

Tenacious Tortoises: A Formalism for Argument over Rules of Inference

Peter McBurney and Simon Parsons
Department of Computer Science
University of Liverpool
Liverpool L69 7ZF U.K.

{P.J.McBurney,S.D.Parsons}@csc.liv.ac.uk

June 2, 2000

Abstract

As multi-agent systems proliferate and employ different and more sophisticated formal logics, it is increasingly likely that agents will be reasoning with different rules of inference. Hence, an agent seeking to convince another of some proposition may first have to convince the latter to use a rule of inference which it has not thus far adopted. We define a formalism to represent degrees of acceptability or validity of rules of inference, to enable autonomous agents to undertake dialogue concerning inference rules. Even when they disagree over the acceptability of a rule, two agents may still use the proposed formalism to reason collaboratively.

1 Introduction

In 1895, the logician Charles Dodgson (aka Lewis Carroll) famously imagined a dialogue between Achilles and a tortoise, in which the application of Modus Ponens (MP) was contested as a valid rule of inference [4]. Given arbitrary propositions P and Q , and the two premises P and $(P \rightarrow Q)$, one can only conclude Q from these premises if one accepts that Modus Ponens is a valid rule of inference. This the tortoise refuses to do, much to the exasperation of Achilles. Instead, the tortoise insists that a new premise be added to the argument, namely: $(P \wedge (P \rightarrow Q)) \rightarrow Q$. When Achilles does this, the tortoise still refuses to accept Q as the conclusion, insisting on yet another premise: $(P \wedge (P \rightarrow Q) \wedge ((P \wedge (P \rightarrow Q)) \rightarrow Q)) \rightarrow Q$. The tortoise continues in this vein, *ad infinitum*.

Eighty years later, philosopher Susan Haack [9] took up the question of how one justifies the use of MP as a deductive rule of inference. If one does so by means of examples of its valid application, then this is in essence a form of induction, which (as she remarks) seems too weak a means of justification for a rule of deduction. If, on the other hand, one uses a deductive means of justification, such as demonstrating the preservation of truth across the inference step in a truth-table, one risks using the very rule being justified. So how can one person convince another of the validity of a rule of deductive inference?

That rules of inference may be the subject of fierce argument is shown by the debate over Constructivism in pure mathematics in the twentieth century [21]: here the rule of inference being contested was double negation elimination in a *Reductio Ad Absurdum* (RAA) proof:

FROM $(\neg P \rightarrow Q)$ and $(\neg P \rightarrow \neg Q)$
INFER $\neg\neg P$
FROM $\neg\neg P$
INFER P

Although the choice of inference rules in purely formal mathematics may be arbitrary,¹ the question of acceptability of rules of inference is important for Artificial Intelligence for a number of reasons. Firstly, it is relevant to modeling scientific reasoning. Constructivism, for example, has been proposed as a formalism for modern physics [3], as have other, non-standard logics. In the propositional calculus proposed for quantum mechanics by Birkhoff and von Neumann [2], for example, the distributive laws did not hold:

¹Goguen [8], for example, argues that standards of mathematical proof are socially constructed.

$$(A \vee B) \wedge (A \vee C) \vdash A \vee (B \wedge C)$$

$$(A \vee B) \wedge C \vdash (A \wedge C) \vee (B \wedge C)$$

Indeed, it is possible to view scientific debates over alternative causal theories as concerned with the validity of particular modes of inference, as we have shown with regard to claims of carcinogenicity of chemicals based on animal evidence [13]. Intelligent systems which seek to formally model such domains will need to represent these arguments [14].

Secondly, it is not obvious that one logical formalism is appropriate for all human reasoning, a subject of much past debate in philosophy (e.g. see [10]). A many-valued logic proposed for quantum physics, for instance, has also been suggested to describe religious reasoning in Azande and Nuer societies, reasoning which appeared to contravene MP [5]. Indeed, some anthropologists have argued that formal human reasoning processes are culturally-dependent and hence different across cultures [18]. To the extent that this is the case, systems of autonomous software agents acting on behalf of humans will need to reflect the diversity of formal processes used by humans. Thus, it is possible that interacting agents may be using logics with different rules of inference, as is possible in the agent negotiation system of [15]. If one agent seeks to convince another of a particular proposition then that first agent may have to demonstrate the validity of a rule of inference used to prove the proposition. Our objective in this work is to develop a formalism in which such a debate between agents could be conducted.

2 Arguments over rules of inference

We begin by noting that a dialogue between two agents in which one only asserts, and the other only denies, a rule of inference will not likely lead very far. A dialogue between agents concerning a rule of inference will need to express more than simply their respective positions if either agent is to be persuaded to change its position. What more may be expressed?

Suppose we have two agents, denoted A and B, and that A seeks to convince B of a proposition θ . For example, this may be a joint intention which A desires both agents to adopt. B asks for a proof of θ . Suppose that A provides a proof which commences from axioms which are all accepted by B. Assume, however, that this proof uses a rule of inference \mathcal{R} which B says its logic

does not include. For example, \mathcal{R} may be the use of the contrapositive or RAA. There are three ways in which the dialogue between A and B could then proceed.

First, A could attempt to demonstrate that \mathcal{R} can be derived from the rules of inference which are contained in B's logic. Similarly, A could attempt to demonstrate that \mathcal{R} is admissible in B's logic [20], i.e. that \mathcal{R} is an element of that set of inference rules under which the theorems of B's logic remain unchanged.² In either of these two cases, it would then be rational for B to accept θ , being a proposition whose proof commences from agreed assumptions and which uses inference rules equivalent (in the sense of derivability or admissibility) to those B has adopted. In such a case, the difference of opinion is resolved, to the satisfaction of both agents.

Suppose then that A is unable to prove that \mathcal{R} is derivable from or admissible in B's logic. The second approach which A may pursue is to attempt to give non-deductive reasons for B to adopt \mathcal{R} . Examples of such reasons could include: scientific evidence for the causal mechanism possibly represented by \mathcal{R} , where the reasoning is in a scientific domain; instances of its valid application (e.g. the use of precedents in legal arguments); the (possibly non-deductive) positive consequences for B of adopting \mathcal{R} (e.g. that doing so will improve the welfare of B, of A and/or of third parties); the (possibly non-deductive) negative consequences for B of not adopting \mathcal{R} (e.g. that not doing so will be to the detriment of B, of A and/or of third parties); or empirical evidence which would impact the choice of a particular logic.³ The precise nature of such arguments will depend upon the domain represented by the multi-agent system, and the nature of the proposition θ . Moreover, for A to successfully convince B using such arguments, B would require some formal means of assessing them, perhaps using a logic of values as outlined in [7]. Although currently being explored, these ideas are not pursued further here.

Suppose, however, that A exhausts all such arguments, and still fails to convince B to adopt either \mathcal{R} or θ . Then, a third approach which A could pursue is to represent B's misgivings over the use of \mathcal{R} in an appropriate formalism and use this to seek compromise

²Note that \mathcal{R} could be admissible in B's logic yet not derivable from the axioms and inference rules of that logic. All derivable rules are admissible, however [20].

³Theory change in logic on the basis of empirical evidence has been much discussed in philosophy, typically in a context of holist epistemology [17].

between the two of them. We term such a formalism an Acceptability Formalism (AF) and see it as akin to formalisms for representing uncertainty regarding the truth of propositions. Note that while B’s misgivings concerning rule \mathcal{R} may arise from uncertainty as to its validity, they need not: B may be quite certain in rejecting the rule.

What would be an appropriate formalism for representing degrees of acceptability of a rule of inference? At this point, A has adopted \mathcal{R} and B has not, so that, in effect, A (or, strictly, A’s designer) has decided that the rule is an acceptable rule and B has not so decided. In other words, A has assigned \mathcal{R} the label *Acceptable* to \mathcal{R} , and B has not assigned this label. Thus, a very simple representation of their views of \mathcal{R} would be assigning labels from the qualitative dictionary: $\{\textit{Acceptable}, \textit{Unacceptable}\}$ or from the dictionary $\{\textit{Acceptable}, \textit{No opinion}, \textit{Unacceptable}\}$. Such simple dictionaries leave little room for compromise; so it behooves A to request B to assign a label from a more granular dictionary, such as the five-element set:

\{Always acceptable, Mostly acceptable but sometimes unacceptable, Acceptable and unacceptable to the same extent, Sometimes acceptable but mostly unacceptable, Always unacceptable\}.

Were B to assign any but the final label, *Always unacceptable*, then A has the opportunity to demonstrate to B that the current use of \mathcal{R} in the proof of θ is an acceptable application of the rule, and thus achieve some form of compromise between the two.

To formalize this third approach we therefore assume that A and B agree a dictionary \mathcal{D} of labels to be assigned to rules of inference. The elements of such an AF dictionary could be linguistic qualifiers, as in the examples above, but they need not be. For example, \mathcal{D} may be the set of integers between 1 and 100 (inclusive), where larger numbers represent greater relative acceptability of the rule. It is possible to view standard statistical hypothesis-testing procedures, Neyman-Pearson theory [6], in this way. Here, for a proposition θ concerning unknown parameters, the inference rule is:

FROM θ is true of a sample

INFER θ is true of the population from which that sample arises.

Under assumptions regarding the manner in which the sample was obtained from the population (e.g. that it was randomly selected) and assumptions regarding the distribution of the parameters of interest in the population, Neyman-Pearson theory estimates an upper bound

for the probability that the application of the inference rule is invalid. Thus, we cannot say that the application of the inference rule is valid in any one case, but we can say that, if applied to repeated samples drawn from the same population, it will be invalidly applied (say) at most 5% of the time. Thus, the calculation of p -values for statistical hypothesis tests, which is common practice in the biological and medical sciences [19], effectively associates each inference with a value from the set $\{p : p \in (0, 1)\}$. The label “100(1 – p)%” is thus a measure of confidence in the validity of application of the inference rule.⁴

Once the two agents have agreed to adopt such a dictionary, the labels could then be applied to multiple con-tested rules of inference, and used in successive proofs. To do this will require a calculus for combining labels for different rules, and for propagating labels through chains of reasoning, which is the subject of the next Section.

3 Terrapin Logic TL

3.1 Formalization

We now present a formal description of the logic, which we call TL (for “Terrapin Logic”, from the Algonquian for tortoise), to enable reasoning about acceptability labels for rules of inference. Our formalization is similar to that for the Logic of Argumentation LA presented in [7], itself influenced by labelled deductive systems and earlier formalizations of argumentation.

We start with a set of atomic propositions including \top and \perp , the ever true and ever false propositions. We assume this set of well-formed formulae (*wffs*), labeled \mathcal{L} , is closed under the connectives $\{\neg, \rightarrow, \wedge, \vee\}$. \mathcal{L} may then be used to create a database Δ whose elements are 4-tuples, $(\theta : G : R : \vec{d})$, in which θ is a *wff*, $G = (\theta_0, \theta_1, \dots, \theta_{n-1})$ is an ordered sequence of *wffs*, with $n \geq 1$, and where $R = (\vdash_1, \vdash_2, \dots, \vdash_n)$ is an ordered sequence of inference rules, such that:

$$\theta_0 \vdash_1 \theta_1 \vdash_2 \theta_2 \dots \theta_{n-1} \vdash_n \theta.$$

In other words, each element $\theta_k \in G$ is derived from the preceding element θ_{k-1} as a result of the application of the k -th rule of inference, \vdash_k , ($k = 1, \dots, n - 1$). The rules of inference in any such sequence may be non-distinct. The element $\vec{d} = (d_1, d_2, \dots, d_n)$ is an ordered sequence of elements from a Dictionary \mathcal{D} , being an assignment of AF labels to the sequence of inference

⁴This interpretation is akin to Pollock’s statistical syllogism [16].

rules R . We also permit wffs $l \in \mathcal{L}$ to be elements of Δ , by including tuples of the form $(l : \emptyset : \emptyset : \emptyset)$, where each \emptyset indicates a null term. Note that the assignment of AF labels may be context-dependent, i.e. the d_i assigned to \vdash_i may also depend on θ_{i-1} . This is the case for statistical inference, where the p -value depends on characteristics of the sample from which the inference is made, such as its size.

With this formal system, we can take a database Δ and use the consequence relation \vdash_{TCR} defined in Figure 1 to build arguments for propositions of interest. This consequence relation is defined in terms of rules for building new arguments from old. The rules are written in a style similar to standard Gentzen proof rules, with the antecedents of the rule above the horizontal line and the consequent below. In Figure 1, we use the notation $G \otimes H$ to refer to that ordered sequence created from appending the elements of sequence H after the elements of sequence G , each in their respective order. The rules are:

- The rule Ax says that if the tuple $(\theta : G : R : \tilde{d})$ is in the database, then it is possible to build the argument $(\theta : G : R : \tilde{d})$ from the database. The rule thus allows the construction of arguments from database items.
- The rule \wedge -I says that if the arguments $(\theta : G : R : \tilde{d})$ and $(\phi : H : S : \tilde{e})$ may be built from the database, then an argument for $\theta \wedge \phi$ may also be built. The rule thus shows how to introduce arguments about conjunctions; using it requires an inference of the form: $\theta, \phi \vdash (\theta \wedge \phi)$, which we denote $\vdash_{\wedge-I}$ in Figure 1. This inference is then assigned an AF dictionary value of $d_{\wedge-I}$.
- The rule \wedge -E1 says that if it is possible to build an argument for $\theta \wedge \phi$ from the database, then it is also possible to build an argument for θ . Thus the rule allows the elimination of one conjunct from an argument, and its use requires an inference of the form: $\theta \wedge \phi \vdash \theta$. This inference is denoted by $\vdash_{\wedge-E1}$, and is assigned an AF value of $d_{\wedge-E1}$. The rule \wedge -E2 is analogous to \wedge -E1 but allows the elimination of the other conjunct.
- The rule \vee -I1 allows the introduction of a disjunction from the left disjunct. The rule \vee -I2 allows the introduction of a disjunction from the right disjunct.

- The rule \vee -E allows the elimination of a disjunction and its replacement by tuple when that tuple is a TL-consequence of each disjunct.
- The rule \neg -I allows the introduction of negation. The rule \neg -E allows the derivation of \perp , the ever-false proposition, from a contradiction. The rule $\neg\neg$ -E allows the elimination of a double negation, and thus permits the (possibly contested) assertion of the Law of the Excluded Middle (LEM).
- The rule \rightarrow -I says that if on adding a tuple $(\theta : \emptyset : \emptyset : \emptyset)$ to a database, where $\theta \in \mathcal{L}$, it is possible to conclude ϕ , then there is an argument for $\theta \rightarrow \phi$. The rule thus allows the introduction of \rightarrow into arguments.
- The rule \rightarrow -E says that from an argument for θ and an argument for $\theta \rightarrow \phi$ it is possible to build an argument for ϕ . The rule thus allows the elimination of \rightarrow from arguments and is analogous to MP in standard propositional logic.

Our purpose in this paper is to propose a formal syntax and proof rules for argument over rules of inference, and so we do not consider semantic issues. Interpretations of TL would be defined with respect to a specified AF dictionary or dictionary-class, and may assign \rightarrow to represent a relationship between propositions other than material implication. A virtue of our initial focus on syntactical elements is that, once defined, the proof rules may be applied in different semantic contexts. We are currently exploring alternative semantic interpretations for TL, along with the issue of its consistency and completeness relative to these.

3.2 Negotiation within TL

Given the formalism TL just defined, how may this be used by two agents, A and B, in dialogue over a contested rule of inference? We assume the agents have agreed a common set of assumptions to which they both adhere, and have agreed a common AF dictionary \mathcal{D} of labels to assign to inference rules. We assume the elements of \mathcal{D} are partially ordered under a relation denoted $<$. We further assume that \mathcal{D} contains an element $d_{-\infty}$ such that for all other $d \in \mathcal{D}$, we have $d_{-\infty} < d$, and that the assignment of $d_{-\infty}$ to a rule of inference by an agent marks it as always and completely unacceptable.

$$\begin{array}{c}
\text{Ax} \frac{(\theta : G : R : \tilde{d}) \in \Delta}{\Delta \vdash_{TCR} (\theta : G : R : \tilde{d})} \\
\wedge\text{-I} \frac{\Delta \vdash_{TCR} (\theta : G : R : \tilde{d}) \text{ and } \Delta \vdash_{TCR} (\phi : H : S : \tilde{e})}{\Delta \vdash_{TCR} (\theta \wedge \phi : G \otimes H \otimes (\theta \wedge \phi) : R \otimes S \otimes (\vdash_{\wedge\text{-I}}) : \tilde{d} \otimes \tilde{e} \otimes (d_{\wedge\text{-I}}))} \\
\wedge\text{-E1} \frac{\Delta \vdash_{TCR} (\theta \wedge \phi : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\theta : G \otimes (\theta) : R \otimes (\vdash_{\wedge\text{-E1}}) : \tilde{d} \otimes (d_{\wedge\text{-E1}}))} \\
\wedge\text{-E2} \frac{\Delta \vdash_{TCR} (\theta \wedge \phi : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\phi : G \otimes (\phi) : R \otimes (\vdash_{\wedge\text{-E2}}) : \tilde{d} \otimes (d_{\wedge\text{-E2}}))} \\
\vee\text{-I1} \frac{\Delta \vdash_{TCR} (\theta : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\theta \vee \phi : G \otimes (\theta \vee \phi) : R \otimes (\vdash_{\vee\text{-I1}}) : \tilde{d} \otimes (d_{\vee\text{-I1}}))} \\
\vee\text{-I2} \frac{\Delta \vdash_{TCR} (\phi : H : S : \tilde{e})}{\Delta \vdash_{TCR} (\theta \vee \phi : H \otimes (\theta \vee \phi) : S \otimes (\vdash_{\vee\text{-I2}}) : \tilde{e} \otimes (e_{\vee\text{-I2}}))} \\
\vee\text{-E} \frac{\Delta \vdash_{TCR} (\theta \vee \phi : G : R : \tilde{d}) \text{ and } \Delta, (\theta : \emptyset : \emptyset : \emptyset) \vdash_{TCR} (\gamma : H : S : \tilde{e}) \text{ and } \Delta, (\phi : \emptyset : \emptyset : \emptyset) \vdash_{TCR} (\gamma : J : T : \tilde{f})}{\Delta \vdash_{TCR} (\gamma : G \otimes H \otimes J \otimes (\gamma) : R \otimes S \otimes T \otimes (\vdash_{\vee\text{-E}}) : \tilde{d} \otimes \tilde{e} \otimes \tilde{f} \otimes (d_{\vee\text{-E}}))} \\
\neg\text{-I} \frac{\Delta, (\theta : \emptyset : \emptyset : \emptyset) \vdash_{TCR} (\perp : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\neg\theta : G \otimes (\neg\theta) : R \otimes (\vdash_{\neg\text{-I}}) : \tilde{d} \otimes (d_{\neg\text{-I}}))} \\
\neg\text{-E} \frac{\Delta \vdash_{TCR} (\theta : G : R : \tilde{d}) \text{ and } \Delta \vdash_{TCR} (\neg\theta : H : S : \tilde{e})}{\Delta \vdash_{TCR} (\perp : G \otimes H \otimes (\perp) : R \otimes S \otimes (\vdash_{\neg\text{-E}}) : \tilde{d} \otimes \tilde{e} \otimes (d_{\neg\text{-E}}))} \\
\neg\neg\text{-E} \frac{\Delta \vdash_{TCR} (\neg\neg\theta : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\theta : G \otimes (\theta) : R \otimes (\vdash_{\neg\neg\text{-E}}) : \tilde{d} \otimes (d_{\neg\neg\text{-E}}))} \\
\rightarrow\text{-I} \frac{\Delta, (\theta : \emptyset : \emptyset : \emptyset) \vdash_{TCR} (\phi : G : R : \tilde{d})}{\Delta \vdash_{TCR} (\theta \rightarrow \phi : G \otimes (\theta \rightarrow \phi) : R \otimes (\vdash_{\rightarrow\text{-I}}) : \tilde{d} \otimes (d_{\rightarrow\text{-I}}))} \\
\rightarrow\text{-E} \frac{\Delta \vdash_{TCR} (\theta : G : R : \tilde{d}) \text{ and } \Delta \vdash_{TCR} (\theta \rightarrow \phi : H : S : \tilde{e})}{\Delta \vdash_{TCR} (\phi : G \otimes H \otimes (\phi) : R \otimes S \otimes (\vdash_{\rightarrow\text{-E}}) : \tilde{d} \otimes \tilde{e} \otimes (d_{\rightarrow\text{-E}}))}
\end{array}$$

Figure 1: The TL Consequence Relation

We then assume the two agents agree to construct a logical language \mathcal{L} which adopts all inference rules in the union of their two respective sets of rules (i.e. \mathcal{L} contains all those rules which either agent has adopted).⁵ We next assume that two databases, Δ^A and Δ^B , of 4-tuples are constructed from \mathcal{L} as outlined above, with Δ^A containing agent A's assignments of dictionary labels in the fourth place of each tuple, while Δ^B contains B's assignments. Thus, the elements of the two databases may potentially only differ in the fourth places of the tuples each contains, since \mathcal{L} uses all inference rules of both agents. One can readily imagine cases where such differences may arise. For example, we noted in the previous section that the TL double negation elimination rule, $\neg\neg$ -E, may be used to assert LEM. If one agent does not agree with the use of this rule in this way they may assign it an AF value of $d_{-\infty}$. As mentioned, this assignment can be context-specific, i.e. an agent could assign the value $d_{-\infty}$ only when $\neg\neg$ -E is used for certain propositions and not for others. Similarly, agents may assign differential dictionary values to the use of inference rules which are derived from those in Figure 1, such as the two distributive laws mentioned in Section 1 in relation to Birkhoff and von Neumann's logic for quantum mechanics.

As in Section 2, assume there is a claim θ which A asserts but which B contests since its proof uses an inference rule which B has not adopted, nor which is derivable from, nor admissible in, B's logic. For simplicity, we first assume there is only one such rule and that it is deployed only once in A's proof of θ . Suppose the tuple which contains A's proof of θ is $(\theta : G : R : \tilde{d}^A)$, and that the contested rule is \vdash_k , for some k . B's assignment of labels to the inference rules used in the proof of θ is the fourth element of the tuple $(\theta : G : R : \tilde{d}^B)$. Since the k -th rule is contested by B, we should expect the k -th elements of \tilde{d}^A and \tilde{d}^B to differ, i.e. that $d_k^A \neq d_k^B$.

If $d_k^B = d_{-\infty}$, then B has assigned the contested rule a label which indicates its use is completely unacceptable to B. This would eliminate any possibility of compromise between the two agents over the use of the rule. The dialogue could proceed only by the second of the two approaches outlined in Section 2, i.e. by means of a discussion of the implications of adopting or not adopting the contested rule or the proposition θ .⁶ Suppose

⁵We assume for simplicity that the axioms of the logics of the two agents are not inconsistent.

⁶Agent A could seek to contest the assignment by B of the label $d_{-\infty}$, an approach we do not pursue here. As Heathcote has demonstrated [11], to justify an assertion that the rule represented an invalid

instead then that $d_k^B \neq d_{-\infty}$. In this circumstance, although B's logic does not include \vdash_k , B may be willing to accept \vdash_k some of the time. For instance, if the labels in \mathcal{D} had a probabilistic interpretation, B may agree to use \vdash_k a proportion of the times it is asked to do so, analogously with statistical confidence values. Alternatively, B may accept the use of contested rules on the basis of the label assigned to them being above some threshold value; such thresholds may differ according to the identity of the requesting agent, A, for example, with contested rules being accepted more readily from trusted agents than from others.

Our approach so far has assumed that A is seeking to persuade B to adopt a proposition θ , and hence an inference rule \vdash_k . However, if the two agents are engaged in some joint task, for instance agreeing common intentions or prioritizations, both A and B may be simultaneously seeking to persuade each other to adopt propositions and thus inference rules. In these circumstances, it may behoove the two agents to agree common acceptability labels for contested inference rules, as a means of ranking or prioritizing propositions. How might this be done? Suppose, as above, that database Δ^A contains the tuple $(\theta : G : R : \tilde{d}^A)$, while Δ^B contains the tuple $(\theta : G : R : \tilde{d}^B)$. We can readily construct a common database Δ of tuples $(\theta : G : R : \tilde{d})$, where the labels \tilde{d} are defined from \tilde{d}^A and \tilde{d}^B by some agreed method. For instance, A and B may agree to define each element d_i of \tilde{d} by $d_i = \min\{d_i^A, d_i^B\}$.

It would also be straightforward to define a function which maps a sequence \tilde{d} to a single value d^* , to provide some form of summary assessment of a chain of inferences. For instance, the mapping $d^* = \min_{i=1, \dots, n}\{d_i\}$ would be equivalent in this context to saying that "A chain is only as strong as its weakest link." If AF dictionary values were real numbers between 0 and 1 (e.g. statistical p -values), then an appropriate mapping may be $d^* = 1 - \prod_{i=1}^n (1 - d_i)$. With such a mapping agreed, the two agents could then readily define a rank order of propositions. For instance, if the weakest-link mapping $d^* = \min_i\{d_i\}$ was used, and Δ contains the tuples $(\theta : G : R : \tilde{d})$ and $(\phi : H : S : \tilde{e})$, then we could define θ to be ranked higher than ϕ whenever $\min_j\{e_j\} < \min_i\{d_i\}$. This may be of value if the propositions represent, for example, alternative joint intentions, or competing allocations of resources. Recent work in AI has explored

form of argument B may ultimately require some form of abduction, which thus provides the possibility of continuing contestation by A.

methods for combining preferences of different agents in argumentation systems [1].

Note also that the AF labels and the summary mapping d^* could be used to define an uncertainty formalism value for the proposition θ at the conclusion of the chain of inference. Again, statistical inference provides an example: consequent statements (about population parameters) are assigned labels *TRUE* or *FALSE* in a statistical inference according to the relative size of the sample p -value compared to some pre-determined threshold value, typically 0.05. Such an assignment of uncertainty values to propositions would provide another way for the two agents to jointly prioritize propositions. If the two agents do agree to use a common database Δ constructed as described here, then the Terapin Logic provides a means for them to do so. This is because the TL Consequence Relation rules of Figure 1 are a calculus for propagation and manipulation of the 4-tuple elements of Δ .

4 Conclusion

We have presented a formalism in which degrees of acceptability of rules of inference can be represented, so that two agents may undertake dialogue and negotiation over contested rules. The formalism also permits agents in disagreement to collaborate on joint tasks. Our approach is related to certain ideas of defeasible reasoning, such as Pollock's notion of undercutting of defeasible rules [16], and we are currently exploring these connections.

Our initial formalization has assumed that both agents establish a common set of assumptions, whose truth neither questions. An obvious extension is to combine the AF with an uncertainty formalism expressing degrees of belief in these assumptions. Another area of exploration is to extend the TL formalism to permit expression by agents of their arguments for and against particular inference rules. Such a logic of argumentation [7, 12] would enable the two agents to express their reasons for their assignment of acceptability labels, which TL does not permit, and thus provide opportunity for richer negotiation between the two agents.

Acknowledgments

This work was partially funded by the UK EPSRC under grant GR/L84117 and a studentship. We are also grateful for comments from Trevor Bench-Capon, Mark

Colyvan, Susan Haack, Grant Malcolm, Vladimir Rybakov & Bart Verheij, and from the anonymous reviewers.

References

- [1] L. Amgoud, S. Parsons, and L. Perrussel. An argumentation framework based on contextual preferences. In *Submission*, 2000.
- [2] G. Birkhoff and J. von Neumann. The logic of quantum mechanics. *Annals of Mathematics*, 37:823–843, 1936.
- [3] D. S. Bridges. Can Constructive Mathematics be applied in physics? *Journal of Philosophical Logic*, 28:439–453, 1999.
- [4] L. Carroll. What the tortoise said to Achilles. *Mind n.s.*, 4 (14):278–280, 1895.
- [5] D. E. Cooper. Alternative logic in “primitive thought”. *Man n.s.*, 10:238–256, 1975.
- [6] D. R. Cox and D. V. Hinkley. *Theoretical Statistics*. Chapman and Hall, London, UK, 1974.
- [7] J. Fox and S. Parsons. Arguing about beliefs and actions. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms*, pages 266–302. Springer Verlag (LNAI 1455), Berlin, Germany, 1998.
- [8] J. Goguen. An introduction to algebraic semiotics, with application to user interface design. In C. L. Nehaniv, editor, *Computation for Metaphors, Analogy, and Agents*, pages 242–291. Springer Verlag (LNAI 1562), Berlin, Germany, 1999.
- [9] S. Haack. The justification of deduction. *Mind*, 85:112–119, 1976.
- [10] S. Haack. *Deviant Logic, Fuzzy Logic: Beyond the Formalism*. University of Chicago Press, Chicago, IL, USA, 1996.
- [11] A. Heathcote. Abductive inferences and invalidity. *Theoria*, 61(3):231–260, 1995.
- [12] P. Krause, S. Ambler, M. Elvang-Gøransson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11 (1):113–131, 1995.

- [13] P. McBurney and S. Parsons. Truth or consequences: using argumentation to reason about risk. *Symposium on Practical Reasoning, British Psychological Society, London, UK*, 1999.
- [14] P. McBurney and S. Parsons. Risk Agoras: using dialectical argumentation to debate risk. *Risk Management*, 2(2):17–27, 2000.
- [15] 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.
- [16] J. L. Pollock. *Cognitive Carpentry: A Blueprint for How to Build a Person*. The MIT Press, Cambridge, MA, USA, 1995.
- [17] W. V. O. Quine. Two dogmas of empiricism. In *From a Logical Point of View*, pages 20–46. Harvard University Press, Cambridge, MA, USA, 1980.
- [18] D. Raven. The enculturation of logical practice. *Configurations*, 3:381–425, 1996.
- [19] K. J. Rothman and S. Greenland. *Modern Epidemiology*. Lippincott-Raven, Philadelphia, PA, USA, second edition, 1998.
- [20] V. V. Rybakov. *Admissibility of Logical Inference Rules*. Elsevier, Amsterdam, The Netherlands, 1997.
- [21] A. S. Troelstra and D. van Dalen. *Constructivism in Mathematics: An Introduction*. North-Holland, Amsterdam, The Netherlands, 1988.