Proof Burdens and Standards

Thomas F. Gordon and Douglas Walton

1 Introduction

This chapter explains the role of proof burdens and standards in argumentation, illustrates them using legal procedures, and surveys the history of research on computational models of these concepts. It also presents an original computational model which aims to integrate the features of these prior systems.

The 'mainstream' conception of argumentation in the field of artificial intelligence is monological and deductive [6]. Argumentation is viewed as taking place against the background of an inconsistent knowledge base, where the knowledge base is a set of propositions represented in some formal logic. Argumentation in this conception is a method for deducing warranted propositions from an inconsistent knowledge base. Which statements are warranted depends on attack relations among the arguments [10] which can be constructed from the knowledge base.

The notions of *proof standards* and *burden of proof* become relevant only when argumentation is viewed as a dialogical process for making *justified* decisions. The input to the process is an initial claim or issue. The goal of the process is to clarify and decide the issues, and produce a justification of the decision which can withstand a critical evaluation by a particular *audience*. The role of the audience could be played by the respondent or a neutral-third party, depending on the type of dialogue. The output of this process consists of: 1) a set of claims, 2) the decision to accept or reject each claim, 3) a theory of the generalizations of the domain and the facts of the particular case, and 4) a *proof* justifying the decision of each issues, showing how the decision is supported by the theory.

Notice that a theory or knowledge-base is part of the *output* of argumentation dialogues, not, as in the deductive conception, its input. This is because, as has been

Thomas F. Gordon
Fraunhofer FOKUS, Berlin, Germany, e-mail: thomas.gordon@fokus.fraunhofer.de

Douglas Walton
University of Windsor, Windsor, Canada, e-mail: dwalton@uwindsor.ca

repeatedly recognized [32, 34, 21], the generalizations (rules) of some domain and the particular facts of a problem or case are dependent on one another and need to be constructed together, in an iterative process. For exmample, one of the founders of the field of computer science and law, Jon Bing, wrote in 1982:

> Legal reasoning is not primarily deductive, but rather a modeling process of shaping an understanding of the facts, based on evidence, and an interpretation of the legal sources, to a construct a theory for some legal conclusion. [7]

The concept of proof in argumentation is weaker than it is in mathematics. The proof need not demonstrate that a proposition is *necessarily* true, given a set of axioms assumed to be true. Rather, as in law, a proof in argumentation is a structure which demonstrates to a particular audience that a proposition statisfies its applicable proof standard. Since expressive logics are undecidable or intractable, the theory constructued during the dialogue cannot usually serve as a proof. A *burden of proof* is not discharged if the audience must solve a hard problem to construct the proof for themselves from the theory.

There are several kinds of proof burdens. The distinctions between them can only be understood with a deeper analysis of particular argumentation processes. There are many kinds of argumentation processes, each regulated by its own procedural rules, usually called 'protocols' in AI. Walton has developed a typology of dialogue types, classifying persuasion dialogues, negotiation , and deliberation, among other types [38].

For our purpose of illustrating different kinds of proof burdens, it is sufficient to use a simplified description of civil procedure, roughly based on the law of California [35]. A civil case begins by the plaintiff filing a *complaint*, stating a claim against the defendant. The complaint is the first step in the pleadings phase of the case. It contains, in addition to the claim, assertions about the facts of the case which the plaintiff contends are sufficient, if true, to prove the defendant has breached some obligation legally entitling the plaintiff to some remedy or compensation. The defendant then has several options for responding the complaint. For the sake of brevity we will mention just one, filing an *answer* in which the factual allegations are each denied or conceded and asserting additional facts, called an *affirmative defense*, which may be useful for defeating or undercutting later arguments put forward by the plaintiff. The final step in the pleadings phase gives the plaintiff an opportunity to file a *reply* in which he concedes or denies the additional facts alleged by the defendant in his answer. The next phase of the process provides the parties various methods to *discover* evidence, for example by interviewing witnesses under oath, called *taking depositions*. At the *trial*, this evidence is presented to the judge, and possibly a jury, and further evidence is produced by examining and cross-examining witnesses during the trial. At the end of the trial, the evidence is passed on to the *trier-of-fact*, either the judge or the jury, if there is one. If there is a jury, the judge first instructs the jury about the relevant law, since the jury is only responsible for finding the facts. After the jury has completed its deliberations, it reports its *verdict* to the judge, who then enters his *judgment* upon the verdict. The judgment may be appealed by the losing party, but we will end our exposition of legal procedure here.

Our account of legal burdens of proof below is based in part on [31]. The first kind of burden of proof is called the *burden of claiming*. A person who feels he has a right to some legal remedy has the burden of initiating the proceeding by filing a complaint, which must allege facts sufficient to prove the *operative facts* of legal rules entitling him to some remedy. The second type of burden of proof is called the *burden of questioning* or contesting. During pleading, any allegations of fact by either party are implicitly conceded unless they are denied. The third type of burden is called the *burden of production*. It is the burden to discover and bring forward evidence supporting the contested factual allegations in the pleadings. The fourth type of burden of proof is the *burden of persuasion*. In a civil proceeding, this burden becomes operative only at the end of the trial, when the evidence and arguments are put to the jury to decided the factual issues. In a civil proceeding, the plaintiff has the burden of persuasion for all operative facts of his complaint and the defendant has the burden of persuasion for all affirmative defenses, i.e. exceptions. In criminal cases this is different. The prosecution has the burden of persuasion for all facts of the case, whether or not they are the operative facts of the elements of the alleged crime, or defenses, such as self-defense in a murder case. The fifth type of burden is called the *tactical burden of proof*. During the trial, arguments are put forward by both parties, pro and con the various claims at issue. At a finer level of granularity, the argumentation phase can be broken down conceptually into a sequence of stages, where each stage consists of all the arguments which have been put forward by both parties so far in the proceeding. The parties take turns putting forward arguments, by introducing new evidence. The next stage is constructed by adding the arguments put forward during this turn to all the previous arguments. The tactical burden arises from considering whether the arguments of a stage would be sufficient to meet the burden of persuasion with regard to some issue, if hypothetically the trial were to end at the stage and issues where immediately put to the jury. The tactical burden of proof is the only burden of proof which, strictly speaking, can shift back and forth between the parties during the proceeding.

How does the burden of persuasion operate? Essentially the jury has the task of *weighing* the arguments pro and con each proposition at issue. If the pro arguments are not deemed to sufficiently outweigh the con arguments, then the jury must reject the alleged fact by deciding that the alleged fact is not true. Because of the way the burden of persuasion is allocated, this amounts to accepting the default truth value of the proposition at issue.

When do pro arguments 'sufficiently' outweigh con arguments to meet the burden of persuasion? This lead us to our final topic, *proof standards*. The question is how to aggregate or 'accrue' [25] arguments pro and con some claim. In the legal domain, four proof standards for factual issues exist, at least in common law jurisdictions. The *scintilla of evidence* proof standard is met if there is "any evidence at all in a case, even a *scintilla*, tending to support a material issue ..." [8, p. 1207] The *preponderance of evidence* proof standard is met by "evidence which as a whole shows that the fact sought to be proved ... is more credible and convincing to the mind." [8, p. 1064]. The *clear and convincing evidence* proof standard is the "measure or degree of proof which will produce in mind of trier of facts a

firm belief or conviction as to allegations sought to be established; it is intermediate, being more than preponderance, but not to extent of such certainty as is required beyond reasonable . . . " [8, p. 227]. Finally, the *beyond reasonable doubt* standard is the strongest legal proof standard, applicable in criminal cases. It requires evidence which leaves the trier of fact "fully satisfied, entirely convinced, . . . to a moral certainty" [8, p. 147].

In our view proof standards cannot and should not be interpreted probabilistically. The first and most important reason is that probability theory is applicable only if statistical knowledge about prior and conditional probabilities is available. Presuming the existence of such statistical information would defeat the whole purpose of argumentation about factual issues, which is to provide methods for making justified decisions when knowledge of the domain is lacking. Another argument against interpreting proof standards probabilistically is more technical. Arguments for and against some proposition are rarely independent. What is needed is some way to accruing arguments which does not depend on the assumption that the arguments or evidence are independent.

Prakken has identified three principles any formal account of accrual must satisfy [25]: 1) Combining several arguments pro or con some proposition can not only strengthen one's position, but also weaken it. 2) Once several arguments have been accrued, the individual arguments, considered separately, should have no impact on the acceptability of the proposition at issue, and 3) Finally, any argument which is 'flawed' may not take part in the aggregation process. The models of proof standards presented in the section are designed to respect these principals.

2 Formal Model

Our goal in this section is to define an abstract formal model of argumentation as a theory and proof construction process for making justified decisions. Inspired by Dung's model of abstract argumentation frameworks, the model shall be as abstract and simple as possible while being sufficient for capturing the distinctions between the various types of proof burdens and proof standards identified in the introduction and meets other known requirements. It is not intended to be a comprehensive formal model of argumentation. We will also take care to abstract from the details of the legal domain and, in particular, the law of civil procedure.

We begin with the concept of an argument. Unlike Dung, we cannot leave this concept fully abstract, since our aim is to model burden of proof and proof standards. The proponent of an argument has the burden of production for its ordinary premises; while the respondent has the burden or production for any exceptions. Moreover, since the task of proof standards is to aggregate arguments pro and con some proposition at issue, the model must represent not only the premises of arguments, and distinctions between types of premises, but also their conclusions. These considerations lead us to the following definition of argument.

Definition 1 (argument). Let $\mathscr{L}$ be a propositional language. An **argument** is a tuple $\langle P, E, c\rangle$ where $P \subset \mathscr{L}$ are its **premises**, $E \subset \mathscr{L}$ are its **exceptions** and $c \in \mathscr{L}$ is its **conclusion**. For simplicity, c and all members of P and E must be literals, i.e. either an atomic proposition or a negated atomic proposition. Let p be a literal. If p is c, then the argument is an argument **pro** p. If p is the complement of c, then the argument is an argument **con** p.

Since all conclusions of arguments are literals according to this definition, the *axioms* of the theory constructed during argumentation consists only of literals. Other propositions of the theory can be derived from these axioms using the inference rules of classical logic and the argumentation schemes of the domain.

To model the distinctions between the various kinds of burden of proof, we must model argumentation as a process, consisting of several phases. It is sufficient to distinguish three phases, the opening, argumentation and closing phases of the process. Since typically argumentation takes place in dialogues, we will use the term 'dialogue' as the generic name for argumentation processes.

Definition 2 (dialogue). A **dialogue** is a tuple $\langle O, A, C\rangle$, O, A and C, the **opening**, **argumentation**, and **closing** phases of the dialogue, respectively, are each sequences of **stages**. A stage is a tuple $\langle \mathit{arguments}, \mathit{status}\rangle$, where *arguments* is a set of arguments and *status* is a function mapping the conclusions of the arguments in *arguments* to their **dialectical status** in the stage, where the status is a member of $\{\mathit{claimed}, \mathit{questioned}\}$. In every chain of arguments, $a_1, \ldots a_n$, constructable from *arguments* by linking the conclusion of an argument to a premise of another argument, a conclusion of an argument a_i may not be a premise of an argument a_j, if $j < i$. A set of arguments which violates this condition is said to contain a *cycle* and a set of arguments which complies with this condition is called *cycle-free*.

Notice that the cycles defined here are not the same as cycles in a Dung argumentation framework. Whereas the links (arcs) between arguments in the directed graph induced by a Dung argumentation framework model the attack relation, the links in the directed graph induced by arguments in our system model the premise and conclusion relations. Notice that, in our system, arguments both pro and con some proposition can be included in a set of arguments without causing a cycle.

Constraining the arguments of a stage to be cycle-free is intended to simplify the evaluation of arguments. The set of arguments is intended to model the current state of the proof being constructed by the parties in the dialogue, not a 'pool of information' for constructing proofs. Intuitively, proofs should not contain cycles.

Next we need a structure for evaluating arguments, to assess the acceptability of propositions at issue. As in value-based argumentation frameworks [4, 5] arguments are evaluated with respect to an *audience*, such as the trier-of-fact (judge or jury) in legal trials..

Definition 3 (audience). An *audience* is a structure $\langle \mathit{assumptions}, \mathit{weight}\rangle$, where $\mathit{assumptions} \subset \mathscr{L}$ is a consistent set of literals assumed to be acceptable by the audience and *weight* is a partial function mapping arguments to real numbers in the

range $0.0 \ldots 1.0$, representing the relative weights assigned by the audience to the arguments.

Whereas in value-based argumentation frameworks the audience is defined by a partial-order on a set of values, which is then used to constrain the attack relation on arguments in a Dung argumentation framework, the audience in our system models the relative strength of arguments for this audience. Intuitively, a stronger argument does not necessarily attack a weaker argument. Both arguments could be arguments pro the same proposition, for example. Thus, these two conceptions of an auidence are not directly comparable.

An argument evaluation structure associates an audience with a stage of dialogue and assigns proof standards to propositions, providing a basis for evaluating the acceptability of propositions to this audience.

Definition 4 (argument evaluation structure). An **argument evaluation structure** is a tuple $\langle stage, audience, standard \rangle$, where *stage* is a stage in a dialogue, *audience* is an audience and *standard* is a total function mapping propositions in $\mathscr{L}$ to their applicable proof standards in the dialogue. A **proof standard** is a function mapping tuples of the form $\langle issue, stage, audience \rangle$ to the Boolean values *true* and *false*, where *issue* is a proposition in $\mathscr{L}$, *stage* is a stage and *audience* is an audience.

Given an argument evaluation structure, the acceptability of a proposition can be defined as follows.

Definition 5 (acceptability). A literal p is **acceptable** in an argument evaluation structure $\langle stage, audience, standard \rangle$ if and only if $standard(p, stage, audience)$ is *true*.

The argument evaluation structure is the component of this formal model which is most like a Dung abstract argumentation framework. The role of the attack relation in abstract argumentation frameworks is played by competing pro and con arguments aggregated by proof standards, using the relative weights assigned the arguments by an audience.

Obviously much of the work of argument evaluation has been delegated to the proof standards. We cannot say anything about the computational properties of acceptability in an argument evaluation structure until these standards have been defined. All the proof standards make use of the concept of argument applicability, so let us define this concept first.

Definition 6 (argument applicability). Let $\langle stage, audience, standard \rangle$ be an argument evaluation structure. An argument $\langle P, E, c \rangle$ is **applicable** in this argument evaluation structure if and only if

- the argument is a member of the arguments of the *stage*,
- every proposition $p \in P$, the premises, is an assumption of the *audience* or, if neither p nor $\overline{p}$ is an assumption, is acceptable in the argument evaluation structure and

- no proposition $p \in E$, the exceptions, is an assumption of the *audience* or, if neither p nor $\overline{p}$ is an assumption, is acceptable in the argument evaluation structure.

Now we are ready to define the proof standards, beginning with scintilla of the evidence. A proposition satisfies the scintilla standard, in our model, if it is supported by at least one applicable pro argument.

Definition 7 (scintilla of evidence). Let $\langle stage, audience, standard \rangle$ be an argument evaluation structure and let p be a literal in $\mathscr{L}$. $scintilla(p, stage, audience) = true$ if and only if there is at least one applicable argument pro p in *stage*.

Scintilla is the weakest of the proof standards we will define and is the only one which can be met by complementary literals in the same argument evaluation structure. That is, if p is an atomic proposition, both p and $\neg p$ can be acceptable in an argument evaluation structure using the scintilla of evidence standard. This would be the case if p has an applicable con argument, as well as an applicable pro argument.

Let us now turn our attention to the three most important legal proof standards: preponderance of the evidence, clear and convincing evidence and beyond reasonable doubt. Intuitively, preponderance is satisfied if the pro arguments outweigh the con arguments, by however much. The issue we have to face when formalizing preponderance is how to aggregate the weights of a set of arguments for the purpose of this comparison. The clear and convincing evidence standard requires more proof than the preponderance standard: not only must the pro arguments outweigh the con arguments, the weight of the pro arguments and the difference in weight of the pro and con arguments both must exceed some thresholds. Finally, the beyond a reasonable doubt standard goes further. Not only must the arguments be clear and convincing, but, as the name of the standard suggests, the weight of the con arguments must be below the threshold of 'reasonable doubt'.

Definition 8 (preponderance of the evidence). Let $\langle stage, audience, standard \rangle$ be an argument evaluation structure and let p be a literal in $\mathscr{L}$. $preponderance(p, stage, audience) = true$ if and only if

- there is at least one applicable argument pro p in *stage* and
- the maximum weight assigned by the audience to the applicable arguments pro p is greater than the maximum weight of the applicable arguments con p.

The preponderance of the evidence standard was called the *best argument* standard in [18].

Definition 9 (clear and convincing evidence). Let $\langle stage, audience, standard \rangle$ be an argument evaluation structure and let p be a literal in $\mathscr{L}$. $clear\text{-}and\text{-}convincing(p, stage, audience) = true$ if and only if

- the preponderance of the evidence standard is met,
- the maximum weight of the applicable pro arguments exceeds some threshold α, and

- the difference between the maximum weight of the applicable pro arguments and the maximum weight of the applicable con arguments exceeds some threshold β.

Definition 10 (beyond reasonable doubt). Let $\langle stage, audience, standard \rangle$ be an argument evaluation structure $\langle stage, audience, standard \rangle$ and let p be a literal in $\mathscr{L}$. $beyond\text{-}reasonable\text{-}doubt(p, stage, audience) = true$ if and only if

- the clear and convincing evidence standard is met and
- the maximum weight of the applicable con arguments is less than some threshold γ.

We assume the α, β and γ thresholds used by the clear and convincing evidence and beyond a reasonable doubt standards are set by the applicable protocol of the dialogue.

At first glance, using maximum weights to aggregate pro and con arguments might seem unintuitive. One might be inclined to compare the *sums* of the weights of the applicable pro and con arguments. However, since arguments cannot be assumed to be independent, summing weights would risk taking the same information or reasons into account multiple times. When several weak arguments can be combined to make a stronger argument, this can be achieved by joining their premises together into a single argument, as discussed further next. We leave it up to the audience to judge the effect of any possible interdependencies among the premises on the weight of the argument. Both alternatives, summing weights or taking their maximum weight, have the property of taking all arguments into account.

Assuming arguments are stated in their strongest form, aggregating arguments using their maximum weight satisfies all three of Prakken's principles of accrual [25]: 1) Aggregated arguments can be evaluated by the audience to be stronger or weaker than the arguments considered separately; 2) Once several arguments have been accrued, the individual arguments, considered separately, have no effect on the acceptability of the proposition at issue; and 3) Any argument which is 'flawed' does not take part in the aggregation process.

By 'stating arguments in their strongest form' we mean the following. Let p and q be two propositions which, when they are true, are evidence pro a third proposition, r. This can be expressed as either two *convergent* arguments or as a *linked* argument [39]. The convergent arguments would be:

a_1: r since p.
a_2: r since q.

The linked form of this argument 'accrues' p and q into a single argument:

a_3: r since p and q.

If the linked argument, a_3, is more persuasive, that is if the party putting forward this argument estimates it would be given more weight by the audience than either a_1 or a_2, then we assume a_3 will be put forward.

To illustrate this more concretely, let's return to Prakken's example about jogging when it is both hot and rainy. The strongest arguments con jogging are the convergent arguments:

a_4: not jogging since hot.
a_5: not jogging since rainy.

The strongest argument pro jogging is the following linked argument:

a_6: jogging since hot and rainy.

Returning to Prakken three principals of accrual: 1) The audience can decide whether to give a_6 greater or lesser weight to each of the arguments a_4 and a_5; 2) If the accrued argument, a_6, is given greater weight, then a_6 will defeat both a_4 and a_5, rendering them ineffective, using any of the proofs standards which aggregate arguments by weight; and 3) Inapplicable arguments are not be taken into consideration using any proof standard.

In [18] one further proof standard was defined, called *dialectical validity*. For the sake of completeness we include its definitions here.

Definition 11 (dialectical validity). Let $\langle stage, audience, standard \rangle$ be an argument evaluation structure $\langle stage, audience, standard \rangle$ and let p be a literal in $\mathscr{L}$. $dialectical\text{-}validity(p, stage, audience,) = true$ if and only if there is at least one applicable argument pro p in $stage$ and no argument con p in $stage$ is applicable.

The dialectical validity standard is suitable for aggregating arguments from general rules and exceptions, where any applicable exception is enough to override the general rule.

One of the requirements identified in the introduction is that checking proofs should be an easy task. In more computational terms, using our formal model, the issue is whether the acceptability of a proposition in an argument evaluation structure is tractably decidable. We conjecture that this is the case, but will not try to prove this formally here. The reasons for this conjecture are many. The argument evaluation structure has been designed in part to achieve this goal, by making several restrictions: 1) The language is propositional, not first-order; 2) premises, exceptions and conclusions of arguments must be literals; and 3) the set of arguments of a stage is finite and, by definition, acyclic. Of course the computational complexity of acceptability also depends on the complexity of the proof standards applied.

No we are ready to turn to modeling the various kinds of burden of proof. The burdens of claiming and questioning must be met during the opening phase of the dialogue. The burden of production and the tactical burden of proof are relevant only during the argumentation phase. Finally, the burden of persuasion comes into play in the closing phase, but is also used hypothetically to estimate the tactical burden of proof during the argumentation phase.

The purpose of the opening stage of a dialogue is to frame the issues. Arguments put forward in the argumentation stage must be relevant to the issues raised in the opening stage. Depending on the protocol of the dialogue, a proposition claimed in the opening stage may be deemed conceded unless it is questioned, requiring the audience to assume it is true, following the principal of "silence implies consent."

Definition 12 (burdens of claiming and questioning). Let $s_1, \ldots, s_n$ be the stages of the opening phase of a dialogue. Let $\langle arguments_n, status_n \rangle$ be the last stage, s_n,

of the opening phase. A party has met the **burden of claiming** a proposition p if and only if $status_n(p) \in \{claimed, questioned\}$, that is, if and only if $status_n(p)$ is defined. The **burden of questioning** a proposition p has been met if and only if $status_n(p) = questioned$.

Notice that a questioned proposition satisfies the burden of claiming, since it is assumed that only propositions which have been claimed in an earlier stage are questioned.

This simple model defines only minimal requirements for raising issues in the opening phase of a dialogue. The argumentation protocol of a dialogue may state additional requirements. For example, according to the law of civil procedure in California, the plaintiff must state a *cause of action*: the facts claimed must be sufficient to give the plaintiff a right to judicial relief, as a matter of law.

The burden of production is relevant only during the argumentation phase of a dialogue. The burden of production for some proposition is satisfied if it is acceptable at the end of the argumentation phase using the the weakest proof standard, scintilla of the evidence. The party who puts forward an argument has the burden of production for its premies. Similarly, the respondent to an argument has the burden of production for each exception.

The audience used to assess the burden of production depends on the protocol of the particular dialogue. In civil proceedings in California, the judge is the audience during the argumentation phase, i.e. the trial.

Definition 13 (burden of production). Let $s_1, \ldots, s_n$ be the stages of the argumentation phase of a dialogue. Let $\langle arguments_n, status_n \rangle$ be the last stage, s_n, of the argumentation phase. Let *audience* be the relevant audience for assessing the burden of production, depending on the protocol of the dialogue. Let *AES* be the argument evaluation structure $\langle s_n, audience, standard \rangle$, where *standard* is a function mapping every proposition to the scintilla of evidence proof standard. The **burden of production** for a proposition p has been met if and only if p is acceptable in *AES*.

Even though the weakest proof standard, scintillia of the evidence, is used to test whether the burden of production has been met, an arbitrary, or silly, argument would not be sufficient, since only applicable arguments are taken into consideration by all proof standards. A silly argument can be defeated by questioning or attacking its premises. Arguments can be undercut in our system by first revealing, if necessary, an implicit premise about the applicability of the warrant underlying the argument to this case and then attacking this premise [18].

Since arguments put forward to met the burden of production can be defeated by further arguments, the burden of production may be met at some stage, s_i, of a dialogue, but not met at some later stage, s_j, where $j > i$. If the burden of production is not met at the end of the argumentation phase, the audience in the closing phase may be required, depending on the dialogue type, to assume that the proposition is false. In this case, the burden of persuasion for this proposition becomes irrelevant.

The burden of persuasion plays a role only in the closing phase of the dialogue. The burden of persuasion is met only if at the end of the closing phase the proposition at issue is acceptable to the audience. The way proof standards are assigned

to propositions depends on the type of dialogue and is regulated by the argumentation protocol. In legal proceedings in California, the proof standards are assigned by the judge, since this is a question of law, not fact. In civil proceedings, the usual proof standard is preponderance of the evidence. In criminal proceedings, the proof standard is beyond reasonable doubt.

Definition 14 (burden of persuasion). Let $s_1, \ldots, s_n$ be the stages of the closing phase of a dialogue. Let $\langle \mathit{arguments}_n, \mathit{status}_n \rangle$ be the last stage, s_n, of the closing phase. Let *audience* be the relevant audience for assessing the burden of persuasion, depending on the dialogue type and its protocol. Let *AES* be the argument evaluation structure $\langle s_n, \mathit{audience}, \mathit{standard} \rangle$, where *standard* is a function mapping every proposition to its applicable proof standard for this type of dialogue. The **burden of persuasion** for a proposition p has been met if and only if p is acceptable in *AES*.

In some cases, the party which has the burden of production in the argumentation phase may not have the burden of persuasion in the closing phase. This is the case, for example, in criminal law proceedings. The defendant, as usual, has the burden of production for exceptions, such as self-defense in murder cases, but once this burden has been met, the burden of persuasion is passed to the prosecution. If any evidence of self-defense has been brought forward, satisfying the burden of persuasion, the prosecution has the burden of persuading the trier of fact, beyond a reasonable doubt, that the defendant did not act in self-defense.

This can be achieved in our model by making the exception an ordinary premise after the burden of production has been met in the argumentation phase. For example, let $\langle P, E, c \rangle$ be an argument and e be a proposition in E, meaning "self defense". After the burden of production for e has been met, the other side can be given the burden of persuasion by removing e from E and adding $\neg e$ to P. It may seem odd to modify the argument in this way, but keep in mind the arguments of a stage do not represent the speech acts of the parties, but rather the state of the proof being constructed collaboratively by all parties, according the protocol of the dialogue type. The stage must be modified in some way to reflect this change, and modifying the arguments of the stage is one way to accomplish this.

One kind of burden of proof remains to be defined formally, the tactical burden of proof. The tactical burden is the only one which can shift back and forth between the parties. It is relevant only during the argumentation phase of the dialogue. We defined the burden of persuasion first, even though it is applicable only in the later closing stage, because the tactical burden of proof requires the burden of persuasion to be estimated. At each stage of the argumentation phase, a party must decide whether stronger arguments might be necessary to persuade the audience. In some dialogue types, the audience may reveal its assumptions and evaluations (weight assignments) during the argumentation phase, at least provisionally. This will be the case, for example, in two-party dialogues where the audience to be persuaded is the same as the respondent. In legal proceedings this is not the case, since the respondent is the defendant and the audience is the judge or jury. In such cases it will be necessary to make assumptions about the audience.

Definition 15 (tactical burden of proof). Let $s_1, \ldots, s_n$ be the stages of the argumentation phase of a dialogue. Assume *audience* is the audience which will assess the burden of persuasion in the closing phase. Assume *standard* is the function which will be used in the closing phase to assign a proof standard to each proposition. For each stage s_i in $s_1, \ldots, s_n$, let AES_i be the argument evaluation structure $\langle s_i, audience, standard \rangle$. The **tactical burden of proof** for a proposition p is met at stage s_i if and only if p is acceptable in AES_i.

Both sides in a dialogue can have a tactical burden. Intuitively, a party has a tactical burden of proof for a proposition p at some stage s_i only if p is not acceptable in s_i and the party has an interest in proving p, either because proving p is relevant for proving some claim of the party or disproving some claim of the other party, given the arguments which have been put forward. A fuller, more complete account of the tactical burden of proof would require the parties, their claims and the concept of relevance to be modeled.

This completes our formal model of proof standards and burden of proof. Again, we do not claim this is a comprehensive dialogical model of argumentation. Many important elements of argumentation have been abstracted out for the sake of brevity, such as the parties, argumentation schemes and their critical questions, dialogue types, argumentation protocols and commitment stores. Our aim here was to define the simplest, most abstract possible model which is sufficient for distinguishing the various kinds of proof standards and burdens of proof which have been discussed in the computational models of argument literature. Of course, we cannot prove that we have achieved this goal. We leave it up to others in the field to try to develop a simpler model with this scope.

In the introduction we formulated our view of argumentation as a kind of process for making justified decisions, where the input to the process is an initial claim or issue and the output consists of a set of claims, the decision to accept or reject each claim, a theory and a proof. Unlike assumption-based instantiations of Dung's model of abstract argumentation frameworks [9], in which a theory or knowledge base is presumed as part of the input, in our model a theory and a proof are constructed together during argumentation and are part of the output. The theory constructed is the deductive closure, in classical propositional logic, of the set of all propositions which have been assumed by the audience or, if neither the proposition or its complement has been assumed, are acceptable in the final stage of the closing phase of the dialogue. The proof constructed is represented by the argument evaluation structure of the final stage.

3 Survey of Prior Research

Early work in the field of Artificial Intelligence and Law recognized the utility of defeasible rules, subject to exceptions, as a tool for allocating the burden of proof and developed nonmonotonic logics for reasoning with such rules [13, 36]. But

different proof standards or kinds of proof burdens were not yet distinguished in these models.

In the Pleadings Game [14], inspired by legal proceedings, argumentation was modeled as dialogical process consisting of several phases, in which theories and arguments are constructed dynamically by the parties during the process. The Pleadings Game modeled the burdens of claiming, questioning and production, but it did not explicitly use this terminology. Proof standards and the burden of persuasion, were not modeled, as they do not play a role in the opening phase.

To our knowledge, the first effort to develop a computational model of proof standards was by Freeman and Farley in 1996 [12]. They modeled argumentation as a dialectical process during which an acyclic argument graph is constructed by putting forward pro and con arguments constructed ('invented') from a propositional rule-base. The model of proof standards comes into play when evaluating the arguments in the graph. An argument is *defendable* if each premise satisfies its proof standard. Five proof standards were defined (scintilla of evidence, preponderance, dialectical validity, beyond a reasonable doubt and beyond a doubt). The relative weights of arguments were not assigned by an audience, but rather computed from certainty factors assigned to premises and the kind of argumentation scheme used to construct the argument, using the weakest premise principle [33]. The model of argumentation developed in this paper has borrowed much from Freeman and Farley, but there are some important differences. By restricting the arguments which can be put forward to those which can be constructed from a static rule-base and model of the facts, provided as input to the dialectical process, Freeman and Farley's model is more of a deductive model of argumentation than a theory construction model. The relative strength of arguments is determined by the rules and facts, rather than by the intersubjective judgment of an audience. Finally, only the burden of persuasion has been modeled by this work. Finally, no attempt was made to model proof burdens other than the burden of persuasion in this work.

The Zeno system [17] was inspired by Freeman and Farley's work. Zeno's model of the structure of argument graphs was based on Kunz and Rittel's [22] concept of issue-based information systems (IBIS). Zeno supported arguments about both factual issues and issues about which action to take to solve some problem or achieve some goal (practical reasoning). For factual issues, three proof standards were modeled (scintilla of evidence, preponderance of the evidence and beyond reasonable doubt). For issues about actions, two further proof standards were provided (no better alternative and best choice). The relative strength of arguments in Zeno was computed from qualitative constraints (equations and inequalities) over propositions. The qualitative constraints were issues which could be argued about. Only the burden of persuasion was modeled in Zeno. As in Freeman and Farley's work, Zeno did not explicitly model an audience. On the contrary, in Zeno it was assumed the parties would argue about the evaluation of their own arguments, by putting forward and arguing about constraints.

Prakken formulated the three principles of argument accrual [25] explained in the introduction, which our model has been designed to satisfy. He compared automatic and manual approaches to accrual. In the manual approach, arguments are

accrued by changing their representation in the model. In the automatic approach, arguments are accrued by the argument evaluation process, without having to modify the representation of the arguments. Prakken illustrated this approach with a novel system, using a rule-based instantiation of Dung's model of an abstract argumentation framework, in which the conclusions of rules are labeled with the set of their premises. The attack relation of the argumentation framework is defined so that an argument *A* does not attack an argument *B* if the set of premise of the conclusion of the argument *A* is a subset of the (accrued) set of premises of the argument *B*. This leads to a bottom-up inference process which, Prakken notes, is similar to Reason-Based Logic [20]: first all arguments pro and con some leaf proposition are combined into two competing sets of accruals; next the conflict between these accruals is resolved; and finally the process iterates moving up the tree, with only the winning defeasible conclusion being used. The way arguments are evaluated in our model has much in common with the bottom-up inference approach taken by both Rule-Based Logic and Prakken's system. Nonetheless, our approach to the accrual problem is manual, not automatic, as responsibility for accruing arguments is allocated to the parties who put them forward in dialogues. But unlike Bayesian Networks, which require a complete conditional probabilities table to be provided as input to the process, our approach does not require all possible arguments to be formulated. Accruing arguments when possible to strengthen one's case is part of the burden of proof allocated to the parties. Prakken points out that there is a tradeoff between automatic and manual approaches to accrual: in the former unwanted inferences must be expressly blocked while in the latter accruals must be expressly formulated. So neither approach is clearly superior.

In 2006, Prakken and Sartor published the first formal account of the distinctions between the burden of production, the burden of persuasion and the tactical burden of proof [29]. Their model is based on an interpretation of presumptions as default rules, formalized using an extended version of their Inference System (IS) defeasible logic [28], called the Litigation Inference System (LIS) [24], which includes as part of the input, together with the defeasible rules, an assignment of the burden of persuasion for literals to either the plaintiff or defendant in the proceeding. Prakken and Sartor's model of the distinctions between these three proof burdens is more concrete than our model, as it is limited to arguments constructed from defeasible rules. Our model abstracts away the process of constructing arguments and can be instantiated with models of various argumentation schemes, for example for arguments from cases as well as defeasible rules. As mentioned above, we have done some research recently in this direction [16]. Prakken and Sartor did not attempt to model the burdens of claiming or questioning or support the use of various proof standards.

Continuing work began with Reed and Walton in 2005 on dialogues about the burden of proof [27], Prakken and Sartor in 2007 presented a model of arguments about the allocation of the burden of persuasion [30]. They note that 'the argumentation games we define in this paper are not intended as a model of actual legal dialogue but as a proof theory for a nonmonotonic logic. ... All we claim is that our games draw the correct defeasible inferences from a given body of information

and an associated allocation of the burden of persuasion. It remains to be seen how the present logical model can be integrated with dialogical and procedural models of legal argument". Clearly, their work is a deductive model of argumentation, not a theory construction model. In their conclusion, Prakken and Sartor express concern that their system "lacks an extension-based semantics in the style of [10]" and note, citing [24], that this "raises questions about the adequacy of 'mainstream' nonmonotonic logics for representing legal reasoning".[1]

The idea of using audiences in our model was inspired by work by Bench-Capon, Doutre and Dunne [5]. Bench-Capon's focus is on modeling practical reasoning, i.e. the process of making decisions about which action to take in order to achieve goals which promote an agent's values. Since different agents can have different values, as well as different priority orderings on their set of values, arguments about action can only be evaluated against the values of a particular agent, or 'audience'. Bench-Capon contrasted practical reasoning with argumentation about "what is true in a situation", i.e. the facts of a case, and appears to consider audiences to be relevant only for the former. As illustrated by legal trials, however, audiences can be important in argumentation dialogues about factual issues as well, since different persons will judge the probative weight of evidence, such as witness testimony, differently. Our model of an audience is more abstract, since it orders arguments by their strength, regardless of the kind of argumentation scheme which has been applied to construct the argument. Formally, Bench-Capon's model is an extension of Dung's concept of an abstract argumentation framework. Although a formal dialogue game is defined, the game serves as a proof theory for a deductive model of argument. The input to the system is a Value-Based Argumentation Framework (VAF) consisting of a set of arguments, an attack relation over these arguments, a set of values, and a mapping from arguments to values. An audience is defined to be a binary relation over these values, modeling the preference ordering over these values of the audience. The output of the dialogue game is the set of arguments which are objectively or subjectively acceptable. Arguments which are objectively acceptable are, or should be, acceptable by all (rational) audiences. Conversely, arguments which are subjectively acceptable are acceptable to a particular, given audience. Bench-Capon compares the distinction between objective and subjective acceptability with the concept of credulous and skeptical inference familiar from nonmonotonic logic and Dung argumentation frameworks. The distinction between objective and subjective acceptability, like the distinction between credulous and skeptical inference, possibly can be viewed as a simple two-level model of proof standards, but it is questionable that this distinction is of much use outside of the procedural context of dialogues in which the allocation of the burden of proof matters, and such dialogues were outside the scope of this work.

Atkinson and Bench-Capon recently modeled a variety of proof standards using Dung-style argumentation frameworks [2]. This paper also recognizes the need for audiences in dialogues about factual issues, as well as teleological issues about values promoted by alternative courses of action. The basic idea of this paper elab-

[1] But see [37] for a defense of nonmonotonic logic for legal reasoning, even when the burden of persuasion is divided among the parties.

orates on the idea discussed above, of using the distinction between various kinds of acceptability in Dung argumentation frameworks to define proof standards. The scintilla of evidence standard is modeled as membership in some preferred extension (credulous acceptability). Preponderance of the evidence corresponds to membership in all preferred extensions (skeptical acceptability). And, finally, Beyond reasonable doubt corresponds to membership in the grounded extension, if there is one. Ways of modeling proof standards between preponderance of the evidence and beyond reasonable doubt, such as the clear and convincing evidence standard, are proposed which make use of an assignment of 'probabilities', i.e. weights, to arguments. However this idea is discarded with the argument that this information is not usually available. In our model this information is provided by the intersubjective judgment of the audience. Later in their paper, Atkinson and Bench-Capon suggest the use of audiences to derive preferences over arguments, starting with Bench-Capon's value-based argumentation frameworks for arguing about teleological issues. This idea is extended to arguments about factual issues by ordering evidence, for example by ordering witness testimony using information about witness credibility.

An advantage of Bench-Capon and Atkinson's line of research is that it elaborates rationality constraints on an audience's assignment of strengths to arguments, using for example its value preferences or its assessment of the relative credibility of witnesses. Although our model leaves the weights assigned by audiences to arguments completely unconstrained, it has the advantage of being more general, applying to arguments from any argumentation scheme. Perhaps it is possible to combine the benefits of these two approaches. This may be an interesting topic for future research.

A disadvantage of modeling proof standards using Dung argumentation frameworks is the computational complexity of testing whether the proof standard has been met. We have argued that, intuitively, to satisfy a burden of proof, the party with the burden must present the proof in a form which is easy to check. The other party shouldn't have to solve an undecidable or intractable problem to check the proof. As discussed previously, in Section 2, this criterion could be satisfied for the proposed scintilla of evidence proof standard, modeled as credulous acceptability, by requiring the proponent to produce an admissible set of arguments. The admissible set could serve as a representation of the proof, since checking whether the set is admissible and whether an argument is a member of this set are both tractable problems. But it is not clear what structures could serve as proofs for the models proposed by Atkinson and Bench-Capon for the stricter proof standards. A related issue is whether or not such proofs could be presented in a form which is comprehensible to human users, using for example some kind of argument mapping or other visualization techniques. Existing argument mapping methods have been developed for presenting proofs (argument graphs) of the kind we have presented in this paper [3, 11, 1, 15]. While Dung argumentation frameworks are often displayed as graphs, these graphs do not represent proofs of the acceptability of any of the arguments in the graph.

Prakken recently published a formal model which explicates the role of judges in deciding issues regarding the admissibility of evidence and the allocation of the burdens of production and persuasion [26]. Interestingly, dialogues are divided into two, rather than tree phases, in the model, called the *pleadings phase* and the *decision phase*. The pleadings phases encompasses the opening and argumentation phases of our model. The decision phase corresponds to our closing phase. The judge plays a role in the decision phase comparable to the audience in the closing phase of our model, but arguments are evaluated by having the judge put forward further arguments, which are then evaluated using Prakken's LIS nonmonotonic logic [24], discussed above. Arguments are not weighed and proof standards are not part of the model.

Prakken and Sartor's chapter on a "Logical Analysis of Burdens of Proof" in [31] highly influenced our model of the distinctions among the various kinds of proof burdens. The main difference between our systems is that they use their nonmonotonic logic [28, 24], a rule-based instantiation of Dung's abstract argumentation framework, to evaluate the arguments at each stage of the dialogue. We have already expressed our reservations about this approach, which does not take into consideration that a burden of proof entails an obligation to put forward the proof in a form which can be tractably checked. In their conclusion they point out that their use of a nonmonotonic logic for evaluating arguments in a stage could be replaced by any formalism "which accepts as input a description of an evidential problem and produces as output a fallible assessment whether a claim has been proven". This is one way looking at we have done here, by replacing their nonmontonic logic with a structure designed to make argument evaluation tractable for a variety of common proof standards. Proof standards are not given much attention in their chapter, except to suggest that they could be handled using some mechanism for ordering arguments by their strength. Finally, although they recognize the role of the finder-of-fact, they did not explictly model audiences or their impact on assessing proof burdens. For example, the evalution of the tactical burden of proof in their model does not require a party to estimate an audience's assessment of the weight of arguments. Rather, this information appears to be assumed as input to the dialogue, available as common knowledge to both parties.

The model of argumentation developed in this paper is most closely related to the Carneades system, which consists of both a mathematical model of argumentation and software tools for supporting argumentation tasks based on this model.[2] Carneades is work in progress and thus prior publications about the system differ in their details. For example, the version of the model in [19] gave the term 'presumption' a technical meaning which is confusing for those familar with the concept of a presumption in the legal domain. This was pointed out by Prakken and Sartor [29] and corrected when Prakken joined us for the next version of Carneades [18]. Similarly, the use of weights to order arguments in [19] was replaced by a partial order in [18], but now we have come full circle, by again using numeric weights. The initial models of the prepondernace of the evidence and beyond reasonable doubt proof

[2] http://carneades.berlios.de

standards of [19] were removed from the [18] version, since we had our doubts about their adequacy as models of these legal standards. We feel confident enough about the models of these legal standards in this paper to want to publish them in order to obtain critical feedback. Aside from improved models of various proof standards, the main contribution of the new version presented here is that it more clearly and explicitly models argumentation as a theory and argument contruction process, consisting of a sequence of stages divided into opening, argumentation and closing phases. Although we had already suggested in [18] how to use Carneades to model the distinction between the burden of production and the burden of persuasion, the version in this article models these distinctions more explictly and extends the model to cover the burdens of claiming and questioning in the opening phase and the tactical burden of proof. This version also introduces an explicit model of audiences. Previous versions had used the concept of an argumentation *context* to model argument strengths or priorities, with no reference to an audience.

4 Concluding Remarks

Viewing argumentation as dialogical process for making justified decisions raises a number of issues which have no place in deductive, monological accounts of argumentation, proof burdens and standards among them. Thus it should come as no surprise that the concept of *proof* has thus far received little attention in mainstream accounts of argumentation in artificial intelligence. An argument may be acceptable in a Dung-style argumentation framework, or a proposition may be warranted by a default theory in some nonmonotonic logic, but what mathematical structures are adequate as models of proofs of these or other inference relations? We have argued that checking a proof should be an easy, tractable problem. Argumentation frameworks and default theories do not, in general, meet this requirement. And other, less computational requirements could also be formulated, such as transparency and comprehensibility for the intended class of audiences.

Another distinction between deductive and dialogical conceptions of argumentation concerns their input/out relations. Whereas in deductive accounts an argumentation framework or default theory is provided as input and the task is to derive acceptable arguments or warranted propositions, argumentation dialogues begin with a claim or issue and construct, as part of their output, theories and proofs. Argumentation dialogues includes synthetic as well as analytic tasks.

Abstract argumentation frameworks, which focus on attack relations among arguments, are not well-suited to modeling proof standards, at least not the familiar proof standards from the legal domain. The intuitive idea of legal proof standards since Roman times, symbolized by the scales of the goddess Justitia, involves the weighing of arguments or evidence pro and con some claim. This simple idea cannot be directly represented using abstract argumentation frameworks. Attempts to model proof standards as variations of credulous and skeptical acceptability in an argumentation framework are not very promising, since they largely leave open the

question of how to represent proofs in an understandable form which can be easily checked.

Research to date on modeling proof standards and burdens calls into question the common research strategy in the field of computational models of argumentation which presumes that valid dialogical models of argumentation as a process can be constructed on the foundation of deductive models of argument as a nonmonotonic inference relation. We recommend a research strategy which begins with a task and requirements analysis of argumentation dialogs in a variety of application domains.

Acknowledgements We'd like to thank Trevor Bench-Capon and Henry Prakken for their helpful comments on various drafts of this article.

References

1. T. Anderson, D. Schum, and W. Twining. *Analysis of Evidence*. Cambridge University Press, 2nd edition, 2005.
2. K. Atkinson and T. Bench-Capon. Argumentation and standards of proof. In *ICAIL '07: Proceedings of the 11th international conference on Artificial intelligence and law*, pages 107–116, New York, NY, USA, 2007. ACM.
3. M. C. Beardsley. *Practical Logic*. Prentice Hall, New York, 1950.
4. T. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.
5. T. J. Bench-Capon, S. Doutre, and P. E. Dunne. Audiences in argumentation frameworks. *Artificial Intelligence*, 171(42-71), 2007.
6. P. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, 2008.
7. J. Bing. Uncertainty, decisions and information systems. In C. Ciampi, editor, *Artificial Intelligence and Legal Information Systems*. North-Holland, 1982.
8. H. C. Black. *Black's Law Dictionary*. West Publishing Co., 1979.
9. A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1-2):63–101, 1997.
10. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.
11. J. B. Freeman. *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure*. Walter de Gruyter, Berlin / New York, 1991.
12. K. Freeman and A. M. Farley. A model of argumentation and its application to legal reasoning. *Artificial Intelligence and Law*, 4(3-4):163–197, 1996.
13. T. F. Gordon. Some problems with prolog as a knowledge representation language for legal expert systems. In C. Arnold, editor, *Yearbook of Law, Computers and Technology*, pages 52–67. Leicester Polytechnic Press, Leicester, England, 1987.
14. T. F. Gordon. *The Pleadings Game; An Artificial Intelligence Model of Procedural Justice*. Springer, New York, 1995. Book version of 1993 Ph.D. Thesis; University of Darmstadt.
15. T. F. Gordon. Visualizing Carneades argument graphs. *Law, Probability and Risk*, 6(1-4):109–117, 2007.
16. T. F. Gordon. Hybrid reasoning with argumentation schemes. In *Proceedings of the 8th Workshop on Computational Models of Natural Argument (CMNA 08)*, pages 16–25, Patras, Greece, July 2008. The 18th European Conference on Artificial Intelligence (ECAI 2008).
17. T. F. Gordon and N. Karacapilidis. The Zeno argumentation framework. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, pages 10–18, Melbourne, Australia, 1997. ACM Press.

18. T. F. Gordon, H. Prakken, and D. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10-11):875–896, 2007.
19. T. F. Gordon and D. Walton. The Carneades argumentation framework — using presumptions and exceptions to model critical questions. In P. E. Dunne and T. J. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, pages 195–207, Amsterdam, September 2006. IOS Press.
20. J. Hage. A theory of legal reasoning and a logic to match. *Artificial Intelligence and Law*, 4(3-4):199–273, 1996.
21. H. L. A. Hart. *Essays in Jurisprudence and Philosophy*. Oxford University Press, 1983.
22. W. Kunz and H. W. Rittel. Issues as elements of information systems. Technical report, Institut für Grundlagen der Planung, Universität Stuttgart, 1970. also: Center for Planning and Development Research, Institute of Urban and Regional Development Research. Working Paper 131, University of California, Berkeley.
23. P. Lorenzen and K. Lorenz. *Dialogische Logik*. Wissenschaftliche Buchgesellschaft,, Darmstadt, 1978.
24. H. Prakken. Modeling defeasibility in law: Logic or procedure? *Fundamenta Informaticae*, 48:253–271, 2001.
25. H. Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, pages 85–94, New York, 2005. ACM Press.
26. H. Prakken. A formal model of adjudication. In S. Rahman, editor, *Argumentation, Logic and Law*. Springer Verlag, Dordrecht, 2008.
27. H. Prakken, C. Reed, and D. Walton. Dialogues about the burden of proof. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, pages 85–94, Bologna, 2005. ACM Press.
28. H. Prakken and G. Sartor. A dialectical model of assessing conflicting argument in legal reasoning. *Artificial Intelligence and Law*, 4(3-4):331–368, 1996.
29. H. Prakken and G. Sartor. Presumptions and burden of proof. In T. van Engers, editor, *Legal Knowledge and Information Systems. JURIX 2006: The Nineteenth Annual Conference*, pages 21–30, Amsterdam, 2006. IOS Press.
30. H. Prakken and G. Sartor. Formalizing arguments about the burden of persuasion. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 97–106, New York, 2007. Stanford University, ACM Press.
31. H. Prakken and G. Sartor. A logical analysis of burdens of proof. In H. Kaptein, H. Prakken, and B. Verheij, editors, *Legal Evidence and Proof: Statistics, Stories, Logic*, Applied Legal Philosophy Series. Ashgate Publishing, 2009.
32. J. Rawls. Outline of a decision procedure for ethics. *Philosophical Review*, 1951.
33. N. Rescher. *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge*. State University of New York Press, 1977.
34. H. W. Rittel and M. M. Webber. Dilemmas in a general theory of planning. *Policy Science*, 4:155–169, 1973.
35. M. Rosenberg, J. B. Weinstein, H. Smit, and H. L. Korn. *Elements of Civil Procedure*. Foundation Press, 1976.
36. G. Sartor. Defeasibility in legal reasoning. In *Informatics and the Foundations of Legal Reasoning*, Law and philosophy library, pages 119–157. Kluwer Academic Publishers, Dordrecht, 1995.
37. K. Satoh, S. Tojo, and Y. Suzuki. Formalizing a switch of burden of proof by logic programming. In *Proceedings of the First International Workshop on Juris-Informatics (JURISIN 2007)*, pages 76–85, Miyazaki, Japan, 2007.
38. D. Walton. The new dialectic: A method of evaluating used for some purpose in a given case. *ProtoSociology*, 13:70–91, 1999.
39. D. Walton. *Fundamentals of Critical Argumentation*. Cambridge University Press, Cambridge, UK, 2006.
40. D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, 2008.