2005

Dialogues About the Burden of Proof

Henry Prakken

Chris Reed

Douglas Walton
University of Windsor

Dialogues about the burden of proof *

Henry Prakken
Institute of Information and Computing Sciences, Utrecht University
Faculty of Law, University of Groningen
The Netherlands

Chris Reed
Department of Applied Computing, University of Dundee
Dundee, Scotland

Douglas Walton
Department of Philosophy, University of Winnipeg
Winnipeg, Canada

ABSTRACT

This paper analyses the phenomenon of a shift of the burden of proof in legal persuasion dialogues. Some sample dialogues are analysed of types of situations where such a shift may occur, viz. reasoning with defeasible rules, reasoning with argumentation schemes and reasoning with mere presumptions. It is argued that whether a shift in the burden of proof occurs can itself become the subject of dispute and it is shown how a dialogue game protocol for persuasion can be extended to let it regulate persuasion dialogues about the burden of proof. It is also shown that dialogues about the burden of proof are often implicitly about the precise form of the rules used in an argument.

1. INTRODUCTION

One subfield of AI & Law is the formal study of dialogue systems for argumentation [5, 9, 8, 2, 13, 3]. Formal systems have been designed that regulate the exchange of arguments and counterarguments and expression of propositional attitudes towards their premises and conclusions (such as claiming, disputing, conceding or retracting them). This field is a further development of so-called 'formal dialectics', a field of philosophical logic and argumentation theory [10, 22], which aims to formalise principles of coherent dialogue. A main assumption of formal dialectics is that coherence depends on the goal of a dialogue and that different types of dialogue have different goals. AI & Law has been especially concerned with so-called 'persuasion dialogue', which aims at fair and effective resolution of a conflict of opinion.

This paper[1] studies the role of burden of proof in legal persuasion dialogues. In other research, the burden of proof is usually hardwired into the protocol. For example, [8] uses the following rule: "whoever advances a standpoint is obliged to defend it if asked to do so". This rule is also implicit in the dialogue systems of [10, 5, 2] and [22]. However, in the law the allocation of the burden of proof can become a legal issue and therefore a dialogue may shift into a meta-level dialogue on who has the burden of proof (see also [7]). To account for this phenomenon, we propose a formal dialogue system for persuasion dialogues in which two opposing parties can argue about who has the burden of proof for a certain claim. In [13] a dialogue game was proposed for three-party disputes where a referee has the authority to distribute the burden of proof for specific assertions over the adversaries. However, in that game allocations of the burden of proof could not be disputed. To our knowledge, the present paper is the first to address the modelling of two-party persuasion dialogues about the burden of proof.

It is not our aim to model actual legal procedures but to formulate rational principles of coherent dialogue about burden of proof. The resulting theory can be used in several ways, for instance, for rationally reconstructing legal disputes, for assessing the quality of legal procedures or for designing interaction protocols for intelligent software agents.

This paper is organised as follows. We first introduce a distinction between global and local burden of proof. Then we semi-formally analyse three types of cases which give rise to disputes about the burden of proof, viz. the use of defeasible rules, the use of argumentation schemes and the use of legal presumptions. We then propose a formal dialogue game for dialogues in which the burden of proof can be argued about and we illustrate it with some examples.

2. GLOBAL AND LOCAL BURDEN OF PROOF

Burden of proof is important at the global level of a dialogue as well as at the local level. In persuasion dialogues there are two participants, called the *proponent* and the *opponent*. There are two subcases. In the *dispute*, the proponent has proposition A as her designated thesis. Her goal is to prove A, while the opponent's goal is to prove the opposite of A. Thus each has a burden of proof. In the *dissent*, the proponent's goal is to prove A, while the goal of the opponent is merely to show that the proponent's attempt is not successful. Thus in a dissent, the proponent has a burden of proof, but the opponent does not. We think that legal disputes usually are of the dissent type and therefore we will in this paper exclusively focus on this type of persuasion dialogue. In a dissent, proponent will have what is called in law an ultimate *probandum*, which constitutes proponent's

*This research was partially supported by the EU under IST-FP6-002307 (ASPIC).

[1]This paper is an extended and revised version of [15].

ICAIL'05 June 6-11, Bologna, Italy

global burden of proof and which pertains to a participant's ultimate goal in a dialogue.

While the global burden of proof is fixed during a dialogue, at the local level the burden of proof may change. Moves in which a proposition is asserted (either as a claim or as a premise of an argument) usually carry a burden with them to defend the assertion or else retract it when challenged. However, in exceptional cases the burden shifts to the other party to provide evidence that the proposition does not hold. This happens, for instance, in legal disputes when a legal presumption is invoked as a premise of an argument put forward to meet the global burden of proof. Then the local burden of proof with respect to the presumption shifts to the other party. For example, if in a civil case plaintiff provides evidence in the form of an affidavit, then according to Dutch civil procedure the content of the affidavit is presumed true unless the defendant proves it is false. So the local burden of proof with respect to the claim that the content of the affidavit is false is on the defendant. The same happens with legally recognised exceptions to legal rules. For instance, if the plaintiff in a civil case claims that a contract exists between him and defendant, plaintiff has the burden of proving that an offer was made and accepted. Once plaintiff has proven this, the local burden of proof shifts to the defendant to prove, for instance, that the offeree was insane when accepting the offer.

3. TYPES OF SITUATIONS WITH SHIFTS IN THE BURDEN OF PROOF

We next discuss three types of cases in which shifts of the local burden of proof may arise, viz. the use of defeasible rules, of argumentation schemes and of mere presumptions.

3.1 Dialogues with defeasible rules

It is commonly accepted in AI & Law that the rules invoked by legal arguments are often defeasible. This both holds for 'official' legal rules, such as statutory rules, and for 'unofficial' rules, such as rules on subsuming a case under a legal concept (interpretation rules) or on when a certain piece of evidence proves a certain claim (evidential rules). With statutory rules it is usually clear when a shift in the burden of proof occurs (although even here disputes about the burden of proof may arise). However, with interpretation and evidential rules it is often less clear what is a normal condition and what is an exception, especially since these rules are often left implicit in legal arguments, so that their precise formulation is not always apparent.

In reasoning about evidence this situation especially arises with the empirical generalisations needed to 'glue' evidence and probandum together. Consider again our contract example and assume that plaintiff proves the existence of an offer and acceptance with two witness testimonies. Then he has in mind the following generalisation

(1) If two witnesses say that P then usually P is the case

Suppose now there is reason to think the witnesses may have conferred. Defendant could then respond in two ways. The first is by saying

> "But the witnesses conferred, so your evidence does not prove your claim".

Arguably, defendant thus agrees that two witnesses usually speak the truth but argues that there is an exception to this rule in case the witnesses conferred, namely:

(2) If two witnesses say that P and they conferred, then it is not so that usually P is the case

However, an alternative attack of defendant is to say

> "but how do you know that the witnesses did not confer?".

Thus defendant question reveals that she thinks plaintiff's generalisation should have an additional condition 'and the witnesses did not confer':

(1') If two witnesses say that P and they did not confer, then usually P is the case

As for the burden of proof the crucial difference between (1) and (1') is that if (1) is the correct evidential rule then defendant must prove that the witnesses conferred while if (1') is the correct rule then plaintiff must prove that the witnesses did not confer. In [3] defendant's second reply was identified as one of two ways ("refining" and "unpacking") in which in dialogues about evidence a generalisation can be gradually reshaped in a process of critical examination, so that a higher-quality theory of a legal case may result. Here we are instead concerned with how disputes about the precise form of generalisations affect issues of burden of proof.

An example with an interpretation rule is a Dutch Supreme Court decision in a labour dispute (HR 19 September 1980, NJ 1981, 131). A music band called *Los Gatos* was hired to work at a cruise ship of the *Holland-America Line* (HAL). At some point the manager told the band to perform for the crew while the ship was waiting for repair in a harbour without passengers. The band refused to play, after which they were immediately dismissed. According to Dutch law such a dismissal is valid if and only if there was a "pressing ground" for dismissal. One such pressing ground is when the employee persistently refuses to obey reasonable orders of the employer (Section 1639p,10 Dutch Civil Code). Los Gatos sued HAL on the ground that this pressing ground would not apply to their case. Their main argument was that the HAL managers had not wanted to listen to to the reason why the band had refused to play. This fact was not disputed. What was disputed is how much had to be proven by Los Gatos to claim that in their case their refusal to obey the orders of HAL was not a pressing ground for dismissal. In particular, the dispute revolved around the issue whether Los Gatos had to prove that they had a good reason to refuse to play or that HAL had to prove the opposite. The Supreme Court decided that HAL had the burden of proof since its managers had made it impossible for Los Gatos to explain their reasons for not wanting to play. Arguably, the underlying dispute was whether the interpretation rule is

(3) If employees were not heard then refusal of work is not a pressing ground for dismissal.

(with an exception for when the employees had no good reason for their refusal), or

(3') If employees were not heard and they had a good reason to refuse work, then refusal of work is not a pressing ground for dismissal.

It is important to note that, although the dispute can be reconstructed in terms of the underlying (implicit) generalisations, the dispute as it took place was about the burden of proof of certain specific facts and did not mention any general interpretation or evidence rule.[2]

3.2 Dialogues with argumentation schemes

Arguments are often constructed according to argumentation schemes, which are stereotypical patterns of arguments. Such schemes are a crucial concept of argumentation theory and they have recently been studied in AI & law [6, 21, 17, 4]. As studied by [19], they technically have the form of an inference rule. However, they are not based on the meaning of logical operators, but on epistemological principles or principles of practical reasoning. Argumentation schemes come with a set of critical questions which, when asked, have to be answered before the scheme can be applied.

For present purposes the main issue is how asking a critical question influences the burden of proof. Suppose an argument instantiating some scheme has been put forward and a critical question matching the scheme for that argument has been asked. Does merely asking the question make the argument default so that the other party has to answer the question positively, or is the burden on the questioner to provide evidence that the answer is negative? We take the view that the answer to this question depends on domain-specific issues and can become the subject of debate.

To provide a focus for the investigation, we consider the argumentation scheme from expert opinion. As proposed in [20, p. 258], it takes the following form (where A is a proposition, E is an expert, and D is a domain of knowledge).

Scheme for Argument from Expert Opinion

E is an expert in domain D
E asserts that A is known to be true
A is within D
Therefore, A may plausibly be taken to be true

The premises in the scheme together warrant drawing a reasonable inference to the conclusion. Argument from expert opinion is, however, a defeasible form of argument that holds on a presumptive basis, subject to the asking of appropriate critical questions by the other party in a dialogue. The opponent can ask any of the following six critical questions:.

Critical Questions Matching Argument from Expert Opinion

1. *Expertise Question*: How credible (knowledgeable) is E as an expert source?
2. *Field Question*: Is E an expert in the field that A is in?
3. *Opinion Question*: What did E assert that implies A?
4. *Trustworthiness Question*: Is E personally reliable as a source, e.g. is E biased?
5. *Consistency Question*: Is A consistent with what other experts assert?
6. *Backup Evidence Question*: Is A's assertion based on evidence?

[2]At least not as far as the published Supreme-Court decision reveals the dispute.

A question of current interest is whether these critical questions should be seen as implicit premises of an argument from expert opinion or as implicit exceptions to the argument. Let us look at them, one at a time. 1: when you put forward an appeal to expert opinion, you assume, as part of the argument, that the source is credible, or has knowledge in some field. 2: you assume that the expert is an expert in the field of the claim made. 3: you assume that the expert said something from which the claim can be extracted by inference or direct quoting. 6: you assume that the expert's assertion was based on some evidence within the field of his/her expertise. The argument does not make much sense without these assumptions being part of it. 4 and 5 seem to be a little different. If the expert is claimed to be biased, or to be dishonest, then if there is evidence for such claims, that attacks the argument. But to mount such an attack, it looks like the critic should have to produce some fairly substantial evidence. If the claim can be shown not to be consistent with what other experts in the same field say, then that is an argument against the claim. But that needs to be shown by telling us what the other experts have in fact said, and showing how these statements conflict with what our expert said.

So critical questions 4 and 5 seem to have a positive burden of proof attached, while the remaining critical questions do not. Once asked, the latter questions must be given an appropriate answer or the original argument is refuted. Asking 4 and 5 is a harder task, if you want the question to get the original argument to default. Merely asking the questions is not enough; evidence has to be provided that, if accepted by the fact finder, gives rise to a counterargument that defeats the argument from expert opinion. Thus there are two possible roles for critical questions. Questions like 1,2,3 and 6 serve to locate missing premises that can be questioned. Questions like 4 and 5, by contrast, serve to seek out points of possible attack in an argument, which can be followed up or not, depending on what is available as evidence in the case. In conclusion, some critical questions seem to shift a burden of proof, some do not.

However, a further issue arises: is the fact that a particular critical question shifts a burden of proof context-dependent or not? There is a Gricean default, perhaps (e.g. Expert Opinion - usually, there is presumption for the honesty of the expert), but it can be overridden by particular circumstances. Accordingly, the dialogue protocol should prevent the infinite regress past 'Why A?; why not A?', by recognising that the argument has shifted to an embedded persuasion dialogue on who has the burden of proof. Individual agents will have defaults covering application of burden of proof rules. Discrepancies in these rules, or unilateral belief in exceptions to agreed-upon, common rules, could both give rise to embedded burden of proof meta-dialogues.

Let us illustrate these points with a small schematic dialogue, between a proponent P and an opponent O.

P_1:	C since E says that C	(Argument from Expert Opinion)
O_2:	Is E unbiased?	(critical question $\neg B$?)
P_3:	You think he is biased?	(B?)
O_4:	Er, no.	

At P_1, P states the main argument for his claim, using the argumentation scheme from Expert Opinion (leaving two of its premises implicit). With O_2, O implicitly takes on

a propositional commitment to the effect that P has the burden of proof with respect to B. Thus O regards 'E is unbiased' as a third implicit premise of the scheme from Expert Opinion. At the next step, P implicitly takes on a commitment that it is O that has the burden of proving $\neg B$, thus rejecting O's interpretation of the scheme from Expert Opinion. The locution O_4 simply concedes and thus stops any opportunity for a burden of proof dialogue.

An alternative completion is $O_{4'}$:

$O_{4'}$: Yes, because E has an investment in C

So again, there is no disagreement about the burden of proof, but this time, O substantiates his challenge. Note that in both dialogues there is an separate, implicit dialogue being conducted concurrently about the burden of proof.

But one further possible completion is $O_{4''}$:

$O_{4''}$: Isn't it for you to show that he's not?

This is a move to an explicit dialogue about who has the burden of proof with respect to B.

The following pictures illustrate how these dialogues dynamically construct and reshape a dialectical theory of the case and thereby illustrate how the (implicit) burden of proof dialogue is related to an (also implicit) dialogue about the precise form of the argument scheme from Expert Opinion. The first two theories both consist of a single argument, displayed as inference trees with the conclusion at the top and where each vertical link represents an application of an inference rule. The final theory also contains the beginning of a counterargument; the (undercutting) defeat relation between the two arguments is displayed with a horizontal link between their conclusions. Figure 1 displays the theory after move P_1. With O_2, opponent tries to reshape the theory by arguing that the argument has a second premise (Figure 2). With move P_3, proponent disagrees with opponent's reshaping and instead regards 'E is biased' as an exception which, if proven by opponent, defeats his original argument. Together, these pictures illustrate that for argumentation schemes the same holds as for defeasible rules: when their use leads to a dispute about the burden of proof, this dispute can be interpreted as a dispute about the precise form of the rule or scheme. Accordingly, we will in the remainder of the paper treat both categories in the same way.

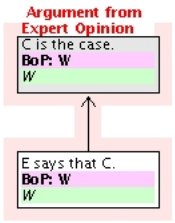

Figure 1: The theory after P_1.

3.3 Dialogues with mere presumptions

Sometimes issues of burden of proof do not concern a defeasible rule or scheme but a particular statement. The issue then is whether that statement must be presumed true or false. A recent American example is the case Weast v. Schaffer, 41 IDELR 176 (4th Cir. 2004) where parents of a special education student sued the school district to seek reimbursement for private school tuition on the grounds that the individualised education programme (IEP) for their son provided by the district was inappropriate. Even though the parents explicitly stated that the IEP was inappropriate, they claimed that the district had the burden of proving that it was inappropriate. A district court agreed with the parents but in appeal the majority of the 4th US circuit Court of Appeals held that the parents had the burden of proving that the IEP was inappropriate, since deciding otherwise would violate the policy of the relevant law to rely on the professional expertise of the local educators. The dissent instead argued that the burden should be on the district since they were obliged to provide an education programme to students and they had an inherent advantage over the parents in assessing the feasibility and likely benefit of alternative educational arrangements. This example illustrates that sometimes a party explicitly makes a claim but the burden of proof can still be on the party who disputes the claim (even though in this case it was ultimately decided otherwise).

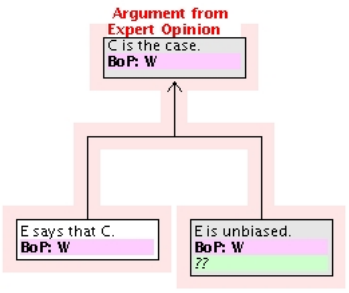

Figure 2: The theory after O_2.

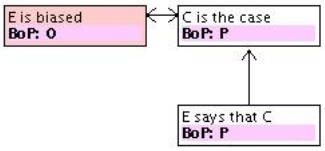

Figure 3: The theory after P_3.

4. A PROTOCOL FOR BURDEN-OF-PROOF DIALOGUES

In this section we formalise a protocol for the embedding of burden-of-proof (BoP) dialogues in 'conventional' persuasion dialogues of the dissent type. The protocol will be formulated as a formal dialogue game, as an instance of the framework of [12, 14]. Dialogue games have a *topic language* L_t with a logic $\mathcal{L}$, and a *communication language* L_c with a *protocol* P. Dialogue games also have *commitment rules*, which specify the effects of an utterance from L_c on the propositional commitments of the dialogue participants. In persuasion dialogues of the dissent type a *proponent* and an

Table 1: Speech acts and replies in L_c (1).

Acts	Attacks	Surrenders
claim p	*why p*	*concede p*
p since Q	*why* $q_i (q_i \in Q)$	*concede* q_i $(q_i \in Q)$
	q since Q' (q *since* Q' defeats p *since* Q)	*concede p*
why p	*p since Q*, *why* $-p$	*retract p*
concede p		
retract p		

opponent argue about a single dialogue topic $t \in L_t$. Proponent aims to make opponent concede t while opponent aims to make proponent give up t.

We present the protocol in two stages. First we present a basic persuasion protocol in which the burden of proof is hardwired (Section 4.2). This is in fact a revised and simplified version of the two-player protocols presented in [12, 13]. Then we extend the basic protocol to let it regulate dialogues about the burden of proof (Section 4.3). But first we briefly describe the underlying logic of both protocols.

4.1 The underlying logic

Since our main concern is not inference but dialogue, we only make some minimal assumptions on the topic language and its logic. The topic language L_t is assumed to be a logical language closed under classical negation $\neg$. As for notation, $-p$ denotes the *converse* of p: the converse of p is $\neg p$ and the converse of $\neg p$ is p (where $\neg$ is classical negation). As in [4, 3], the logic $\mathcal{L}$ is assumed to be an argument-based logic in the style of [11, 18], in which arguments can be constructed by chaining inference rules into trees. For any argument A, the sets $prem(A)$ and $conc(A)$ denote the premises, respectively conclusion of A. Inference rules are of two kinds, the *strict* rules S, consisting of all deductively valid inference rules and the *defeasible* rules D. Arguments can be defeated in three ways: with an argument for the negation of a premise (*premise defeat*), with an argument for a contradictory conclusion (*rebutting defeat*) or with an argument that a defeasible inference rule cannot be applied in a certain case (*undercutting defeat*). These notions of defeat can be refined with notions of argument strength, but to keep things simple we disregard strength below.

4.2 A basic protocol for persuasion dialogue

We now present a dialogue game protocol for persuasion dialogues with hardwired burden of proof. The game is based on the following ideas. Each dialogue move except the initial one replies to one earlier move of the other party in the dialogue (its *target*). Thus a dialogue can be regarded in two ways: as a sequence (reflecting the order in which the moves are made) and as a tree (reflecting the reply relations between the moves). Each replying move is either an *attacker* or a *surrender*. For instance, a *claim p* move can be attacked with a *why p* move and surrendered with a *concede p* move. And a *why p* move can be attacked with a *p since Q* move and surrendered with a *retract p* move. The **communication language** L_c is specified in Table 1. In this table, p, q, q_i and the elements from Q and Q' are from L_t. Furthermore, *since* moves express an argument A constructible in $\mathcal{L}$. Finally, defeat relations between arguments are determined according to $\mathcal{L}$.

The protocol for L_c is defined in terms of the notion of a **dialogue**, which in turn refers to the notion of a **move**:

Definition 4.1.

- *The set M of moves is defined as $\mathbb{N} \times \{P, O\} \times L_c \times \mathbb{N}$, where the four elements of a move m are denoted by, respectively:*
  - *$id(m)$, the identifier of the move,*
  - *$pl(m)$, the player of the move,*
  - *$s(m)$, the content of the move,*
  - *$t(m)$, the target of the move.*
- *The set of dialogues, denoted by $M^{\leq\infty}$, is the set of all sequences $m_1, \ldots, m_i, \ldots$ from M such that*
  - *each i^{th} element in the sequence has identifier i,*
  - *$t(m_1) = 0$;*
  - *for all $i > 1$ it holds that $t(m_i) = j$ for some m_j preceding m_i in the sequence.*

  The set of finite dialogues, denoted by $M^{<\infty}$, is the set of all finite sequences that satisfy these conditions. For any dialogue $d = m_1, \ldots, m_n, \ldots$, the sequence $m_1, \ldots, m_i$ is denoted by d_i, where d_0 denotes the empty dialogue.

When $t(m) = id(m')$ we say that m replies to m' in d and also that m' is the target of m in d. We will sometimes slightly abuse notation and let $t(m)$ denote a move instead of just its identifier. When $s(m)$ is an attacking (surrendering) reply to $s(m')$ we will also say that m is an attacking (surrendering) reply to m'.

The semantics for L_c is defined in axiomatic style as a set of precondition-postcondition rules. In fact, as we will see below, the only precondition for each move is that it is legal according to the **protocol**.[3]

Definition 4.2. *A protocol on M is a function P with domain a nonempty subset of $M^{<\infty}$ taking subsets of M as values. The elements of $dom(P)$ (the domain of P) are called the legal finite dialogues. The elements of $P(d)$ are called the moves allowed after d. If d is a legal dialogue and $P(d) = \emptyset$, then d is said to be a terminated dialogue.*

Furthermore, for all moves m it holds that $m \in P(d)$ if and only if m satisfies all of the following conditions:

- *R_1: $pl(m) \in T(d)$;*[4]
- *R_2: If $d \neq d_0$ and $m \neq m_1$, then $s(m)$ is a reply to $s(t(m))$ according to L_c;*
- *R_3: If m replies to m', then $pl(m) \neq pl(m')$;*
- *R_4: If there is an m' in d such that $t(m) = t(m')$ then $s(m) \neq s(m')$.*
- *R_5: If $d = d_0$, then $s(m)$ is of the form claim p.*
- *R_6: If m is a replying move, then m is relevant in d.*

[3]The first part of this definition is adapted from [1].
[4]$T(d)$ denotes the player whose turn it is to move in d.

R_1 says that the player of a move must be to move. R_2-R_4 formalise the idea of a dialogue as a move-reply structure that allows for alternative replies. R_5 says that each dialogue begins with a claim; the initial claim is the *topic* of the dialogue. Finally, R_6 says that each replying move must be relevant in a structural sense to be explained below.

These rules are the minimal conditions for legality of a move. Further legality conditions could be defined, such as for respecting one's commitments (see e.g. [22]) or for preventing circular dialogues (see e.g. [10]). However, to keep the protocol simple, we will below only add rules that are essential for regulating the embedding of BoP subdialogues in a persuasion dialogue.

Relevance is defined in terms of the *dialogical status* of a move, which in turn is recursively defined in terms of the nature of its replies. A move is *in* iff it is surrendered or else if all its attacking replies are out. (This implies that a move without replies is in). And a move is *out* if it has an attacking reply that is in. Furthermore, a move is *definitely in (out)* if it is in (out) and its status cannot change any more. Finally, a move is *surrendered* if it is a *p since Q* move and it has a reply *concede p* or else if it has any surrendering reply. With this concept of dialogical status a structural notion of relevance can be defined. A move is *relevant* if it replies to a relevant target. And a move is a *relevant target* if making it out changes the dialogical status of the initial move of the dialogue. Together with Definition 4.2 these definitions imply that a move is a relevant target for proponent (opponent) if making it out makes the initial move in (out). Accordingly we say that *P currently wins d* if m_1 is in and *O currently wins* if m_1 is out.

Figure 4 (with only attacking replies) illustrates the notion of relevance. A move labelled + is in and a move labelled − is out. P_5 is not a relevant target for O: although making P_5 out makes O_4 in, P_3 was already out because of O_6 and therefore O_2 stays out because of P_7, so that P_1 stays in. However, P_7 is a relevant target for O: making P_7 out makes O_2 in since its only attacking reply is now out; then P_1 is out since it now has an attacking reply that is in.

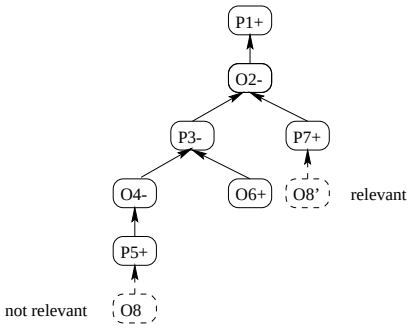

Figure 4: Dialogical status of moves

The requirement that each move be relevant allows the players maximal freedom on issues such as backtracking and postponing replies while yet ensuring a strong focus of a dispute. It should be noted that our notion of relevance is structural, defined in terms of the reply structure of a dialogue. Our notion says nothing about substantial relevance or arguments. For instance, an argument "The suspect is guilty since the moon is made of green cheese" may very well be structurally relevant in our sense. Its substantial irrelevance should be argued for in the dialogue with an undercutting counterargument or a *why* attack.

The requirement of relevance comes with a **turntaking rule** T that the turn switches as soon as a player has changed the dialogical status of the initial move (below pl is a variable ranging over $\{P, O\}$ and $\overline{p}$ denotes O if $pl = P$ and P if $pl = O$). Formally, T is a function

- $T : M^{<\infty} \longrightarrow \{P, O\}$

such that $T(d_0) = P$ and if $d \neq d_0$ then $T(d) = pl$ iff $\overline{pl}$ currently wins d.

The rationale of this rule is that as soon as a player has changed the dialogical status of m_1, he has no relevant moves any more so to avoid premature termination the turn should shift to the other party.

The **commitment rules** are defined as a function of the following type:

- C: $\{P, O\} \times M^{<\infty} \longrightarrow \mathcal{P}(L_t)$.

$C_p(d)$ denotes the commitments of player p in the dialogue d. The following commitment rules for L_c seem uncontroversial and can be found throughout the literature. (Below s denotes the speaker of the move; effects on the other parties' commitments are only specified when a change is effected; finally, d, m stands for the dialogue starting with dialogue d and continuing with move m.)

- If $s(m) =$ *claim* p then $C_s(d, m) = C_s(d) \cup \{p\}$
- If $s(m) =$ *why* p then $C_s(d, m) = C_s(d)$
- If $s(m) =$ *concede* p then $C_s(d, m) = C_s(d) \cup \{p\}$
- If $s(m) =$ *retract* p then $C_s(d, m) = C_s(d)/\{p\}$
- If $s(m) = p$ *since* Q then $C_s(d, m) = C_s(d) \cup Q \cup \{p\}$

The **axiomatic semantics** of the system then is as follows: for each move m and dialogue d:

precondition: $m \in P(d)$
postcondition: as specified by $C_s(d, m)$.

To give a feel for how dialogues evolve in this system, we now list a few properties of the system (cf. [14]). Firstly, a *turn* of a player always consists of zero or more surrenders followed by a single attack. Further, the turn shifts to the opponent if the initial move is made in while it shifts to the proponent if the initial move is made out. A dialogue terminates only if the status of the initial move is *against* the player to move (out for the proponent and in for the opponent). So if a dialogue terminates when player pl is to move, pl can be said to have *lost* the dialogue. Moreover, it can be shown that a dialogue terminates if and only if either proponent has surrendered to opponent's first move by retracting the dialogue topic or opponent has surrendered to proponent's first move by conceding the dialogue topic.

4.3 Extending the protocol for BoP dialogues

We now extend our protocol to let it regulate dialogues about the burden of proof. First we make some further assumptions about the topic language L_t. It is now defined as $L_0 \cup L_1$, where L_0 is some propositional or first-order language and L_1 consists of all formulas of the form

Table 2: Speech acts and replies in L_c (2).

Acts	Attacks	Surrenders
claim p	*why* p	*concede* p
p *since*$_A$ Q	*why* $q_i (q_i \in Q \cup A)$ q *since*$_A$ Q' (q *since*^c_A Q' defeats p *since*$^c_{A'}$ Q)	*concede* q_i $(q_i \in Q \cup A)$ *concede* p
why p	p *since*$_A$ Q, *why* $-p$, $BoP(-p)$ *since* R	*retract* p
concede p		
retract p		

$\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ such that $\varphi_i, \varphi \in L_0$. Formulas $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ informally read as "if $\varphi_1, \ldots, \varphi_n$ then normally φ". There is just one defeasible inference rule, viz. modus ponens for $\Rightarrow$. Furthermore, we make the notion of undercutting more precise as follows. For each $\varphi \in L_1$ we define a function $c(\varphi)$, returning a set $S \in L_0$ (the critical questions of φ). The set $ass(B)$ of *assumptions* of an argument B then consists of all critical questions of any L_1-premise of B. Below we will denote the assumptions of an argument with a subscript A of *since* (dropping the subscript when there is no danger for confusion). An undercutter of B then is any argument that concludes to the negation of an assumption of B.

We also want to account for the fact that arguments often leave their general rules implicit. For present purposes this is very important since, as we have seen in Section 3, burden-of-proof dialogues are often about the precise form of the rules implicit in an argument. In some other systems (e.g. [22]) an incomplete argument commits the speaker to a material implication *premises* $\rightarrow$ *conclusion*. However, for defeasible arguments this is not appropriate while, moreover, usually it is clear from the dialogue context what is left implicit. To respect these observations, we assume that each dialogue has a *context* $\mathcal{C} \subseteq L_t$. We next refine the definition of *since* moves in L_c as follows. For each p *since* Q move it now holds that

1. p *since* Q is constructible in $\mathcal{L}$; or else
2. there is a $r = Q \Rightarrow p$ in $\mathcal{C}$ such that p *since* $Q \cup \{r\}$ is constructible in $\mathcal{L}$.

Note that case (2) implies that the other premises are all from L_0. We next refine the notions $ass(A)$ and $prem(A)$ for arguments that satisfy (2), by including $c(r)$ in $ass(A)$ and r in $prem(A)$. Thus it becomes possible to challenge or concede implicit premises of an argument. Also, we refine the commitment rules by including r in $C_s(d, m)$ if m moves an argument that satisfies case (2). Finally, whenever we need to refer to the completed version of an argument, we denote it with *since*^{c} (see Table 2).

Next we extend L_c with two new replies to *why* p moves, viz. *why* $-p$ and $BoP(-p)$ *since* R, and with a *why* p reply to a *since* move for each assumption p of the moved argument. Table 2 displays the communication language resulting from the changes made so far. In this table, p, q, q_i are from L_0 and the elements from Q and Q' are from $L_0 \cup L_1$. Note that formulas of the form $BoP(p)$ are propositional atoms with some syntactic sugar so that a $BoP(p)$ *since* Q move can be replied to as any other *since* move.

As for the new protocol rules, they include rules $(R_1 - R_6)$ above. What remains is to regulate the embedding of BoP dialogues in other dialogues (possibly also BoP dialogues). First we add three obvious rules.

- R_7: if $s(m) = BoP(p)$ *since* Q, then $\overline{p}$ is not the dialogue topic.
- R_8: If m is a *why* reply to a *why* move m', then $t(m')$ is not a *why* move.
- R_9: If m is a *why* reply to a *why* move m', then m' has not yet been replied to.

R_7 makes the players respect that the global burden of proof remains fixed during a dialogue. R_8 avoids an infinite regress "why?, why not?, why?, ...". Finally, R_9 avoids that a player first meets a *why* attack with an argument and then challenges that he has the burden of proof.

Next, in order to ensure orderly and well-structured dialogues, we want that each BoP dialogue is completed before a participant may jump back to a surrounding dialogue. This can be achieved by assigning to each move a dialogue level and requiring that a target of a move is of the highest possible level. (Below we say that a level l is higher than a level l' if $l > l'$; the symbols $<, =$ and $\neq$ are defined as usual for natural numbers). The level of a move in a dialogue $m_1, \ldots, m_n$ is defined as follows:

- $\text{level}(m_1) = 1$
- $\text{level}(m_{i+1}) =$
  - $1 + \text{level}(t(m_{i+1}))$ if m_{i+1} is a $BoP(p)$ *since* Q reply to a *why* $-p$ move;
  - $\text{level}(t(m_{i+1}))$ otherwise.

Now we add to the above protocol conditions $(R_1 - R_9)$ the following final condition:

- R_{10}: There is no $m' \neq m$ that satisfies protocol rules $R_1 - R_9$ and such that $\text{level}(t(m)) < \text{level}(t(m'))$.

Taken together, our protocol rules imply that a jump back to a lower level is legal only if the current BoP dialogue is 'terminated', that is if the move that started the current BoP dialogue has obtained a definitive status. To see this, observe that otherwise there are still legal moves that can change its status and so by R_{10} no reply to a higher-level target is legal. Also, our protocol implies that a jump to a lower level that is not equal to level 1 will always be to the immediately preceding surrounding dialogue.

To illustrate this with a schematic example, suppose that a legal dialogue starts with a sequence $m_1, \ldots, m_3$ at level 1, then jumps to level 2 at $m_4, \ldots, m_9$, then jumps back to level 1 at $m_{10}, \ldots, m_{17}$, then jumps to level 2 at $m_{18}, \ldots, m_{23}$, and jumps higher to level 3 at $m_{24}, \ldots, m_{27}$. Our protocol implies that the BoP subdialogue $m_4, \ldots, m_9$ is 'terminated' at m_9 by making m_4 definitely in or out, so it contains no relevant targets for subsequent moves. This implies that if $m_{24}, \ldots, m_{27}$ 'terminates' at m_{27}, then the only relevant targets for m_{28} that are of level 2 are in $m_{18}, \ldots, m_{23}$. Note, however, that this phenomenon does not hold for moves that jump back to level 1: suppose that $m_{28}, \ldots, m_{32}$ continues

$m_{18}, \ldots, m_{23}$ at level 2 and m_{33} 'terminates' this BoP dialogue; then m_{34} may reply to any level-1-move in the dialogue that is a relevant target, even to m_1 (by conceding or retracting its claim), since the level-1-dialogue is not terminated until the claim of m_1 is conceded or retracted.

5. EXAMPLE DIALOGUES

In this section we illustrate the protocol with the examples of Section 3. The target of a move is given between square brackets. Below each move the dialogical status of some moves of interest is listed. To avoid too much formal detail, the logical form of propositions is sometimes informally paraphrased and we treat $BoP(p)$ and $BoP(-p)$ as contradictory propositions. Also, for ease of readability the relevant player is added to BoP expressions.

5.1 A dialogue with an argumentation scheme

To begin with, we illustrate the use of an argumentation scheme with the schematic example from Section 3.2.

P_1: *claim* C
- P_1 is in

$O_2[P_1]$: *why* C
- P_1 is out, O_2 is in

$P_3[O_2]$: C *since* E says so and E is an expert about C
- O_2 is out, P_1 is in

$O_4[P_3]$: *why* $\neg E$ biased
- P_3 and P_1 are out

$P_5[O_4]$: *why* E biased
- P_1 is in, O_4 is out

$O_6[P_5]$: $BoP(\neg E$ biased, $P)$ *since* only experts proven to be unbiased can be trusted.
- P_5 and P_1 are out, O_4 is in

$P_7[O_6]$: *why* only experts proven to be unbiased can be trusted.
- O_6 is out, P_1 is in

$O_8[P_7]$: *why* $\neg$ only experts proven to be unbiased can be trusted.
- P_7 is out, O_6 is in, P_1 is out

$P_9[O_8]$ $\neg$ only experts proven to be unbiased can be trusted *since* experts may be presumed to be unbiased.
- O_6 is out, P_1 is in

$O_{10}[P_7]$: *retract* only experts proven to be unbiased can be trusted.
- O_6 is definitively out, P_1 is in

$O_{11}[P_5]$: E is biased *since* his previous research was paid by the company he testifies for.
- O_{11} is in, P_5 and P_1 are out

$P_{12}[O_{11}]$: *concede* E is biased.
- O_{11} is definitely in, P_5 and P_1 are out

$P_{13}[O_2]$: *retract* C
- P_1 is definitively out

At this point P remains to move but has no legal moves any more since P_1 is definitively out, so the dialogue terminates with a loss for P. Apparently, the proponent could not find another argument for his initial claim C and therefore retracted it. As for the dialogue levels, at O_6 the dialogue jumps from level 1 to level 2. Consequently, after O_6 no replies to moves before O_6 are allowed until the level-2 dialogue is 'terminated'. This happens at O_{10}, when O retracts the only premise of his BoP argument, thus making O_6 definitively out so that no reply to any level-2 move can change the status of P_1 anymore. So after O_{10} the dialogue jumps back to level 1 so that new replies to moves before O_6 are allowed. This explains why O can reply at O_{11} to P_5.

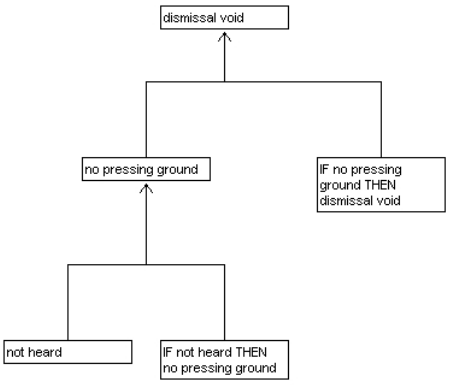

Figure 5: The theory in *Los Gatos* after P_5.

5.2 The Los Gatos case: generalisations

Next we reconstruct the BoP debate from the *Los Gatos* case in our system.

P_1: *claim* dismissal-void
- P_1 is in

$O_2[P_1]$: *why* dismissal-void
- P_1 is out, O_2 is in

$P_3[O_2]$: dismissal-void *since* $\neg$ pressing-ground
- O_2 is out, P_1 is in

$O_4[P_3]$: *why* $\neg$ pressing-ground
- P_3 and P_1 are out

$P_5[O_4]$: $\neg$ pressing-ground *since* $\neg$ heard
- O_4 is out, P_1 is in.

$O_6[P_5]$: *why* good-reason-for-refusal
- P_5 and P_1 are out.

$P_7[O_6]$: *why* $\neg$ good-reason-for-refusal
- P_1 is in, O_6 is out

$O_8[P_7]$: BoP(good-reason-for-refusal, P) *since* plaintiff must prove his main claim.
- P_7 and P_1 are out, O_6 is in

$P_9[O_8]$: $BoP(\neg$ good-reason-for-refusal, $O)$ *since* employer made expressing reasons for refusal impossible.
- O_8 is out, P_1 is in

$O_{10}[P_9]$: *concede* $BoP(\neg$ good-reason-for-refusal, $O)$.
- P_9 is definitely in, O_8 is definitely out, P_1 is in

At this point, the BoP-dialogue that started with O_8 has terminated, so that level-1 moves can be replied-to again. All of P's moves except P_9 are a relevant target for O. Opponent can fulfil his newly incurred burden of proof by attacking P_5 with a $\neg$ good-reason-for-refusal *since* Q move.

We next illustrate how during this dialogue implicitly a dialectical theory of the case is built and reshaped. Figure 5 depicts the theory after move P_5. It consists of a single two-step argument constructed by proponent in his first three moves. Figure 6 depicts the theory as modified by opponent with O_6. It still contains a single two-step argument but the interpretation rule for *no pressing ground* has received an additional condition and the argument has received a

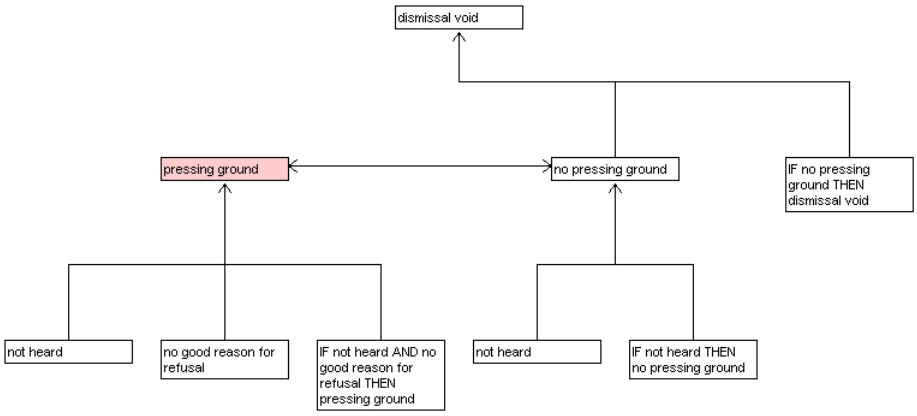

Figure 7: The theory in *Los Gatos* after P_7.

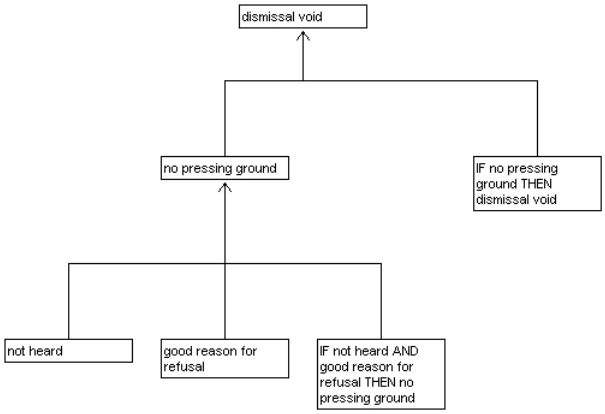

Figure 6: The theory in *Los Gatos* after O_6.

corresponding third premise. Finally, Figure 7 shows how proponent has modified his initial theory in an alternative way. The condition *good reason for refusal* that was added by O_6 is now negated and has become the condition of an exceptional rule for *pressing ground*.

5.3 The school tuition case: mere presumptions

We end with formalising the school tuition example, which illustrates BoP-debates involving mere presumptions. Below we list a dialogue as it could have taken place in this case (the case summary does not contain enough detail to reconstruct an actual dialogue).

P_1: *claim* district should pay
- P_1 is in

$O_2[P_1]$: *why* district should pay
- P_1 is out, O_2 is in

$P_3[O_2]$: district should pay *since* the IEP is inappropriate and if IEP the is inappropriate then it denies free appropriate public education.
- O_2 is out, P_1 is in

$O_4[P_3]$: *why* the IEP is inappropriate
- P_1 is out

$P_5[O_4]$: BoP(the IEP is appropriate, O)
- O_4 is out, P_1 is in

$O_6[P_5]$: *why* BoP(the IEP is appropriate, O)
- P_1 is out

$P_7[O_6]$: BoP(the IEP is appropriate, O) *since* district can better assess IEP and district has obligation to provide education programme.
- O_6 is out, P_1 is in

$O_8[P_7]$: BoP(the IEP is inappropriate, P) *since* deciding otherwise violates the policy of the relevant law to rely on the professional expertise of the local educators.
- P_1 is out

$P_9[O_8]$: *retract* BoP(the IEP is appropriate, O)
- P_5 is definitely out, P_1 remains out.

$P_{10}[O_4]$ the IEP is inappropriate *since* ...

P_9 makes the initial move of the BoP-dialogue definitely out, so that no move in the BoP dialogue is a relevant target; because of this, P can jump back to level 1 and continue the dialogue about his main claim. Note that if opponent had conceded P_5 at O_8, then opponent would have needed to provide an argument for 'the IEP is appropriate' in reply to P_3; such an argument is an example of a premise attack. The notion of a premise attack is crucial in capturing shifts in the burden of proof concerning mere presumptions.

6. CONCLUSION

In this paper we have developed a formal dialogue game for persuasion dialogues in which the burden of proof can become the topic of dispute. We have shown with some examples how our model can handle several kinds of situations, viz. situations arising from disagreement on the precise form of defeasible rules and situations where the issue is whether a proposition or its negation must be presumed true.

It has been noted before that the notion of burden of proof is related to nonmonotonic reasoning (cf. e.g. [16]). This paper has shed more light on this relation. In particular,

the reasoning modelled in this paper conceptually *precedes* nonmonotonic reasoning. That kind of reasoning needs a division into general rules and exceptions as *input* before the nonmonotonic inference machinery can be applied. For example, Verheij [17], when distinguishing various logical roles of critical questions of argumentation schemes, defines these roles in terms of a given distinction between conditions of and exceptions to a scheme. Dialogues about the burden of proof, by contrast, often have such a division as (implicit) *output* since they are often implicitly about whether something is a condition of or an exception to a general rule.

We have seen that an explicit dialogue often has implicit elements. Sometimes these implicit elements are about the allocation of the burden of proof, sometimes about the precise form of a rule used in an argument and sometimes about both. This is a very frequent situation; in reality dialogues rarely make everything that is relevant explicit. Therefore inferring implicit elements from explicit elements of a dialogue is a very important research topic. However, in the present paper we have refrained from including this in our formal model since we believe this is a very intricate issue which deserves detailed study in its own right. As for the implicit elements of BoP dialogues we currently see two ways to formally capture them. One is to define implicit commitments of explicit moves. For instance, a *why* p move could be regarded as committing the speaker to $BoP(p)$. Another is to allow that explicit dialogues have associated implicit dialogues. In this approach a *why* p move could be regarded as also making an implicit *claim* $BoP(p)$ move. We leave the investigation of these two approaches for future research.

7. REFERENCES

[1] J. Barwise and L. Moss. *Vicious Circles*. Number 60 in CSLI Lecture Notes. CSLI Publications, Stanford, CA, 1996.

[2] T.J.M. Bench-Capon, T. Geldard, and P.H. Leng. A method for the computational modelling of dialectical argument with dialogue games. *Artificial Intelligence and Law*, 8:233–254, 2000.

[3] F.J. Bex and H. Prakken. Reinterpreting arguments in dialogue: an application to evidential reasoning. In *Legal Knowledge and Information Systems. JURIX 2004: The Seventeenth Annual Conference*, pages 119–129, Amsterdam etc, 2004. IOS Press.

[4] F.J. Bex, H. Prakken, C. Reed, and D.N. Walton. Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. *Artificial Intelligence and Law*, 12:125–165, 2003.

[5] T.F. Gordon. *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*. Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.

[6] K. Greenwood, T.J.M. Bench-Capon, and McBurney. Towards a computational account of persuasion in law. In *Proceedings of the Ninth International Conference on Artificial Intelligence and Law*, pages 22–31, New York, 2003. ACM Press.

[7] R.E. Leenes. Burden of proof in dialogue games and Dutch civil procedure. In *Proceedings of the Eighth International Conference on Artificial Intelligence and Law*, pages 109–118, New York, 2001. ACM Press.

[8] A.R. Lodder. *DiaLaw. On Legal Justification and Dialogical Models of Argumentation*. Law and Philosophy Library. Kluwer Academic Publishers, Dordrecht/Boston/London, 1999.

[9] R.P. Loui and J. Norman. Rationales and argument moves. *Artificial Intelligence and Law*, 3:159–189, 1995.

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

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

[12] H. Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA'2000)*, number 1919 in Springer Lecture Notes in AI, pages 224–238, Berlin, 2000. Springer Verlag.

[13] H. Prakken. Modelling reasoning about evidence in legal procedure. In *Proceedings of the Eighth International Conference on Artificial Intelligence and Law*, pages 119–128, New York, 2001. ACM Press.

[14] H. Prakken. Coherence and flexibility in two-person dialogue games for argumentation. Technical report, Institute of Information and Computing Sciences, Utrecht University, Utrecht, 2005.

[15] H. Prakken, C. Reed, and D.N. Walton. Argumentation schemes and burden of proof. In *Proceedings of the ECAI-2004 Workshop on Computational Models of Natural Argument*, pages 81–86, 2004.

[16] G. Sartor. Defeasibility in legal reasoning. In Z. Bankowski, I. White, and U. Hahn, editors, *Informatics and the Foundations of Legal Reasoning*, Law and Philosophy Library, pages 119–157. Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.

[17] B. Verheij. Dialectical argumentation with argumentation schemes: an approach to legal logic. *Artificial Intelligence and Law*, 11:167–195, 2003.

[18] G.A.W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.

[19] D.N. Walton. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[20] D.N. Walton. *Appeal to Expert Opinion*. Penn State Press, University Press, 1997.

[21] D.N. Walton. Is there a burden of questioning? *Artificial Intelligence and Law*, 11:1–43, 2003.

[22] D.N. Walton and E.C.W. Krabbe. *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany, NY, 1995.