Reasoning across scenarios in planning under uncertainty

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

Planning under uncertainty requires the adoption of assumptions about the current and future states of the world, and the preparation of conditional plans based on these assumptions. In any realistic domain, however, there will be an exponential explosion in the number of conditional plans required. One approach to this problem is to articulate a set of scenarios, which together are representative of the possible and/or likely futures. Doing this then creates the challenge of reasoning across the scenarios to decide a course of action. We present an argumentation-based formalism for representing different assumptions in a scenario framework and for reasoning across the resulting scenarios.

Introduction

Consider an agent operating in some complex domain, perhaps a robot with the goal of collecting and delivering objects in a factory (Parsons *et al.* 2000), or a telecommunications operator considering how to best provide future services (McBurney & Parsons 2001c). In both these cases, and in many others involving decisions about what to do and how to do it, the decision-making entity is faced with what is essentially a planning problem—building a plan from a set of options available to it—but one in which the best plan (and indeed the best goal or set of goals, though we will say little about this matter here) is very dependent upon not just the initial state of the world, but also on how the world evolves over time. The plan that is initially best for the robot may turn out to be sub-optimal when a corridor is found to be blocked, and the plan that is initially best for the telecommunications operator may turn out to be sub-optimal when global demand for wireless services turns out to fall below projections, or when competing technologies emerge unexpectedly.

There are two difficulties involved in identifying a good plan. First is the problem of dealing with the fact that any agent only has approximate knowledge of the state of the world in which it operates. This can be tackled by the application of an appropriate uncertainty handling formalism, such as probability theory, to express the degree of belief that an agent has in certain facts being true. However, in many important domains, absence of objective data or the presence of conflicting perceived interests, makes deciding the quantification of uncertainty difficult. Argumentation formalisms have been proposed for the qualitative representation of uncertainty in these circumstances (Krause *et al.* 1995) and have found application in intelligent systems, for example in medical and safety analysis domains (Carbogim, Robertson, & Lee 2000; Fox & Das 2000).

In (McBurney & Parsons 2000), we proposed a formalism using dialectical argumentation for representing and resolving the arguments for and against a claim in a given domain. This representation was grounded in specific philosophies of rational human discourse and was centered on an electronic space for presentation of arguments, which we termed an *Agora*. In subsequent work (McBurney & Parsons 2001a), we extended this formalism and showed that it had several desirable properties when used for inference and decision-making.[1] With this apparatus we believe it is possible to handle the types of uncertainty inherent in the planning problems outlined above. However, doing this, as with using probability theory in conventional planning systems, does not deal with a second difficulty in planning under uncertainty.

This second difficulty is the problem of handling the conditional nature of plans under uncertainty. Any plan is built based on a set of assumptions about how the world will change (or stay the same) over time, and linear planners basically assume that the initial state of the world is only changed by the deterministic actions of the planning agent. For a more sophisticated approach, which can incorporate both non-determinism of actions and the operation of other agents, it is common to use conditional planning,[2] where the branching structure of the conditional plan covers all the ways in which the world may evolve as time passes, and gives a solution

[1] Preprints of these papers are at: www.csc.liv.ac.uk/~peter/pubs.html.

[2] One can, of course, consider the policy solutions of MDPs and POMDPs as implicit conditional plans (Boutilier, Dean, & Hanks 1999).

for each (Warren 1976; Peot & Smith 1992; Pryor & Collins 1996). The problem with this approach is the sheer number of conditional plans that may need to be generated, and the subsequent computational overhead for even modestly complex environments.

One response to this complexity problem is to build plans for a number of likely or representative scenarios rather than for all possible futures. The notion of scenario has found most widespread application in business forecasting (Schwartz 1991). In these applications scenarios are usually treated as alternative possible futures, and they typically differ according to the propositions assumed true in each. The different implications of each scenario are then explored in order to guide the development and selection of business strategy or public policy.

The use of scenario-based reasoning is also found in science. For instance, in statistical mechanics a key question is the extent to which properties of a physical system, such as its entropy at a given time, depend on the initial state of the system. Boltzmann (Boltzmann 1872) explored this question by comparing the given system to a set of other, imaginary systems each having different initial conditions; one could thereby assess the extent to which the property of interest was independent of the initial system state. In this paper, we extend our work on dialectical argumentation to allow reasoning across scenarios, providing a qualitative formalism which could be used for decision making in conditional planning applications.

The next section presents brief summaries of our previous work on Agoras and the comparison of different scenarios, which this papers extends. The following section then proposes a framework in which to compare the results of debates on the same topic, but conducted under different scenario assumptions or inference mechanisms. We also explore the formal properties of our framework. The subsequent section presents an example and the last section compares our approach to related work and discusses possible future research.

Agoras and Scenarios

Agoras

In this section we briefly summarize the Agora framework for the qualitative representation of uncertainty presented in (McBurney & Parsons 2000; 2001a). In this framework, arguments for and against claims are articulated by participants in an electronic space, called an *Agora*. Claims are expressed as formulae in a propositional language, typically denoted by lower-case Greek letters. By means of defined locutions, participants in the Agora can variously posit, assert, contest, justify, rebut, undercut, qualify and retract claims, just as happens in real discourse. For example, a debate participant $\mathcal{P}_i$ could demonstrate her argument $\mathcal{A}(\rightarrow \theta)$ supporting a claim θ, an argument to which she was committed with strength D, by means of the locution:

$$\mathbf{show\_arg}(\mathcal{P}_i : \mathcal{A}(\rightarrow \theta, D)).$$

The rules governing the use of each permitted locution are expressed in terms of a formal dialogue-game between the participants (Hamblin 1971). We assume that the Agora participants begin a debate with a set of agreed facts, or assumptions, and a set of inference rules. Because we want to model many forms of reasoning, these rules need not be deductive and may themselves, in our Agora formulation, be the subject of debate.

We demonstrated the use of this framework for the representation of uncertainty by defining a set of uncertainty labels, which are assigned to claims on the basis of the arguments presented for and against them in the Agora. Essentially, one could say that claims have more credibility (and hence less uncertainty) the fewer and the weaker are the arguments against them. While any set of labels could be so defined, we drew on earlier work in argumentation (Krause *et al.* 1998) and used the set: {*Accepted, Probable, Plausible, Supported, Open*}, with the elements listed in decreasing order of certainty. For example, a claim was regarded as *Probable* at a particular time if at least one consistent argument had been presented for it in the Agora by that time, but no arguments for its negation (rebuttals) nor for the negation of any of its assumptions (undercuts) had been presented by then. We defined a claim as *well-defended* if there was an argument for it and any rebuttals or undercuts were themselves subject to counter-rebuttals or -undercuts. *Accepted* claims were defined as those which are well-defended.

We then defined the truth valuation of a claim θ at time t, denoted $v_t(\theta)$, to be 1 if θ had the label *Accepted* at this time, otherwise it was 0. Such a valuation summarizes the knowledge of the community of participants at the particular time, since it incorporates, via the definitions of the labels, all the arguments for and against θ articulated to that time. Consequently, assessing the truth-status of a claim at a particular time can be viewed as taking a *snapshot* of an Agora debate. Of course, because these definitions are time-dependent, and arguments may be articulated in the Agora at any time, such an assignment of uncertainty labels and truth valuation must be defeasible.

In using the Agora framework to represent uncertainty, attention will focus on the truth valuation function over the long-run.[3] The sequence $(v_t(\theta) \mid t = 1, 2, \ldots)$ may or may not converge as $t \rightarrow \infty$. Suppose that it does converge to a finite limit, and denote its limit value by $v_\infty(\theta)$. What will the value of a snapshot taken at time t tell us about this infinite limit value? Of course, any finite snapshot risks being overtaken by subsequent events, such as new and relevant information being becoming known to the participants, or new arguments being presented. Thus, we cannot infer with complete accuracy from the finite snapshot to

[3] Strictly, we are assuming throughout that time in the Agora is discrete, and can be represented by a set isomorphic to the positive integers.

the infinite value. However, we have shown that, under certain conditions, we can place a bound on the likelihood that such an inference is in error (McBurney & Parsons 2001a). The conditions essentially require that: (a) the snapshot is taken at a time after commencement sufficient for all the arguments using the initial information to be presented, and (b) there is a bound on the probability that new information arises following the snapshot. This result is proved as Proposition 7 of (McBurney & Parsons 2001a), which we reproduce here.

First, some notation. We write LE_θ for the statement: *"The function $v_t(\theta)$ converges to a finite limit as $t \rightarrow \infty$."* Also, we write $\mathcal{X}_{t,\theta}$ for the statement: *"New information relevant to θ becomes known to an Agora participant after time t."* In general, at any time s, we do not know whether new evidence will become available to Agora participants at a later time t or not. Consequently, the variables $\mathcal{X}_{t,\theta}$, for t not in the past, represent uncertain events. Also uncertain for the same reason are statements concerning the future values of $v_t(\theta)$ for any θ. Because these events are uncertain, we 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. We thus write Pr for a probability function defined over statements of the form $\mathcal{X}_{t,\theta}$ and statements concerning the values of $v_t(\theta)$, for any given formula θ.

Proposition 1: (Proposition 7 of (McBurney & Parsons 2001a)) *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 $v_{t_m}(\theta) = 1$ for some $t_m \geq s$. Also, assume that $Pr(\mathcal{X}_{t_m,\theta}) \leq \epsilon$, for some $\epsilon \in [0, 1]$. Then the following inequalities hold:*
(a) $Pr(LE_\theta$ and $v_\infty(\theta) = 1 \mid v_{t_m}(\theta) = 1) \geq 1 - \epsilon$ and
(b) $Pr(LE_\theta$ and $v_\infty(\theta) = 0 \mid v_{t_m}(\theta) = 1) \leq \epsilon$. □

As with the standard procedures for statistical hypothesis testing, this proposition provides us with some confidence in our use of finite snapshots to make inferences about the long-run truth-valuation function for a debate. While such inference is not deductively valid, at least its likelihood of error may be bounded. We next consider the notion of a Scenario.

Scenarios

The framework we have just outlined provides a means to represent the diverse arguments that may arise from a given set of assumptions, and using a given set of inference rules (deductive or otherwise). If we were to start with a different set of assumptions, and/or permit the use of a different set of inference rules, the arguments presented in the Agora may well be different. As a result, the uncertainty labels and truth values assigned to formulae may well also be different, both when taken at finite snapshots and in the limit. In (McBurney & Parsons 2001b), we defined each combination of assumptions and inference rules as a scenario:

Definition 1: *A **Scenario** for a given domain consists of a set of assumptions and a set of inference rules, with which participants are equipped at the commencement of an Agora debate over formulae in that domain. We denote scenarios for a given domain by $\mathcal{S}^1, \mathcal{S}^2, \ldots$, etc. For each scenario, $\mathcal{S}^i$, an Agora debate undertaken with the assumptions and inference rules of that scenario, is said to be the **associated Agora**, denoted $\mathcal{A}^i$. We assume only one Agora debate is conducted in association with any scenario.*

In this paper, we will be assuming that all scenarios, and all the resulting Agora debates, relate to the same planning domain. For this domain, suppose we are interested in a particular proposition θ. We imagine we have a number of scenarios in parallel, each with a different set of starting assumptions and possibly also different inference mechanisms. We now allow the associated Agora debates to proceed up to a certain time t, when we take a finite snapshot of each debate. It would be expected that the truth status of θ would be different under different scenarios. Not only are the assumptions and inference mechanisms different, but not all arguments may have been presented to each Agora debate at the time of the snapshot. We are thus faced with the question: *Given these different truth assignments under different scenarios, what overall truth-label should be assigned to θ?* Our problem is thus one of aggregation across multiple scenarios, and we present a method for doing this below. We permit scenarios to be weighted differentially, for example, according to their relative importance, their likelihood, or their feasibility, etc. However, to aggregate results arising from multiple scenarios, we need to ensure that each distinct scenario is only counted once, i.e. that no "double-counting" of scenarios takes place. In other words, we need a means to decide whether two scenarios are the same or not. In (McBurney & Parsons 2001b), we proposed a decision rule for determining whether two scenarios were distinct or not. We do not repeat the rule here, but simply assume that such a rule exists; consequently, we assume that any two scenarios may be assessed to be distinct or non-distinct.

Definition 2: *An **Ensemble** $\mathcal{S}$ is a finite collection of distinct Scenarios $\{\mathcal{S}^1, \ldots, \mathcal{S}^m\}$ relating to a common domain.*

We have adopted this terminology following its use in statistical mechanics by Gibbs (Gibbs 1902), who formalized Boltzmann's (Boltzmann 1872) notion of a possibly-infinite collection of hypothetical systems varying in their initial states; Gibbs' term is now standard in statistical mechanics (Gallavotti 1999).

Aggregating across Scenarios

Dialectical status and measures of support

Assuming we are only dealing with distinct scenarios, in this section we present a formalism for considering claims across multiple scenarios. We assume we have an ensemble $\mathcal{S} = \{\mathcal{S}^1, \mathcal{S}^2, \ldots, \mathcal{S}^m\}$ of m distinct scenarios, each $\mathcal{S}^i$ with an associated Agora debate $\mathcal{A}^i$. As before, we are interested in the dialectical-argumentation status of a formula θ in the domain, but across all m debates of the ensemble, not simply in one debate. We assume that at each time t, associated with each scenario $\mathcal{S}^i$ is a real-number, $a_t^i \in [0,1]$, called its scenario weight. We call $\tilde{a}_t = (a_t^1, a_t^2, \ldots, a_t^m)$ the *ensemble weights vector* at time t. We assume that the sum of the weights is constant for all t. If the weights are probabilities, this constant sum will be unity. Note that although we have allowed the weights to vary with time, we assume that their assignment to scenarios at any time t is independent of the dialectical status of propositions in the corresponding debates at t; in other words, the weights are assigned without knowledge of the arguments presented for and against claims in the debates.

What interpretation we give this measure depends upon the meanings applied to the logical language, to the ensemble, its scenarios and their weights, and to arguments for claims in the corresponding Agora debates. Several interpretations are possible, e.g.:

- The assumptions and claims may represent objects in the physical world, and the inference rules physical manipulations of these objects, such as actual construction of new objects from existing ones. Scenarios can thus be interpreted as different sets of resourcing assumptions, with claims being well-defended in an Agora debate when the objects they represent are able to be constructed with the assumed resources. In this interpretation, the ensemble weights may be the relative costs or benefits of different resources, or their likelihoods of occurence.
- The scenarios may represent alternative sets of rules of procedure for interaction between a group of participants, for example in a legal domain or in automated negotiation. Here the rules of inference represent different allowable modes of reasoning, such as reasoning by analogy or from authority. The ensemble weights may represent the extent of compliance with some set of principles of rational discourse, e.g. (Hitchcock 1991), or with some normative economic theory.
- The scenarios could represent different descriptions of some uncertain domain, for example different scientific theories, with propositions being statements about the domain, and the inference rules representing different causal mechanisms. The ensemble weights could be relative likelihoods of occurrence, or valuations of relative consequence or utility.

Each of these interpretations may be appropriate for particular planning domains. For example, the first interpretation may be appropriate for robotic planning when the robots are uncertain of their own resources and capabilities. In this case, if the scenarios are mutually exclusive and comprehensive of the possible outcomes of the domain, then it would be reasonable to assume that the ensemble weights, whether they be relative costs or relative likelihoods of occurrence, to sum to unity across the m scenarios.

Arguments presented for and against a claim θ in an Agora debate $\mathcal{A}^i$ are presented at discrete time-points. Therefore, at any one time-point, t, various situations are possible regarding the arguments for θ. We can consider four such situations, which together are mutually-exclusive and exhaustive: (a) At time t, no arguments have been articulated for θ; (b) At time t, arguments have been articulated for θ, but no rebuttals or undercuts have yet been presented; (c) At time t, arguments have been articulated for θ, and these have been rebutted or undercut; (d) At time t, arguments have been articulated for θ, for which rebuttals or undercuts have been presented, but these have themselves faced rebuttals or undercuts, i.e. θ is well-defended.

In any one scenario debate, of course, arguments may be presented for both θ and for $\neg\theta$, and, indeed, it is possible for both to be well-defended simultaneously. In assessing the strength of our belief in an uncertain proposition θ, we usually take into account arguments for and against the proposition, along with arguments for and against its negation. In our formalism, therefore, we have combined the four possible dialectical argumentation situations for θ with the equivalent four situations for $\neg\theta$ in all possible combinations; this gives sixteen mutually-exclusive and exhaustive dialectical states. We now list these in increasing order of support for θ, with each group of four states corresponding to one of the situations just listed, and the status of $\neg\theta$ cycling through the same list of four situations within each group. For subsequent reference, we label the states $1, \ldots, 16$, and for reasons of space we present only some of these:

1: At time t, no arguments have been articulated for θ, and $\neg\theta$ is well-defended.

2: At time t, no arguments have been articulated for θ, and arguments have been articulated for $\neg\theta$ which have also been rebutted or undercut.

3: At time t, no arguments have been articulated for θ, and arguments have been articulated for $\neg\theta$ which have not yet been rebutted or undercut.

4: At time t, no arguments have been articulated for θ, and no arguments have yet been articulated for $\neg\theta$.

5: At time t, arguments have been articulated for θ, but these have not yet been rebutted or undercut, and $\neg\theta$ is well-defended.

6: At time t, arguments have been articulated for θ, but these have not yet been rebutted or undercut, and arguments have been articulated for $\neg\theta$ which have also been rebutted or undercut.

⋮

16: At time t, θ is well-defended, and no arguments have yet been articulated for $\neg\theta$.

We refer to the numeric labels as *dialectical status labels.* Note that the order in which these states are listed, although increasing in the degree of support for θ, is *not* decreasing in support for $\neg\theta$. We now define a set of sixteen dialectical status functions $v^i_{d,t}(.)$ which together characterize the status of the arguments for θ in Agora debate i at time t, as follows:

Definition 3: *Given an ensemble $\mathcal{S} = \{\mathcal{S}^i | i = 1, \ldots, m\}$, a time t and a claim θ, the d-th* dialectical status valuation $v^i_{d,t}(\theta)$ *takes the value* 1 *precisely when θ is assigned the status label d at time t in Agora debate i, and zero otherwise, for $d = 1, 2, \ldots, 16$ and $i = 1, 2, \ldots, m$.*

Given this function, an obvious question is to what extent does a claim have d-level support across all the scenarios in the ensemble. Because we have weighted the scenarios by the ensemble weights vector $\tilde{a}$, it makes sense to weight the answer to this question also.

Definition 4: *At a given time t and for dialectical status label $d = 1, 2, \ldots, 16$, the d-*Support Function *$E_{d,t}(.)$ on the space of formulae is defined as:*

$$E_{d,t}(\theta) = \frac{\sum_{i=1}^{m} a^i_t v^i_{d,t}(\theta)}{\sum_{i=1}^{m} a^i_t}.$$

We say that $E_{d,t}(\theta)$ represents the (weighted) d-level support for θ at t, and we call $\tilde{E}_t(\theta) = (E_{1,t}(\theta), E_{2,t}(\theta), \ldots, E_{16,t}(\theta))$ the *$\mathcal{S}$-Support Vector* for θ at time t. We next discuss the properties of these support functions.

Properties of support functions

The $\mathcal{S}$-Support Vector shows the weighted dialectical status of θ at time t, across the m debates conducted under the scenarios in the ensemble $\mathcal{S}$. We have the following properties:

Proposition 2: *Given an ensemble $\mathcal{S}$, for any time t and claim θ, and for all dialectical status labels $d = 1, 2, \ldots, 16$, the d-Support Functions satisfy:*

1. $0 \leq E_{d,t}(\theta) \leq 1$
2. $\sum_{d=1}^{16} E_{d,t}(\theta) = 1$
3. $E_{d,t}(\theta) = E_{k,t}(\neg\theta)$*, where* $k = 20 - 4d + 15[\frac{d-1}{4}]$*, with* $[x]$ *the integer part of* x*.*

Proof. The first two properties follow immediately from the definition of the d-Support Functions. The third property follows from the fact the dialectical status labels, although listed in order of increasing strength for θ, could be readily re-arranged in increasing order of strength for $\neg\theta$. This would result in the sixteen categories being placed in the order: $16, 12, 8, 4, 15, 11, 7, 3, 14, 10, 6, 2, 13, 9, 5, 1$. The required re-arrangement sends the d-th category to the k-th category, where $k = 20 - 4d + 15[\frac{d-1}{4}]$, with $[x]$ the integer part of x. □

Note that these properties hold even when the ensemble weights are not probabilities, e.g. even if the weights do not sum to unity. Property 3 arises from the manner in which we have defined the 16 dialectical status categories for θ, definitions which include statements of the status of both θ and $\neg\theta$. As a consequence, it is not necessarily the case that $E_{d,t}(\theta) + E_{d,t}(\neg\theta) = 1$. For, by property 3, the left-hand side of this equation is equal to: $E_{d,t}(\theta) + E_{k,t}(\neg\theta)$, where $k = 20 - 4d + 15[\frac{d-1}{4}]$. If this sum is equal to 1 for some d, then, by Property 2, we must have $E_{l,t}(\theta) = 0$ for all $l \neq d, k$. This will only occur if there are either no scenarios in which θ is l-supported, or when the weights assigned to such scenarios are each zero.

The vector $\tilde{E}_t(\theta) = (E_{1,t}(\theta), \ldots, E_{16,t}(\theta))$ describes the dialectical status of a claim θ at time t across the m scenarios in the ensemble $\mathcal{S}$. If we were to plot these 16 values as a histogram, with the values $d = 1, 2, \ldots, 16$ along the horizontal axis, and the value of $E_{d,t}(\theta)$ along the vertical, we would have a function with the appearance of a probability distribution over the values of d at time t. With this perspective in mind, an obvious question is how can we summarize this information. In other words, given these 16 values, what single value provides a summary of the dialectical status of θ across the m scenario debates at time t? There are a number of alternatives:

Mode: The most common value(s), i.e. $\hat{d}_{mode,t} = arg\ \max_{d=1(1)16} E_{d,t}(\theta)$.

Median: The value(s) of d around which the E-mass is most evenly distributed, i.e.

$$\hat{d}_{median,t} = arg \min_{k=1(1)16} \left| \sum_{d=1}^{k-1} E_{d,t}(\theta) - \sum_{d=k+1}^{16} E_{d,t}(\theta) \right|$$

Mean: The average value of d, i.e. $\hat{d}_{mean,t} = \sum_{d=1}^{16} E_{d,t}(\theta) d$.

As with any statistical estimation, which of these estimators is most appropriate will depend upon the application. And like all summary statistics, these estimators potentially eliminate or obscure important information. If there are relatively large values of $E_{d,t}(\theta)$ for both small and large values of d, then there are highly-weighted scenarios in the ensemble where θ has strong support and others where it does not. What will be important in such a case will be identifying the differences between these scenarios. In other words, there is no reasons to believe that the vector $\tilde{E}_t(\theta)$, considered as a probability distribution over the values of d at time t, will be well-behaved. Its mass may be distributed unevenly (skewness), it may be multi-modal, and/or it may exhibit peakedness and large mass in the tails (kurtosis). Moreover, there is also no reason why such properties should not persist as t increases: in general, we may expect that differences in the scenario assumptions or inference mechanisms would lead to differential impacts in the corresponding Agora debates, and that these Agora differences would persist, rather than disappear, over the long run, if the scenarios in an ensemble remain distinct.

Let us assume, then, that such differential impacts may persist across the m Agora debates. However, *within* each debate, assume that the arguments regarding some claim θ eventually "stabilize" over the long-run; i.e. that each of the md dialectical status values, $v^i_{d,t}(\theta)$, converges to a finite limit as $t \to \infty$. Denote this limit by $v^i_{d,\infty}(\theta)$. Now, at each time t, we can calculate the $\mathcal{S}$-support vector $\tilde{E}_t(\theta) = (E_{1,t}(\theta), \ldots, E_{16,t}(\theta))$ from the values $\{v^i_{d,t}(\theta) | d = 1, \ldots, 16 \text{ and } i = 1, \ldots, m\}$. Will the values of this vector also converge to a finite limit as $t \to \infty$? We next show that this is the case, provided the ensemble weights $\tilde{a}_t = (a^1_t, \ldots, a^m_t)$ also converge with t.

Proposition 3: *Suppose $\mathcal{S} = \{\mathcal{S}^i | i = 1, \ldots, m\}$ is an ensemble, with weights $\tilde{a}_t = (a^1_t, \ldots, a^m_t)$. Assume $\tilde{a}_t$ converges to a vector of finite limits $\tilde{a}_\infty = (a^1_\infty, \ldots, a^m_\infty)$, as $t \to \infty$. Assume further that, for $d = 1, \ldots, 16$, each d-th dialectical status value, $v^i_{d,t}(\theta)$, converges to a finite limit, $v^i_{d,\infty}(\theta)$, as $t \to \infty$. Then, each d-Support Function, $E_{d,t}(\theta)$ also converges to a finite limit as $t \to \infty$, and this limit is:*

$$E_{d,\infty}(\theta) = \frac{\sum_{i=1}^m a^i_\infty v^i_{d,\infty}(\theta)}{\sum_{i=1}^m a^i_\infty}.$$

Proof. For simplicity of notation, we omit the argument θ. Suppose the sequence $(E_{d,t} | \; t = 1, 2, \ldots)$ does not converge, as $t \to \infty$. Then, there exists $\delta \in (0, 1]$ such that $\forall t, \exists s > t$ with $|E_{d,t} - E_{d,s}| \geq \delta$. Replacing the support functions with their definitions gives:

$$\left| \frac{\sum_{i=1}^m a^i_t v^i_{d,t}}{\sum_{i=1}^m a^i_t} - \frac{\sum_{i=1}^m a^i_s v^i_{d,s}}{\sum_{i=1}^m a^i_s} \right| > \delta$$

Recall that the weights sum to a constant, say a, across all time-values, and so:

$$\left| \sum_{i=1}^m a^i_t v^i_{d,t} - \sum_{i=1}^m a^i_s v^i_{d,s} \right| > \delta a$$

Choose $\epsilon \in (0, \min\{\delta a, 1\})$. Now, both the ensemble weights a^i_k and the dialectical status values $v^i_{d,k}$ converge as $k \to \infty$. Hence, we can choose t_i so that $|a^i_{t_i} - a^i_k| < \frac{\epsilon}{m}$, $\forall k > t_i$. Let $u_1 = \max\{t_1, \ldots, t_m\}$. Likewise, we can also choose s_i so that $|v^i_{d,s_i} - v^i_{d,k}| < \epsilon$, $\forall k > s_i$. Let $u_2 = \max\{s_1, \ldots, s_m\}$. But $\epsilon < 1$ and the status values $v^i_{d,k}$ are zero-one variables; so we must have either $v^i_{d,u_2} = v^i_{d,k} = 0$ or $v^i_{d,u_2} = v^i_{d,k} = 1$, for each i. Now choose $t \geq \max(u_1, u_2)$. Consequently, $\exists s$ with:

$$\delta a < \left| \sum_{i=1}^m a^i_t v^i_{d,t} - \sum_{i=1}^m a^i_s v^i_{d,s} \right| = \left| \sum_{i=1}^m (a^i_t v^i_{d,t} - a^i_s v^i_{d,s}) \right|$$

$$\leq \left| \sum_{i=1}^m (a^i_t - a^i_s) \right| < \epsilon$$

This contradicts our choice of ϵ and so our initial assumption of non-convergence of the d-Support functions must be false. A similar argument shows that the infinite limit value for each sequence $(E_{d,t}(\theta) | t = 1, 2, \ldots)$ of d-Support Functions is that expressed in the statement of the proposition. □

Similarly, we have convergence of the three estimators mentioned above:

Proposition 4: *Under the same assumptions as for Proposition 3, the mean, median and mode estimators defined above also converge to finite limits as $t \to \infty$.*
Proof. By reasoning similar to that for the proof of Proposition 3. □

We defined the dialectical status valuation functions $v^i_{d,t}(.)$ and the d-Support Functions $E_{d,t}(.)$ in terms of the 16 categories we identified for the dialectical argumentation status of a claim. Our categories were motivated by our intuitions regarding arguments and the relationships between them, and the circumstances under which different dialectical relationships constitute greater or lesser support for a claim. However, these categories were not essential to our subsequent definitions. Indeed, any mutually-exclusive and exhaustive partition of the space of possible arguments could have been used for our valuation and support functions. Thus, our framework is quite general, permitting a diversity of instantiations according to different intuitions and objectives.

Example

Given space limitations, our example is very simplified, illustrating only the core aggregation idea and not the application to conditional planning. We consider the situation facing an intending operator of global mobile satellite-based telecommunications services (GMSS) in 1990 (McBurney & Parsons 2001c). Demand for these services was predicted to depend heavily on the extent to which terrestrial mobile communications services would expand, both in terms of customer numbers and the geographic area under coverage. One could imagine a number of scenarios for the future, under each of which there would be arguments for and against the claim that demand for GMSS would be large. We consider the following ensemble, with arguments articulated at time t as indicated:

Scenario 1: Terrestrial mobile services expand rapidly and customers wish to use their phone everywhere. Argument: Large numbers of terrestrial customers leads to high demand for GMSS outside terrestrial coverage.

Scenario 2: Terrestrial mobile services expand rapidly and customers are happy with the terrestrial coverage. Arguments: 1. Large numbers of terrestrial customers leads to high demand for GMSS outside terrestrial coverage. 2. However, large geographic coverage for terrestrial services leads to lower demand for GMSS, as most areas have terrestrial coverage.

Scenario 3: Terrestrial mobile services do not expand. Argument: Small geographic coverage means high demand for GMSS.

Let θ be the claim: *"GMSS experiences high demand."* In each Scenario, we have an argument for θ at time t. In Scenario 2, however, we also have an argument against θ, but this counter-argument is not itself countered by this time. Thus, θ is well-defended only in Scenarios 1 and 3 at time t, and in neither scenario have arguments yet been presented for $\neg\theta$ at this time. Now, recall that scenario weights are assigned independently of the arguments under each, and assume this ensemble is assigned the weights vector $(0.7, 0.7, 0.3)$ at time t. We could therefore calculate the 16-Support Function for θ at t as follows: $E_{16,t}(\theta) = (0.7 + 0.3)/(0.7 + 0.7 + 0.3) = 0.59$. This value is the weighted mass of the ensemble in which θ has the strongest position, i.e. where θ is well-defended and where no arguments have yet been articulated for $\neg\theta$. This example has not illustrated the working of the argumentation apparatus within the Agora debates under each scenario, but has simply shown the basic mechanism for aggregation across scenarios.[4]

Discussion

Building on prior work using dialectical argumentation as a qualitative representation of uncertainty, we have presented an argumentation-based formalism for reasoning across multiple future scenarios. Our formalism provides a basis for the comparison of different plans, when these are conditional on assumptions about the current or future states of the world. The state-space explosion problem precludes the complete articulation and comparison of all possible conditional plans in any realistic domain. Instead, a resource-constrained planner may reason only about a subset of all the plans possible, by considering only some of the possible futures. Our formalism enables a planner to group future states with common assumptions into a bundle called a scenario, and then to reason across a collection of scenarios in a coherent manner.

Our previous work had articulated a set of dialogue-game rules for the conduct of debates over scientific or other domains, drawing on theories of rational human discourse (McBurney & Parsons 2000; 2001a). We had also previously defined scenarios in terms of the assumptions and inference mechanisms available to participants in such debates, and proposed a decision rule to assess whether two given scenarios are distinct or not (McBurney & Parsons 2001b). With such a decision rule, we are able to ensure that any procedure for aggregation across scenarios counts only distinct scenarios. In this paper, we have proposed a vector measure of the degree of support generated for a claim at a given time in a finite collection of debates conducted under different scenarios. This vector assesses the weighted proportion of scenarios in which the claim is supported at that time, for different degrees of support. We have explored some of the properties of this vector measure, and have found sufficient conditions for its convergence to finite values, as time increases to infinity. These conditions include the requirement that the debates under each scenario individually stabilize in their degree of support for the claim in question, even though there may be great differences from one debate to another.

It is possible to view the different scenarios as different possible world-states, and to view the weight attached to each scenario as the probability of its occurrence. This is one interpretation of the Boltzmann-Gibbs notion of *ensemble* in statistical mechanics, which is why we adopted this terminology for our collections of scenarios. Under this interpretation of our framework, assessment of the weighted degree of argumentation support for a claim across all scenarios is analogous to assessment of the "probability of provability" of the claim (Pearl 1988). In our framework, "provability" of a claim corresponds to saying the claim *"has a defined degree of argumentation support in an Agora debate conducted under some scenario"*, and "probability" corresponds to *"the relative weight of the scenarios contained in some ensemble in which this is the case,"* given an ensemble weighting. Similarly, our approach may be seen as an argumentation analog of the *Ents* model of belief of Paris and Vencovská (Paris & Vencovská 1993), in which an agent's belief in a claim is determined by imagining possible worlds in which the claim is decided, either true or false, and then setting its belief in the claim equal to the proportion of possible worlds in which the claim is true.

There are several possible extensions to the work presented here. Firstly, it will be valuable to implement our framework in conjunction with a conditional planning algorithm, such as CNLP (Peot & Smith 1992) or Cassandra (Pryor & Collins 1996). Secondly, we have assumed that scenarios are based on different sets of assumptions concerning beliefs and inference rules, but we have not discussed how such assumptions should be made. This issue is related to the problem of selection of contingencies in conditional planning (Onder & Pollack 1996). Finally, our definition of the strength of support for a claim assesses this status *within* each Agora debate and then aggregates across all the debates. Alternatively, assessment procedures could scan across all scenarios initially, before aggregation. For example, a

[4] For the record, in reality the very high growth of terrestrial mobile services witnessed worldwide over the last decade did not lead to high demand for GMSS, a fact which contributed to the business failures of the main intending GMSS providers, Iridium, ICO and Globalstar.

claim may be defined as having the strongest support if an argument for it is well-defended in at least one debate, and if arguments presented for it in other debates where it not well-defended face no attackers or rebuttals.

Acknowledgments We are grateful for partial funding received from the British Engineering and Physical Sciences Research Council (EPSRC) under grant GR/L84117 and a PhD studentship.

References

Boltzmann, L. 1872. Weitere studien über das wärmegleichgewicht unter Gasmolekülen. *Wissenschaftliche Abhandlungen* 1:316–402.

Boutilier, C.; Dean, T.; and Hanks, S. 1999. Decision theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research* 11:1–94.

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

Fox, J., and Das, S. 2000. *Safe and Sound: Artificial Intelligence in Hazardous Applications.* Cambridge, MA, USA: MIT Press.

Gallavotti, G. 1999. *Statistical Mechanics: A Short Treatise.* Berlin, Germany: Springer.

Gibbs, J. W. 1902. *Elementary Principles in Statistical Mechanics, Developed with Especial Reference to the Rational Foundation of Thermodynamics.* New Haven, CT, USA: Yale University Press. (Reprinted by Dover, New York, 1962.).

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

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

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

Krause, P.; Fox, J.; Judson, P.; and Patel, M. 1998. Qualitative risk assessment fulfils a need. In Hunter, A., and Parsons, S., eds., *Applications of Uncertainty Formalisms*, LNAI 1455. Berlin, Germany: Springer. 138–156.

McBurney, P., and Parsons, S. 2000. Risk agoras: Dialectical argumentation for scientific reasoning. In Boutilier, C., and Goldszmidt, M., eds., *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI-2000)*, 371–379. San Francisco, CA, USA: Morgan Kaufmann.

McBurney, P., and Parsons, S. 2001a. Representing epistemic uncertainty by means of dialectical argumentation. *Annals of Mathematics and Artificial Intelligence* (in press).

McBurney, P., and Parsons, S. 2001b. Strawmen and eidolons: using argumentation to reason across scenarios. Preprint, Department of Computer Science, University of Liverpool, UK.

McBurney, P., and Parsons, S. 2001c. Using Belief Functions to forecast demand for Mobile Satellite Services. In Srivastava, R. P., and Mock, T., eds., *Belief Functions for Business Decisions.* New York, NY, USA: Physica (in press).

Onder, N., and Pollack, M. E. 1996. Contingency selection in plan generation. In Pryor, L., ed., *Plan Extension: Problems and Issues: Proceedings of the 1996 AAAI Fall Symposium.* Menlo Park, CA, USA: AAAI Press. Technical Report.

Paris, J., and Vencovská, A. 1993. A model of belief. *Artificial Intelligence* 64:197–241.

Parsons, S.; Pettersson, O.; Saffiotti, A.; and Wooldridge, M. 2000. Intention reconsideration in theory and practice. In Horn, W., ed., *Proceedings of Fourteenth European Conference on Artificial Intelligence (ECAI-2000)*, 378–382. Berlin, Germany: IOS Press.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Francisco, CA, USA: Morgan Kaufmann.

Peot, M., and Smith, D. 1992. Conditional nonlinear planning. In *Proceedings of the First International Conference on AI Planning*, 189–197.

Pryor, L., and Collins, G. 1996. Planning for contingencies: a decision-based approach. *Journal of Artificial Intelligence Research* 4:287–339.

Schwartz, P. 1991. *The Art of the Long View.* New York, NY, USA: Doubleday.

Warren, D. H. D. 1976. Generating conditional plans and programs. In *Proceedings of the Summer Conference of the Society for Artificial Intelligence and the Simulation of Behaviour (AISB-76)*, 344–354. Edinburgh, UK: University of Edinburgh.