statistic value obtained from the sample.[6]

This approach avoids the scientific researcher having to choose between competing hypotheses using a possibly-inappropriate test procedure. Public policy makers do not have the same luxury. Ultimately, policy decisions need to be made, and, in the risk domain, these will hinge on the acceptance or rejection of scientific hypotheses [16, 39]. The primary danger is that the statistical tests of hypotheses are applied using the standard approach in a mechanistic fashion

[5] Of course, this would be the result were one to reverse the null and alternative hypotheses, but standard practice is to make the null hypothesis the hypothesis of no effect.

[6] Although the use of p-values is also not uncontroversial. See, for example: [15].

(e.g. using a 5% acceptance vs. rejection threshold for p-values), and without regard to the subjective judgments inherent in this standard approach.

3 Dialectical Argumentation

Faced with the difficulties of quantitative risk assessment, consideration in recent years has turned to qualitative approaches, particularly using methods developed over the last 25 years in Artificial Intelligence.[7] Leading these developments has been the team under John Fox at the Advanced Computation Laboratory of the Imperial Cancer Research Fund (ICRF), in London.[8]

One approach adopted there has been the use of argumentation to reason about the possible risks of a substance or a course of action [9, 10, 24, 25, 34]. Using a generic model first proposed by the philosopher Stephen Toulmin to represent an argument [49], this approach formalises concepts such as an argument's premises, its force, rebuttals, etc, and then develops an algebraic calculus for argument manipulation [11, 35]. Once formalised, argumentation can then be deployed in intelligent computer systems designed to undertake autonomous reasoning about some knowledge domain. For example, intelligent systems now exist for predicting the risks of: pesticide toxicity [5]; chemical carcinogenicity [48]; food chemical toxicity [23]; and breast cancer [8]. These computer systems are intended to assist human decision-makers, rather than to replace them, by collating, clarifying and aggregating the arguments relevant to a particular decision.

These systems have generally used argumentation as if from the perspective of an omniscient observer, rationally comparing cases for and against a claim, and combining these to produce a single, coherent argument. But science policy debates are usually contentious, with different participants having different interests, values, and preferences, and even different modes of reasoning [17, 22, 53]. For example, a recent research project in Britain sought explicitly to identify the criteria which different groups of people believed to be important in any risk assessment of Genetically-Modified Organisms (GMOs) [45]. A related branch of philosophy, dialectical argumentation, has sought to develop frameworks for the conduct of debates on contentious issues between reasonable, consenting participants [3, 12, 38].

One influential framework for dialectical argumentation has been that proposed by the German philosopher, Jürgen Habermas [18]. Originally seeking to understand how ethical norms could be agreed between different people, and building on Toulmin's work, Habermas proposed a framework in which consenting members of a community can engage in a civil discourse. A key feature

[7] For example, a recent U.S. conference organised by the Society for Risk Analysis had the title "Uncertainty: Its Nature, Analytical Treatment and Interpretation," and included contributions from researchers in artificial intelligence, philosophy and statistics, as well as senior Government risk assessors and policy-makers. (See: http://www.ramas.com/feb1011.htm.) For an introduction to qualitative reasoning, see: [33].

[8] The second author is a Visiting Research Fellow at the ICRF.

of a discourse is that each proponent presents his or her assertions to an audience. An audience needs to be persuaded, and listeners may withhold their agreement to the claims being advanced. Indeed members of an audience may advance counter-claims of their own, or rebuttals or undercutting arguments, or may question the premises or the modes of inference used by a proponent. Habermas sought to identify formal and agreed rules under which such debate could occur in a civil manner and so that all reasonable participants would feel satisfied with the conduct of the debate, even when they disagreed with the conclusions reached. Examples of the types of rules he proposed were: "Different speakers may not use the same expression with different meanings" and "Everyone is allowed to question any assertion whatever."

Habermas applied his framework to debate in politics, law and the social sciences [19, 20] and others have applied it to scientific discourses (e.g. [37]). We have recently proposed the use of dialectical argumentation for intelligent systems which can reason about claims of chemical carcinogenicity when evidence sources conflict, and have taken initial steps to its formalization [29]. Our formalization revolves around the notion of an "agora,"[9] a formal mathematical structure in which arguments can be constructed, compared and combined, and their acceptability assessed. We believe there is potential for the use of a similar approach, a risk agora, as a normative framework for debates about the risks of new substances and technologies. We feel that this would be particularly useful for debates regarding the potential consequences of errors of statistical inference. With appropriate modification for this arena, the risk agora could enable the elucidation and articulation of consequences of alternative courses of action, and thereby better enable their comparison.

What sort of rules would be contained in a risk agora? Table 1 below presents an indicative list. Many of these rules will require further modification for the risk domain. For example, the framework needs to describe precisely the types of claims permitted. These will include: claims that something is a potential hazard; claims about a potential consequence of a previously-claimed hazard or an inference error; claims about the incidence (who suffers or gains) of a previously-claimed consequence; and claims about the duration of a consequence. Development of a full and comprehensive list of proposed rules, together with complete explanation and justification, is a major undertaking, and is one aspect of our ongoing work.

Table 1: Indicative List of Rules for The Risk Agora: A Framework for Debate about Risk

- Rules as to who may participate, when, in what guise.
  - For example: independent expert, corporation, lay-person, policy-maker.
  - Identifying roles against participants will better enable others to assess their contributions.

[9] From the Greek for "meeting place."

- Rules as to withdrawal from the debate (who, when, how).
- General rules as to the form that participation may take, e.g.
  - Everyone must speak/write in English
  - New technical terms must be accompanied by a definition
  - Everyone must use agreed definitions of terms.
- Specific rules about the "moves" each participant may make, e.g.
  - Assertion of a definition
  - Assertion of a claim
  - Assertion of a value or a preference
  - Assertion of an open-question
  - Assertion of a research agenda (e.g. to resolve an open question)
  - Presentation of an argument and premises for a claim
  - Query for information
  - Query for argument for a claim
  - Challenge to a definition
  - Challenge to claim (If reasons are required for a challenge, this would be a rebuttal.)
  - Challenge to a sub-ordinate claim (i.e. If with reasons, then this would be an undercut.)
  - Challenge to a mode of reasoning used in an argument
  - Challenge to a premise used in an argument
  - Acceptance of a claim, etc
  - Retraction of a claim, etc.
- Rules for what moves may follow what others, when.
- Rules for aggregation, consolidation, resolution, etc of arguments.
- Rules for ending the debate.

We see a number of benefits from adoption of the risk agora. Foremost among these is that it would provide a rigorous framework for identifying the consequences of potentially risky decisions, somewhat analogous to an Environmental Impact Statement which some countries (e.g. Australia) now require to be undertaken before any major change to the physical environment. Because the agora would force proponents in a debate to specify the detailed assumptions and the exact modes of reasoning that they use, the precise differences between contending theories will be exposed along with any weaknesses in the modes of reasoning used.[10] This will, in turn, increase the transparency of public policy decision making, so that all the arguments for and against are presented; this may also help to facilitate resolution.[11] In addition we hope that it will provide a means to widen the knowledge sources used in scientific debate, including incorporation of lay (non-scientific) knowledge. As a result we believe that the approach has the potential to aid and broaden democratic participation in debates about environmental and health risk, something we see as crucial. Wynne [53] has argued that lay people often have specific local knowledge which scientists — in their rush to generalise and standardise knowledge — often ignore at their peril, and we agree with him that it is important to incorporate such knowledge into the decision-making process. In this respect, the risk agora approach may help to dethrone science from its modern-day pedestal of rationalistic and objective perfection, where the picture it presents to the non-scientific world is one we might call "scientific realism" (after the term "socialist realism," from art criticism) — an idealised and false depiction of scientific research reality.[12]

4 Discussion

As engineers, our interest in this area is in developing intelligent computer systems which can support human decision-making in complex domains. In the case of the potential risks associated with new technologies and substances, the decisions involved are public policy ones rather than corporate or personal decisions. Even so, we believe those involved in the decision can benefit from the support of intelligent systems. Hence, once an effective risk agora has been formalised, we intend to embody it in just such an intelligent computer system. As an example of the possibilities of this approach, the U.S. Government is currently developing a similar intelligent system to assist national defence agencies manage global security crises [42]. This system, called GENOA, will seek to predict geopolitical risks and then support the operations of the inter-agency teams of national security personnel tasked with crisis management. It is important to recognise that no such system should, or indeed could, replace human decision-makers, but will only act to support them.

We believe that an intelligent system that embodies a risk agora would have many of the advantages anticipated for GENOA. Such a system would facilitate clarity in debates about risk, ensuring that disputants were clear about each other's premises, arguments, counter-arguments, and modes of inference. Also, the system would be able to keep track of the complexity of the domain in which the debate was taking place more effectively than could the human participants. For example, in assessing risk of carcinogenicity, data on tens of thousands of

[10] For example, any fallacious modes of reasoning should be exposed, a situation we term "syllogism abuse."

[11] A philosopher of science, Steve Fuller, has proposed three roles for someone seeking to assist resolution of a conflict in a knowledge domain: Facilitator, Negotiator and Arbitrator [14]. Our proposed system could potentially assist in all three of these roles.

[12] Although, in Britain, science probably fell off the pedestal in the course of the BSE-CJD debate.

chemicals and scores of experiments may be relevant, far more than even the most expert human can reasonably deal with. Such a system would also be able to manage the interactions between the various components and issues under debate, allowing updated evidence to be applied consistently, and allowing the running of "what if" scenarios (effectively, a form of sensitivity analysis). Thus, such systems could support environmental and health-safety agencies in identifying and modeling chemical and technological risks, and in developing appropriate regulatory responses. Made publicly available, such systems could also be used by ordinary people (not those involved in the regulatory process) to decide the issues for themselves, by personally judging the relevant arguments presented in the Agora.

A related issue here is arguing from authority. Traditionally, logicians and philosophers have treated arguments from authority — for instance, saying that homosexuality should be outlawed because the Bible says so[13] — as fallacies. However, Charles Willard [52], without wanting to argue for the validity of arguments like the homosexuality/Bible one, argues that life and its problems are now so complex that even experts are non-experts outside their own sphere. Consequently, all of us must, at times, accept arguments on grounds of authority. As an everyday example, most people do this unthinkingly with medicine, since they are usually happy to accept a doctor's opinion rather than asking to see the original medical research papers which support the choice of particular treatments. As a result, Willard argues, a key task we all face is that of judging experts. The risk agora system could also help us judge an expert in a particular risk domain on the basis of the arguments s/he presents in the system.

To implement the ideas presented here for a risk agora in an intelligent computer system will require advances in the relevant artificial intelligence technologies. To this end, we have recently proposed a high-level architecture for such computer systems [29, 36], and have formalised an argumentation system for handling dialogues between two disputants [1]. We believe this formalization can be generalised and adopted as the mechanism underlying the risk agora, and we are pursuing this line of research. It is important to restate, however, that we see such systems as supporting, not replacing, human decision-making. Were they to be used by risk assessment agencies for policy development, then appropriate safeguards, accountabilities and appeal processes would need to be established to ensure that decisions were not delegated to machines.

The fact that all of us are potentially impacted by environmental and health hazards means that we all should have a role in determining appropriate responses to them. This means that public policy in risk domains must be transparent and accountable, which means, in turn, that all of us should have access to, and understanding of, the scientific conclusions underpinning public policy decisions. The extent that such conclusions depend upon subjective analysis and interpretation by scientists and policy-makers, and the extent they involve assumptions, modes of inference or conclusions that are contested by other sci-

[13] This argument is, of course, doubly questionable, since finding such an argument in the Bible in the first place requires some fairly careful interpretation.

entists, should be known to - or, at least, knowable by - all of us. It is for this reason that we see great value in a formal and agreed framework for articulating, comparing and assessing arguments on matters of environmental and health risk.[14]

Acknowledgments

This work was partly supported by the British Engineering and Physical Sciences Research Council (EPSRC) under grant GR/L84117, and the first author is supported by an EPSRC studentship. This support is gratefully acknowledged. We are also grateful for discussions with Chip Heathcote on issues of statistical inference, although the views expressed here are solely those of the authors.

References

[1] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In E. Durfee, editor, *Proceedings of the 4th International Conference on Multi-Agent Systems (ICMAS-2000)*, Boston, MA, USA, 2000. IEEE Press.

[2] L. Atkins and D. Jarrett. The significance of "significance tests". In I. Miles J. Irvine and J. Evans, editors, *Demystifying Social Statistics*. Pluto Press, London, UK, 1979.

[3] E. M. Barth and E. C. W. Krabbe. *From Axiom to Dialogue: A Philosophical Study of Logics and Argumentation*. Walter de Gruyter, Berlin, Germany, 1982.

[4] U. Beck. Risk society and the provident state. In S. Lash, B. Szerszynski, and B. Wynne, editors, *Risk, Environment and Modernity: Towards a New Ecology*, pages 27–43. Sage, London, UK, 1996.

[5] R.D. Coombes and P. N. Judson. Artificial intelligence systems for predicting toxicity. *Pesticide Outlook*, 7 (4):11–15, 1996.

[6] D. R. Cox and D. V. Hinkley. *Theoretical Statistics*. Chapman and Hall, London, UK, 1974.

[7] U.K. Economic and Social Research Council (ESRC) Global Environmental Change (GEC) Programme. The Politics of GM Food: Risk, Science and Public Trust. Special Briefing 5, University of Sussex, Brighton, UK, 1999.

[14] Arguably, the controversy in Britain over GMOs, which led, earlier in 1999, to the Government re-organizing its advisory structures and processes, and planning a new programme of scientific experiments, is indicative of the absence of an agreed, formal process for conducting public debates on these issues [7].

[8] J. Emery, R. Walton, A. Coulson, D. W. Glasspool, S. Ziebland, and J. Fox. Computer support for recording and interpreting family histories of breast and ovarian cancer in primary care (RAGs): qualitative evaluation with simulated patients. *British Medical Journal*, 319:32–36, 1999.

[9] J. Fox. Will it happen? Can it happen? A new approach to formal risk analysis. *Risk, Decision and Policy*, 4, August 1999.

[10] J. Fox, P. Krause, and S. Ambler. Arguments, contradictions and practical reasoning. *Proceedings of the European Conference on Artificial Intelligence 1992, Vienna, Austria*, 1992.

[11] J. Fox and S. Parsons. Arguing about beliefs and actions. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms*, pages 266–302. Springer Verlag (Lecture Notes in Artificial Intelligence 1455), Berlin, Germany, 1998.

[12] J. B. Freeman. *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure.* Foris, Berlin, Germany, 1991.

[13] J. A. Freiman, T. C. Chalmers, H. Smith, and R. R. Kuebler. The importance of beta, the Type II error, and sample size in the design and interpretation of the randomised control trial. Survey of 71 "negative" trials. *New England Journal of Medicine*, 299 (13):690–694, 1978.

[14] S. Fuller. *Philosophy, Rhetoric, and the End of Knowledge: The Coming of Science and Technology Studies.* University of Wisconsin Press, Madison, WI, USA, 1993.

[15] S. N. Goodman and R. Royall. Evidence and scientific research. *American Journal of Public Health*, 78:1568–1574, 1988.

[16] J. D. Graham, L. C. Green, and M. J. Roberts. *In Search of Safety: Chemicals and Cancer Risk.* Harvard University Press, Cambridge, MA, USA, 1988.

[17] R. Grove-White. Environmental knowledge and public policy needs: on humanising the research agenda. In S. Lash, B. Szerszynski, and B. Wynne, editors, *Risk, Environment and Modernity: Towards a New Ecology*, pages 269–286. Sage, London, UK, 1996.

[18] J. Habermas. *The Theory of Communicative Action: Volume I: Reason and the Rationalization of Society.* Heinemann, London, UK, 1984. *(Translated by T. McCarthy.).*

[19] J. Habermas. *Justification and Application: Remarks on Discourse Ethics.* MIT Press, Cambridge, MA, USA, 1993. *(Translated by C. P. Cronin.).*

[20] J. Habermas. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy.* MIT Press, Cambridge, MA, USA, 1996 *(Translated by W. Rehg.).*

[21] L. Hogben. *Statistical Theory.* W. W. Norton, 1957.

[22] D. Jamieson. Scientific uncertainty and the political process. *Annals of the American Academy of Political and Social Science*, 545:35–43, 1996.

[23] P. N. Judson. Expert systems for hazard evaluation. In D. Tennant, editor, *Food Chemical Risk Analysis*, pages 109–132. Chapman and Hall, London, UK, 1997.

[24] P. Krause, J. Fox, and P. Judson. An argumentation based approach to risk assessment. *IMA Journal of Mathematics Applied in Business and Industry*, 5:249–263, 1994.

[25] P. Krause, J. Fox, P. Judson, and M. Patel. Qualitative risk assessment fulfils a need. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms*, pages 138–156. Springer Verlag (Lecture Notes in Artificial Intelligence 1455), Berlin, Germany, 1998.

[26] C. Lau. Neue risiken und gesellschaftliche konflikte. In U. Beck, editor, *Politik in der Risikogesellschaft.* Surhkamp, Frankfurt-am-Main, Germany, 1991.

[27] E. L. Lehmann. *Testing Statistical Hypotheses.* Springer Verlag, New York, NY, USA, second edition, 1997.

[28] D. Mayo. Uncertainty and values in the scientific assessment of risk. In *SRA Forum on Uncertainty, February 2000*, Arlington, VA, USA, 2000. Society for Risk Analysis.

[29] P. McBurney and S. Parsons. Truth or consequences: using argumentation to reason about risk. *Symposium on Practical Reasoning, British Psychological Society, London, UK*, 1999.

[30] D. E. Morrison and R. E. Henkel, editors. *The Significance of Significance Tests: A Reader.* Butterworths, London, UK, 1970.

[31] J. Neyman and E. S. Pearson. *Joint Statistical Papers.* Cambridge University Press, Cambridge, UK, 1967.

[32] T. Page. A generic view of toxic chemicals and similar risks. *Ecology Law Quarterly*, 7 (2):207–244, 1978.

[33] S. Parsons. *Qualitative Approaches to Reasoning Under Uncertainty.* MIT Press, Cambridge, MA, USA, 2000 (Forthcoming).

[34] S. Parsons, J. Fox, and A. Coulson. Argumentation and risk assessment. In *AAAI Spring Symposium on Predictive Toxicology.* AAAI Press, Stanford, CA, USA, 1999.

[35] S. Parsons and S. Green. Argumentation and qualitative decision making. In A. Hunter and S. Parsons, editors, *The 5th European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (EC-SQARU99)*, pages 328–339. Springer Verlag (Lecture Notes in Artificial Intelligence 1638), Berlin, Germany, 1999.

[36] S. Parsons and P. McBurney. Risky business. *Scenario and Strategy Planning*, 1 (3):15–17, 1999.

[37] W. Rehg. Argumentation theory and the philosophy of science since Kuhn: the rationality of theoretical advance. Paper presented to Seminaret i Vitenskapsteori, Universitetet i Oslo, Oslo, Norway, August 1998.

[38] N. Rescher. *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge*. State University of New York Press, Albany, NY, USA, 1977.

[39] L. R. Rhomberg. A survey of methods for chemical health risk assessment among Federal regulatory agencies. *Human and Ecological Risk Assessment*, 3 (6):1029–1196, 1997.

[40] K. J. Rothman and S. Greenland. *Modern Epidemiology*. Lippincott-Raven, Philadelphia, PA, USA, second edition, 1998.

[41] S. Ryle. Insurers balk at risks of phones. *The Observer (London, UK)*, 11 April 1999.

[42] G. I. Seffers. Crisis system for President: DARPA's Project Genoa would speed U.S. decision-making. *Defense News*, 13 (13):1, 30 March 1998.

[43] T. Seidenfeld. *Philosophical Problems of Statistical Inference*. D. Reidel, Dordrecht, The Netherlands, 1979.

[44] M. E. Shere. The myth of meaningful environmental risk assessment. *Harvard Environmental Law Review*, 19 (2):409–492, 1995.

[45] A. Stirling and S. Mayer. *Rethinking Risk: A Pilot Multi-Criteria Mapping of a Genetically Modified Crop in Agricultural Systems in the UK*. Report, SPRU, University of Sussex, Brighton, UK, 1999.

[46] H. Teff and C. R. Munro. *Thalidomide: The Legal Aftermath*. Saxon House, Westmead, Farnborough, Hampshire, UK, 1976.

[47] J. E. Toll. Elements of environmental problem-solving. *Human and Ecological Risk Assessment*, 5(2):275–280, 1999.

[48] C. A. G. Tonnelier, J. Fox, P. N. Judson, P. J. Krause, N. Pappas, and M. Patel. Representation of chemical structures in knowledge-based systems: the StAR system. *Journal of Chemical Information and Computer Sciences*, 37:117–123, 1997.

[49] S. E. Toulmin. *The Uses of Argument.* Cambridge University Press, Cambridge, UK, 1958.

[50] A. Wald. *Statistical Decision Functions.* Wiley, New York, NY, USA, 1950.

[51] N. Watt. Millions still at risk from CJD. *The Guardian (London, UK)*, 22 September 1999.

[52] C. A. Willard. Authority. *Informal Logic*, 12:11–22, 1990.

[53] B. Wynne. May the sheep safely graze? A reflexive view of the expert-lay knowledge divide. In S. Lash, B. Szerszynski, and B. Wynne, editors, *Risk, Environment and Modernity: Towards a New Ecology*, pages 44–83. Sage, London, UK, 1996.