Representing Epistemic Uncertainty by means of Dialectical Argumentation

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

Abstract

We articulate a dialectical argumentation framework for qualitative representation of epistemic uncertainty in scientific domains. The framework is grounded in specific philosophies of science and theories of rational mutual discourse. We study the formal properties of our framework and provide it with a game theoretic semantics. With this semantics, we examine the relationship between the snaphots of the debate in the framework and the long run position of the debate, and prove a result directly analogous to the standard (Neyman-Pearson) approach to statistical hypothesis testing. We believe this formalism for representating uncertainty has value in domains with only limited knowledge, where experimental evidence is ambiguous or conflicting, or where agreement between different stakeholders on the quantification of uncertainty is difficult to achieve. All three of these conditions are found in assessments of carcinogenic risk for new chemicals.

SHORTITLE: Argumentation and Uncertainty

KEYWORDS: Dialectical Argumentation, Qualitative Reasoning, Uncertainty Representation.

1 Introduction

We seek to build intelligent systems which can reason autonomously about the risk of carcinogenicity of chemicals, drawing on whatever theoretical or experimental evidence is available. Claims of carcinogenicity may be based on a several different types of evidence [27, 28, 76]: experimental results of the chemical on tissue cultures; bioassay experiments on animals; human epidemiological studies; analytical comparisons with known carcinogens; or explication of biomedical causal pathways.[1] Evidence from these different sources may conflict, and carcinogen risk assessment usually involves the comparison and resolution of multiple evidence. In a celebrated case, exposure to formaldehyde was shown in animal bioassays to cause significant increases in the incidence of nasal cancers in rats, but not in mice. Retrospective epidemiological studies of humans whose work exposed them to the chemical, such as morticians, yielded no such increases. However, these human studies did reveal a statistically significant increase in the incidence of brain cancers, for which there were no plausible bio-medical causal mechanisms [27].[2]

How should such epistemic uncertainty be represented? Any attempt to generate a quantified measure of uncertainty runs into three major difficulties. Firstly, relationships between evidence and conclusions are not straightforward. As we have shown in previous work [53], to assert a carcinogenic risk to humans from animal bioassay evidence, for example, may require as many as a dozen distinct types of inference, none of which is conclusive. Consequently, each will potentially introduce some uncertainty into the final assertion. Although there has been some effort to quantify the effects of this (e.g. [24, 29]), this area is still poorly understood, and there is no reason to believe that chemicals yet to be tested or even invented will follow the patterns of past chemicals. Secondly, how does one combine evidence from different sources, such as animal bioassay and epidemiological evidence? The U.S. Environmental Protection Agency guidelines for carcinogenic risk assessment [76] and proposed revisions [77] may be seen as rules for the combination of different types of evidence, but these still leave a great deal of freedom of interpretation to the risk assessor. These first two difficulties demonstrate that any quantification of uncertainty requires the adoption of many subjective assumptions and assessments. This leads to the third difficulty facing quantification efforts, that of achieving agreement between different people. Debates over the potential health risk of chemicals are often contentious, no doubt due to the high stakes and conflicting interests involved, and typically agreement is not readily forthcoming. It has even been argued that much uncertainty may be a political artefact, established, maintained and propagated by participants in environmental health debates to serve their political or other interests [42]. Even without differing interests, reasonable people may disagree on the interpretation of ambiguous or conflicting

[1] Automated prediction of chemical properties, such as carcinogenicity, on the basis of chemical structure and analytic comparisons with other chemicals is an active area of research, e.g. [15, 73]. However this work has not looked at combining such different types of evidence for properties.

[2] Subsequent epidemiological studies have provided statistically-significant evidence for human nasal and other cancers from exposure to formaldehyde [82].

evidence. In the formaldehyde case, for instance, the US Environmental Protection Agency produced, within a six month period in 1981, two opposite assessments of formaldehyde's human carcinogenic risk from precisely the same data [52].[3]

Given these problems with quantification, we seek a qualitative representation of uncertainty. Moreover, because claims of carcinogenicity may be based on multiple types of evidence, an argumentation framework would seem appropriate, as such a framework may permit the combination of disparate categories of data. Argumentation formalisms have previously been used to represent uncertainty in intelligent systems (e.g. [43]), but typically using a monolectical approach, where arguments for and against a proposition are combined in some manner to produce an overall summary case. However, debates in the carcinogenic risk domain are usually polyphonic, with different participants arguing for or against a proposition from different perspectives and assumptions, and even with different views of what constitutes valid reasoning. To model this rational cacophony, and represent uncertainty within it, we have therefore adopted a dialectical argumentation framework.

Formal models of dialectical argument were proposed by philosophers Charles Hamblin [33, 34] and Jim MacKenzie [50, 51] to study fallacious reasoning.[4] These formal approaches have since formed the basis for studies of dialogue from the perspective of linguistics [11] and from argumentation theory [79]. Within Artificial Intelligence, formal dialogue models have been applied to modeling legal argument [7, 25], to debates over local urban planning decisions [26], to the design of software components [74], and to interactions between intelligent software agents, such as persuasion dialogues and negotiations [4, 5, 61].[5] However, one difference between dialectical argumentation systems for legal applications and those for multi-agent systems is that the former sometimes assume a common knowledge base, either from the outset, or constructed by means of the dialogue. This is neither desirable nor possible for multi-agent systems, where autonomous agents may have many reasons not to share or pool their knowledge, including legal privacy requirements, national security concerns or plain self-interest. Our application is in between these two extremes. We do not assume the participants commence with a common knowledge base, nor that one is necessarily constructed in the course of the dialogue. Yet, claims may, as will be shown, be accepted by *"the community"*

[3]The changed assessment occurred after a change in management, which reinforces the point being made here.

[4]A formal model of dialectical argumentation was also developed by Paul Lorenzen and his colleagues [47, 48] to provide a game-theoretic semantics for intuitionistic logic. These models have since been applied, for example, to logics for quantum physics [57].

[5]Recent reviews of argumentation in AI include [13, 66], which discuss theoretical aspects, and [10], which reviews applications. In addition, an international symposium of philosophers, linguists and computer scientists met recently to identify open questions at the interface of argumentation and computation [68].

on the basis of the arguments presented for and against them by the participants in the dialogue. However, this may happen without a single participant expressing a personal commitment to the claim. This is as should be: the republic of science is a democracy not a dictatorship; individual scientists may express strong reservations with currently accepted theory while still using it as the best available.

This paper presents our detailed framework for an intelligent system for the carcinogenic risk assessment domain, a framework we have previously termed a Risk Agora [53]. We ground our framework in a specific philosophy of science, based on work of philosophers Marcello Pera and Paul Feyerabend, and a formal philosophical model of rational dialogue, from work of philosophers David Hitchcock, Jürgen Habermas and Robert Alexy. These models are presented in the next Section. Section 3 defines our formalism, and Section 4 examines its properties. Section 5 presents an example of its application, while Section 6 concludes with a discussion of future work.

2 A Dialectical Model of Scientific Inquiry

2.1 Scientific discourse

Nicholas Rescher [71], a philosopher of logic and argumentation, claims to have been the first to propose a dialectical framework for the progress of scientific inquiry. Similarly, James Freeman [21], another argumentation theorist, discusses scientific discourse in his study of generic argument structure. Both these approaches are from an argumentation theory perspective rather than from the philosophy of science, and so neither is grounded in, nor engages with, a detailed understanding of actual scientific practice. As a consequence, the frameworks proposed could easily be applied to other, non-scientific, domains.

One novel approach from a philosophy of science perspective is the dialectical model of scientific discourse proposed by Marcello Pera [62]. Pera views the enterprise of science as a three-person dialogue, involving a scientific investigator, Nature and a skeptical scientific community. In this model, the investigator proposes theoretical explanations of scientific phenomena and undertakes scientific experiments to test these. The experiments lead to “replies” from Nature in the form of experimental evidence. However, Nature’s responses are not given directly or in a 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 favor or against proposed theoretical explanations. We have adopted Pera’s model for our application, and provided Nature with a formal role, manifested through the contributions of the other participants.

Although more specific than Rescher’s or Freeman’s models, Pera’s model of modern science as a dialogue game could still be applied to domains which do not share science's success in explaining and predicting natural phenomena. We believe, therefore, that our model requires an explanation of the success of science. Some philosophers of science believe this is due to the application of universal principles of assessment of proposed scientific theories, such as the confirmationism of Rudolph Carnap or the falsificationism of Karl Popper [64]. However, we do not share these views, instead believing, with Paul Feyerabend [17], that the standards of assessment used by any scientific community are domain-, context- and time-dependent. This view, that there are neither universal nor objective standards by which scientific theories can be judged, was called "epistemological anarchism" by Imre Lakatos [45]. Moreover, there is a methodological problem with falsificationism in our chosen domain of carcinogenicity. As many have argued (e.g. Hansson [35]), it is not possible to falsify statements of the form "*Chemical $\mathcal{X}$ has carcinogenic effects*," because one can never completely eliminate the possibility of very weak effects. For instance, if the effects of a carcinogen at the levels of typical exposure are very small or its actions are long delayed, sample sizes in the millions or billions may be required to have reasonable confidence of identifying the effects of typical exposure levels in a statistical experiment [72].

Instead of the application of universal principles of assessment of theories, we believe science's success arises in part from applying two normative principles of conduct: firstly, that every theoretical explanation proposed by a scientific investigator is contestable by anyone;[6] and secondly, that every theoretical explanation adopted by a scientific community is defeasible.[7] In other words, all scientific theories, no matter how compelling, are always tentative, being held only until better explanations are found, and anyone may propose these. Note that in saying all conclusions are always defeasible, we are not specifying the manner by which they may be overthrown: defeasibility is thus a more general concept than falsificationism. Contestability distinguishes science from, say, extreme political ideologies, such as Nazism or the Juche philosophy of Kim Il Sung. Defeasibility distinguishes science from, say, traditional religion or creationism. On the other hand, both principles apply to human endeavours commonly thought of as scientific but which may fail criteria such as predictive capability (e.g. paleontology; climatology; macro-economics) or falsifiability (e.g. sociobiology; Freudian psychology).

To build an intelligent system based on these principles, we therefore require a (normative) model of scientific discourse which enables contestation and defeasi-

[6] At least, by anyone from within the scientific community concerned. While an argument may only be given serious consideration by a scientific community when it arises from a member of that community, there usually are no formal barriers to anyone seeking to join the community. Double-blind reviewing of research papers reinforces this openness.

[7] These two principles are each necessary to explain science's success, but not sufficient.

bility of claims. Our model has several components. At the highest level, we are attempting to model a discourse between reasonable, consenting scientists, who accept or reject arguments only on the basis of their relative force. To model debates of this type we draw on two sources: firstly, we utilize certain principles of rational mutual inquiry proposed by philosopher David Hitchcock [38]. These provide a series of high-level *desiderata* for the conduct of a debate, consistent with our epistemological anarchist standpoint on the philosophy of science. Secondly, we draw upon the philosophy of Discourse Ethics developed by Jürgen Habermas [32] for debates in ethical and moral domains. Habermas's rules of discourse were first fully articulated by Robert Alexy [3], and these are at a lower level than Hitchcock's principles.[8] Together they form the basis of the desired properties of the Agora formalism.

Next, within this structure, we wish to be able to model dialogues in which different participants variously posit, assert, contest, justify, qualify and retract claims. To represent such activity requires a model of an argument, and we use Stephen Toulmin's model [75], within a dialectical framework. To embody our belief in epistemological anarchism, we permit participants to contest any component of a scientific argument: its premises; its rules of inference (Toulmin's "warrants"); its degrees of support (his "modalities"); and its consequences. We believe this is exactly what real scientists do when confronted with new theoretical explanations of natural phenomena [17]. When a scientific claim is thus contested, its proponent may respond, not only by retracting it, but by qualifying it in some way, perhaps reducing its scope of applicability. Arne Naess [59] called this process "precizating", and we seek to enable such responses in the system. We thus ground our formalism for the Agora in a model of scientific discourse as dialectical argumentation.

In an influential typology, Doug Walton and Erik Krabbe [79] identified several types of dialogue, distinguished by their initial situations, the goals of their participants, and the goals of the dialogue itself (which may differ from those of its participants). The dialogue types were: Persuasion dialogues; Negotiations; Inquiries; Deliberations; Information-seeking dialogues; and Eristic (strife-ridden) dialogues. Scientific dialogues may have elements of several of these categories: one view would see scientific activity as a pure Inquiry dialogue, where participants collaborate to prove or disprove some hypothesis of interest. However, this assumes an hypothesis has been explicitly stated, and prior work — involving data collection, data analysis, theory development and much thinking, especially counterfactual thinking — may be needed to induce or form an hypothesis. All these activities may be undertaken or supported through dialogue. Moreover, once a scientist adopts a position on an open issue, the debate which then occurs is more

[8]Alexy's rules have some similarity with Grice's Maxims for Conversation [30].

like a Persuasion dialogue, where each side seeks to convince the other sides of the correctness of its views. These exchanges can be quite emotionally charged, so that some may view them as Eristic dialogues. Because scientific practice does not fit neatly into the categories of Walton and Krabbe we have not used this typology in the work reported here.

2.2 Desired Agora properties

As mentioned, we desire our Agora formalism to satisfy Hitchcock's principles of rational mutual inquiry, and Alexy's lower-level rules for a reasoned discourse [3]. We begin by listing Hitchcock's Principles, adapted for multiple participation dialogues, and numberered H1 through H18. The linguistic labels are those of Hitchcock.

H1 Externalization: The rules should be formulated in terms of verifiable linguistic behaviour.

H2 Dialectification: The content and methods of dialogue should be subject to the agreement of participants, without any prior imposition.

H3 Mutuality: No statement becomes a commitment of a participant unless he or she specifically accepts it.

H4 Turn-taking: At most one person speaks at a time.

H5 Orderliness: One issue is raised at a time and is dealt with before proceeding to others.

H6 Staging: An inquiry dialogue should proceed by a series of stages, from initial clarification of the question at issue and on the methods of resolving it, through data gathering and intepretation, to formation of arguments.

H7 Logical Pluralism: Arguments should permit both deductive and non-deductive forms of inference.

H8 Rule-consistency: There should be no situation where the rules prohibit all acts, including the null act.

H9 Semantic Openness: The rules should not force any participant to accept any statement, even when these follow by deduction from previous statements.

H10 Realism: The rules must make agreement between participants a realistic possibility.

H11 Retraceability: Participants must be free at all times to supplement, change or withdraw previous tentative commitments.

H12 Role reversal: The rules should permit the responsibility for initiating suggestions to shift between participants.

H13 Experiential Appeal: The rules should permit direct mutual appeal to experience.

H14 Openness: There should be no restrictions on the content of contributions.

H15 Tentativeness: Participants should be free to make tentative suggestions as well as assertions.

H16 Tracking: The rules should make it possible to determine at any time the cumulative commitments, rights and obligations of each participant.

H17 Termination: There should be rules for the orderly termination of the dialogue. Hitchcock proposes that an inquiry terminate as soon as (a) a participant declares an intention to abandon it, (b) in two successive turns neither participant has a suggestion for consideration, or (c) there is agreement on the conclusion of the discussion.

H18 Allocation of Burden of Proof: The burden of proof remains with the participant who makes a suggestion, even after contestation by another Participant.

Note that Principles H5 (Orderliness) and H6 (Staging) may conflict with Principle H2 (Dialectification). Since we believe the conduct of the dialogue, including the content, nature, duration and sequencing of any stages should be a matter for the participants to decide as part of the debate, we do not impose any external structure or content-requirements on the dialogue. Accordingly, our formalism does not implement these two principles. For the same reason, we do not impose any Termination Rules (Principle H17). In any case, the ultimate defeasibility of all scientific claims means that no scientific debate is ever completed. As will be seen below, the manner in which we implement Principle H16 (Tracking) will enable an observer at any time to obtain a "snap-shot" of the status of a debate, including an assessment of those statements adopted by the community concerned as (defeasibly) true. Finally, we assume that our formalism is to be implemented on a sequential processor, so that Principle H4 (turn-taking) will be guaranteed, and so not require specific implementation.

In contrast to Hitchcock's generic principles, Alexy's rules for reasoned discourse were intended to guide discussion of ethical and moral matters and are at a lower level of specification. In restating them, we have re-ordered them, and have ignored rules specific to ethical questions. We have also ignored Alexy's rules regarding the relevance of utterances, since our formalism is intended for debate regarding only one chemical at a time. Moreover, we have modified the rules slightly to conform to Hitchcock's Principles. For instance, Alexy's rules require participants to assert only claims for which they have an argument, a condition which cannot be verified directly, thus conflicting with Principle H1. Instead, we permit participants to assert any statement (Property A3), but also permit any other participant to request the argument justifying this assertion (Property A5).[9] We have also added a property (A8) concerning precization, and added linguistic labels to the presentation.

In articulating these rules as desirable properties of our formalism, we have adopted a terminology which will be given precise definition in Section 3. In essence, a grounded argument for a claim is an argument which begins from some premises and proceeds according to some specified rules of inference. A consequential argument from a claim is an argument from a claim to some consequence, again using specified rules of inference. A valued argument is one to which the participant has assigned degrees of support (in the form of modality labels) to premises, inference rules and conclusions.

A1 Freedom of Assembly: Anyone may participate in the Agora, and they may execute dialogue moves at any time, subject only to move-specific conditions (defined in Section 2 below).

A2 Common Language: Participation entails acceptance of the semantics for the logical language used, and of the associated modality (degrees of support) dictionaries.

A3 Freedom of Speech: Any participant may assert any claim or consequence of a claim.

A4 Freedom to Challenge Claims: Any participant may question or challenge any claim or any consequence of a claim.

[9]Note that some argumentation formalisms for multi-agent systems insist that agents making assertions must first verify that they have an argument for the statement in their own knowledge base, e.g. [4]. Such conditions are analogous to the *sincerity* requirements of agent languages, such as the *feasibility pre-condition* in the FIPA Agent Communications Language [18, p.48]. Conditions such as these may be suitable for some applications, e.g. information-seeking dialogues, but not for others, e.g. deliberations. In deliberative dialogues, for example, it may be in the interests of every agent to consider suggestions inconsistent with or unprovable from their own prior and partial knowledge.

A5 Arguments required for Claims: Any participant who asserts a claim (respectively, a consequence of a claim) must provide a valued grounded argument for that claim (respectively, a valued consequential argument from the claim) if queried or challenged by another participant.

A6 Freedom to Challenge Arguments: Any participant may question or challenge the grounds, the rules of inference or the modalities for any claim.

A7 Freedom of Modal Disagreement: Whenever a participant asserts a valued grounded argument for a claim (or a valued consequential argument from a claim), any other participant may assert a valued grounded argument (respectively, a valued consequential argument) for the same claim with different dictionary values.

A8 Precization: A participant who has provided a grounded argument for a claim which has been challenged should be able to respond by qualifying (precizating) the original claim or argument.

A9 Proportionate Defence: Any participant who provides a grounded argument for, or a consequential argument from, a claim is not required to provide further defence if no counter-arguments are provided by other participants.

A10 No Contradictions: No participant may contradict him or herself.

3 The Risk Agora Formalism

3.1 Preliminary definitions

We begin by assuming the system is intended to represent debate regarding the carcinogenicity of a specific chemical, and that statements concerning this can be expressed in a propositional language $\mathcal{L}$, whose well-formed formulae (wffs) we denote by lower-case Greek letters. Subsets of $\mathcal{L}$ (i.e. sets of wffs) are denoted by upper-case Greek letters, and $\mathcal{L}$ is assumed closed under the usual connectives. We assume multiple modes of inference (warrants) are possible, these being denoted by $\vdash_i$. These may include non-deductive modes of reasoning, and we make no presumptions regarding their validity in any truth model.[10] We assume a finite set of debate participants, denoted by $\mathcal{P}_i$, who are permitted to introduce new wffs and new modes of inference at any time. We denote Nature, in its role in the

[10]This liberal view allows our rules of inference to be used to represent scientific causal mechanisms, debate over which is arguably the origin of all scientific dialogue.

debate, by the italicized name *Nature* or by the symbol $\mathcal{P}_N$. Likewise, the scientific community as a whole in the Agora is called the *Agora* or denoted $\mathcal{P}_A$.

Definition 1: *A* grounded argument for a claim θ, *denoted* $\mathcal{A}(\rightarrow \theta)$*, or* $\mathcal{A}$*, is a 3-tuple* (G, R, θ)*, where* $G = (\Theta_0, \theta_1, \Theta_1, \theta_2, \ldots, \Theta_{n-2}, \theta_{n-1}, \Theta_{n-1})$ *is an ordered sequence of wffs* θ_i *and possibly-empty sets of wffs* Θ_i*, with* $n \geq 1$ *and with* $R = (\vdash_1, \vdash_2, \ldots, \vdash_n)$ *an ordered sequence of inference rules such that:*

$$\Theta_0 \ \vdash_1 \ \theta_1,$$

$$\Theta_1, \ \theta_1 \vdash_2 \ \theta_2,$$

$$\vdots$$

$$\Theta_{n-1}, \ \theta_{n-1} \vdash_n \ \theta.$$

In other words, each θ_k $(k = 1, \ldots, n-1)$ is derived from the preceding wff θ_{k-1} and set of wffs Θ_{k-1} as a result of the application of the k-th rule of inference, $\vdash_k$. The rules of inference in any argument may be non-distinct. We call the set $\{\theta_{k-1}\} \cup \Theta_{k-1}$ the *grounds* (or *premises*) for θ_k. Also, an argument (G', R', θ_k), where $G' = (\Theta_0, \theta_1, \Theta_1, \theta_2, \ldots, \Theta_{k-1})$ and $R' = (\vdash_1, \vdash_2, \ldots, \vdash_k)$, is called a *subsidiary argument* of $\mathcal{A}(\rightarrow \theta)$.

Definition 2: *A* consequential argument from a claim θ, *denoted* $\mathcal{A}(\theta \rightarrow)$*, is a 3-tuple* (θ, R, C)*, where* $C = (\Theta_0, \theta_1, \Theta_1, \theta_2, \ldots, \Theta_{n-2}, \theta_{n-1}, \Theta_{n-1}, \theta_n)$ *is an ordered sequence of wffs* θ_i *and possibly-empty sets of wffs* Θ_i*, with* $n \geq 1$*, and with* $R = (\vdash_1, \vdash_2, \ldots, \vdash_n)$ *an ordered sequence of inference rules such that:*

$$\Theta_0, \ \theta \vdash_1 \ \theta_1,$$

$$\Theta_1, \ \theta_1 \vdash_2 \ \theta_2,$$

$$\vdots$$

$$\Theta_{n-1}, \ \theta_{n-1} \vdash_n \ \theta_n.$$

In other words, the wffs θ_k in C are derivations from θ arising from the successive application of the rules of inference in R, and we call each θ_k in C a *consequence* of θ. We also say write $\mathcal{A}(\theta \to \theta_n)$, which is a *consequential argument of* θ_n *from* θ.[11]

In order that participants may effectively state and contest degrees of commitment to claims, we require a common dictionary of degrees of commitment or support (what Toulmin called "modalities"). Our formalism will support any agreed dictionary, whether quantitative (such as a set of probability values or belief measures) or qualitative (such as non-numeric symbols or linguistic qualifiers), provided there is an agreed partial order on its elements. We define dictionaries for modalities for claims, grounds, consequences and rules of inference.

Definition 3: *Four* modality dictionaries *are defined as follows, each being a (possibly infinite) set of elements having a partial order. The* claims dictionary *is denoted by* $\mathcal{D}_C$*, the* grounds dictionary *by* $\mathcal{D}_G$*, the* consequences dictionary *by* $\mathcal{D}_Q$*, and the* inference dictionary *by* $\mathcal{D}_I$.

Because claims, grounds and consequences are all elements of the same language $\mathcal{L}$, two or more of the dictionaries $\mathcal{D}_C$, $\mathcal{D}_G$ and $\mathcal{D}_Q$ may be the same. However, a distinct dictionary will generally be required for $\mathcal{D}_I$.[12] Because of our belief in epistemological anarchism, we do not specify rules of assignment of dictionary labels by participants in the Agora. In particular, the labels assigned to the conclusions and consequences of arguments are not constrained by those assigned to premises or rules of inference.

Example 1: *The generic argumentation dictionary defined for assessment of risk by [44] is an example of a linguistic dictionary for statements about claims, grounds or consequences, comprising the set:* {Certain, Confirmed, Probable, Plausible, Supported, Open}*. The elements of this dictionary are listed in descending order, with each successive label indicating a weaker belief in the claim.*

Example 2: *Two examples of Inference Dictionaries are* $\mathcal{D}_I$ = {Valid, Invalid}, $\mathcal{D}_I$ = {Acceptable, Sometimes Acceptable, Open, Not Acceptable}*. The dictionary* {Conclusive, Probabilistic, Presumptive, Suggestive, None} *could also be used, provided participants first agreed a partial order on its elements.*

[11] Note that these definitions assume the Cut rule. It would be possible to formulate the definitions without Cut, by creating a grand set of premises $\Theta = \Theta_0 \cup \Theta_1 \cup \ldots \cup \Theta_{n-1}$ and then using this set as a common antecedent premise for each inference $\vdash_i$. We have not done this in order to emphasize the context-dependence of label assignments allowed under Definition 4. In other words, expressing arguments in the way we have done in Definitions 1 and 2 makes clear that the assignment of a label to each inference rule $\vdash_i$ depends on the specific values of the antecedent premises, θ_{i-1} and Θ_{i-1}, but not on other elements of Θ.

[12] In [55], we define a formalism for arguments over acceptability of rules of inference.

Definition 4: A valued grounded argument for a claim θ, *denoted* $\mathcal{A}(\rightarrow \theta, D)$, *is a 4-tuple* (G, R, θ, D), *where* (G, R, θ) *is a grounded argument for* θ *and* $D = (\tilde{d_0}, \tilde{d_1}, \ldots, \tilde{d}_{n-1}, d_\theta, r_1, r_2, \ldots, r_n)$ *is an ordered sequence of labels and vectors of labels, with each* $\tilde{d_i}$ *a vector of dictionary labels from* $\mathcal{D}_C$ *(for* $i = 0, \ldots, n-1$*), with* $d_\theta \in \mathcal{D}_C$ *and with* $r_i \in \mathcal{D}_I$ *(for* $i = 1, \ldots, n$*). Each vector* $\tilde{d_i}$ *comprises those values of the Claims Dictionary assigned to grounds* $\{\theta_i\} \cup \Theta_i$*, the element* d_θ *is that value of the Claims Dictionary assigned to* θ *and each element* r_i *is that value of the Inference Dictionary assigned to* $\vdash_i$. A valued consequential argument from a claim θ, *denoted* $\mathcal{A}(\theta \rightarrow, D)$ *or* $\mathcal{A}(\theta \rightarrow \theta_n, D)$, *is defined similarly.*

Note that modality labels assigned by participants will be revisable in the course of a debate.

3.2 Utterance rules

We next define the rules for discourse participants, building on the definitions above. Moves are denoted by 2-ary or 3-ary functions of the form *name(*$\mathcal{P}_i$*: .)*, where the first argument denotes the participant executing the move. If the move responds to an earlier move by another participant, that earlier move is the second argument. Arguments are separated by colons. We present each move M in the following format:

> *Precons:* Any moves required before M can be executed.
> *Move:* The syntax of M.
> *Meaning:* A textual description of move M.
> *Response:* Any responses required following execution of M.
> *CS Update:* Any amendments to the commitment stores of participants.

The definition and updating rules for participant commitment stores will be given in Section 3.3. In the rules of the Agora, we make a distinction between proposed claims and asserted claims, with the latter, but not the former, leading to Agora commitments on behalf of the participant making them. Likewise, commitments are incurred by a participant who accepts proposed or asserted claims made by other participants. Consequently, only asserted claims made or accepted need to be retracted if the participant desires to express a change of opinion to the Agora. Proposed claims can thus be made or accepted both for a claim and for its negation, without contradicting oneself (Rule 3.10).

In Section 4, we will consider to what extent these rules give operational effect to the Desired Properties listed in Section 2.

Rule 1: Query and Assertion Moves

Move 1.1 Pose Claim:

Precons: None.

Move: *pose*($\mathcal{P}_i :\rightarrow \theta ?$)

Meaning: Participant $\mathcal{P}_i$ asks the Agora if there is a grounded argument for θ.

Response: If any participant $\mathcal{P}_j$ has such an argument, she may present it with: *show_arg*($\mathcal{P}_j : \mathcal{A}(\rightarrow \theta, D)$).

CS Update: None.

Move 1.2 Propose Claim:

Precons: None.

Move: *propose*($\mathcal{P}_i : (\theta, d_\theta)$)

Meaning: Participant $\mathcal{P}_i$ informs the Agora that she has a grounded argument for θ, and has assigned it a modality of d_θ, where $\theta \in \mathcal{L}$ and $d_\theta \in \mathcal{D}_C \cup \{\ \}$. The use of the empty set for d_θ indicates that the participant has not assigned a modality label to the claim.

Response: None required.

CS Update: None.

Move 1.3 Assert Claim:

Precons: None.

Move: *assert*($\mathcal{P}_i : (\theta, d_\theta)$)

Meaning: Participant $\mathcal{P}_i$ informs the Agora that she has a valued grounded argument for θ, and has assigned it a modality of d_θ, where $\theta \in \mathcal{L}$ and $d_\theta \in \mathcal{D}_C$, which she believes is compelling.

Response: No moves required.

CS Update: (θ, d_θ) inserted into $\mathcal{CS}(\mathcal{P}_i)$.

Move 1.4 Query Proposed Claim:

Precons: *propose*($\mathcal{P}_i : (\theta, d_\theta)$)

Move: *query*($\mathcal{P}_j$:*propose*($\mathcal{P}_i : (\theta, d_\theta)$))

Meaning: Participant $\mathcal{P}_j$ asks participant $\mathcal{P}_i$ for her valued grounded argument for claim θ, following the latter's Propose Claim move.

Response: $\mathcal{P}_i$ must respond with: *show_arg(*$\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)$*)*.

CS Update: None.

Move 1.5 Query Asserted Claim:

Precons: assert($\mathcal{P}_i : (\theta, d_\theta)$*)*

Move: *query(*$\mathcal{P}_j$ *:assert(*$\mathcal{P}_i : (\theta, d_\theta)$*)).*

Meaning: Participant $\mathcal{P}_j$ asks participant $\mathcal{P}_i$ for her valued grounded argument for claim θ, following the latter's Assert Claim move.

Response: Provided the claim θ has not been retracted under Move 3.8 since its assertion, $\mathcal{P}_i$ must respond with: *show_arg(*$\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)$*)*.

CS Update: None.

Move 1.6 Show Grounded Argument:

Precons: None.

Move: *show_arg(*$\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)$*).*

Meaning: Participant $\mathcal{P}_i$ presents the Agora with her valued, grounded argument for $\theta \in \mathcal{L}$ and D a sequence of labels and vectors of labels defined as in Definition 4.

Response: None required.

CS Update: None.

Move 1.7 Pose Consequence:

Precons: None.

Move: *pose_cons(*$\mathcal{P}_k : \theta \rightarrow ?$*)*

Meaning: Participant $\mathcal{P}_k$ asks the Agora if there is a consequential argument from θ.

Response: None required.

CS Update: None.

Move 1.8 Propose Consequence:

Precons: None.

Move: *propose_cons(*$\mathcal{P}_i : (\theta, \phi, d_\phi)$*)*

Meaning: Participant $\mathcal{P}_i$ informs the Agora that she has a consequential argument of ϕ from θ, and has assigned it a modality of d_ϕ, where $\theta, \phi \in \mathcal{L}$ and $d_\phi \in \mathcal{D}_Q \cup \{\ \}$.

Response: None required.

CS Update: None.

Move 1.9 Assert Consequence:

Precons: None.

Move: *assert_cons(*$\mathcal{P}_i : (\theta, \phi, d_\phi)$*)*

Meaning: Participant $\mathcal{P}_i$ informs the Agora that she has a valued consequential

argument of ϕ from θ, which she believes is compelling, and has assigned it a modality of d_ϕ, where $\theta, \phi \in \mathcal{L}$ and $d_\phi \in \mathcal{D}_Q$.

Response: None required.

CS Update: None.

Move 1.10 Query Proposed Consequence:

Precons: propose_cons($\mathcal{P}_i : (\theta, \phi, d_\phi)$*)*

Move: query_cons($\mathcal{P}_j$ *:propose_cons(*$\mathcal{P}_i : (\theta, \phi, d_\phi)$*))*

Meaning: Participant $\mathcal{P}_j$ asks participant $\mathcal{P}_i$ for her valued consequential argument for ϕ from θ, following the latter's Propose Consequence move.

Response: $\mathcal{P}_i$ must respond with: *show_cons(*$\mathcal{P}_i : \mathcal{A}(\theta \rightarrow \phi, D)$*)*.

CS Update: None.

Move 1.11 Query Asserted Consequence:

Precons: assert_cons($\mathcal{P}_i : (\theta, \phi, d_\phi)$*)*

Move: *query_cons(*$\mathcal{P}_j$ *:assert_cons(*$\mathcal{P}_i : (\theta, \phi, d_\phi)$*))*

Meaning: Participant $\mathcal{P}_j$ asks participant $\mathcal{P}_i$ for her valued consequential argument for ϕ from θ, following the latter's Assert Consequence move.

Response: $\mathcal{P}_i$ must respond with: *show_cons(*$\mathcal{P}_i : \mathcal{A}(\theta \rightarrow \phi, D)$*)*.

CS Update: None.

Move 1.12 Show Consequential Argument:

Precons: None.

Move: *show_cons(*$\mathcal{P}_i : \mathcal{A}(\theta \rightarrow \phi, D)$*)*.

Meaning: Participant $\mathcal{P}_i$ presents the Agora with her valued consequential argument for ϕ from θ, where $\theta \in \mathcal{L}$ and D a sequence of labels and vectors of labels defined as in Definition 4.

Response: None required.

CS Update: None.

Move 1.13 Propose Mode of Inference:

Precons: None.

Move: *propose_inf*($\mathcal{P}_i : (\vdash_t, r_t)$)

Meaning: Participant $\mathcal{P}_i$ informs the Agora that she believes that $\vdash_t$ is a mode of inference of strength at least r_t, where $r_t \in \mathcal{D}_I \cup \{\ \}$.

Response: None required.

CS Update: None. □

Note that the query and assertions rules are not symmetric between grounded and consequential arguments: participants may only propose or assert claims for which they have grounded arguments, but they need not necessarily have considered the consequences of these claims. Next, we explicitly define the *Contest Proposed Claim* move, the *Contest Asserted Claim* move and the associated query moves. However, for reasons of brevity, we state only the syntax of the other contestation moves, and omit their associated query moves.

Rule 2: Contestation Moves

Move 2.1 Contest Proposed Claim:

Precons: *propose*($\mathcal{P}_i : (\theta, d_\theta)$)

Move: *contest*($\mathcal{P}_j$:*propose*($\mathcal{P}_i : (\theta, d_\theta)$))

Meaning: Participant $\mathcal{P}_j$ informs the Agora that she contests either $\mathcal{P}_i$'s conclusion θ and/or the modality d_θ assigned to θ in the latter's Proposed Claim.

Response: None required.

CS Update: None.

Move 2.2 Contest Asserted Claim:

Precons: *assert*($\mathcal{P}_i : (\theta, d_\theta)$)

Move: *contest*($\mathcal{P}_j$:*assert*($\mathcal{P}_i : (\theta, d_\theta)$))

Meaning: Participant $\mathcal{P}_j$ informs the Agora that she contests either $\mathcal{P}_i$'s conclusion θ and/or the modality d_θ assigned to θ in the latter's Asserted Claim.

Response: None required.

CS Update: None.

Move 2.3 Query Contested Proposed Claim:

Precons: *contest(*$\mathcal{P}_j$ *:propose(*$\mathcal{P}_i : (\theta, d_\theta)$*))*

Move: *query(*$\mathcal{P}_k$ *:contest(*$\mathcal{P}_j$ *:propose(*$\mathcal{P}_i : (\theta, d_\theta)$*)))*

Meaning: Participant $\mathcal{P}_k$ queries the contestation by $\mathcal{P}_j$ of $\mathcal{P}_i$'s Proposed Claim.

Response: *propose(*$\mathcal{P}_j : (\theta, d'_\theta)$*)* (where $d'_\theta \neq d_\theta$) OR *propose(*$\mathcal{P}_j : (\neg\theta, d'_\theta)$*)*, (where $d'_\theta > d_\theta$).

Meaning: Participant $\mathcal{P}_j$ must respond to the query either with an assignment of an alternative modality d'_θ for claim θ, OR with with a stronger assertion of the negation of θ.

CS Update: None.

Move 2.4 Query Contested Asserted Claim:

Precons: *contest(*$\mathcal{P}_j$ *:assert(*$\mathcal{P}_i : (\theta, d_\theta)$*))*

Move: *query(*$\mathcal{P}_k$ *:contest(*$\mathcal{P}_j$ *:assert(*$\mathcal{P}_i : (\theta, d_\theta)$*)))*

Meaning: Participant $\mathcal{P}_k$ queries the contestation by $\mathcal{P}_j$ of $\mathcal{P}_i$'s Proposed Claim.

Response: *assert(*$\mathcal{P}_j : (\theta, d'_\theta)$*)* (where $d'_\theta \neq d_\theta$) OR *assert(*$\mathcal{P}_j : (\neg\theta, d'_\theta)$*)*, (where $d'_\theta > d_\theta$).

Meaning: Participant $\mathcal{P}_j$ must respond to the query either with an assignment of an alternative modality d'_θ for claim θ, OR with with a stronger assertion of the negation of θ.

CS Update: None.

Move 2.5 Contest Ground:

Move: *contest_ground(*$\mathcal{P}_j$ *: show_arg(*$\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D) : (\theta_t, d_{\theta_t})$*))*

Move 2.6 Contest Inference:

Move: *contest_inf(*$\mathcal{P}_j$ *:show_arg(*$\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D) : \vdash_t$*))*

Move 2.7 Contest Modality:

Move: *contest_mod*($\mathcal{P}_j$:*show_arg*($\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)$))

Move 2.8 Contest Consequence:

Move: *contest_cons*($\mathcal{P}_j$:*show_cons*($\mathcal{P}_i : \mathcal{A}(\theta \rightarrow \phi, D) : (\theta_t, d_{\theta_t})$)) □

We next define moves for acceptance, modification and retraction of claims and modalities. As before, where a move is similar to earlier moves, we state this and present only the syntax of the later move. Note that Move 3.3 *(Change Modalities)* allows a participant to revise her assignment of modalities to a valued argument. Similarly, declarations of modal beliefs expressed in other moves (e.g. in *accept_assert*) may also be revised by subsequently executing the same move with a different set of dictionary values.

Rule 3: Resolution Moves

Move 3.1 Accept Proposed Claim:

Precons: Both *propose*($\mathcal{P}_i : (\theta, d_\theta)$) and *show_arg*($\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)$)

Move: *accept_prop*($\mathcal{P}_j$:*show_arg*($\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)$))

Meaning: Participant $\mathcal{P}_j$ informs the Agora that she accepts the claim θ proposed earlier by participant $\mathcal{P}_i$. This move is equivalent to executing the following two moves in sequence:

propose($\mathcal{P}_j : (\theta, d_\theta)$) and *show_arg*($\mathcal{P}_j : \mathcal{A}(\rightarrow \theta, D)$),

except that $\mathcal{P}_j$ does not incur the obligation to respond to any query under Move 1.4 that is incurred by the *propose* move.

Response: None required.

CS Update: None.

Move 3.2 Accept Asserted Claim: As for the previous move, but for asserted claims.

Move: *accept_assert*($\mathcal{P}_j$:*show_arg*($\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)$))

CS Update: (θ, d_θ) inserted into $\mathcal{CS}(\mathcal{P}_j)$.

Move 3.3 Change Modalities:

Precons: *show_arg*($\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)$) and not *retract*($\mathcal{P}_i : assert(\mathcal{P}_i : (\theta, d_\theta))$)

Move: *show_arg*($\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D')$)

Meaning: Participant $\mathcal{P}_i$ informs the Agora that she wishes to revise the modalities assigned in an earlier valued argument for θ, from D to D', where $D' \neq D$.

Response: None required.

CS Update: (θ, d'_θ) replaces (θ, d_θ) in $\mathcal{CS}(\mathcal{P}_i)$.

Move 3.4 Accept Mode of Inference: As for Move 3.1 (*Accept Proposed Claim*), but for modes of inference:

Move: *accept_inf($\mathcal{P}_j$:propose_inf($\mathcal{P}_i$: $(\vdash_t, r_t)$))*

Move 3.5 Accept Consequence: As for Move 3.1 (*Accept Proposed Claim*), but for consequences:

Move: *accept_cons($\mathcal{P}_j$:show_cons($\mathcal{P}_i$: $\mathcal{A}(\theta \rightarrow \phi, D)$))*

Move 3.6 Precizate Proposed Claim:

Precons: Both *propose($\mathcal{P}_i$: (θ, d_θ))* and *show_arg($\mathcal{P}_i$: $\mathcal{A}(\rightarrow \theta, D)$)* and not *retract($\mathcal{P}_i$: $assert(\mathcal{P}_i : (\theta, d_\theta))$)*

Move: *precizate($\mathcal{P}_i$:show_arg($\mathcal{P}_i$: $\mathcal{A}(\rightarrow \theta, D)$): $\mathcal{A}'(\rightarrow \theta, D')$)*

Meaning: Participant $\mathcal{P}_i$ informs the Agora that she wishes to qualify her earlier argument $\mathcal{A}(\rightarrow \theta, D)$ with the argument $\mathcal{A}'(\rightarrow \theta, D')$, where these two arguments are identical except that: (a) the latter begins from ground $\Phi \cup \Theta_0$ instead of Θ_0, with $\Phi \cap \Theta_k = \{\}$ $(\forall k = 0, 1, \ldots, n-1)$ and $\Phi \cap \theta = \{\}$, and (b) D' may be different to D.

Response: None required.

CS Update: None.

Move 3.7 Precizate Asserted Claim:

Precons: Both *assert($\mathcal{P}_i$: (θ, d_θ))* and *show_arg($\mathcal{P}_i$: $\mathcal{A}(\rightarrow \theta, D)$)* and not *retract($\mathcal{P}_i$: $assert(\mathcal{P}_i : (\theta, d_\theta))$)*

Move: *precizate($\mathcal{P}_i$:show_arg($\mathcal{P}_i$: $\mathcal{A}(\rightarrow \theta, D)$): $\mathcal{A}'(\rightarrow \theta, D')$)*

Meaning: As for the previous move, but for asserted rather than proposed claims.

Response: None required.

CS Update: If $d'_\theta) \neq d_\theta)$, then (θ, d'_θ) replaces (θ, d_θ) in $\mathcal{CS}(\mathcal{P}_i)$

Move 3.8 Retract Asserted Claim:

Precons: *assert($\mathcal{P}_i$: (θ, d_θ))*

Move: *retract*($\mathcal{P}_i : assert(\mathcal{P}_i : (\theta, d_\theta))$)

Meaning: Any participant $\mathcal{P}_i$ who has earlier asserted a claim for θ may withdraw it at any time. This move releases $\mathcal{P}_i$ from the obligation of responding to any query under Move 1.5.

Response: None required.

CS Update: (θ, d_θ) removed from $\mathcal{CS}(\mathcal{P}_i)$

Move 3.9 Retract Accepted Claim:

Precons: *accept_assert*($\mathcal{P}_j$:*show_arg*($\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)$))

Move: *retract*($\mathcal{P}_j : assert(\mathcal{P}_i : (\theta, d_\theta))$)

Meaning: As for the previous rule, but for accepted asserted claims.

Response: None required.

CS Update: (θ, d_θ) removed from $\mathcal{CS}(\mathcal{P}_j)$

Move 3.10 No contradiction: Any participant $\mathcal{P}_i$ who asserts θ may not at any time subsequently assert or accept an assertion for $\neg\theta$, unless they have in the interim moved: *retract*($\mathcal{P}_i$:*assert*($\mathcal{P}_i : (\theta, d_\theta)$)). Similarly, any participant $\mathcal{P}_j$ who has accepted an assertion for θ may not at any time subsequently assert or accept an assertion for $\neg\theta$, unless they have in the interim moved: *retract*($\mathcal{P}_j : assert(\mathcal{P}_i : (\theta, d_\theta))$). □

This last rule, 3.10, prohibits explicit contradictions. An interesting question is to what extent we should prohibit implicit contradictions, for example, those arising as consequences — perhaps many inference steps removed — from a claim. We have decided not to prohibit these. In real scientific debates, when a claim is shown to lead, after one or more steps of reasoning, to a contradiction, this is normally brought to the attention of the claim's proponent. She may then ignore this, or may retract the claim or may present counter-arguments against the argument asserting the implicit contradiction. The Agora formalism as we have defined it permits each of these options and any resulting dialogues to be represented.[13]

3.3 Dialogue rules

We next define a dialogue and a rule which precludes infinite regression by malevolent participants. We then define sets called Commitment Stores, as in [4, 34, 79].

[13]Note that our approach differs from that of MacKenzie [49], who distinguished between immediate inference, whose consequences a proponent of a claim must accept, and those arising from multiple inference steps, which need not be accepted.

These stores record the proposals and assertions made by participants, both individually and for the Agora as a community, and track these as they change.

Definition 5: *A* Dialogue *is a finite sequence of discourse moves by participants in the Agora, in accordance with the rules above.*

Rule 4: Moves may only be executed once by any participant with respect of the same participant and claim, ground, inference, consequence or modality. □

Thus, this rule permits the following three moves by participant $\mathcal{P}_j$:

$$\textit{contest\_mod}(\mathcal{P}_j : \textit{show\_arg}(\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)))$$

$$\textit{contest\_mod}(\mathcal{P}_j : \textit{show\_arg}(\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D')))$$

$$\textit{contest\_mod}(\mathcal{P}_j : \textit{show\_arg}(\mathcal{P}_k : \mathcal{A}(\rightarrow \theta, D))),$$

but not both the following two moves:

$$\textit{contest\_mod}(\mathcal{P}_j : \textit{show\_arg}(\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)))$$

$$\textit{contest\_mod}(\mathcal{P}_j : \textit{show\_arg}(\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D))).$$

This rule is intended to prevent the dialogue degenerating into an infinite regress, as when a child repeatedly asks "Why?", or mindless repetition. However, the rule does not prevent genuine objections. For example, the successive objections to the use of Modus Ponens and its variants voiced by the Tortoise in Lewis Carroll's dialogue with Achilles [12] would not be prohibited by Rule 4 because each objection contests the use of a different rule of inference, even though the differences may be seen by some as marginal.

Definition 6: *The* commitment store of player $\mathcal{P}_i$, $i = 1, 2, \ldots$, *denoted* $\mathcal{CS}(\mathcal{P}_i)$, *is a possibly empty set* $\{(\theta, d_\theta) \mid \theta \in \mathcal{L},\ d_\theta \in \mathcal{D}_C\}$, *where each* θ *is an asserted claim made or accepted by* $\mathcal{P}_i$, *and each corresponding* d_θ *is the claim dictionary value assigned by* $\mathcal{P}_i$ *to* θ.

The values in participants' stores are updated by the following rule:

Rule 5: Participant Commitment Store Update: All commitment stores of all participants are initially empty. Whenever participant $\mathcal{P}_i$ executes either of the moves 1.3 or 3.2:

$$\textit{assert}(\mathcal{P}_i : (\theta, d_\theta)),$$

$$\textit{accept\_assert}(\mathcal{P}_i : assert(\mathcal{P}_j : (\theta, d_\theta)))$$

or their equivalents, then the tuple (θ, d_θ) is inserted into $\mathcal{CS}(\mathcal{P}_i)$. Whenever participant $\mathcal{P}_i$ subsequently executes a retraction move (3.8 or 3.9) for (θ, d_θ), the tuple (θ, d_θ) is removed from $\mathcal{CS}(\mathcal{P}_i)$. Similarly, whenever $\mathcal{P}_i$ executes a change modalities move (move 3.3) for (θ, d_θ), the value of (θ, d_θ) in $\mathcal{CS}(\mathcal{P}_i)$ is revised accordingly. □

3.4 Experiment and Nature's responses

Uncertainty in scientific domains, such as that of carcinogenic risk assessment, is normally only resolved by gathering further evidence, typically in the form of experimental results. We may think of experiments as means to test predictions, themselves the hypothesized consequences of some theory. Because our argumentation framework includes consequences from claims, we have a means by which to represent a prediction from a theory. Then, once an experiment is undertaken (outside the Agora), Pera's three-person dialogue model gives us a means to represent the manner in which the Agora community assesses the import of the experimental results.

In this section, we show how this can be done. Let θ be a wff which expresses some disputed claim, for example, "*Chemical $\mathcal{X}$ causes brain cancers in humans.*" Let ϕ be a consequence of this statement, for example "*Humans exposed in a specified manner and to a specified extent to $\mathcal{X}$ will show statistically-significant increased incidences of brain cancers in a properly designed and conducted epidemiological study of the effects of $\mathcal{X}$.*" A participant $\mathcal{P}_i$ may then move the consequential assertion:

$$\textit{assert\_cons}(\mathcal{P}_i : (\theta, \phi, d_\phi)),$$

where $d_\phi \in \mathcal{D}_Q$. When queried, we can assume she presents a valued consequential argument from θ leading to ϕ, say $\mathcal{A}(\theta \rightarrow \phi, D)$. However, she may or may not have a grounded argument *for* θ from premises, and thus may not necessarily have proposed or asserted the claim θ.

Suppose now that an epidemiological study of the effect of $\mathcal{X}$ on humans is undertaken. Under Pera's model, the results of such a study correspond to *Nature's* move in the three-person dialogue game, so let us denote the experimental results by ϕ^N. Often in scientific dialogues of this type, there is then a debate in the scientific community as to whether or not the study was designed or conducted properly.[14] We may consider such a debates to be about the statement: "*ϕ^N is a valid instantiation of ϕ.*" Arguments for and against this statement are arguments

[14] Witness the heated debate in Britain during 1999 over the validity of a study undertaken to assess potential adverse impacts of feeding Genetically-Modified potatoes to rats [6, 58].

for and against the validity, and hence the acceptability, of the experimental evidence arising from the epidemiological study; such arguments are likely to involve statistical and methodological issues rather than, say, issues of chemical analysis or biomedical causality. In addition, there may be debate over whether or not the results tend to confirm or refute ϕ, that is whether they are statistically-significant or not, especially if the statistical testing procedures yield values near the critical region boundary values of the test. This debate concerns the statement: "ϕ^N *provides statistically-significant evidence for ϕ, at a specified level of significance.*" This discussion, too, often involve statistical and methodological issues. Both these debates form part, therefore, of the process of mediation of *Nature's* experimental responses, a role which Pera assigns to the scientific community concerned. In our model, that community is the set of participants in the Agora; the Agora can readily accommodate this debate, since we make no prior specification of the content of propositions, premises or inference rules.

Suppose, after due debate, participants in the community accept that ϕ^N is a valid instantiation of, and provides confirming evidence for, ϕ, at a specified level of statistical significance. In other words, the epidemiological study of $\mathcal{X}$ is believed to show statistically-significant increased incidences of brain cancers in humans exposed to the chemical. This statement, ϕ, may then form the basis for an argument for the claim θ, that $\mathcal{X}$ causes brain cancers in humans, additional to whatever arguments may have been advanced for θ earlier. This new argument would run as follows:

$$\mathcal{A}(\theta \rightarrow \phi)$$

$$\phi \vdash_{Abd} \theta$$

where $\mathcal{A}(\theta \rightarrow \phi)$ is participant $\mathcal{P}_i$'s earlier consequential argument from θ to ϕ, and $\vdash_{Abd}$ represents inference by abduction. Whether this use of abduction in this context at this moment is acceptable or not would be a matter for the Agora participants, and may depend on the weight of other arguments or evidence for θ. In practice, at least in the environmental health domain, unexpected findings of carcinogenic or toxic effects are not typically accepted as providing firm evidence of the claim until found repeatedly. Thus, one argument using abduction may not be accepted, but several arguments doing so, proceeding from separate and independent experimental results to the same conclusion, may well be. For the latter to happen, of course, all such arguments need to be articulated.

Definition 7: *A* testable prediction *ϕ of a wff θ is a consequence of θ whose truth-status may potentially be verified (at least to some level of statistical significance) by means of a scientific experiment involving a change to the material world (i.e not a thought experiment). The outcome of such an experiment is denoted by ϕ^N.*

Definition 8: *Let ϕ be a testable prediction of a wff θ. If the community has agreed a level of statistical significance for the conduct of a scientific experiment testing ϕ and then, following that experiment, agreed that ϕ^N is both a valid instantiation of ϕ and that it provides statistically-significant evidence for ϕ, at the specified level of significance, then we say that ϕ^N is a* confirming instance *of ϕ.*

Definition 9: *Let ϕ be a testable prediction of a wff θ. If the community has agreed a level of statistical significance for the conduct of a scientific experiment testing ϕ and then, following that experiment, agreed that ϕ^N is a valid instantiation of ϕ but that it does not provide statistically-significant evidence for ϕ, at the specified level of significance, then we say that ϕ^N is a* disconfirming instance *of ϕ.*

Of course, these definitions beg the question as to what constitutes "agreement" of statements by the Agora community. We make this notion precise in the next subsection. In a real scientific debate, there may be many proposed tests of a theory, i.e. possible consequences which are testable predictions of a claim. An advantage of a computerized system is that it is straightforward to track these. Accordingly, analogously with Commitment Stores, we could define a Consequence Store for each claim θ for which consequences are proposed or asserted, in order to maintain a current list of all such consequences. The status of any experimental tests undertaken for these consequences could also be tracked. We also leave formalization of this idea to the next section, after we have formalized the notion of acceptance.

3.5 The *Agora's* Commitment Store

Analogous to the concept of Commitment Store for each participant, we now define a commitment store for the *Agora*, the community as a whole. Claims in this store are labeled with modalities on the basis of the Agora debate at that point and the weight of any experimental evidence. This could be achieved in a number of ways. For example, a skeptical community could define the *Agora's* modality for a claim θ to be the minimum claim modality assigned by any of those Participants claiming or supporting θ. A credulous community could instead assign to the *Agora* the maximum claim modality assigned by any of the participants to θ. Variations on these approaches could utilize majority opinion or weighted voting schemes. In the real world, for instance, the opinions of senior scientists generally carry more weight than do those of younger scientists, no doubt because of their greater perceived understanding and intuition.

Because we wish to model dialectical discourse, we have instead chosen to assign the *Agora's* modalities on the basis of the existence of arguments for and against the claim. To do this, we draw on the generic argumentation dictionary for debates about carcinogenicity of chemicals presented in [44], which is also derived from Toulmin's [75] schema. However, we modify this work to allow for responses to counter arguments to claims. We begin by defining certain relationships between arguments followed by the Claims Dictionary for the *Agora*, and our definitions assign the *Agora* a status in the debate which is above that of any one participant.[15]

Definition 10: *Let* $G = (\Theta_0, \theta_1, \Theta_1, \theta_2, \ldots, \Theta_{n-2}, \theta_{n-1}, \Theta_{n-1})$. *Then an argument* $\mathcal{A}(\rightarrow \theta) = (G, R, \theta)$ *is* consistent *if* G *is consistent, that is if there do not exist* $\alpha, \beta \in \Theta_0 \cup \{\theta_1\} \cup \Theta_1 \cup \{\theta_2\} \cup \ldots \cup \Theta_{n-1}$ *such that* $\neg\beta$ *is a consequence of* α *under any combination of the rules of inference contained in* R.

Definition 11: *Let* $\mathcal{A}(\rightarrow \theta) = (G, R, \theta)$ *and* $\mathcal{B}(\rightarrow \neg\theta) = (H, S, \neg\theta)$ *be two arguments. We say that* $\mathcal{B}(\rightarrow \neg\theta)$ rebuts $\mathcal{A}(\rightarrow \theta)$, *and that* $\mathcal{B}$ *is a* rebuttal *for* $\mathcal{A}$.

Definition 12: *Let* $G = (\Theta_0, \theta_1, \Theta_1, \theta_2, \ldots, \theta_{n-1}, \Theta_{n-1})$ *and let* $\mathcal{A}(\rightarrow \theta) = (G, R, \theta)$ *be an argument. Suppose* $\mathcal{B}(\rightarrow \neg\theta_k) = (H, S, \neg\theta_k)$ *is another argument, for some* k. *We say* $\mathcal{B}(\rightarrow \neg\theta_k)$ undercuts $\mathcal{A}(\rightarrow \theta)$ *and that* $\mathcal{B}$ *is an* undercutting *argument or an* undercutter *for* $\mathcal{A}$.

We may think of an undercutter as a rebuttal for one of the intermediate wffs within an argument. The reverse is also true: the subsidiary argument of $\mathcal{A}(\rightarrow \theta)$ for θ_k is a rebuttal of any undercutter of θ_k.

Definition 13: *Let* $\mathcal{A}(\rightarrow \theta) = (G, R, \theta)$ *be an argument for* θ, $\mathcal{B}(\rightarrow \neg\theta)$ *a rebuttal, and* $\mathcal{C}(\rightarrow \neg\theta_k)$ *an undercutter. We say that* $\mathcal{B}$ *and* $\mathcal{C}$ attack $\mathcal{A}$ *and call them* attackers. *An argument* $\mathcal{A}' \neq \mathcal{A}$, *where* $\mathcal{A}'$ *is not a subsidiary argument of* $\mathcal{A}$ *and which undercuts either* $\mathcal{B}$ *or* $\mathcal{C}$, *is called a* counter-argument *or a* counter-attacker *for* $\mathcal{A}$. *An argument* $\mathcal{A}$ *for which counter-attackers exist for each of its attackers is said to be* well-defended.

Definition 14: *The claims dictionary for the* Agora *is the set* $\mathcal{D}_{C,A} = \{$Accepted, Probable, Plausible, Supported, Open$\}$.

We next define claim modality labels and the Commitment Store for the *Agora*. As with the modality labels for individual participants, the *Agora's* assignment of labels may change over time. In exploring the semantics of the formal system in Section 4, we will find it useful to incorporate the temporal element explicitly in our notation and definitions. We assume time is discrete, countable and partially

[15]Note that the work we draw on, [43, 44], because it allows for arguments of more than one inference step, uses a definition of *undercut* which differs from that of Pollock [63], whose earlier work introduced the term. Our definition follows that of [43].

ordered under $\leq$, and that only one Agora move is executed at each time point. We use the notation $t, s, u, \ldots$ and $t_1, t_2, \ldots$ to denote time points. For simplicity of expression we sometimes omit the time index on the modality symbols, where there is no likelihood of confusion.

Definition 15: *The commitment store of the* Agora *at time* t*, denoted* $\mathcal{CS}_t(\mathcal{P}_A)$*, is a possibly empty set* $\{(\theta, d_{\theta,t,A}) \mid \theta \in \mathcal{L},\ d_{\theta,A} \in \mathcal{D}_{C,A}\}$*, where each* $d_{\theta,t,A}$ *is the claim modality assigned by the Agora community to* θ *at time* t*, in accordance with the next two rules.*

Rule 6: The Agora's Claim Modalities: The modality $d_{\theta,t,A}$ of the *Agora* for the wff θ is assigned values at time t as follows:

- If, at time t, θ is a wff for which no grounded argument has yet been provided by an Agora participant, then $d_{\theta,t,A}$ is assigned the value *Open*.
- If, at time t, θ is a wff for which at least one grounded argument has been provided by an Agora participant by an execution of Move 1.6, then $d_{\theta,t,A}$ is assigned the value *Supported*.
- If, at time t, θ is a wff for which a consistent grounded argument has been provided to the Agora by a participant through an execution of Move 1.6 then $d_{\theta,t,A}$ is assigned the value *Plausible*.
- If, at time t, θ is a wff for which a consistent grounded argument has been provided to the Agora by a participant through an execution of Move 1.6, and for which no rebuttals or undercutters have yet been presented by Agora participants via Move 1.6, then $d_{\theta,t,A}$ is assigned the value *Probable*.
- If, at time t, θ is a wff for which a consistent grounded argument has been provided to the Agora by a participant through an execution of Move 1.6 and this argument is well-defended, then $d_{\theta,t,A}$ is assigned the value *Accepted*.
- The modality label definitions here are listed in ascending order of strength, and the label assigned is always the highest applicable label. $\Box$

Rule 7: Update of The Agora Commitment Store: The commitment store of the *Agora* is initially empty. Let θ be a wff. Suppose time s is the first time that an Agora participant executes a pose claim, propose claim or an assert claim move (moves 1.1, 1.2 and 1.3) regarding θ. Then at time s the element $(\theta, d_{\theta,s,A})$ is inserted into $\mathcal{CS}_s(\mathcal{P}_A)$, with $d_{\theta,s,A} = Open$. These elements are then updated according to the previous Rule after each move by Agora participants. □

These rules assign the *Agora's* claim modalities according to the status of debate within the Agora. As such, there is nothing to prevent claims being accepted without any reference to external experimental evidence, which is consistent with the epistemological anarchist philosophy of science we have adopted. However, to allow for reference to experiment, we now define two further sets of modalities, relating to the experimental test status of a particular predicted consequence of a claim, and to the summary status across all predicted consequences. In both cases, our definitions encode Pera's model of the scientific community acting to mediate the results of experiments.

Definition 16: *The test status modality dictionary for a testable prediction ϕ is the set $\mathcal{D}_{Test} = \{$*Confirming instance, Disconfirming instance, Inconclusive test, Invalid test, Open$\}$.

Definition 17: *Let θ be a wff. The Consequence Store for θ at time t, denoted $\mathcal{QS}_t(\theta)$, is a possibly empty set of 3-tuples, $\{(\phi, \phi^N, e_{\theta,\phi,\phi^N}) \mid \phi, \phi^N \in \mathcal{L},\ e_{\theta,\phi,\phi^N} \in \mathcal{D}_{Test}\}$, where each ϕ is a testable prediction for θ proposed or asserted by means of an execution of Move 1.8 or Move 1.9 by time t, each ϕ^N is the outcome of an experiment undertaken to test ϕ, and e_{θ,ϕ,ϕ^N} is the test status modality value assigned by the Agora community to ϕ at time t, in accordance with the next two rules.*

Rule 8: The Agora's Test Modalities: The modalities e_{θ,ϕ,ϕ^N} of the *Agora* for experimental outcomes ϕ^N of testable predictions ϕ of claims θ are assigned values at time $t+1$ as follows:

- For each ϕ where, by time t, no scientific experiment to test ϕ has been undertaken, then ϕ^N is assigned the value $\{\ \}$, and e_{θ,ϕ,ϕ^N} is assigned the value *Open* at time t.

- For each ϕ and ϕ^N where, by time t, (a) a scientific experiments to test ϕ has been undertaken and has resulted in outcome ϕ^N, and where (b) the modality label, $d_{\phi^V,A}$, for the *Agora* at time t assigned by Rules 6 and 7 to the statement $\phi^V =$ "*ϕ^N is a valid instantiation of ϕ*" is not *Accepted*, then e_{θ,ϕ,ϕ^N} is assigned the value *Invalid test* at time t.

- For each ϕ and ϕ^N where, by time t, (a) a scientific experiments to test ϕ has been undertaken and has resulted in outcome ϕ^N, and where (b) the modality label, $d_{\phi^V,A}$, for the *Agora* at time t assigned by Rules 6 and 7 to the statement ϕ^V = "ϕ^N *is a valid instantiation of* ϕ" is *Accepted*, and (c) the modality label, $d_{\phi^D,A}$, for the *Agora* at time t assigned by Rules 6 and 7 to the statement ϕ^D = "ϕ^N *is a disconfirming instance of* ϕ" is not *Accepted*, and (d) the modality label, $d_{\phi^C,A}$, for the *Agora* at time t assigned by Rules 6 and 7 to the statement ϕ^C = "ϕ^N *is a confirming instance of* ϕ" is not *Accepted*, then e_{θ,ϕ,ϕ^N} is assigned the value *Inconclusive test* at time t.

- For each ϕ and ϕ^N where, by time t, (a) a scientific experiments to test ϕ has been undertaken and has resulted in outcome ϕ^N, and where (b) the modality label, $d_{\phi^D,A}$, for the *Agora* at time t assigned by Rules 6 and 7 to the statement ϕ^D = "ϕ^N *is a disconfirming instance of* ϕ" is *Accepted*, then e_{θ,ϕ,ϕ^N} is assigned the value *Disconfirming instance* at time t.

- For each ϕ and ϕ^N where, by time t, (a) a scientific experiments to test ϕ has been undertaken and has resulted in outcome ϕ^N, and where (b) the modality label, $d_{\phi^C,A}$, for the *Agora* at time t assigned by Rules 6 and 7 to the statement ϕ^C = "ϕ^N *is a confirming instance of* ϕ" is *Accepted*, then e_{θ,ϕ,ϕ^N} is assigned the value *Confirming instance* at time t.

- The modality label definitions here are listed in ascending order of strength, and the label assigned is always the highest applicable label. □

Rule 9: Consequence Store Update: Let θ be a wff. The Consequence Store for θ, $\mathcal{QS}_t(\theta)$, for $t \geq 0$, is initially empty. Let s be the time at which the first proposal or assertion of a consequence, say ϕ, from θ, is undertaken by an Agora participant executing Moves 1.8 or 1.9. Then, at time s the element $(\phi, \phi^N, e_{\theta,\phi,\phi^N})$ is inserted into $\mathcal{QS}_{s+1}(\theta)$. For $t \geq s$, the elements of $\mathcal{QS}_t(\theta)$ are updated according to the previous rule after each move in the Agora. □

We now define rules which accumulate across all the experiments undertaken to test a claim θ, so as to provide an aggregate view of the experimental evidence for and against θ. In this way, we can draw summary conclusions about the evidential status of claims. Note that there are many ways in which the experimental modalities and update rules could be defined, and we present one set of definitions as an illustration of our approach. Our definitions assign empirical support modalities on the basis of the proportion of experiments which confirm or disconfirm a claim.

Definition 18: *The empirical support modality dictionary for a wff* θ *is the set* $\mathcal{D}_{Emp} = \{\text{Confirmed, Refuted, Inconclusive, Untested}\}$.

Definition 19: *The empirical support store of the* Agora *at time* t*, denoted* $\mathcal{ES}_t(\mathcal{P}_A)$*, is a possibly empty set* $\{(\theta, f_{\theta,t,A}) \mid \theta \in \mathcal{L},\ f_{\theta,t,A} \in \mathcal{D}_{Emp}\}$*, where each* $f_{\theta,t,A}$ *is the empirical support modality assigned by the Agora community to* θ *at time* t*, in accordance with the next two rules.*

Rule 10: The Agora's Empirical Support Modalities: The empirical support modality $f_{\theta,t,A}$ of the *Agora* for the wff θ is assigned values at time t as follows:

- If, at time t, either (a) θ is a wff for which no testable predictions have been proposed by means of moves 1.8 or 1.9, or (b) for all such predictions ϕ which have been proposed, either (i) no scientific experiment of ϕ has yet been conducted, or (ii) for all such experiments undertaken, the test modality label e_{θ,ϕ,ϕ^N} has the value *Invalid test* at time t, then $f_{\theta,t,A}$ is assigned the value *Untested* at time t.

- If, at time t, (a) θ is a wff for which at least one testable prediction ϕ proposed by executions of Rules 1.8 or 1.9, and (b) at least one scientific experiment of ϕ has been conducted, and, (c) for these experiments, either (i) the majority have a test modality e_{θ,ϕ,ϕ^N} with the value *Invalid* at time t, or (ii) at least a substantial minority have a test modality e_{θ,ϕ,ϕ^N} with the value *Disconfirming instance* at time t and at least a substantial minority have a test modality e_{θ,ϕ,ϕ^N} with the value *Confirming instance* at time t, then $f_{\theta,t,A}$ is assigned the value *Inconclusive* at time t.

- If, at time t, (a) θ is a wff for which at least one testable prediction ϕ proposed by executions of Rules 1.8 or 1.9, and (b) at least one scientific experiment of ϕ has been conducted, and (c) for the overwhelming majority of such experiments, the test modality e_{θ,ϕ,ϕ^N} has the value *Disconfirming instance* at time t, then $f_{\theta,t,A}$ is assigned the value *Refuted* at time t.

- If, at time t, (a) θ is a wff for which at least one testable prediction ϕ proposed by executions of Rules 1.8 or 1.9, and (b) at least one scientific experiment of ϕ has been conducted, and (c) for the overwhelming majority of such experiments, the test modality e_{θ,ϕ,ϕ^N} has the value *Confirming instance* at time t, then $f_{\theta,t,A}$ is assigned the value *Confirmed* at time t.

- The modality label definitions here are listed in ascending order of strength, and the label assigned is always the highest applicable label. $\square$

Rule 11: Update of the Agora's Empirical Support Store: The empirical support store, $\mathcal{ES}_t(\mathcal{P}_A)$, of the *Agora* is initially empty. Let θ be a wff. Suppose time s is the first time that an element is inserted into the Consequence Store for θ, $\mathcal{QS}_s(\theta)$. Then, at time s the element $(\theta, f_{\theta,s,A})$ is inserted into $\mathcal{ES}_s(\mathcal{P}_A)$, with $f_{\theta,s,A}$ assigned the value *Untested*. □

Note that experimental results may give contrary indications for an hypothesis. Indeed, if the experiment involves statistical inference (from a sample to a population), we would expect contrary results for some proportion of tests undertaken. Accordingly, we have defined confirmation or refutation of claims in terms of the direction of results from an overwhelming majority of tests. One could place more stringent requirements into this Definition, for instance that confirmation requires at least (say) 95% of tests to be confirmations; this would be appropriate if the Agora participants had agreed a uniform statistical significance level to apply to all experiments. Moreover, if all the experiments relate to the same population (e.g. human adults), then these definitions could invoke statistical meta-analysis [81]; we have not done this, so as not to constrain all the tests to be from the same population. By distinguishing acceptability of claims from their evidential support we are modeling what happens in science — theories may be accepted by scientists before all the evidence is in,[16] or even despite overwhelming contrary evidence [17, 45].

3.6 Architecture and user interface

Our main purpose in this paper is to define and study the dialectical argumentation formalism we present, so issues of system architecture, interface design and deployment are mentioned only briefly. We anticipate the Risk Agora system being used to represent a completed or on-going scientific debate, but not in real-time. Once instantiated with a specific knowledge base, the Agora could be used in a number of ways:

1. To understand the logical implications of the scientific knowledge relating to the particular issue, and the arguments concerning the consequences and value-assignments of alternative regulatory options.

2. To consider the various arguments for and against a particular claim (including regulatory options), how these arguments relate to each other, their

[16] As an example, we quote from a newspaper report announcing the recent detection for the first time of the tau neutrino, a sub-atomic particle, at Fermilab [23, p. 2]: "Although their existence had been suspected for 25 years, tau neutrinos had escaped detection because it takes a large amount of energy to create them and because neutrinos pass through most matter without a trace. '*It's just been accepted that this guy exists,*' said Regina Rameika, a physicist at Fermilab and a member of the team."

respective degrees of certainty, and their relative strengths and weaknesses.

3. To develop an overall case for a claim, combining all the arguments for it and against it.

4. To enable interested members of the public to gain an overview of the debate on an issue.

5. To support group deliberation on the issue, for example in Citizens Panels.

6. To support risk assessment and regulatory determination by government regulatory agencies.

For these different functions we believe a layered architecture is appropriate, and in earlier discussion [53] we proposed a three-layered model, drawing on Habermas and Aristotle. By contrast, computational argumentation systems for legal applications have commonly adopted a three-, four- or five-layered model (see [8] for a review). Such applications have different purposes to those listed here, and further analysis is needed to assess to what extent our proposed functions can be accommodated within any of these structures. Recent work in designing negotiation spaces for multi-agent systems in electronic commerce applications may also be relevant [46]. With regard to user-interfaces, argumentation systems provide novel challenges for designers, as there are both static and dynamic elements to any dialogue, and to the relationships between arguments within it. Some researchers (e.g. [22, 78]) have therefore argued that new approaches are needed for these systems. This is also an issue requiring further analysis and prototyping as part of the implementation of the Agora system.

4 Agora Properties and Semantics

In this section we examine the formal properties of the system defined in Section 3. We first consider to what extent Hitchcock's Principles and Alexy's Rules are satisfied by our formalism, and to to what extent the formalism operationalizes the desired Agora properties of Section 2.2. Propositions 1 and 2 below relate to these questions. We then develop a Game Theoretic Semantics for the Agora formalism, motivated in part by the Ehrenfeucht-Fraïssé games of model theory.

4.1 System properties

Proposition 1: *The Agora system defined in Section 3 satisfies fifteen of Hitchcock's Principles of rational mutual inquiry: H1 through H4, H7 through H16, and H18.*

Proof. We consider each Principle in turn:

H1 Externalization: As can be seen from an examination of the Agora rules listed in Section 3, all rules are formulated in terms of observable linguistic behaviour.

H2 Dialectification: The Agora formalism neither proscribes or prohibits any particular content, knowledge base or mode of inference. These are all open to suggestion and agreement from the participants at any time.

H3 Mutuality: Under Rule 5, commitments are only incurred when a participant explicitly asserts a claim (Move 1.3) or accepts an asserted claim (Move 3.2). Thus no statement becomes a commitment of a participant unless they specifically desire it.

H4 Turn-taking: Because the Agora is intended to represent scientific discourses, rather than be used for real-time debating, ensuring turn-taking is straightforward, especially with implementation on a sequential processor.

H7 Logical Pluralism: Participants are free to propose, and to use, any forms of inference they find acceptable. The use of a dictionary of acceptability labels for inference rules in fact can facilitate agreement between participants who may otherwise disagree over use of a certain rule, as we have shown elsewhere [55].

H8 Rule-consistency: Only two rules prohibit utterances: Rule 3.10 which outlaws contradictions, and Rule 4 which prohibits precise repetitions of query and contestation moves. Neither of these rules prohibits all utterances, and neither prohibits the null utterance (remaining silent). So the Agora is rule-consistent.

H9 Semantic Openness: The rules of the formalism do not force any participant to accept any statement, even those following via deductive inference from previously-accepted statements. Nor is any participant presumed by silence to have accepted any statement or to have accepted a particular mode of inference.

H10 Realism: The rules of the formalism do not inhibit agreement between participants. Indeed the rules relating to acceptance by one participant of statements made by another (Moves 3.1, 3.2, 3.4 and 3.5) facilitate such agreement.

H11 Retraceability: Participants are free at any time to amend (Move 3.3), supplement (1.8 following 1.2 or 1.3), precizate (3.6, 3.7), withdraw (3.8, 3.9) or replace (successive uses of 1.2 or of 1.8) earlier statements.

H12 Role reversal: The formalism permits any participant at any time to initiate suggestions, with Moves 1.1 (Pose Claim), 1.2 (Propose Claim), 1.7 (Pose Consequence), 1.8 (Propose Consequence) and 1.13 (Propose Mode of Inference).

H13 Experiential Appeal: The formalism permits direct mutual appeal to experience, in particular to the results of scientific experiments as outlined in Section 3.4.

H14 Openness: Examination of the rules shows that there are no restrictions on the content of contributions.

H15 Tentativeness: Participants are free to make tentative suggestions, via Moves 1.1 (Pose Claim), 1.2 (Propose Claim), 1.7 (Pose Consequence), 1.8 (Propose Consequence) and 1.13 (Propose Mode of Inference).

H16 Tracking: Commitments made by Participants are tracked via the set of commitment stores. Tracking of the complete history of a dialogue is also readily implemented in a computerized system such as this.

H18 Allocation of Burden of Proof: A participant who poses or asserts a claim or a consequence must, when queried, provide an argument for the statement made (Moves 1.2–1.5, 1.8–1.11). Thus the initial burden of proof lies with the participant making a suggestion. This burden of proof is retained throughout the debate whilesoever the claim is unretracted, since any subsequent query by another participant must also be answered until the claim is retracted. Likewise, a participant contesting a proposal or assertion also must provide an argument or dictionary labels supporting that intervention (Moves 2.1–2.8). In this case, the burden of proof for the contestation (but not the claim itself) then lies with the contesting participant. □

As mentioned in Section 2.2, we believe that Principles H5 (Orderliness), H6 (Staging) and H17 (Termination) should be subject to discussion by and agreement of the participants, in accordance with Principle H2 (Dialectification). Thus, we have not encoded these principles in our design of the Agora. Our formalism could be extended in order to facilitate participant discussion of these issues, a subject we mention in the discussion of future work in Section 6 below. We next consider satisfaction of Alexy's rules.

Proposition 2: *The Agora system defined in Section 3 satisfies all ten of Alexy's rules of Discourse Ethics, A1 through A10.*
Proof. We consider each property in turn:

A1 Freedom of Assembly: This property is fulfilled by the overall Agora design, which allows anyone to participate and to execute dialogue moves at any time (subject only to the rules of Section 3).

A2 Common Language: This is also fulfilled by the overall Agora design, which assumes participants accept the logical language and the modality dictionaries used.

A3 Freedom of Speech: Participants are able to assert any claim by virtue of Move 1.3 and assert any consequence of a claim by Move 1.9. Participants are also able to pose and propose claims (Moves 1.1, 1.2) and consequences (1.7, 1.8), speech acts which have less force and fewer obligations than do assertions. Thus, Property A3 is satisfied.

A4 Freedom to Challenge Claims: Participants may question a claim (via Moves 1.4 and 1.5) or a consequence of a claim (1.10, 1.11), or contest a claim, ground, consequence, modality or rule of inference (2.1, 2.2, 2.5–2.8). Participants may also query a contestation (2.3, 2.4 and query rules associated with 2.5–2.8).

A5 Arguments required for Claims: Rules 1.4 and 1.5 ensure that participants who propose or assert a claim provide a grounded argument for that claim if subsequently queried. They may do so with an execution of Move 1.6. Likewise, Rules 1.10 and 1.11 encode the same requirement with regard to consequences from claims, with Rule 1.12 being used to reveal a consequential argument from a claim.

A6 Freedom to Challenge Arguments: Participants may contest the grounds of a claim (by Move 2.5), a mode of inference (2.6) or a modality assignment (2.7).

A7 Freedom of modal Disagreement: A participant $\mathcal{P}_j$ may accept a claim previously proposed or asserted by another participant $\mathcal{P}_i$ by means of Moves 3.1 or 3.2, respectively. If $\mathcal{P}_j$ wishes to accept the proposed or asserted claim with different modality assignments, she may immediately follow the acceptance move with Move 3.3 (Change Modalities). Likewise, a conseqential argument may be accepted with different modalities by executing in succession the two Moves 3.5 (Accept Consequence) and 3.3 (Change Modalities).

A8 Precization: Participants may precizate earlier statements made to the Agora by execution of Moves 3.6 or 3.7. These moves can happen in response to a challenge by another participant, but such a challenge is not a precondition for their use.

A9 Proportionate Defence: Rules 1.4, 1.5, 1.10 and 1.11 require a participant $\mathcal{P}_i$ proposing or asserting claims or consequences to respond to queries from other participants by revealing her supporting arguments to the Agora (using 1.6 to reveal a grounded argument for a claim and 1.12 to reveal a consequential argument from a claim). No further response from $\mathcal{P}_i$ is required, even if further contestations are made, or rebutting and undercutting arguments presented by other participants.

A10 No Contradictions: Rule 3.10 prohibits contradiction. It may be avoided by use of Moves 3.8 and 3.9 (Retraction of asserted and accepted claims). $\square$

Finally, we note that the Agora operationalizes our two key principles of scientific discourse, contestability and defeasibility of claims.

Proposition 3: *The Agora enables contestation and defeasibility of scientific claims.*
Proof. Contestation of claims and the arguments supporting them occurs through Moves 2.1 through 2.8. Defeasibility of claims occurs through the assignment of the *Agora* claim modalities, as defined by Rules 6 and 7. For any statement θ, the value of $d_{\theta,A}$ may change according to the current status of the arguments for and against θ in the Agora. Clearly, with the exception of tautologies, $d_{\theta,A}$ is non-monotonic, as claims previously considered to be, say, *Confirmed* will be re-labeled when new rebuttals or undercutting arguments are proposed. $\square$

4.2 Game-theoretic semantics

We next consider semantic issues, using the definition of the claim modalities for the *Agora* provided by Rule 6 to construct a valuation function on formulae. Because we earlier assumed time to be discrete and countable, we can denote it by the symbols $t_0, t_1, t_2, \ldots$. Throughout this section, we assume that participants obey all the rules of the Agora, in particular Rule 4, so that they do not repeat themselves. We further assume that at commencement of the Agora dialogue the information available to the participants is finite and they have only a finite number of possible inference rules.

Definition 20: *The* Agora community valuation at time t *is a function* $v_{t,A}$ *defined from the set of wffs of* $\mathcal{L}$ *to the set* $\{0, 1\}$*, such that* $v_{t,A}(\theta) = 1$ *precisely when* $d_{\theta,t,A} = Accepted$*; otherwise,* $v_{t,A}(\theta) = 0$.

Proposition 4: *Let θ be a wff, and suppose $\mathcal{A}(\rightarrow \theta)$ is a consistent, well-defended argument for θ. Suppose further that all arguments pertaining to θ using the initial information and inference rules are eventually articulated by participants within the Agora, and that no new information concerning θ is received by participants following commencement. Then:*

$$\lim_{k\rightarrow\infty} v_{t_k,A}(\theta) \ = 1$$

Proof. Because all arguments pertaining to θ are assumed to be articulated in the Agora eventually, then there is some time-point at which $\mathcal{A}(\rightarrow \theta)$ is articulated through an execution of move 1.6. Likewise, there are time-points where every rebuttal and every undercut of $\mathcal{A}$ is presented, and, other, possibly later, time-points where every counter-attacking argument against these is presented. Because we are assuming there is no new information beyond that at commencement, then there is a time-point t_m which is the last of these time-points, i.e. m is the maximum of the indexes on the time-points at which any attacks and counter-attacks for $\mathcal{A}(\rightarrow \theta)$ are presented to the Agora. Thus, from that time onwards, $\mathcal{A}(\rightarrow \theta)$ is well-defended, by definition. Hence, $d_{\theta,t_m,A} = Accepted$, by Rule 6. Thus, by Definition 20, $v_{t_m,A}(\theta) = 1$.

However, since all attacks and counter-attacks have been presented by this time, then $d_{\theta,t_k,A} = Accepted$ for all $k \geq m$. Hence, $v_{t_k,A}(\theta) = 1$, also for all $k \geq m$, and so we have:

$$\lim_{k\rightarrow\infty} v_{t_k,A}(\theta) \ = 1.$$

□

Proposition 5: *Let θ be a wff. Suppose that:*

$$\lim_{k\rightarrow\infty} v_{t_k,A}(\theta) \ = 1$$

Then there is a consistent, well-defended argument for θ which is presented to the Agora at some time.

Proof. The antecedent limit implies that there for any time-point t_m there are infinitely-many $k \geq m$ such that $v_{t_k,A}(\theta) = 1$. Choose any t_m and let $s \geq t_m$ be such a time-point. Then, $v_{s,A}(\theta) = 1$. Hence, by Definition 20, $d_{\theta,s,A} = Accepted$. Therefore, by Rule 6, a well-defended consistent argument for θ has been presented by a participant to the Agora through an execution of move 1.6 at or before time s.□

The model of science we have adopted asserts that scientific claims are regarded as "defeasibly true" when and only when the relevant scientific community agrees to so regard them.[17] Our definition of community valuation is in effect a proxy for the scientific community's opinion on the truth of a claim. Accordingly, Propositions 4 and 5 say that the debate within the Agora, when considered as a procedure for generating claims, neither under-generates nor over-generates defeasibly true claims, provided all relevant arguments are eventually asserted. These two propositions demonstrate a form of consistency and completeness for dialogues in Agora, relative to the Agora community valuation function. This valuation function is not "truth", because we regard scientific claims as defeasible. We can think of an Agora community valuation equal to 1 as signifying *"Currently Accepted as True"* (or *"Defeasibly True"*) and a valuation equal to 0 as signifying *"Not Currently Accepted as True."* If further evidence is found against a claim that is *Currently Accepted as True*, and this evidence is accepted by the Agora community, then the valuation function for this claim will change.

What we have just articulated is essentially a Game-Theoretic semantics, in the sense of Jaako Hintikka [37].[18] In this approach, the truth of a statement is understood to mean that a player in an associated game has a winning strategy. Here the associated game is the dialogue in the Agora, and the winning strategy for a player asserting the statement as a claim is that the claim survives all attempts to defeat it by rebutting and undercutting arguments.

However, the limit function of Propositions 4 and 5 is a statement of an infinite property rather than a finite one. Given any argument $\mathcal{A}$, and any future time t_k which is a finite distance from the time t_0 at which the Agora dialogue commences, there is no guarantee that $\mathcal{A}$ will be articulated by an Agora participant before t_k. This is true no matter how large the difference between t_k and t_0.

It is useful here to consider Ehrenfeucht-Fraïssé games from model theory [40]. In considering whether two infinite mathematical structures were isomorphic, Roland Fraïssé [20] was led to develop finite approximations of an isomorphism, namely isomorphisms between finite subsets of the structures. Andrzej Ehrenfeucht [16] extended this idea by constructing games of $2n$ moves between two players, Duplicator (D) and Spoiler (S), who alternate in choosing elements of the two structures. The objective of D is to show that the two structures are isomorphic, while S has the objective of showing they are not. D wins a game of $2n$ moves if the two subsets of n elements selected by the players from the two parent structures are isomorphic. S wins otherwise. If the two infinite structures are in

[17] It is debatable whether or not a transcendent truth exists. Even if it does, however, science has no privileged means to access it.

[18] Although Hintikka is its modern proponent, similar ideas have been found in the work of nineteenth-century American logician Charles Sanders Peirce [36]. Game theoretic semantics have found application in theoretical computer science, for example in programming language theory [1, 2], as well as in AI [4, 65].

fact isomorphic, then a winning strategy exists for D for the game of size $2n$, for every finite n. However, the converse need not be true, as one could readily imagine (finite) isomorphisms existing between every pair of equal-sized finite subsets without there being an (infinite) isomorphism between the parent structures.[19]

With this motivation, we now consider finite approximations to the infinite case, asking the question *How informative is the finite snapshot at any one time of the infinite situation?*[20] Unfortunately, the answer is not consoling, as the next proposition shows.

Proposition 6: *Let θ be a wff and let $s = t_m$ be some fixed time after commencement of the Agora. Suppose further that all arguments pertaining to θ are eventually articulated by participants within the Agora. Then the value of $v_{s,A}(\theta)$ does not tell us anything about the value of $\lim_{k\rightarrow\infty} v_{t_k,A}(\theta)$, or even if this limit exists.*

Proof. At time s, the Agora valuation function $v_{s,A}(\theta)$ may be equal to one or zero. We consider the two cases in turn.

1. If $v_{s,A}(\theta) = 1$ then, by Rule 6, a consistent grounded argument for θ, say $\mathcal{A}(\rightarrow \theta)$, has been articulated to the Agora, and this argument is well-defended at time s. This means that counter-attackers have been articulated for all attackers of $\mathcal{A}$ by this time. However, there may be other rebuttals or undercutters yet to be articulated (even confining information to that known to participants at commencement of the dialogue) and it is possible that not all of these will have counter-attackers. Thus, we cannot conclude that:

$$\lim_{k\rightarrow\infty} v_{t_k,A}(\theta) = 1.$$

2. If $v_{s,A}(\theta) = 0$ then, by Rule 6, either no consistent grounded argument for θ has yet been presented to the Agora, or such an argument has been presented but it is not well-defended. In either case, the absence of such an argument at time s does not mean no such argument will appear at a future time. Thus we cannot conclude that:

$$\lim_{k\rightarrow\infty} v_{t_k,A}(\theta) = 0.$$

In either case, there is also nothing to preclude the valuation function $v_{t,A}(\theta)$, for $t > s$, having the opposite value to $v_{s,A}(\theta)$ infinitely often. Thus we cannot even conclude that the sequence $v_{t_k,A}(\theta)$ converges as $k \rightarrow \infty$. □

[19]For example, every finite subset of the Real numbers is countable, and so isomorphic to every equal-sized finite subset of the positive integers, but $\mathcal{R}$ is not isomorphic to $\mathcal{Z}^+$.

[20]This question calls to mind Jorge Luis Borges' short story, "The Library of Babel," about the library of all possible books in all possible languages [9].

This result is not surprising, as nothing in our formalism precludes new information being presented to the Agora at any time, no matter how long the value of $v_{t,A}(\theta)$ has been stable. This is a consequence of our principle that all scientific claims are always defeasible. However, under certain conditions, we can draw probabilistic conclusions about the relationship between a snapshot and the long run. We first define some notation.

Definition 21: *Let θ be a wff. We write LE_θ for the statement:* "The function $v_{t_k,A}(\theta)$ converges to a finite limit as $k \to \infty$." *For any non-negative integer k, we write $\mathcal{X}_{k,\theta}$ for the statement:* "New evidence concerning θ becomes known to an Agora participant after time t_k."

In general, at any time t_j, we do not know whether new evidence will become available to Agora participants at a later time t_k or not. Consequently, the variables $\mathcal{X}_{k,\theta}$, for t_k not in the past, represent uncertain events. Also uncertain for the same reason are statements concerning the future values of $v_{t,A}(\theta)$ for any θ. Because these events are uncertain, we may assume the existence of a probability function over them, i.e. a real-valued measure function mapping to $[0, 1]$ which satisfies the axioms of probability.

Definition 22: *For any wff θ, Pr is a probability function defined over statements of the form $\mathcal{X}_{k,\theta}$ and statements concerning the values of $v_{t,A}(\theta)$.*

Proposition 7: *Let θ be a wff and suppose that all arguments pertaining to θ and using the information available at commencement are articulated by participants by some time $s > 0$. Suppose further that $v_{t_m,A}(\theta) = 1$ for some $t_m \geq s$. Also, assume that $Pr(\mathcal{X}_{m,\theta}) \leq \epsilon$, for some $\epsilon \in [0, 1]$. Then:*

$$Pr(LE_\theta \ and \ \lim_{k\to\infty} v_{t_k,A}(\theta) \ = 1 \mid v_{t_m,A}(\theta) = 1) \geq 1 - \epsilon$$

and

$$Pr(LE_\theta \ and \ \lim_{k\to\infty} v_{t_k,A}(\theta) \ = 0 \mid v_{t_m,A}(\theta) = 1) \leq \epsilon.$$

Proof. We have that $v_{t_m,A}(\theta) = 1$. For simplicity of expression we omit this conditioning event in the following probabilities until the final statements. We have the following:

$$Pr(LE_\theta \ and \ \lim_{k\to\infty} v_{t_k,A}(\theta) = \ 1)$$

$$= 1 - \ Pr(\neg LE_\theta \ or \ (LE_\theta \ and \lim_{k\to\infty} v_{t_k,A}(\theta) \ \neq 1))$$

$\geq 1 \;-\; Pr(\forall k \geq m, \exists j > k \; such \; that \; v_{t_j,A}(\theta) = 0)$.

This is because both disjuncts in the second probability expression imply that the sequence $v_{t_k,A}(\theta), v_{t_{k+1},A}(\theta), \ldots$ does not end with an infinite sequence of "1's", i.e. that

$$\forall k \geq m, \exists n > k \;\; such \; that \; v_{t_n,A}(\theta) = 0.$$

(Note that, for the first implication, the converse is not necessarily true.) Hence, by Definition 20, we have:

$Pr(\lim_{k\to\infty} v_{t_k,A}(\theta) = 1)$

$\geq 1 \;-\; Pr(\forall k \geq m, \exists n > k \; such \; that \; d_{\theta,t_n,A} \neq Accepted)$.

Now, take $k = m$ and choose $t_n > t_m$ so that $d_{\theta,t_n,A} \neq Accepted$. By Rule 6, this means that either no consistent, grounded argument for θ has been articulated in the Agora by time t_n or else all such arguments which have been articulated are not well-defended. Clearly at least one consistent, grounded argument has been articulated, since we had $v_{t_m,A}(\theta) = 1$. Select one of these arguments. This argument was well-defended at time t_m, but at time t_n it is not well-defended. By Definition 13, this can only be because a rebuttal or undercutter for θ has been articulated in the Agora between times t_m and t_n for which no counter-attacker has yet been advanced at t_n. But $t_m > s$, which was the time at which all arguments based on the initial information and inference rules of participants were advanced. Hence, this rebuttal or undercutter for $\mathcal{A}$ must involved new information, since, by Rule 4, repetition is not permitted. Thus, we have that:

$$\forall k \geq m, \exists n > k \; such \; that \; d_{\theta,t_n,A} \neq Accepted$$

implies that:

"New evidence concerning θ becomes known to an Agora participant after time t_m."

Hence, $Pr(\forall k \geq m, \exists n > k \; such \; that \; d_{\theta,t_n,A} \neq Accepted) \;\leq Pr(\mathcal{X}_{m,\theta})$.

But $Pr(\mathcal{X}_{m,\theta}) \leq \epsilon$. Thus,

$$Pr(\forall k \geq m, \exists n > k \; such \; that \; d_{\theta,t_n,A} \neq Accepted) \;\leq \epsilon$$

Therefore,

$$Pr(LE_\theta \ and \ \lim_{k\to\infty} v_{t_k,A}(\theta) = 1 \mid v_{t_m,A}(\theta) = 1) \ \geq 1 \ - \ \epsilon.$$

Consequently,

$$Pr(\neg LE_\theta \ or (LE_\theta \ and \ \lim_{k\to\infty} v_{t_k,A}(\theta) \neq 1 \ ) \mid v_{t_m,A}(\theta) = 1) \leq \epsilon$$

and so:

$$Pr(LE_\theta \ and \ \lim_{k\to\infty} v_{t_k,A}(\theta) = 0 \mid v_{t_m,A}(\theta) = 1) \leq \epsilon.$$

□

This result is directly analogous to the standard (Neyman-Pearson) approach to statistical hypothesis testing [14]. Statistical inference — that is, inference from sample to population — is an unsound form of inference, in that statements true of a sample are not necessarily true of a population from which the sample is drawn. The key achievement of mathematical statistics in the twentieth century was to place a bound on the extent of this unsoundness: if we know (or can approximate) the probability distribution of the variable of interest in the population, and we know 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. 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. Although we cannot know whether any particular instance is invalid, we are able to estimate an upper bound for the proportion of times that our reasoning is invalid, when the same inference procedure is used repeatedly.

The properties we have demonstrated with the Agora are qualitative analogies of this. Proposition 6 tells us that we cannot infer the long run result of an Agora debate from a current snapshot of that debate. But Proposition 7 says that, if we are able to place a bound on the possibility of new information becoming known, then after some point in the debate, we have a bound on the error of incorrectly inferring the long-run result of the debate from a snapshot at that particular time. This analogy with standard statistical hypothesis testing theory is not surprising when one realizes that a snapshot of an Agora debate is a finite sample of the infinite population of time-points through which the debate may run. Of course, in reality, we can never know the value of ϵ, the probability that new information relevant to θ may become known in the future. But scientists working with long-standing and widely-accepted theories usually assume that this probability is zero or extremely small.

We also have the following converse of Proposition 7.

Proposition 8: *Let θ be a wff and suppose that all arguments pertaining to θ and using the information available at commencement are articulated by participants by some time $t_m > 0$. Suppose that we also have that $Pr(\mathcal{X}_m) \leq \epsilon$, for some $\epsilon \in [0, 1]$. Assume that LE_θ and that $\lim_{k\to\infty} v_{t_k,A}(\theta) = 1$. Then we have:*

$$Pr(v_{t_m,A}(\theta) = 1 \mid LE_\theta \ and \ \lim_{k\to\infty} v_{t_k,A}(\theta) = 1) \geq 1 - \epsilon.$$

Proof. As with the previous proposition, we omit the conditioning event in probability expressions until the final statement. We have that $\lim_{k\to\infty} v_{t_k,A}(\theta) = 1$. Then, $v_{t_k,A}(\theta) = 1$ for infinitely-many and infinitely-large k. Choose one such time-point $t_n > t_m$. Hence, $v_{t_n,A}(\theta) = 1$ and so, by Definition 20 and Rule 6, at or before time t_n a grounded, consistent argument for θ has been articulated which is well-defended. Choose one such argument and denote it $\mathcal{A}$.
Now, if $v_{t_m,A}(\theta) = 0$, then either $\mathcal{A}$ had not been articulated in the Agora by time t_m or else it had been presented but was not well-defended at this time. But, by assumption, all arguments pertaining to θ and using the initial information are articulated by time t_m. Thus, $v_{t_m,A}(\theta) = 0$ implies that some new information is received by an Agora participant between times t_m and t_n, in order to account for the change in status of $\mathcal{A}$ between these times. Thus, we have:

$$(v_{t_m,A}(\theta) = 0) \ \Rightarrow \ \mathcal{X}_{m,\theta}.$$

Hence,

$$Pr(v_{t_m,A}(\theta) = 0) \ \leq Pr(\mathcal{X}_{m,\theta}) \ \leq \epsilon.$$

Therefore,

$$1 \ - \ Pr(v_{t_m,A}(\theta) = 0) \ \geq 1 \ - \epsilon$$

and so:

$$Pr(v_{t_m,A}(\theta) = 1 \mid LE_\theta \ and \ \lim_{k\to\infty} v_{t_k,A}(\theta) = 1) \geq 1 - \epsilon. \qquad \square$$

Proposition 8 says that, subject to certain conditions, the long-run truth of a proposition will probably be reflected in its status at finite snapshots, provided these are taken after the articulation of all arguments based on the initial information. The degree of confidence we can attach to this conclusion is a function of the probability of new evidence arising in the future, as was the case with Proposition 7. Thus, both Propositions 7 and 8 demonstrate probabilistic relationships between finite snaphots of Agora debates and their long-run outcomes. While not as strong as deterministic relationships would be, these probabilistic relationships do demonstrate a form of soundness of the dialogues in the Agora framework relative to the *Agora* valuation.

5 Example

To illustrate these ideas we present a simple and hypothetical example of an Agora debate. In a real debate, participants would be free to introduce supporting evidence and modes of inference at any time. To aid understanding, in this example we first list the assumptions and modes of inference to be used in assertions and proposals. The dialogue concerns a chemical $\mathcal{X}$, and the various statements to be used as grounds are labeled K1 through K4:

K1: $\mathcal{X}$ is produced by the human body naturally (i.e. it is endogenous).

K2: $\mathcal{X}$ is endogenous in rats.

K3: If $\mathcal{X}$ is an endogenous chemical then it is not carcinogenic.

K4: Bioassay experiments applying $\mathcal{X}$ to rats result in significant carcinogenic effects.

The modes of inference used by participants are labeled R1 through R3:

R1 (And Introduction): Given a wff ϕ and a wff θ, we may infer the wff $(\phi \wedge \theta)$.

R2 (Modus Ponens): Given a wff ϕ and the wff $(\phi \rightarrow \theta)$, we may infer the wff θ.

R3: If a chemical is found to be carcinogenic in an animal species, then we may infer it to be carcinogenic in humans.

We now give an example of an Agora dialogue concerning the statement: $\mathcal{X}$ *is carcinogenic to humans,* which we denote by ϕ. The moves are numbered M1, M2,..., in sequence. We assume that participants use the claims dictionary of Example 1, namely $\{$*Certain, Confirmed, Probable, Plausible, Supported, Open*$\}$, and the inference dictionary $\mathcal{D}_I = \{$*Valid, Invalid*$\}$. Through the course of the dialogue, we show the contents of the *Agora's* commitment store $\mathcal{CS}_0(\mathcal{P}_A)$ and the value of the relevant Agora valuation functions only as these change, in steps numbered according by the relevant time-point, ACS0, ACS1,. . ., and Val0, Val1,. . ., respectively. To assist understanding of the example, each move is followed by an annotation, in italics.

ACS0: $\mathcal{CS}_0(\mathcal{P}_A) = \{\}$.

> *Note that the Agora's Commitment Store is empty at commencement of the dialogue.*

Val1: $v_{0,A}(\phi) = 0$ and $v_{0,A}(\neg\phi) = 0$.

> *Since neither ϕ nor $\neg\phi$ are entries in the Agora Commitment Store yet, then their Agora claim modalities are not equal to* Accepted. *Thus, by Definition 20, their Agora valuation functions have value 0.*

M1: *assert(*$\mathcal{P}_1 : (\phi, \mathit{Confirmed})$*).*

> *Participant $\mathcal{P}_1$ asserts the claim ϕ, that $\mathcal{X}$ is carcinogenic to humans, which she believes has strength* Confirmed.

ACS1: $\mathcal{CS}_1(\mathcal{P}_A) = \{(\phi, \mathit{Open})\}$.

> *Move M1 which asserted ϕ leads to an entry being made for ϕ in $\mathcal{CS}_1(\mathcal{P}_A)$. The Agora's modality for this is* Open, *as no argument has yet been presented for ϕ.*

Val1: $v_{1,A}(\phi) = 0$ and $v_{1,A}(\neg\phi) = 0$.

> *By Definition 20, the Agora community valuation for ϕ is zero since the Agora claim modality for ϕ is* Open.

M2: *query(*$\mathcal{P}_2 : \mathit{assert}(\mathcal{P}_1 : (\phi, \mathit{Confirmed}))$*).*

> *Participant $\mathcal{P}_2$ asks $\mathcal{P}_1$ for her argument for ϕ.*

M3: *show_arg(*$\mathcal{P}_1 : (K4, R3, \phi, (\mathit{Confirmed}, \mathit{Valid}, \mathit{Confirmed}))$*).*

> *Participant $\mathcal{P}_1$ presents her argument for ϕ, which rests on grounds that bioassay experiments of $\mathcal{X}$ have been shown to produce carcinogenic effects in rats, and that one can infer from these results to humans, by means of Rule R3. $\mathcal{P}_1$ assigns this rule a modality of* Valid.

ACS3: $\mathcal{CS}_3(\mathcal{P}_A) = \{(\phi, Probable)\}$.

Move M3 has presented an argument for ϕ, and so an update to the Agora's Commitment Store is necessary. The Agora's Claim Modality will be at least Supported, *but because the argument presented is consistent, and also because no rebuttals or undercutters have yet been presented against ϕ, then a modality of* Probable *is assigned, according to Rules 6 and 7.*

Val3: $v_{3,A}(\phi) = 0$ and $v_{3,A}(\neg\phi) = 0$.

The Agora valuations remain unchanged.

M4: *contest($\mathcal{P}_2$:assert($\mathcal{P}_1$: $(\phi, Confirmed)$)).*

Participant $\mathcal{P}_2$ contests the assertion of ϕ with modality Confirmed *by $\mathcal{P}_1$.*

M5: *query($\mathcal{P}_3$:contest($\mathcal{P}_2$:assert($\mathcal{P}_1$: $(\phi, Confirmed)$)))*

Participant $\mathcal{P}_3$ asks $\mathcal{P}_2$ for her reasons for the contestation in Move M4.

M6: *propose($\mathcal{P}_2$: $(\neg\phi, Plausible)$).*

Participant $\mathcal{P}_2$ proposes the claim $\neg\phi$, i.e. that $\mathcal{X}$ is not carcinogenic to humans, and says she believes this is Plausible.

ACS6: $\mathcal{CS}_6(\mathcal{P}_A) = \{(\phi, Probable), (\neg\phi, Open)\}$.

The Agora's Commitment Store is updated with a new entry for $\neg\phi$, following Move M6. This is initially given a claim modality of Open.

Val6: $v_{6,A}(\phi) = 0$ and $v_{6,A}(\neg\phi) = 0$.

The Agora valuations remain unchanged.

M7: *query($\mathcal{P}_1$:propose($\mathcal{P}_2$: $(\neg\phi, Plausible)$)).*

Participant $\mathcal{P}_1$ asks $\mathcal{P}_2$ for her argument for $\neg\phi$.

M8: *show_arg($\mathcal{P}_2$: $((K1, K3), R2, \neg\phi, (Confirmed, Probable, Valid, Plausible)$)).*

Participant $\mathcal{P}_2$ presents her argument for $\neg\phi$. This argument starts from the premises that $\mathcal{X}$ is endogenous and that endogenous chemicals are not carcinogenic, and then uses Modus Ponens (R2) to conclude that $\mathcal{X}$ is not carcinogenic to humans.

ACS8: $\mathcal{CS}_8(\mathcal{P}_A) = \{(\phi, Plausible), (\neg\phi, Plausible)\}$.

Because an argument for $\neg\phi$ has now been presented to the Agora, the Agora Commitment Store is again updated. This argument is consistent and the argument for ϕ presented in Move M3 is a rebuttal of it, so by Rule 6, it is assigned an Agora claim modality of Plausible. *Likewise, the argument for $\neg\phi$ is a rebuttal for ϕ, so the Agora claim modality for ϕ changes from* Probable *to* Plausible.

Val8: $v_{8,A}(\phi) = 0$ and $v_{8,A}(\neg\phi) = 0$.

The Agora valuations remain unchanged.

M9: *contest_ground(*$\mathcal{P}_4$: *show_arg(*$\mathcal{P}_2 : ((K1, K3), R2, \neg\phi,$ ($Confirmed,$ $Probable, Valid, Plausible$)*:* $(K3, Probable)$*)).*

Participant $\mathcal{P}_4$ informs the Agora that she contests a grounds of the argument presented by $\mathcal{P}_2$ in Move M8, namely the premise K3, that an endogenous chemical is carcinogenic.

M10: *show_arg(*$\mathcal{P}_4 : ((K2, K4), R1, \neg K3, (Confirmed, Confirmed, Valid,$ $Confirmed)))$

Participant $\mathcal{P}_4$ immediately follows the contestation with a presentation of her own argument for the negation of K3, i.e. an argument for the claim that it is not the case that an endogenous chemical is carcinogenic. This argument uses And Introduction (Rule R1) on the premises K2, that $\mathcal{X}$ is endogenous in rats, and K4, that $\mathcal{X}$ has been shown to cause cancers in rats. This argument is an undercutter for the argument for $\neg\phi$, presented by $\mathcal{P}_2$ in Move M8.

ACS10: $\mathcal{CS}_10(\mathcal{P}_A) = \{(\phi, Accepted), (\neg\phi, Plausible)\}$.

The argument for ¬ K3 presented in Move M10 undercuts the argument for $\neg\phi$ presented in M8. The Agora claim modality for $\neg\phi$ does not change, as a rebuttal already had been presented. However, the undercutter of M10 attacks a rebuttal (M8) of ϕ. This rebuttal is the only rebuttal or undercutter presented which attacks ϕ. Thus, the argument for ϕ presented in M3 is now well-defended. Hence, the Agora claim modality for ϕ changes from Plausible *to* Accepted.

Val10: $v_{10,A}(\phi) = 1$ and $v_{10,A}(\neg\phi) = 0$.

The Agora valuation for ϕ changes to 1, while that for $\neg\phi$ remains unchanged.

⋮

Note that this change in the value of $v_{i,A}(\phi)$ at Move M10 may only be temporary: if further rebuttals or undercutters are presented which attack ϕ and counter-attacks for these are not presented, then the claim ϕ will cease to be accepted according to Rules 6 and 7, and so the Agora valuation for ϕ will revert to zero. If subsequent counter-attacks are presented, so that the argument of Move M3 is once again well-defended, then the value of $v_{i,A}(\phi)$ may change once again.

6 Discussion

As shown in Proposition 3, our formal definition of the Risk Agora enables contestation and defeasibility of scientific claims. Our system therefore operationalizes the two normative principles of conduct for scientific discourses presented in Section 2.1. In addition, as shown in Propositions 1 and 2, our system satisfies certain proposed normative principles of dialogue between reasonable, consenting participants, namely various of Hitchcock's Principles for rational mutual inquiry and Alexy's rules for Discourse Ethics. As so often happens in real scientific debates, argument may begin over claim and then proceed to argument over its grounds, the inferences used to establish it, its degree of support, or its consequences. We believe, therefore, that the Risk Agora is a very expressive formalism for representation of scientific arguments. The propositions proved in Section 4.2 also show that, when used for inference about the defeasible truth-status of claims, the Agora has certain desirable properties. In particular, if the probability of new evidence pertaining to a claim θ being presented in the Agora is small, then the probability is correspondingly high that the current Agora valuation for θ reflects the long-run position.

We are currently exploring a number of refinements to the Agora. Firstly, we believe the Agora syntax can be readily extended to facilitate the discussion and agreement of participants regarding the internal structure of debates, for example discussing theoretical questions before experimental ones, ascertaining the validity of experiments before discussing their implications, etc. Such an extension of the formalism would enable the Agora participants to give operational effect to Hitchcock's Principles H5 (Orderliness) and H6 (Staging) in a manner consistent with Principle H2 (Dialectification). There also may be value in extending the Agora syntax to permit participant discussion over the modality dictionaries, which are presently assumed to have been agreed by the participants. In other work [55], we have presented a formal language in which arguments over the acceptability of rules of inference may be conducted. Secondly, William Rehg [69] has demonstrated the rationality of incorporation of non-deductive and rhetorical devices (such as epideictic speech and appeals to emotions) in dialectical argument and decision-making, and we seek a means to incorporate such devices in the Agora. For example, in scientific domains respected, senior scientists are often believed by others within their community to possess uncommon intuition or insight, and so *ad verecundiam* arguments (arguments from authority) are an appropriate non-deductive argument schema for inclusion in the Agora. This would not be novel: the argumentation system of Chris Reed [67], for example, allows for the modeling of rhetorical devices, although in a monolectical context.

Secondly, in modern societies, scientific risk assessment forms part of a process to determine what legal regulations, if any, should be imposed on the production, sale or use of the chemicals involved. Debate over such issues can be viewed as a deliberative dialogue aimed at deciding what to do [22, 79]. We have commenced the task of modelling such dialogues by first specifying the different types of speech acts which are appropriate to the environmental regulatory domain [56]. This draws on other work of Habermas, namely his philosophy of Communicative Action [31], in which he sought to understand how people collaborate rationally to achieve a common understanding of a situation or a collective action. Speech acts relevant to such dialogues include: statements of value (revealing the speaker's preferences or value assignments); statements of connection (asserting the existence of a relationship between different parties to a dispute); and statements of obligation (asserting some legal or ethical obligations on the part of participants). However, before we can implement a system incorporating such utterances, a coherent formal model of a deliberation dialogue is required, akin to the formal models for negotiation dialogues developed by Hulstijn [41] and for persuasion dialogues developed by Walton and Krabbe [79]. Together with David Hitchcock, we have recently proposed the first formal model of a deliberation dialogue [39], drawing on the philosophy of retroflexive argumentation of Harald Wohlrapp [80]. Using the Agora in a deliberative context would require incorporation of values for the projected consequences and the development of an appropriate qualitative decision-theory, topics which have received some recent attention in AI. For example, in [19, 60] we developed monolectical argumentation formalisms for the articulation and manipulation of statements of qualitative value, as part of calculi for qualitative decision-making.

We believe the Risk Agora has a number of potential benefits. Firstly, as presented in Section 1, the use of an argumentation formalism enables the qualitative representation of epistemic uncertainty which can coherently incorporate multiple types of evidence and support intersubjective agreement between participants. In the carcinogenic risk domain, these are not typically features of quantitative formalisms. Secondly, by articulating precisely the arguments used to assert carcinogenicity, gaps in knowledge and weaknesses in arguments can be identified more readily. Such identification could be used to prioritize bio-medical research efforts for the particular chemical. Thirdly, by exploring the consequences of claims, the Risk Agora can serve a social maieutic function, making explicit knowledge which may only be latent. Moreover, once instantiated with the details of a particular debate, the system could be used for self-education by others outside the scientific community concerned, or, as mentioned above, be used for deliberation over the regulatory or societal decisions involved.[21] Finally, with argumentation increasingly being used in the design of multi-agent systems [4, 61], the formalism presented here could readily be adapted for deliberative dialogues between independent software agents.

Acknowledgments

This article extends work presented at the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI-00), held at Stanford, CA, in June 2000 [54], and we are grateful to the anonymous UAI reviewers for their comments. We also thank David Hitchcock for his detailed and perceptive reading of that version. Our results were also presented to the Department of Computer Science, University of Liverpool, in February 2001, and we thank the seminar participants for their comments, particularly Paul Dunne. Partial funding for this research was received from the British Engineering and Physical Sciences Research Council (EPSRC) under grant GR/L84117 and a PhD studentship.

References

[1] S. Abramsky and R. Jagadeesan. Games and full completeness for multiplicative linear logic. *Journal of Symbolic Logic*, 59(2):543–574, 1994.

[2] S. Abramsky and G. McCusker. Game semantics. In H. Schwichtenberg and U. Berger, editors, *Logic and Computation: Proceedings of the 1997 Marktoberdorf Summer School*. Springer Verlag, Berlin, Germany, 1998.

[3] 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.*

[21] Its use in such a context raises philosophical and other issues about the role of computer systems in public decision-making contexts, issues we have explored in joint work with William Rehg [70].

In: W. Oelmuller, Ed., *Normenbegrundung-Normendurchsetzung*. Paderborn, Germany: Schoningh, 1978).

[4] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In E. Durfee, editor, *Proceedings of the Fourth International Conference on Multi-Agent Systems (ICMAS-2000)*, pages 31–38, Boston, MA, USA, 2000. IEEE.

[5] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proceedings of the Fourteenth European Conference on Artificial Intelligence (ECAI 2000)*, pages 338–342, Berlin, Germany, 2000. IOS Press.

[6] Anon. Editorial: Health risks of genetically modified foods. *The Lancet*, 353(9167), 29 May 1999.

[7] T. J. M. Bench-Capon, F. P. Coenen, and P. Orton. Argument-based explanation of the British Nationality Act as a logic program. *Computers, Law and AI*, 2(1):53–66, 1993.

[8] T. J. M. Bench-Capon, J. Freeman, H. Hohmann, and H. Prakken. Computational models, argumentation theories and legal practice. In C. Reed and T. Norman, editors, *Bonskeid Symposium on Argument and Computation*. 2001 (In press).

[9] J. L. Borges. The Library of Babel. In *Collected Fictions*, pages 112–118. Allen Lane, The Penguin Press, London, UK, 1998.

[10] D. V. Carbogim, D. S. Robertson, and J. R. Lee. Argument-based applications to knowledge engineering. *The Knowledge Engineering Review*, 15(2):119–149, 2000.

[11] L. Carlson. *Dialogue Games: An Approach to Discourse Analysis*. D. Reidel, Dordrecht, The Netherlands, 1983.

[12] L. Carroll. What the tortoise said to Achilles. *Mind*, 4 (14) (n.s.):278–280, 1895.

[13] C. I. Chesnevar, A. G. Maguitman, and R. P. Loui. Logical models of argument. *ACM Computing Surveys*, 2000.

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

[15] S. Dzeroski, H. Blockeel, B. Kompare, S. Kramer, B. Pfahringer, and W. Van Laer. Experiments in predicting biodegradability. In S. Dzeroski and P. Flach, editors, *Inductive Logic Programming (ILP-99)*. Springer Verlag, Berlin, Germany, 1999.

[16] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Mathematicae*, 49:129–141, 1961.

[17] P. Feyerabend. *Against Method*. Verso, London, UK, third edition, 1993. First edition published 1971.

[18] FIPA. Specification part 2 – agent communication language. Technical report, Foundation for Intelligent Physical Agents, 23 October 1997.

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

[20] R. Fraisse. Sur les classifications des systems de relations. *Publications Scientifiques de l'Universite d'Alger*, 1(1), juin 1954.

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

[22] R. Girle, D. Hitchcock, P. McBurney, and B. Verheij. Practical reasoning: an argument and computation perspective. In C. Reed and T. Norman, editors, *Bonskeid Symposium on Argument and Computation*. 2001 (In press).

[23] J. Glanz. Elusive tau neutrino detected by Fermilab. *International Herald Tribune*, 22-23 July 2000.

[24] G. Goodman and R. Wilson. Quantitative prediction of human cancer risk from rodent carcinogenic potencies: a closer look at the epidemiological evidence for some chemicals not definitively carcinogenic to humans. *Regulatory Toxicology and Pharmacology*, 14:118–146, 1991.

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

[26] T. F. Gordon and N. Karacapilidis. The Zeno argumentation framework. In *Proceedings of the Sixth International Conference on AI and Law*, pages 10–18, New York City, NY, USA, 1997. ACM Press.

[27] 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.

[28] 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.

[29] G. M. Gray, P. Li, I. Shlyakhter, and R. Wilson. An empirical investigation of factors influencing prediction of carcinogenic hazard across species. *Regulatory Toxicology and Pharmacology*, 22:283–291, 1995.

[30] H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics III: Speech Acts*, pages 41–58. Academic Press, New York City, NY, USA, 1975.

[31] J. Habermas. *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society*. Heinemann, London, UK, 1984. (Translation by T. McCarthy of: *Theorie des Kommunikativen Handelns, Band I, Handlungsrationalitat und gesellschaftliche Rationalisierung.* Suhrkamp Verlag, Frankfurt, Germany. 1981.).

[32] 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).

[33] C. L. Hamblin. *Fallacies*. Methuen and Co Ltd, London, UK, 1970.

[34] C. L. Hamblin. Mathematical models of dialogue. *Theoria*, 37:130–155, 1971.

[35] S. O. Hansson. What is philosophy of risk? *Theoria*, 62(1-2):169–186, 1996.

[36] R. Hilpinen. On C. S. Peirce's theory of the proposition. In E. Freeman, editor, *The Relevance of Charles Peirce*, pages 264–270. The Hegeler Institute, La Salle, IL, USA, 1983.

[37] J. Hintikka and G. Sandu. Game-theoretical semantics. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 361–410. Elsevier, Amsterdam, The Netherlands, 1997.

[38] D. Hitchcock. Some principles of rational mutual inquiry. In F. van Eemeren et al., editor, *Proceedings of the Second International Conference on Argumentation*, pages 236–243, Amsterdam, The Netherlands, 1991. SICSAT.

[39] D. Hitchcock, P. McBurney, and S. Parsons. A framework for deliberation dialogues. In H. V. Hansen, C. W. Tindale, J. A. Blair, and R. H. Johnson, editors, *Proceedings of the Fourth Biennial Conference of the Ontario Society for the Study of Argumentation (OSSA 2001)*, Windsor, Ontario, Canada, 2001.

[40] W. Hodges. *A Shorter Model Theory*. Cambridge University Press, Cambridge, UK, 1997.

[41] J. Hulstijn. *Dialogue Models for Inquiry and Transaction*. PhD thesis, Universiteit Twente, Enschede, The Netherlands, 2000.

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

[43] P. Krause, S. Ambler, M. Elvang-Gørannson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11 (1):113–131, 1995.

[44] 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*, Lecture Notes in Artificial Intelligence 1455, pages 138–156. Springer Verlag, Berlin, Germany, 1998.

[45] I. Lakatos and P. Feyerabend. *For and Against Method*. University of Chicago Press, Chicago, IL, USA, 1999. Edited by M. Motterlini.

[46] A. R. Lomuscio, M. Wooldridge, and N. R. Jennings. A classification scheme for negotiation in electronic commerce. In F. Dignum and C. Sierra, editors, *Agent-Mediated Electronic Commerce: The European AgentLink Perspective*, Lecture Notes in Computer Science 1991, pages 19–33. Springer Verlag, Berlin, Germany, 2001.

[47] P. Lorenzen. Ein dialogisches konstruktivitatskriterium. In *Infinitistic Methods: Proceedings of the Symposium on the Foundations of Mathematics, Warsaw, 2-9 September 1959*, pages 193–200, Warszawa, Poland, 1961. PWN. Reprinted in [48].

[48] P. Lorenzen and K. Lorenz. *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt, Germany, 1978.

[49] J. D. MacKenzie. How to stop talking to tortoises. *Notre Dame Journal of Formal Logic*, 20 (4):705–717, 1979.

[50] J. D. MacKenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133, 1979.

[51] J. D. MacKenzie. Four dialogue systems. *Studia Logica*, 49 (4):567–583, 1990.

[52] D. G. Mayo. Toward a more objective understanding of the evidence of carcinogenic risk. In A. Fine and J. Leplin, editors, *Proceedings of the Philosophy of Science Association Conference*, volume 2, pages 489–503, East Lansing, MI, USA, 1998. Philosophy of Science Association.

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

[54] 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)*, pages 371–379, San Francisco, CA, USA, 2000. Morgan Kaufmann.

[55] P. McBurney and S. Parsons. Tenacious tortoises: a formalism for argument over rules of inference. In G. Vreeswijk, editor, *Workshop on Computational Dialectics, Fourteenth European Conference on Artificial Intelligence (ECAI2000)*, Berlin, Germany, 2000. ECAI.

[56] P. McBurney and S. Parsons. Intelligent systems to support deliberative democracy in environmental regulation. *Information and Communications Technology Law*, 10(1):33–43, 2001.

[57] P. Mittelstaedt. *Quantum Logic*. D. Reidel, Dordrecht, The Netherlands, 1979.

[58] N. Murray, B. Heap, W. Hill, J. Smith, M. Waterfield, and R. Bowden. Review of data on possible toxicity of GM potatoes. Report 11/99, The Royal Society, London, U.K., 1999.

[59] A. Naess. *Communication and Argument: Elements of Applied Semantics*. Allen and Unwin, London, UK, 1966. (Translation of: *En del Elementaere Logiske Emner.* Universitetsforlaget, Oslo, Norway, 1947).

[60] 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 (ECSQARU99), Lecture Notes in Artificial Intelligence 1638, pages 328–339. Springer Verlag, Berlin, Germany, 1999.

[61] 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.

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

[63] J. L. Pollock. *Cognitive Carpentry: A Blueprint for How to Build a Person*. MIT Press, Cambridge, MA, USA, 1995.

[64] K. R. Popper. *The Logic of Scientific Enquiry*. Hutchinson, London, UK, 1959.

[65] H. Prakken and G. Sartor. Modelling reasoning with precedents in a formal dialogue game. *AI and Law*, 6:231–287, 1998.

[66] H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. In D. Gabbay, editor, *Handbook of Philosophical Logic*. Kluwer Academic, Dordrecht, The Netherlands, second edition, 2000.

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

[68] C. Reed and T. Norman, editors. *Bonskeid Symposium on Argument and Computation*. 2001 (In press).

[69] W. Rehg. Reason and rhetoric in Habermas's Theory of Argumentation. In W. Jost and M. J. Hyde, editors, *Rhetoric and Hermeneutics in Our Time: A Reader*, pages 358–377. Yale University Press, New Haven, CN, USA, 1997.

[70] W. Rehg, P. McBurney, and S. Parsons. Computer decision-support systems for public argumentation: Philosophical issues in deployment and evaluation. In H. V. Hansen, C. W. Tindale, J. A. Blair, and R. H. Johnson, editors, *Proceedings of the Fourth Biennial Conference of the Ontario Society for the Study of Argumentation (OSSA 2001)*, Windsor, Ontario, Canada, 2001.

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

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

[73] 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, 1995.

[74] K. Stathis. A game-based architecture for developing interactive components in computational logic. *Electronic Journal of Functional and Logic Programming*, 2000(5), March 2000.

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

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

[77] U.S.A. Environmental Protection Agency. Proposed guidelines for carcinogen risk assessment. *U.S. Federal Register*, 61:17960–18011, 23 April 1996.

[78] 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.

[79] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Series in Logic and Language. State University of New York Press, Albany, NY, USA, 1995.

[80] H. Wohlrapp. A new light on non-deductive argumentation schemes. *Argumentation*, 12:341–350, 1998.

[81] F. M. Wolf. *Meta-analysis: Quantitative Methods for Research Synthesis*. Sage, Beverly Hills, CA, USA, 1986.

[82] R. Zimmerman. When studies collide: Meta-analysis and rules of evidence for environmental health policy — Applications to Benzene, Dioxins, and Formaldehyde. *Policy Studies Journal*, 23(1):123–140, 1995.