Dialectical Argumentation for Reasoning about Chemical Carcinogenicity

Peter McBurney and Simon Parsons
Department of Computer Science
University of Liverpool
Liverpool L69 7ZF U.K.
{p.j.mcburney,s.d.parsons}@csc.liv.ac.uk

Abstract

We aim to build intelligent systems which can reason autonomously about the carcinogenicity of chemicals. Scientific debates in this area draw on evidence from multiple, and often conflicting sources, both theoretical and experimental, and participants use various modes of inferential reasoning. In seeking to automate such reasoning, we have first articulated precisely the multiple modes of inference used when an assertion of human carcinogenicity is made from experimental animal evidence. Because such inferences are often contested, scientific debate in this domain can be vigorous. To model such debates, we propose the use of a form of dialectical argumentation, drawing on Habermas' philosophy of Discourse Ethics [9] and Pera's philosophy of science [18]. The resulting formalism permits the representation of uncertainty and disagreement regarding the modes of inference used, as well as the claims being asserted.

KEY WORDS: Dialectical Argumentation, Inference, Qualitative Reasoning, Risk Assessment.

1 INTRODUCTION

Claims of chemical carcinogenicity or toxicity of a substance can be based on evidence from a number of sources [8, 28], including: (a) chemical theoretical reasoning, comparing the chemical structure of the substance with that of a known carcinogen; (b) experiments where the substance is applied to tissue-cultures in laboratories (mutagenic tests), or to human or animal cadavars; (c) bioassays, applying the substance to living animals in a laboratory experiment; (d) epidemiological studies of humans; and (e) theoretically-sound descriptions of bio-medical causal pathways.

To construct intelligent systems which can reason from evidence sources such as these, we first need to formalize the reasoning used by scientists when claims of carcinogenicity are made. In Section 2 we present twelve distinct modes of inferences deployed when carcinogenicity claims are made on the basis of bioassay evidence. To our knowledge, these modes of inference have not been comprehensively articulated before. This is surprising, because science policy debates often include attacks by participants on the modes of inference used by others, particularly before a theoretically-sound causal mechanism has been agreed.[1]

Evidence from different sources may conflict, and carcinogen risk assessment usually involves the comparison and resolution of multiple and diverse evidence [7, 28]. In representing this domain, it is therefore appropriate to use some form of argumentation (so that the reasons for claims can be represented in association with the claims themselves), and within a dialectical framework (so that cases for and against a particular claim can be compared). An argumentation formalism also permits the representation of both quantitative and qualitative information in the reasoning process. This paper proposes the use of a dialectical argumentation formalism from moral philosophy, originally due to philosophers Habermas [9] and Alexy [1], along with a model of rational scientific enquiry due to Pera [18], a philosopher of science. Our framework is outlined in Section 3, and an example is presented in Section 4. Section 5 concludes.

2 INFERRING CARCINOGENICITY

2.1 Modes of inference

Evidence of carcinogenicity of a substance can be derived from many sources. For reasons of cost, convenience and speed, laboratory tests of the substance on animal species (bioassays) are a common source of evidence [7, 8, 22]. Because of the difficulties in inferring conclusions about humans on the basis of evidence about animal species, most cautious scientists and policy makers would not *assert* carcinogenicity to humans from a bioassay: they would, at best, only claim that there is a (perhaps high) probability of human carcinogenicity.[2] However, although it is perhaps the most contentious, the animal-to-human inference is not the only inference being deployed in concluding such a probability. It is also not the only inference deployed when quantifying the extent of risk. It therefore behooves us to examine all the modes of inference used. In doing so, we have abstracted from a number of descriptions and critiques of carcinogenic risk assessment processes, including [7, 8, 15, 22, 25, 28], both ideal and actual.

For the purposes of exposition, we therefore suppose an archetypical animal bioassay for a chemical substance $\mathcal{X}$ is undertaken. This will involve the administration of specific doses of $\mathcal{X}$ to selected animal subjects, usually repeatedly, in a laboratory environment. Typically, two or three non-zero dose-levels are applied to the subject animals, along with a zero-dose to the control group. The rates at which cancers of a specific nature develop is then observed in each group until a pre-determined time-point (usually the natural life-span of the animal). Those animals still alive at that time are then killed, and a statistical analysis of the hypotheses that exposure to the substance $\mathcal{X}$ results in increased incidence of cancer is then undertaken. Suppose that, based on this animal bioassay, a claim is then made that $\mathcal{X}$ is carcinogenic to humans at a specified dose. For ease of expression we will notate this claim by ϕ. In asserting ϕ

[1] As an example of debate concerning the possible causes of a cluster of cancer cases, see the summary in [3].

[2] Indeed, the U.S. Environmental Protection Agency guidelines [28] permit one to claim probable human carcinogenicity from (sufficiently strong) animal evidence alone. Although such a claim would be classed in the second of two categories of "probable", it is still above "possible" human carcinogenicity.

from the evidence of the bioassay, a number of subsidiary inferences need to be made, numbered below as R1 to R12. We have expressed these in the form of *"FROM antecedent TO consequent"*. This is short-hand for saying that an act of inference is undertaken whenever one assumes that the consequent is true (or takes a particular value) upon the antecedent being true (or, respectively, having taken a corresponding value).

R1: FROM Administered dose TO Delivered dose. Animal bodies defend themselves against foreign substances. For example, chemicals applied to nasal tissues are initially repelled by defences in the tissues themselves. Larger doses may destroy this first line of defence, thereby permitting proportionately more of the chemical to enter the body's circulatory pathways than would occur for smaller doses. In other words, the dose delivered to the target tissue or organ of the body may not be proportionate to the dose administered to the animal by the experimenter.

R2: FROM A sample of animals TO A population of the same species. Reasoning from a sample to a population from which the sample is drawn in known as statistical inference.

R3: FROM A genetically uniform animal population TO A genetically more diverse population. For reasons of experimental control and of convenience, animal subjects used in laboratory experiments are often more closely related genetically than is the natural population as a whole.

R4: FROM An animal population TO The human population. Animals differ from humans in their physiology and in their body chemistry, so it is not surprising that they also differ from us in reactions to potential carcinogens. Indeed, different species differ from each other. Formaldehyde, for instance, was found to cause significant nasal cancers in rats but not in mice [7], while epidemiological studies of humans whose professions exposed them to high levels of the chemical found no significant increases in such cancers.

R5: FROM A site specificity in bioassay animals TO A possibly different site specificity in humans. Most chemicals are pre-carcinogens which must be altered by the body's metabolic processes into an actively carcinogenic form. This happens differently in different species, because the body-chemistries are different or because the physiology or relative sizes of organs are different.

R6: FROM Localised exposure TO Broader exposure. Bioassays adminster a chemical to a specific site in a specific way to the subject animals, as for example, in bioassays of formaldehyde applied to nasal passages to test for nasal cancer. In contrast, humans exposed to it may receive the chemical in a variety of ways.

R7: FROM Large doses TO Small doses. At typical levels of exposure, the incidences of most individual cancers in the general population are quite small, of the orders of a few percent or much less. At equivalent dose levels, then, bioassays will require very large sample sizes to detect statistically significant increases in cancer incidence. This would be prohibitively expensive, and so most bioassays administer doses considerably greater than the equivalent doses received (allowing for the relative sizes of the animal and human species) in the environment. In order to assert carcinogenicity, then, a conversion model — a dose-response curve — is required to extrapolate back from large to small dose levels. The extent of carcinogenicity asserted can be very sensitive to the dose-response model used. Two theoretically-supported models for the risks associated with aflatoxin peanuts, for example, show human risk likelihood differing by a factor of 40,000 [19].

R8: FROM An animal dose-level TO A human equivalent. The previous paragraph used the phrase "allowing for the relative sizes of the animal and human species". But how is this to be done? Is the dose extrapolated according to relative body weights of the two species (animal and human); or skin surface area (which may be appropriate for chemicals absorbed through the skin); or relative size of the organ affected? What is appropriate if different organs are affected in different species?

R9: FROM Administered doses TO Environmental exposure. In order to expedite response times, bioassays may adminster the chemical in a manner different to that likely to be experienced by humans exposed to it in their environment. For example, the chemical may be fed via a tube directly into the stomach of the animal subject, which is unlikely to be the case naturally.

R10: FROM A limited number of doses TO Cumulative exposure. Some chemicals may only produce adverse health effects after a lifetime of accumulated exposure. Body chemistry can be very subtle, and a small number of large doses of a chemical may have a very different impact from a much larger number of smaller doses, even when the total dose received is the same in each case.

R11: FROM A pure chemical substance TO A chemical compound. Most chemicals to which people are exposed are compounds of several chemicals, not pure substances. Bioassay experiments, however, need to be undertaken with pure substances, so as to eliminate any spurious causal effects.

R12: FROM The human population TO Individual humans. Individuals vary in their reactions to chemical stimuli, due to factors such as their genetic profiles, lifestyles, and personalities. Risks of carcinogenicity may be much higher or much lower than claimed for specific groups or individuals.

To claim human carcinogenicity on the basis of evidence from a bioassay thus depends on a number of different modes of inference, each of which must be valid for the claim to stand. We could write:

"The chemical $\mathcal{X}$ is carcinogenic to humans at dose d based on a bioassay of animal species a if:

- *There is a relationship between administered dose and delivered dose in the bioassay, AND*
- *The sample of animals used for the experiment was selected in a representative manner from the population of animals, AND*
- *The animal population from which the sample was drawn is as genetically diverse as the animal population as a whole,"*

$\vdots$

and so on, through the remaining nine inference steps.

It is important to note that even if all modes of inference were valid in a particular case, our assertion could, strictly speaking, only validly be that the chemical $\mathcal{X}$ is associated with an increase in incidence of the particular cancer. The assertion ϕ does not articulate, nor could a bioassay or epidemiological study prove, a causal pathway from one to the other. There may, for example, be other causal factors leading both to the presence of the chemical in the particular environment and to the observed carcinogenicity.

For the archetypical analysis above, we began with the assumption of just one bioassay being used as evidence to assert a claim for carcinogenicity. In reality, however, there is often evidence from more than one experiment and, if so, statistical meta-analysis may be appropriate. This may involve pooling of results across different animal species, or across both animal and human species. None of these tasks are straighforward, and will generally involve further modes of inference, which we do not explore in this paper.

2.2 The example of statistical inference

Only one of the forms of inference listed in the previous example is Statistical Inference, that is, reasoning about a population on the basis of evidence from a sample of that population. Strictly according to logic, statistical inference is unsound: true antecedents are not guaranteed to generate true consequents. However, the key achievement of mathematical statistics this century has been to place a bound on the extent of unsoundness: if we know the probability distribution of the variable of interest in the population, and that the mechanism which generated the sample was random (or, if not, the extent to which it is not), then we can estimate the probability that the inference from sample to population is incorrect. For example, we may conclude from particular functions of the sample values that there is a 95% chance that a certain interval contains the mean of the population.[3] This form of inference is still unsound (i.e. we still cannot guarantee the truth of a claim about a population parameter, given the truth of a claim about a sample parameter), but we now have an estimate of the upper bound on the extent of unsoundness. If we (as a society) make decisions based on the sample data, we still do not know which decisions are correct and which wrong, but we can estimate how many of the latter there will be at most. We are better off as a result.

[3] There are contending views within statistical theory about the meaning of a statement such as this, a debate we do not enter [2, 24].

The same would be true for the other modes of inference listed above. None of these is sound, in the sense of guaranteeing the preservation of truth (from antecedent to consequent) in all circumstances. But, just as with statistical inference, if we were to have a quantitative bound on the possible error in inference, then we would be better off than without it. Moreover, if such bounds existed for all the inferential modes listed, it may be possible to combine these bounds in an appropriate way, thereby generating a bound for the overall assertion of carcinogenicity from a bioassay. Estimating the soundness of each type of inference could be a matter of detailed examination of all the experimental and theoretical evidence (which may be a considerable undertaking) and then using this to develop a framework for theory development relevant to the mode of inference. Such theories would be analogous to the theories (e.g. Neyman-Pearson, Bayesian, Fiducial) supporting the use of statistical inference.

3 ARGUMENTATION FRAMEWORKS

3.1 Monodic and dialectical argumentation

An argument for a claim may be considered as a tentative proof for the claim. The philosopher Stephen Toulmin [27] proposed a generic framework for the structure of arguments which has been influential in the design of intelligent systems which use argumentation [4, 30]. Our analysis, informed by Toulmin's structure, considers an argument to have the form of a proof, without necessarily its force.

Suppose ϕ is a statement that a certain chemical is carcinogenic at a specified level of exposure. Then an argument for ϕ is a finite, ordered sequence of inferences $G_\phi = (\phi_0, \phi_1, \phi_2, \ldots, \phi_{n-1})$. Each sub-claim ϕ_i is related to the preceding sub-claim ϕ_{i-1} in the sequence as a result of the application of an inference rule, R_i. These rules correspond to warrants in Toulmin's schema. Note that R_i and R_j may be the same rule for i and j different. The modes of inference listed in Section 2.1 are examples of such rules. We may present this sequence graphically as follows:

$$\phi_0 \xrightarrow{R_1} \phi_1 \xrightarrow{R_2} \phi_2 \longrightarrow \ldots \longrightarrow \phi_{n-1} \xrightarrow{R_n} \phi.$$

If any of these rules were rules of inference generally considered valid in deductive logic (Modus Ponens, say), then we would be confident that truth would be preserved by use of the rule. In other words, using a valid rule of inference at step i means that whenever ϕ_{i-1} is true, so too is ϕ_i. If all the rules of inference were valid in this sense, then the argument G_ϕ would constitute a deductive proof of ϕ. The situations of interest to us, however, are when some or all of the inference rules are not valid in this sense, such as those of Section 2.

In pure mathematics in general, once a theorem has been proven true, further proofs do not add to its truth, nor to the extent to which the theorem is believed to be true. With arguments, however, alternative arguments may be of great interest. The greater the number of independent arguments that exist for a claim, the stronger is the case for it, and the stronger may be our belief in its truth. However, in arriving at a considered view as to our belief in the truth of a claim ϕ, we also need to consider the arguments against it, the arguments in favour of its negation $\neg\phi$ (which may be different), and any arguments which attack its supporting sub-claims, ϕ_i.

Given these different arguments and counter-arguments, it is possible to define a symbolic calculus, called a Logic of Argumentation, which enables the combination ("flattening") of arguments for and against a proposition [5]. Since an argument is a tentative proof of a claim, our degree of belief in the claim will likely depend upon the argument advanced for it. Thus, for each pair (ϕ, G_ϕ) consisting of a claim and an argument for it, we can associate a measure α_ϕ of our strength of belief in ϕ given G_ϕ.[4] We represent this as a triple $(\phi, G_\phi, \alpha_\phi)$, which we call an assessed argument. The belief-indicator may be a quantitative measure, such as a probability, or an element from a qualitative dictionary, such as $\{$*Likely, Unlikely*$\}$. In either case, we can define algebraic operations on the set of belief-indicators (the "denotation dictionary for belief") enabling us to generate the degree of belief in a combined argument, when we know the degrees of belief of the subsidiary arguments. In addition to belief-indicators, one can also define other labels for claim-argument pairs, such as the values of world-states and the consequences of actions arising from the claim [5]. With such formal calculi, argumentation can be used in intelligent computer systems, and has been so used, particularly in the medical and legal domains (e.g. see [5, 11, 30]).

This view of argumentation presents the arguments as disembodied cases for and against a proposition. It is as if there were just one person in the debate, weighing the pros and cons with him or herself to arrive at a considered conclusion. Indeed, the carcinogenicity risk assessment guidelines of the U.S. Environmental Protection Agency [28], which are rules for the combination of evidence from disparate sources, have the appearance of an algorithm for the dispassionate weighing of evidence. We term this monodic (single-voice) argumentation. However, in real life, there are usually many voices, each arguing for and against a proposition from differing perspectives, and sometimes with different views as to what constitutes acceptable rules of inference.

In seeking to model this rational cacophony, we have therefore turned to dialectical argumentation, a branch of philosophy dealing with the conduct of debate and discourse [29]. One influential framework for dialectical argumentation has been that proposed by the philosopher Jürgen Habermas [9]. Originally seeking to understand how ethical norms could be agreed between different people, and building on Toulmin's work [27], Habermas proposed a framework in which consenting members of a community can engage in a civil discourse. The key difference between monodic and dialectical argumentation is the presence in the latter of an audience. An audience needs to be persuaded, and may withhold her (or his or their) agreement to the claims being advanced by a proponent. Indeed, members of an audience may advance counter-claims of their own, or rebuttals and undercutting arguments, or may question the premises or modes of inference used by a proponent. Habermas sought to identify rules under which such discourse could occur in a civil manner and so that all reasonable participants would feel satisfied with the process of discourse. In [9], Habermas gave examples of the sort of rules his framework would include, for instance: "Different speakers may not use the same expression with different meanings"

[4] Such degrees of belief are called modalities by Toulmin [27].

and "Everyone is allowed to question any assertion whatever". Legal philosopher Robert Alexy [1] articulated a comprehensive list of such rules appropriate for debates in ethical domains.

Habermas has applied his framework to discourse in political, legal and social science arenas. Dialectical argumentation has also been applied by philosophers of science to the natural sciences. Marcello Pera [18], for example, models scientific discourse as a three-person dialogue, involving the scientific investigator, Nature and a skeptical scientific community. In Pera's model, the investigator proposes theoretical explanations of scientific phenomena and undertakes scientific experiments to test them. These experiments lead to "replies" from Nature in the form of experimental evidence. However, Nature's responses are not given in a direct or pure form, but are mediated through the third participant, the scientific community, which interprets the evidence, undertakes a debate as to its meaning and implications, and eventually decides in favour or against proposed theoretical explanations. Drawing on the work of Pera and Habermas, William Rehg [21] proposes a form of dialectical argumentation for the debate which occurs within the scientific community (and between it and the experimenter), arising from Nature's responses to experiments. One of Rehg's aims is to capture the fact that even though the resolution of scientific questions may be objective and rational, as these terms are commonly understood, such resolution, by its nature as a human activity, takes place within a particular social, cultural and institutional context which invariably influences the course of resolution.[5]

3.2 A framework for dialogue

Motivated by these approaches, we are developing a dialectical argumentation framework for claims of carcinogenicity.[6] As do Habermas and Alexy, we need to codify rules of engagement. As do Pera and Rehg, we desire this to be a realistic model of scientific debate in the natural sciences, at least in our specific domain of chemical carcinogenicity. Our further aim, as was mentioned, is to encode this framework in an intelligent computer system. This requires us to be explicit and comprehensive in the framework structure and rules we propose. We define an "agora" (from the Greek for "meeting place") as a space in which the dialogue will occur, and we use this term also for the dialogue itself. Thus a "ϕ-agora" is a debate about the claim ϕ. A ϕ-agora consists of the following elements:[7]

- A database ∇ of well-formed formulae of a symbolic propositional language, with the usual connectives, in which atomic propositions are denoted ϕ_i.
- A set of different modes of inference, each denoted R_j.

[5] As just one example, Jamieson [10] has argued that, in science policy debates over environmental and health risks, even uncertainty itself is, at least partly, socially constructed, with debate participants establishing, maintaining and using it to further particular agendas.

[6] Note that Verheij [30], building systems for legal appplications, uses the term "dialectical argument" to refer to a monodic argument which incorporates undercutting exceptions. We are using dialectical argumentation not in this sense, but to refer to a debate involving different views.

[7] From the perspective of software functionality, our dialogue space is similar to the negotiation spaces of electronic commerce systems using intelligent agents [12].

- A set of debate participants, each denoted $\mathcal{P}_k$.
- A set of rules for asserting, supporting, questioning, denying, rebutting, undercutting, assumption-denying, mode-of-inference-denying of a claim. (i.e. which argument-moves are valid, when; which responses are valid, when.)
- A set of rules for summarising, comparing and manipulating arguments.
- A set of rules by which all the arguments for and against a proposition can be combined on behalf of the scientific community concerned, in accordance with Pera's model of scientific enquiry.
- A presentation and advice module (so this can be presented to the user).

The agora framework presented here will be embodied in an intelligent computer system which advises a human user or users, allowing different interactions according to the user's purposes. For instance, a user may wish to explore both the arguments for and the against a particular claim, as when exploring the consequences of a particular action. Or, he or she may wish to marshal together all the arguments for the claim, ignoring or rebutting any arguments against it. Our structure is intended to allow for such variety of user purposes, with the system undertaking both autonomous reasoning and argument mediation, in the terminology of Verheij [30]. Following Aristotle, Habermas [9] proposed a three-level structure for his dialectical argumentation framework, and we have adapted this for the Agora:

Logical Level: At this level we seek to understand what are the logical implications of the knowledge base, which is needed because, in general, we do not know the consequences of our own knowledge. At this level, the system would be undertaking automated reasoning generating all the possible arguments that may be constructed from its knowledge in exactly the same way that the argumentation systems described in [5, 16] construct arguments.

Dialectical level: At the dialectical level, we are considering the cases for and against some proposition, the pros and the cons, and we need to be able to combine and flatten arguments generated by the first level in a way that is similar to that described earlier. This may be by simply looking at the strengths of the arguments [5, 16], or by looking at the relationships between them [17]. As indicated earlier, it is important for the carcinogenicity domain that the modes of inference themselves are able to be the subject of argument (for instance, being attacked or denied). Some recent argumentation systems, such as those developed for legal applications by Verheij [30], permit argument about inference rules.

Rhetorical level: At the rhetorical level, we are concerned with convincing an audience of a particular case. In terms of our system, we see this as a presentation layer, where the user can manipulate the activities of the other two layers for self-elucidation or for presentation to others. For example, this layer permits the user, within the permitted rules of the dialectical framework, to interrogate the arguments articulated by the system, to propose rebuttals and undercutting arguments, etc. Reed [20], for example, has explored the use of rhetorical devices for persuasion purposes in argumentation systems.

The user interfaces at these different layers may not necessarily be the same. Verheij [30] has argued that new kinds of user interfaces are required for argumentation systems, and he reports on several approaches in this area.

4 EXAMPLE

To illustrate these ideas we present a simple and hypothetical example of an Agora debate. Suppose we have a knowledge base containing the following agreed facts, labeled K1 through K11, about chemicals $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$:

K1: Elements of the chemical structures of $\mathcal{Y}$ and $\mathcal{Z}$ are similar.

K2: Elements of the chemical structures of $\mathcal{X}$ and $\mathcal{Y}$ are similar.

K3: Chemicals $\mathcal{X}$ and $\mathcal{Z}$ do not have similar chemical structures.

K4: $\mathcal{X}$ is produced by the human body naturally (i.e. it is endogenous).

K5: $\mathcal{X}$ is endogenous in rats, but not in mice.

K6: An endogenous chemical is rarely carcinogenic at normal bodily levels.

K7: Above a threshold of x_0 ppm in the bloodstream, $\mathcal{X}$ is excreted by the human body.

K8: The presence of large quantities of $\mathcal{Z}$ in the human body appear to inhibit the excretion of other chemicals.

K9: Experiments applying $\mathcal{Y}$ to human tissue culture show it to be carcinogenic, although only at high doses.

K10: Bioassay experiments applying $\mathcal{X}$ at doses above level d_r to rats show it to be carcinogenic.

K11: Bioassay experiments applying $\mathcal{X}$ to mice show no evidence for carcinogenicity up to and including dose levels of d_m.

At the **Logical Level**, the Agora develops the logical consequences of the knowledge in the knowledge base. For this example, there are three logical inferences, labeled L1, L2 and L3:

L1: Chemical $\mathcal{Z}$ is possibly carcinogenic to humans, because:

> **K1:** Elements of the chemical structures of $\mathcal{Y}$ and $\mathcal{Z}$ are similar.
>
> **K9:** Experiments applying $\mathcal{Y}$ to human tissue culture show it to be carcinogenic.

L2: If humans have cancer as a result of exposure to $\mathcal{Z}$, then this may indicate high levels of the chemical in the body, because:

K1: Elements of the chemical structures of $\mathcal{Y}$ and $\mathcal{Z}$ are similar.

K9: Experiments applying $\mathcal{Y}$ to human tissue culture show it to be carcinogenic *at high doses.*

L3: If $\mathcal{Z}$ is carcinogenic, then its presence may lead to high bodily levels of $\mathcal{X}$, because:

L2: If humans have cancer as a result of exposure to $\mathcal{Z}$, then this may indicate high levels of the chemical in the body.

K8: The presence of large quantities of $\mathcal{Z}$ in the human body appear to inhibit the excretion of other chemicals.

Suppose ϕ is the statement: *$\mathcal{X}$ is carcinogenic to humans.* Then, at the **Dialectical Level** of the Agora, our knowledge base allows us to construct arguments for and against ϕ, which we list below, numbered A1, A2, etc. These are presented as 2-tuples, where the conclusion is the first element of the tuple, and the supporting grounds for the conclusion are the second element of the tuple.

A1: (ϕ, K10).

A2: (ϕ, (K2, K9)).

A3: (ϕ, (K4, K5, K10, K11)).

A4: (ϕ, (L1, L3)).

A5: ($\neg\phi$, K11).

A6: ($\neg\phi$, (K4, K6, K7)).

A7: ($\neg\phi$, (K3, L1)).

Observe that there are four arguments for the carcinogenicity of $\mathcal{X}$ and three against it, although not all of these will be of the same force. Several arguments implicitly draw on one or more the inference rules listed in Section 2.1 above. Argument A1, which asserts human carcinogenicity on the basis of rat bioassays, potentially draws on all the inference rules, as does Argument A3. Likewise, although Argument A5 asserts non-carcinogenicity to humans on the basis of mice experiments, it draws on the same set of inference rules to make this assertion. Argument A2, which asserts human carcinogenicity on the basis of tissue experiments with a related chemical, draws on an equivalent set of inferences to those for bioassays. Note also that Argument A6, which asserts non-carcinogenicity on the basis of the endogeneity of $\mathcal{X}$, in effect contests Argument A1 via inference rules R9 and R10 regarding the manner in which human exposure to the chemical occurs. Arguments A4 and A7 draw on inferences L1 and L3 derived at the Logical Level of the Agora.

At the **Rhetorical Level**, we could seek to build a case for a particular claim, say ϕ, by examining the various arguments for and against it and the relationships between them. For example, Argument A4 draws on Logical Inferences L1 and L3. Together, these rebut statement K7, and thus undercut argument A6. Likewise, Argument A5 is rebutted by Argument A3, as the latter proposes a plausible mechanism which supports carcinogenicity of $\mathcal{X}$ despite the lack of evidence from mice experiments (statement K11). We are thus left with four arguments (A1 – A4) in favor of ϕ and one (A7) against it. Which side is considered stronger will be a matter of the relative degrees of belief associated with each argument.

5 CONCLUSION

Automated prediction of chemical properties is an active area of artificial intelligence (AI) research [23, 26]. However, most effort to date has been devoted to systems which predict the properties of substances on the basis of chemical theoretical reasoning or by comparison with other chemicals whose properties are known. Yet, scientific claims about the toxicity or carcinogenicity of chemicals are usually based on bioassay or epidemiological evidence, and so we seek to construct intelligent systems able also to reason from such evidence.

This paper has made two original contributions to the design of such systems: First, we have articulated the precise modes of inference used when chemical carcinogenicity is asserted on the basis of animal bioassay experiments. To our knowledge, no such complete list has been developed before. Second, we have taken Pera's three-person model of scientific enquiry and Habermas' Discourse Ethics as the basis for an intelligent system using dialectical argumentation to reason about scientific domains. Such an application is novel, although work in intelligent systems for legal applications has applied similar rules for legal discourse (also due to Alexy) [6]. We are currently developing the specification of the system presented here and studying its formal properties [13]. In addition to representing the cases for and against particular claims, we also seek to incorporate qualitative assessment of the values of claims and their consequences [14], building on recent work extending logics of argumentation, for example, [5, 16].

In this paper, we have considered just one scientific domain, but our approach is clearly applicable more widely. Although our specific agenda here is the development of intelligent systems, the delineation of modes of inferences and the modeling of arguments used should benefit whichever is the community concerned with, and debating, the claim at issue. In the case of carcinogenicity of chemicals that community is all of us.

Acknowledgments

This research was partly funded by the British EPSRC under grant GR/L84117 and a PhD studentship. We are also grateful for discussion with Heather Dickinson, Susan Haack, William Rehg, Martijn Schut and Bart Verheij. An earlier version of this paper was presented at the Symposium on Practical Reasoning organized by the British Psychological Society in London in December 1999.

References

[1] R. Alexy. A theory of practical discourse. In S. Benhabib and F. Dallmayr, editors, *The Communicative Ethics Controversy*, Studies in Contemporary German Social Thought, pages 151–190. MIT Press, Cambridge, MA, USA, 1990. *(Translation by D. Frisby of: Eine Theorie des praktischen Diskurses.*

In: W. Oelmüller, Ed., Normenbegründung-Normendurchsetzung. Paderborn, Germany: Schöningh, 1978).

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

[3] R. Doll. The Seascale cluster: a probable explanation. *British Journal of Cancer*, 81 (1):3–5, 1999.

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

[5] J. Fox and S. Parsons. Arguing about beliefs and actions. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms*, LNAI 1455, pages 266–302. Springer Verlag, Berlin, Germany, 1998.

[6] T. F. Gordon. The Pleadings Game: An exercise in computational dialectics. *Artificial Intelligence and Law*, 2:239–292, 1994.

[7] 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.

[8] J. D. Graham and L. Rhomberg. How risks are identified and assessed. *Annals of the American Academy of Political and Social Science*, 545:15–24, 1996.

[9] J. Habermas. *Moral Consciousness and Communicative Action.* MIT Press, Cambridge, MA, USA, 1991. *(Translation by C. Lenhardt and S. W. Nicholsen of: Moralbewusstsein un kommunikatives Handeln. Suhrkamp Verlag, Frankfurt-am-Main, Germany, 1983.).*

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

[11] 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*, LNAI 1455, pages 138–156. Springer Verlag, Berlin, Germany, 1998.

[12] A. Lomuscio, M. Wooldridge, and N. Jennings. A classification scheme for negotiation in electronic commerce. In *A European Perspective on Agent-Mediated Electronic Commerce*, LNAI. Springer Verlag, Berlin, Germany, 2000 (To appear).

[13] P. McBurney and S. Parsons. Risk agoras: Dialectical argumentation for scientific reasoning. In C. Boutilier and M. Goldszmidt, editors, *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI-2000)*, Stanford, CA, USA, 2000. UAI.

[14] P. McBurney and S. Parsons. Risk Agoras: using dialectical argumentation to debate risk. *Risk Management*, 2(2):17–27, 2000.

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

[16] S. Parsons and S. Green. Argumentation and qualitative decision making. In A. Hunter and S. Parsons, editors, *Proceedings of the 5th European Conference on Quantitative and Symbolic Approaches to Reasoning and Uncertainty (ECSQARU99)*, LNAI 1638, pages 328–339. Springer Verlag, Berlin, Germany, 1999.

[17] S. Parsons, C. Sierra, and N. R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261—292, 1998.

[18] M. Pera. *The Discourses of Science.* University of Chicago Press, Chicago, IL, USA, 1994.

[19] R. A. Pollak. Government risk regulation. *Annals of the American Academy of Political and Social Science*, 545:25–34, 1996.

[20] C. Reed. *Generating Arguments in Natural Language.* PhD thesis, University College, University of London, London, UK, 1998.

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

[22] 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.

[23] A. M. Richard. Application of artificial intelligence and computer-based methods to predicting chemical toxicity. *The Knowledge Engineering Review*, 14(4):307–317, 1999.

[24] T. Seidenfeld. *Philosophical Problems of Statistical Inference.* Reidel, Boston, MA, USA, 1979.

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

[26] A. Srinivasan, S. H. Muggleton, M. J. E. Sternberg, and R. D. King. Theories of mutagenicity: a study in first-order and feature-based induction. *Artificial Intelligence*, 85:277–299, 1996.

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

[28] U.S.A. Environmental Protection Agency. Guidelines for carcinogen risk assessment. *U.S. Federal Register*, 51:33991–34003, 24 September 1986.

[29] F. H. van Eemeren, R. Grootendorst, and F. S. Henkemans. *Fundamentals of Argumentation Theory.* Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1996.

[30] B. Verheij. Automated argument assistance for lawyers. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Law, Oslo, Norway*, pages 43–52, New York City, NY, USA, 1999. ACM.