Strategic Argumentation:
A Game Theoretical Investigation

Bram Roth
CIRSFID, Università di Bologna
Via Galliera 3-I-40141, Bologna, Italy
abrahamcornelis.roth@unibo.it

Antonino Rotolo
CIRSFID, Università di Bologna
Via Galliera 3-I-40141, Bologna, Italy
antonino.rotolo@unibo.it

Régis Riveret
CIRSFID, Università di Bologna
Via Galliera 3-I-40141, Bologna, Italy
regis.riveret@unibo.it

Guido Governatori
School of ITEE, University of Queensland
Brisbane, Queensland, QLD 4072, Australia
guido@itee.uq.edu.au

ABSTRACT

Argumentation is modelled as a game where the payoffs are measured in terms of the probability that the claimed conclusion is, or is not, defeasibly provable, given a history of arguments that have actually been exchanged, and given the probability of the factual premises. The probability of a conclusion is calculated using a standard variant of Defeasible Logic, in combination with standard probability calculus. It is a new element of the present approach that the exchange of arguments is analysed with game theoretical tools, yielding a prescriptive and to some extent even predictive account of the actual course of play. A brief comparison with existing argument-based dialogue approaches confirms that such a prescriptive account of the actual argumentation has been almost lacking in the approaches proposed so far.

Keywords

Argumentation, game theory, predictive force

1. INTRODUCTION

Over the years a lot of dialogue games of legal argument have been proposed [24, 25, 19, 5, 11, 12]. These models have shed light on questions such as which conclusions are (defeasibly) justified, or how legal procedure should be structured to arrive at a fair and just outcome.

Other aspects of legal debate have received no or hardly any attention yet, however. One of these is in the common sense observation that the outcome of a legal debate does not depend solely on the factual premises of a case, their measure of probability and the applicable law, but also on the strategies that parties in a dispute actually adopt. At first sight one may conclude from this observation that the outcome of a dispute cannot be predicted at all, and content oneself with setting the rules of the procedure. As illustrated in the following, however, one can also approach legal argument from a game theoretical angle, and try to apply the powerful mathemati-

cal apparatus of that field. As it turns out, then, game theoretical concepts like strategy dominance can be of great help to predict the actions the parties will actually take. Given these actions, it is then also possible to predict the outcome of the game in terms of the probability that the initial claim is defeasibly provable.

Accordingly, the present approach will start from two principles: (1) the outcome of a dispute depends on the strategies actually adopted by parties, but (2) this does not mean that the outcome can never be predicted because by using game theoretical solution concepts, the actions themselves can often be found.

Let's turn to an example. A worker is dismissed for having caused considerable damage to company property and for having lost credit from his superiors. The worker challenges the dismissal in court, claming that it can be voided. He argues that the working atmosphere has not been affected and that he is highly esteemed as a colleague. The following four legal rules are in place. The first says that a dismissal can be voided if considerable damage was done to company property but the working atmosphere was not affected. The second says that a dismissal can be voided if the worker lost credit from his superiors but is highly esteemed as a colleague. The third says that the dismissal cannot be voided if the working atmosphere was not affected but the worker lost credit from his superiors. The fourth rule says that the dismissal cannot be voided if the worker is highly esteemed as a colleague but did considerable damage to company property. It is commonly accepted by both the worker and the employer that it is very probable that the worker is highly esteemed as a worker, less probable that the working atmosphere has not been affected, even less probable that the damage was considerable, and the least probable that the worker lost credit from his superiors.

Which strategy is best for the worker? Should he use the first rule first and then see what the employer does? And if the employer uses the third rule, should the worker use the second rule or stick to the first? Likewise, what should the employer do? More generally, what will both parties do in the dispute?

For analysing the example it is convenient to introduce atomic sentences that abbreviate the conclusion that is at stake and the reasons that parties adduced in their arguments (see Table 1).

This paper is organised as follows. First, in Section 2 we briefly present the logic on which the argumentation system is based. In Section 3 we define the notion of an argument. In Section 4 we present the dialectical layer. Section 5 provides the procedural layer that prescribes how a dialogue can be conducted. In Section 6 we explain how we weight conclusions in terms of the probability that they are defasible provable. In Section 7 we apply game theory to

literal	meaning
c	dismissal-can-be-voided
a	working-atmosphere-not-affected
b	considerable-damage-to-company
d	lost-credit-from-superiors
e	highly-esteemed-as-colleague

Table 1: Literals with their meaning.

the example, in effect demonstrating a means to address the heuristic layer of legal argument [25]. In Section 8 we suggest an extension of the game that involves, amongst others, strict derivability next to its defeasible counterpart. In Section 9 we compare the present approach to existing and well-known dialogue models of legal argument in AI and Law, concluding that they have less prescriptive use than ours. In the Conclusion the point on predictive use is briefly recapitulated, and ways of generalising the present approach are suggested, for instance by allowing for subjective assignments of probability.

2. LOGIC LAYER: DEFEASIBLE LOGIC

Defeasible Logic [21] is a non-monotonic logic based on a programming-like language. It is a simple but flexible non-monotonic formalism suitable for dealing with many different intuitions of defeasible reasoning. It has been applied in many fields and had been proposed as the appropriate language for executable regulations [2, 3], contracts [13], automated negotiation [14] and for programming cognitive agents [7, 16].

A Defeasible Logic (as formalized by [6]) theory is a structure $D = (F, R, >)$ where F is a finite set of factual premises, R a finite set of rules, and $>$ a superiority relation on R. Factual premises are factual statements, for example, "John is a minor"which is formally written as $minor(John)$. Rules can be strict, defeasible or defeaters. Strict rules are rules in the classical sense; whenever the premises are indisputable (e.g. factual premises) then so is the conclusion. An example of a strict rule is "Minors are persons"which is formally written as $r1 : minor(X) \rightarrow person(X)$. Defeasible rules are rules that can be defeated by contrary evidence. An example of a defeasible rule is "persons have legal capacity"; formally, $r2 : person(X) \Rightarrow legalcapacity(X)$. Defeaters are rules that cannot be used to draw any conclusion. Their only use is to prevent conclusions by defeating defeasible rules. An example of such rules is "minors might not have legal capacity", which is formally expressed as $r3 : minor(X) \rightsquigarrow legalcapacity(X)$. The idea here is that even if we know that an individual is minor this fact is not sufficient evidence for the conclusion that it does not have legal capacity. The superiority relation between rules indicates the relative strength of each rule. That is, stronger rules override the conclusions of weaker rules. For example, if $r3 > r2$ then $r3$ overrides $r2$.

Given a set R of rules, we denote the set of all strict rules in R by R_s, the set of strict and defeasible rules in R by R_{sd}, the set of defeasible rules in R by R_d, and the set of defeaters in R by R_{dft}. $R[q]$ denotes the set of rules in R with consequent q. In the following $\sim p$ denotes the complement of p, that is, $\sim p$ is $\neg q$ if $p = q$, and $\sim p$ is q if p is $\neg q$. For a rule r we will use $A(r)$ to indicate the body or antecedent of the rule and $C(r)$ for the head or consequent of the rule. A rule r consists of its antecedent $A(r)$ (written on the left; $A(r)$ may be omitted if it is the empty set) which is a finite set of literals, an arrow, and its consequent $C(r)$ which is a literal. In writing rules we omit set notation for antecedents.

Conclusions are tagged according to whether they have been derived using defeasible or strict rules. So, a conclusion of a theory D is a tagged literal having one of the following four forms:

1. $+\Delta q$, which means that q is stricly provable in D;

2. $-\Delta q$, which means that q is not stricly provable in D;

3. $+\partial q$, which means that q is defeasibly provable in D;

4. $-\partial q$, which means that q is not defeasibly provable in D.

Provability is based on the concept of a derivation (or proof) in D. A derivation is a finite sequence $P = (P(1), \ldots, P(n))$ of tagged literals. Each tagged literal satisfies some proof conditions. A proof condition corresponds to the inference rules corresponding to one of the four kinds of conclusions we have mentioned above. $P(1..i)$ denotes the initial part of the sequence P of length i. Here we state the conditions for strictly and defeasibly derivable conclusions:

If $P(i+1) = +\Delta q$ then
    (1) $q \in F$, or
    (2) $r \in R_s[q], \forall a \in A(r) : +\Delta a \in P(1..i)$.

If $P(i+1) = +\partial q$ then
    (1) $+\Delta q \in P(1..i)$, or
    (2) (2.1) $\exists r \in R_{sd}[q] \forall a \in A(r) : +\partial a \in P(1..i)$ and
        (2.2) $-\Delta \sim q \in P(1..i)$ and
        (2.3) $\forall s \in R[\sim q]$ either
            (2.3.1) $\exists a \in A(s) : -\partial a \in P(1..i)$ or
            (2.3.2) $\exists t \in R_{sd}[q] \ \forall a \in A(t) :$
                $+\partial a \in P(1..i)$ and $t > s$.

A positive defeasible derivation consists of three phases: an argument in favour of the literal we want to prove is proposed. In the simplest case, this consists of an applicable rule for the conclusion (a rule is applicable if its antecedent has already been proved). Second, all counter-arguments are examined (rules for the opposite conclusion). Third, all the counter-arguments have to be rebutted (the counter-argument is weaker than the pro-argument) or undercut (some of the premises of the counter-argument are not provable).

Observe that all factual premises in F are strictly derivable according to the definition, and are therefore defeasibly derivable as well. Conversely, *basic* factual premises which do not appear as a consequent of any rule (see later on) must be strictly derivable if they are defeasibly derivable. A theory containing among its rules only the rule $a \Rightarrow b$, for instance, must contain a among the factual premises if a is to be defeasibly derivable. In other words, for such basic factual premises the notions of strict and defeasible derivability coincide. Later on we will assign probabilities to the defeasible derivability of basic factual premises, as input information for calculating the probability of the defeasible derivability of conclusions. In light of the equivalence for basic factual premises of strict and defeasible derivability, the probabilities thus assigned to their defeasible derivability also hold for their strict derivability.

Let's illustrate the proof conditions with the following theory $D_{dismissal}$,

$$F = \{a, b, d, e\}$$
$$R = \{r1 : a, b \Rightarrow c,$$
$$r2 : d, e \Rightarrow c,$$
$$r3 : a, d \Rightarrow \neg c,$$
$$r4 : b, e \Rightarrow \neg c\}$$
$$> = \emptyset$$

All the rules are applicable so we get $-\partial c$ and $-\partial c$. Hereafter, the superiority relation between rules is disregarded in order to avoid unnecessary technicalities. This restriction does not affect the generality of the approach: a modular transformation given in [1] enables to empty the superiority relation. The transformation takes a theory $D = (F, R, >)$ as input and builds from it a theory $D' = (F, R', \emptyset)$ with the desired properties. Note that such a transformation may change the game's characteristics in terms of the arguments exchanged, though.

In the following, we briefly present the argumentation system that provides an argumentation semantics of Defeasible Logic. This system is fully studied in [15].

3. THE ARGUMENT LAYER

The argument layer defines arguments. An argument for a literal is a proof tree (or monotonic derivation) of that literal. Nodes are literals and arrows connecting nodes correspond to ground instances of rules.

Definition 1. An argument for a literal γ based on a set of rules R is a (possibly infinite) tree with nodes labelled by literals such that the root is labelled by α and for every node labelled by β:

1. If $\alpha_1, \ldots, \alpha_n$ label the children of β then there is a ground instance of a rule in R with body $\alpha_1, \ldots, \alpha_n$ and head β.

2. If this rule is a defeater then β is the root of the argument.

3. The arcs in a proof tree are labelled by the rules used to obtain them.

Condition 2 specifies that a defeater may only be used at the top of an argument, or in other words, no chaining of defeaters is allowed.

A (proper) sub-argument of an argument A is a (proper) sub-tree of the tree associated to A. A literal γ is a conclusion of an argument A if and only if γ labels a node of A. A more intuitive alternative would be to regard only the root of an argument as the unique conclusion of an argument, but this choice would make the other definitions more complex. Given a defeasible theory D, the set of arguments that can be generated from D is denoted by $Args_D$. For example, the set $Args_{D_{dismissal}}$ of arguments that can be generated from $D_{dismissal}$ is illustrated in Figure 1.
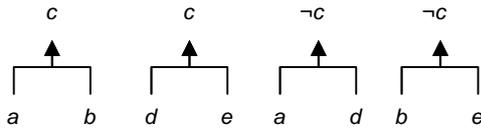

Figure 1: The available arguments in the game.

Sometimes we need to differentiate between arguments, depending on the rules used. A supportive argument is a finite argument in which no defeater is used. A strict argument is an argument in which only strict rules are used. An argument that is not strict is called defeasible.

4. THE DIALECTICAL LAYER

The previous section defined the argument layer and isolated the concept of argument. This section presents the dialectical layer which is concerned with relations amongst arguments. It defines the notion of support and attack, and focuses on the interaction amongst arguments. The dialectical layer is not meant primarily to give an alternative to the proof theory defined earlier. Instead it is meant to introduce some basic notions such as attack, which are used later in the definition of the game that is presented in the next section on the procedural layer. Firstly, we introduce the notion of support:

Definition 2. A set of arguments S supports a defeasible argument A if every proper sub-argument of A is in S.

Note that, in our setting, the atomic arguments, constituted of a factual premise or a rule of the theory, are supported by the empty set. At the opposite of the notion of support, stands the notion of attack. Roughly, an argument attacks another argument if the former supports a literal in conflict with a literal of the latter.

Definition 3. An argument A attacks an argument B iff a conclusion in A is the complement of a conclusion of B, and that conclusion of B is not part of a strict argument of B.

Defeasible reasoning differentiates traditionally between rebuttal and undercutting. We stick to the tradition and define the notion of undercutting as follows:

Definition 4. A defeasible argument B is undercut by a set of arguments S if S supports an argument A attacking a proper sub-argument of B.

Comparing arguments by pairs is not enough since an attacking argument can in turn be attacked by other arguments ('reinstatement', [25]). As a remedy, the notion of the status of arguments is defined on the basis of all ways in which arguments interact. Based on the concept of an acceptable argument, it is possible to define justified arguments and justified conclusions, that is conclusions that may be drawn even taking conflicts into account.

Definition 5. An argument A is acceptable w.r.t. a set of arguments S, if A is finite and

1. A is strict, or

2. every argument attacking A is undercut by S.

Based on this concept we proceed to define justified arguments and justified literals. That an argument A is justified means that it resists every refutation. The following definition is based on [24] definitions of fixed point semantics.

Definition 6. The set of justified arguments in a defeasible theory D is $JArgs_D = \bigcup_{i=0}^{+\infty} J_{D,i}$ with

1. $J_{D,0} = \emptyset$ and

2. $J_{D,i+1} = \{A \in Args_D | A \text{ is acceptable w.r.t. } J_{D,i}\}$.

Definition 7. A literal γ is justified by $JArgs_D$ iff it is the conclusion of an argument in $JArgs_D$.

A literal that is justified means that it is provable $(+\partial)$. However, Defeasible Logic permits to express when a conclusion is not provable $(-\partial)$. Briefly, that a conclusion is not provable means that every possible argument for that conclusion has been refuted. In the following, this notion is captured by assigning the status rejected to arguments that are refuted. Roughly speaking, an argument is rejected if it has a rejected sub-argument or it cannot overcome an

attack from a justified argument. Given an argument A, a set S of arguments (to be thought of as arguments that have already been rejected), and a set T of arguments (to be thought of as justified arguments that may be used to support attacks on A), we assume the following definition of the argument A being rejected by S and T:

Definition 8. Let S and T be two sets of arguments. Then an argument A is rejected by S and T iff A is not strict and either a proper sub-argument of A is in S or it is attacked by an argument supported by T.

Definition 9. The set of rejected arguments in a defeasible theory D w.r.t. T is $RArgs_D(T) = \bigcup_{i=0}^{+\infty} R_{D,i}(T)$ with

1. $R_{D,0}(T) = \emptyset$ and

2. $R_{D,i+1}(T) = \{A \in Args_D | A$ is rejected by $R_{D,i+1}(T)$ and $T\}$.

Definition 10. A literal γ is rejected by T iff there is no argument in $Args_D - RArgs_D(T)$ that ends with a supportive rule for γ.

More generally, we say that an argument is rejected if it is rejected w.r.t. $JArgs_D$. and a literal is rejected if it is rejected by $JArgs_D$. A skeptical argumentation semantics can now be provided, i.e., defeasible and definite conclusions of Defeasible Logic are characterised in argumentation terms:

THEOREM 1. *Let D be a defeasible theory and γ be a literal.*

- $D \vdash +\Delta\gamma$ iff there is a strict argument for γ in $Args_D$;

- $D \vdash -\Delta\gamma$ iff there is no strict argument for γ in $Args_D$;

- $D \vdash +\partial\gamma$ iff γ is justified by $JArgs_D$;

- $D \vdash -\partial\gamma$ iff γ is rejected by $JArgs_D$.

This argumentation semantics is consistent with the proof conditions of Defeasible Logic in the sense that conclusions get similarly tagged as proved in [15]. It follows that for any defeasible theory, no argument is both justified and rejected, and thus no literal is both justified and rejected. Eventually, if the set $JArgs_D$ of justified arguments contains two arguments with conflicting conclusions then both arguments are strict. That is, inconsistent conclusions can be reached only when the strict part of the theory is inconsistent. Therefore we can call a defeasible theory consistent if and only if the strict part of theory is consistent.

5. THE PROCEDURAL LAYER

We now proceed with our definition of the protocol of our game to address the procedural layer. The protocol is not meant to be sound and complete w.r.t. the argumentation semantics of the previous section. Instead it merely aims at representing actual discourse, whereby the logic introduced in the foregoing serves the purpose of calculating the payoffs of the game, that is, probabilities of defeasible derivability.

As in most dialogue models of legal argument (e.g. [19]), there are two players in the game, the proponent (P) of some initially claimed conclusion, and the opponent (O) of the claim. It is the proponent's purpose to maximise the probability that the claimed conclusion is defeasibly provable, while it is the opponent's purpose to minimise this probability. In other words, the payoffs in the game are measured in terms of probability of defeasible derivability (see next section).

First the game theoretic notion of a history is introduced [22]. Informally, a history is a sequence of actions taken by the players. Formally, a history (denoted h) is a sequence $(A_k)_{k=1..n}$ of arguments $A_k \in Args_D$. The usual convention is used that if h denotes a history and A an argument, then (h, A) denotes the history that results if history h is followed by argument A ([22], p. 90).

Moreover, there is a player function (denoted P) that assigns to every history the player whose turn it is after that history. More formally, after each history $h = (A_k)_{k=1..n}$ it is player $P(h)$'s turn to move, whereby $P(h) = P$ (proponent, see above) or $P(h) = O$ (opponent). The proponent starts the game, after which players take turns.

At each subsequent move only arguments are allowed that defeat the latest argument, a requirement that intuitively ensures that as a matter of efficiency, the debate remains focused on the initial claim (cf. [24] on dialogues, pp. 21f.).

In sum, the player function and the allowed histories $h = (A_k)_{k=1..n}$ in the game adhere to the following protocol (cf. the 'dialogues' in [24], pp. 21f.):

1. $A_k \in Args_D$; and

2. $P(h) = P$ if n is odd and $P(h) = O$ if n is even; and

3. A_k attacks A_{k-1}.

By definition the game ends after a terminal history. Informally, a history will be defined terminal if and only if it ends with an action that does not introduce any new premises into the debate. For defining terminal histories formally, let $Prem(h)$ denote the union of all premise sets of all arguments appearing in h:

$$Prem(h) = \bigcup_{X \in h} Prem(X)$$

Formally, a history (h, X) is then terminal if and only if $Prem(h, X) = Prem(h)$. Note that in particular, a history is terminal if the same argument occurs twice in it.

The intuitive reason for defining terminal histories in this way is twofold. First, if no new premises are introduced, the final argument's premises are accepted by both players, so that on the condition that the debate has reached the terminal history, both have to concede its conclusion. Second, a player can thus always end a debate at a point where that player is content with its outcome. In particular, that player may do so by repeating an argument the player made earlier.

Technically it is not necessary to define terminal histories in this way, however. More 'liberal' (cf. [23] on liberal protocols) termination criteria are also possible here, and one may even allow for infinite histories such as $(C, E, C, E, \ldots)$. As it turns out, however, such a termination criterion does not essentially change the solution to the game, because at some point in the game it is always best not to introduce new premises anyway. The intuitive reason for this is that introducing premises gives the other player more opportunities for defeating counterarguments. Accordingly, for keeping the game simple the present account will stick to the termination criterion that an argument did not introduce any new premises.

6. WEIGHTING CONCLUSIONS

As remarked above, the payoffs in the game are measured in terms of the probability that the initially claimed conclusion is defeasibly provable. It is proponent's purpose to maximise this probability while opponent's purpose is to minimise it.

Intuitively, the probability that a statement is defeasibly derivable is the likelihood that it is accepted by people. As a practical

matter this likelihood could be estimated, for instance, by studying case law on the relative frequency with which a premise was considered true by a judge. More formally, a probability is a function p that associates a unique number between 0 and 1 to the state of affairs that some statement γ is defeasibly provable $(+\partial \gamma)$, or not defeasibly provable $(-\partial \gamma)$. The probability of argument premise a, for instance, is denoted $p(+\partial a)$. If $p(+\partial a) = 0.9$, for example, then this intuitively means that there is a 90% chance that premise a is going to be accepted. The mechanism with which the probability of a conclusion is derived from the probability of the premises is based on standard probability calculus. More sophisticated methods for representing uncertainty are, for instance, to be found in possibility theory [9, 29] or the Dempster-Shafer theory [8, 28].

Accordingly, every assignment of probability is assumed to adhere to a probabilistic principle of the excluded middle, to the effect that the probability that some statement is defeasibly provable, and the probability that it is not defeasibly provable, add up to one. More formally, this principle holds that for any conclusion γ, we have $p(+\partial \gamma) + p(-\partial \gamma) = 1$. This is in line with the following result for standard Defeasible Logic [1]:

THEOREM 2. *Let D be a defeasible theory and let # denote any derivability in $\{\Delta, \partial\}$. Then there is no literal γ such that $D \vdash +\#\gamma$ and $D \vdash -\#\gamma$.*

The above theorem in fact states that no literal is simultaneously provable and demonstrably non provable; thus it establishes the coherence of Defeasible Logic.

The probabilities of *different* premises are assumed to behave like probabilities of *independent* events. In other words, the probability of a set of premises is obtained simply as the product of the probabilities of the individual premises. If the probability of premise a equals 0.9 and that of b equals 0.1, for instance, then the probability of both a and b equals $0.9 \times 0.1 = 0.09$. Below this principle of independence is used to obtain the probability of a conclusion under some argument, given the probabilities of the argument's premises. Note that such a principle of independence constrains assignment of probability to literals that do not appear in the head of any rule of the theory. A theory containing the facts a and b and the rule $a \Rightarrow b$, for instance, does not allow an independent assignment of probability to the defeasible derivability of the factual premises a and b. A possible generalisation of the present approach would be to allow for such dependency among premises.

Note also that the approach can be generalised in another way, namely by allowing different players to assign different subjective probability to factual premises, and thus different subjective probability to the main conclusion. To model the situation as a game with complete information regarding payoffs (cf. [4], Ch. 1 and 2), it suffices that these subjective probabilities are common knowledge ([22], p. 73) among the players, that is, each player knows the other player's probabilities, knows that the other player knows the former's probabilities, and so on.

In the course of the game the probability of the main conclusion is updated in accordance with the arguments that have actually been played out by parties. To do so, we reconstruct the theory in Defeasible Logic on which the arguments are based by considering the set $R(h)$ of strict and defeasible rules and the set $FP(h)$ of factual premises (expressed by literals) introduced during the history h. For any subset of the factual premises and the set $R(h)$ of rules one has a theory from which one can build an extension to see which conclusions follow. Note that this can be done in a time linear in the size of the theory [20]. We then consider cases that can occur according to the reconstructed theory by means of the following constructions:

- The defeasible theory $D(h)$ after history h is $(FP(h), R(h), \emptyset)$;
- the set of basic factual premises after history h is $BFP(h) = FP(h) - \{\gamma | r \in R(h),\ \gamma = C(r)\}$;
- $Pow(BFP(h))$ is the power set of $BFP(h)$ and $BFP_i(h)$ denotes an element of $PowBFP(h)$;
- a case after history h is an ordered pair of sets of basic factual premises of the form $(BFP_i(h), (BFP(h) - BFP_i(h)))$;
- for any literal γ, a case $(BFP_i(h), (BFP(h) - BFP_i(h)))$ is a $case^{+\partial \gamma}$ (respectively a $case^{-\partial \gamma}$) if and only if $(BFP_i(h), R(h), \emptyset) \vdash +\partial \gamma$ (respectively $(BFP_i(h), R(h), \emptyset) \vdash -\partial \gamma$).

Note that the definition of a case does not refer to a conclusion. In this respect it is different from Roth's cases [27], where conclusions were treated as facts about cases that were as such included in the case representation.

We define the notion of the probability of a case as follows. A case is a combination of defeasible derivabilities of the basic factual premises of a theory. Furthermore, for these basic factual premises the notions of defeasible and strict derivability coincide and can be replaced just with derivability. Finally, the probability of a case is then simply the probability that all basic factual premises in it are derivable, while all others are not derivable. Given our assumption that basic factual premises behave like independent events, the corresponding probability is obtained as a product:

Definition 11. The probability of a case $(BFP_i(h), (BFP(h) - BFP_i(h))$ is the product of the probabilities of derivability of the basic factual premises in $BFP_i(h)$, and the probabilities of non-derivability of the basic factual premises in $(BFP(h) - BFP_i(h))$.

A complete set of cases w.r.t. a history h is the set of all possible cases after a history h. A complete set of cases w.r.t. a history h is denoted $S(h)$.

When we are interested in examining a complete set of cases w.r.t a history h, it is perhaps more perspicuous to represent it using a table where each column represents a different case. For example, suppose a history h in which the worker has played the argument $a, b \Rightarrow c$ and the employer $b, e \Rightarrow c$. The complete set $S(h)$ of cases w.r.t. to h can be represented in the format of Table 2.

X	C1	C2	C3	C4	C5	C6	C7	C8
a	1	0	1	0	1	0	1	0
b	1	1	0	0	1	1	0	0
e	1	1	1	1	0	0	0	0
c	0	0	0	0	1	0	0	0
$\neg c$	0	1	0	0	0	0	0	0

Table 2: Cases with their conclusions.

The rows indicate the defeasible provability $(+\partial)$ and non-provability $(-\partial)$ of literals respectively by 1 and 0. The columns indicate a particular case composed of the different tagged literals. For example the column $C5$ corresponds to the case $(\{a, b\}, \{e\})$, which is a $case^{+\partial c}$ and a $case^{-\partial \neg c}$.

We now turn our attention to the calculus of the probability of a conclusion w.r.t. a complete set of cases.

Definition 12. The probability $p(\pm \partial^h \gamma)$ of a conclusion $\pm \partial \gamma$ w.r.t. a complete set of cases $S(h)$ is the sum of the probabilities of all cases $case^{\pm \partial \gamma}$ in $S(h)$.

literal	meaning	probability
a	working-atmosphere-not-affected	0.8
b	considerable-damage-to-company	0.7
d	lost-credit-from-superiors	0.6
e	highly-esteemed-as-colleague	0.5

Table 3: Literals with their meaning and probability.

Turning to the example, it was considered highly probable that the worker was highly esteemed as a colleague, somewhat less probable that the working atmosphere had been affected, even less probable that considerable damage was done to company property and the least probable that the superiors lost credit in their worker. Putting this all together in a table and choosing numerical values for these probabilities we get Table 3.

Suppose a history $h = (B, E)$ in which the worker has played the argument B ($r1 : a, b \Rightarrow c$) and the employer E ($r4 : b, e \Rightarrow c$).

According to the complete set $S(h)$ (see Table 2), the probability $p(+\partial^{(B,E)}c)$ that the dismissed can be voided is:

$$
\begin{aligned}
p(+\partial^{(B,E)}c) &= p(+\partial a) \times p(+\partial b) \times p(-\partial e) \\
&= p(+\partial a) \times p(+\partial b) \times (1 - p(+\partial e)) \\
&= 0.8 \times 0.7 \times (1 - 0.9) \\
&= 0.056
\end{aligned}
$$

Note that the algebraic expression of the probability of a conclusion to be (un)provable w.r.t to a set of cases may be manipulated. For example, develop the right hand side and simplify it. However, we constrain those manipulations by forbidding to move a term from the right hand side to the left hand side and vice versa of algebraic expressions in order to keep track of the sense of causality.

Finally, the payoffs ([4], pp. 8f.) for both parties are defined as follows. Since it is proponent's (P) purpose to maximise the probability of the initially claimed conclusion, that player's payoff resulting from some terminal history will be defined as the probability of the conclusion after that history. Likewise, since it is opponent's (O) aim to minimise the probability that the conclusion is defeasibly provable, that player's payoff will be defined as the probability that the conclusion is not defeasibly provable.

Note that this approach can easily be generalised for situations with several possible claims, each having a certain 'value' or 'preference' to the players. In addition, we could use the distinction between strict and defeasible conclusions and assign different preferences or values to the modes through which each conclusion is obtained (see Section 8). In such situations the payoff for proponent could be defined as the expected value of the outcome, under the appropriate probability measure over the different claims. Such an approach could, for instance, also allow for the possibility of strategic choices between different claims that can be made. For present purposes, though, it suffices to consider only the probability of the one main disputed conclusion.

This concludes the discussion of the formalism proper. In the next section the example will be thoroughly analysed using this formalism, in combination with some game theoretical notions such as strategy dominance (cf. [22], pp. 59), which are widely accepted and applied in that field.

7. ANALYSIS OF THE EXAMPLE

The question now is how parties are going to act in the game. In other words, the question is which strategies the players are likely to adopt. Technically, a strategy of a player is a prescription of how to act at each decision node where it is that player's turn to move

([22], pp. 92). In the game tree depicted in Figure 2, for instance, one can see that the worker has a decision node at the beginning of the game but that there are many more, such as the node after the history (C, D). The employer has decision nodes as well, for instance those after the histories (C, D, B) and (B, C, E). How can one isolate the combination of strategies that the players are most likely to adopt? This can be done easily in this example by using backwards induction ([4], pp. 50f.). Briefly, backwards induction means that one starts at a player's final decision nodes to see what a player will do there, and then reason backwards to tell which action is best for the other player. There are a number of nodes at which it is the employer's turn to move, and where the employer can end the game with a maximal payoff for himself. Examples are the nodes after the history (C, D, B) and the history (B, D, C).

Obviously the worker will avoid histories ending in such a node. He can do this by ending the game before such a decision node for the employer is reached, for instance by playing C after (C, D) or B after (B, D).

The employer, in turn, will know this. He will expect, for instance, to end up in history (C, E, C) if he chooses to play E after history (C), and in history (B, D, B) if he chooses to play D after history (B).

The worker, in turn, knows that the employer knows that the game will end in this way. He will therefore infer that the employer will prefer history (B, E, B) to history (B, D, B), the former having a higher payoff (0.944) for the employer than the latter (0.776). Likewise, the employer will prefer history (C, D, C) (payoff 0.892) to history (C, E, C) (payoff 0.838). In sum, the employer will play E after history (B) and D after history (C).

The worker, knowing this, will therefore choose to play C first because that leads to history (C, D, C) with a payoff 0.108 for the worker, higher than the worker's payoff of 0.056 associated with history (B, E, B). In sum, the game will be played following history (C, D, C) with payoffs of 0.108 and 0.892 for worker and employer, respectively.

A number of interesting observations on this result can now be made. First, observe that both the worker and employer will not end up with the best overall result they could get. For the worker this best result is 0.332 associated, for instance, with history (B, D, C, D), for the employer it is 0.944 with history (B, E, B) leading to it. In other words, both players have to compromise to some extent, and in this sense there is no winning strategy for either of them. As the solution of the game shows, however, there are strategies that are dominant in the game theoretical sense, that is, strategies that are at least as good as any other strategy, for each strategy the other player may choose. This point will return later on in the discussion of related work (next section).

A second observation is that the players will not play the arguments that are a priori the strongest, that is, considering the probabilities of their premises in isolation, irrespective of the histories in which they appear. Consider, for instance, the worker's actions C and B. Consider the a priori probability of conclusion c on the basis of each of these arguments:

$$
\begin{aligned}
p(+\partial^{(C)}c) &= p(+\partial d) \times p(+\partial e) = 0.6 \times 0.9 = 0.54 \\
p(+\partial^{(B)}c) &= p(+\partial a) \times p(+\partial b) = 0.8 \times 0.7 = 0.56
\end{aligned}
$$

In other words, this means that argument B seems a priori a stronger argument than argument C since $p(+\partial^{(B)}c)$ exceeds $p(+\partial^{(C)}c)$. However, as the game solution shows, the worker will nevertheless play the latter argument rather than the former. The reason for this is that if the worker plays out the a priori stronger argument B,
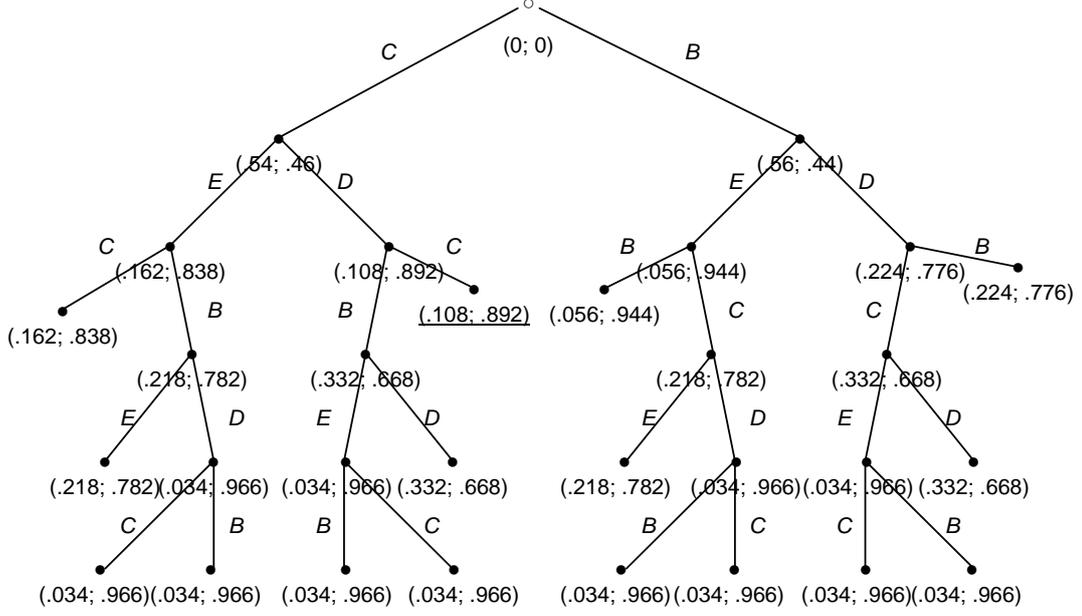
Figure 2: Game tree with moves and probabilities. Note that the game ends if no new premises have been introduced in the last move, in accordance with the termination criterion. The payoffs for Worker, Employer are the probabilities at the end nodes.

then thanks to the introduction of that argument's premise b the employer can respond all the more strongly by playing out argument E, which apart from b only relies on the very probable premise e.

Likewise, the employer's argument E seems a priori stronger than argument D:

$$p(-\partial^{(D)}c) = p(+\partial a) \times p(+\partial d) = 0.8 \times 0.6 = 0.48$$
$$p(-\partial^{(E)}c) = p(+\partial b) \times p(+\partial e) = 0.7 \times 0.9 = 0.63.$$

In other words $p(-\partial^{(D)}c)$ is smaller than $p(-\partial^{(E)}c)$. As the game solution shows, however, the employer will still play argument D rather than E. That is because argument E is only stronger a priori because of its highly probable premise e, which is available anyway in response to the worker's argument C. At the same time, however, argument's D's relative improbable premise d has then also become available to the employer.

What these observations show is that for determining which arguments to play the arguments cannot be assessed in isolation from other arguments. Rather, the complete strategic interaction between the players has to be taken into consideration, as was done to arrive at the game solution. In particular, each player has to take into account which premises are being given away by playing out an argument, because these premises in part determine the opportunities for counterargument.

8. FUTURE EXTENSIONS

In the previous sections we have considered a game of defeasible provability. In other words, we analysed, within a game-theoretical perspective, when payoffs are measured in terms of the probability that the claimed conclusion γ is, or is not, defeasibly provable, i.e., whether $+\partial\gamma$ or $-\partial\gamma$. In isolation, a similar procedure can be devised for strict derivation, namely, when $+\Delta\gamma$ and $-\Delta\gamma$ are

under consideration. The interesting issue is when the two games are jointly analyzed, thus playing a game of generic provability. Consider the following theorems [1]:

THEOREM 3. *Let D be an acyclic defeasible theory. For any literal γ, $D \vdash +\partial\gamma$ and $D \vdash +\partial \sim \gamma$ iff $D \vdash +\Delta\gamma$ and $D \vdash +\Delta \sim \gamma$.*

This theorem gives the consistency of Defeasible Logic. In particular, it affirms that it is not possible to obtain conflicting literals unless the information given about the environment is itself inconsistent. Thus, in the case of consistent theories, the following corollary holds:

COROLLARY 1. *Let D be a consistent defeasible theory. For any literal γ, if $D \vdash +\partial\gamma$, then $D \vdash -\partial \sim \gamma$.*

Another basic theorem is the following:

THEOREM 4. *Let D be a consistent defeasible theory. For any literal γ, if $D \vdash +\Delta\gamma$, then $D \vdash +\partial\gamma$.*

On the basis of Corollary 1 and Theorem 4, we can state that the probability of a strict derivation for a literal γ does not exceed the probability of a defeasible derivation for γ. In addition, the probability of proving that there is no defeasible derivation in favour of $\sim \gamma$ is smaller than the probability of a defeasible derivation for γ and greater than the probability of proving that there is no strict derivation for $\sim \gamma$:

THEOREM 5. *Let D be a consistent defeasible theory. For any literal γ, the following conditions hold:*

$$p(+\Delta\gamma) \leq p(+\partial\gamma) \text{ and } p(-\partial\gamma) \leq p(-\Delta\gamma)$$

However, arguably a rational player should look first for strict arguments in support of the wanted conclusion, as these cannot be defeated. More generally, one can consider strict and defeasible derivability in combination with a conclusion or its opposite. Specifically, if the wanted conclusion is γ, the player's preference relation over these combinations is arguably as follows:

$$+\Delta\gamma \succ +\partial\gamma \succ -\partial \sim \gamma \succ -\Delta \sim \gamma$$

Hence, if we play a game of generic derivability, we have to balance, for example, the fact that probability for $+\Delta\gamma$ is smaller than that for $+\partial\gamma$ with the fact that the goal $+\Delta\gamma$ should be preferred. In the simplest case, this would require to assign a preference value also to the different modes through which conclusions are obtained, so that the payoffs would be measured in terms of the product of the probability that the claimed conclusion is, or is not, provable with the preference value assigned to the mode through which the conclusion is claimed to be proved. This extension is left for future research.

9. RELATED WORK

Prakken and Sartor's Formal Dialogue Game [24] is a dialectical model of legal argument, in the sense that arguments can be attacked with appropriate counterarguments.

The building blocks of arguments are rules, that is, conditional statements that assign a conclusion to a set of conjunctive conditions. Arguments are formed by chaining rules, where each of a rule's conditions is the conclusion of some rule appearing earlier in the chain (p. 10). Arguments can defeat other arguments in a number of ways, the commonest and best known of which are rebutting and undercutting. An argument rebuts another if it contains a rule with a conclusion opposite to that of a rule in the rebutted argument. Arguments can thus rebut each other, but if one of the rules involved has priority over the other, the argument with the higher priority rule may be not defeated (p. 16). The second way in which an argument can defeat another is by undercutting it. This way of defeating an argument comes down to attacking an assumption behind a rule. Unlike a condition, an assumption can normally be left out of consideration for the purpose of applying the rule. Technically, assumptions are represented by weak negation (p. 8), which has the informal meaning that there is no evidence to the contrary of the assumption. If an argument defeats another but not the other way round, then the former is said to defeat the latter strictly.

There are two players in the Dialogue Game, the proponent of a claimed legal conclusion and the opponent. The aim of the proponent is to prove that its initial argument is justified, which roughly means that it can be upheld against all possible defeating counterarguments. The aim of the opponent is to show that the initial argument is not justified because it can be successfully attacked.

A dialogue is a sequence of arguments (p. 21), regulated by a protocol that restricts the set of allowed arguments at each stage of the debate. The proponent of a claimed conclusion starts the dialogue and then the players take turns. To reflect the dialectical asymmetry between proponent and opponent, the former's arguments are required to be strictly defeating while the latter's may be merely defeating.

A dialogue terminates if the set of available arguments becomes empty, so that the player whose turn it is cannot move any more. A player that can leave the other in that position is said to win the dialogue, while the player who cannot move anymore loses it. A dialogue tree is a tree of dialogues, with the property that the opponent seized all its opportunities for defeat at each of that player's turns. In this way it is ensured that the initial claim is tested against all possible attacks, making dialogue trees good candidates for being

dialectical proofs of conclusions. A conclusion is justified, then, if it is a conclusion of the initial argument of some dialogue tree that is won by the proponent. A player is said to win a dialogue tree, finally, if the player wins all the dialogues in the tree.

How does Prakken and Sartor's model compare to the present approach? For a start they regard the factual premises of their arguments as indisputable since they treat them as strict rules with empty antecedents. Accordingly, they do not allow for an approach in which the premises are assigned probabilities that are less than one. Furthermore, the outcome of the dialogue game does not depend on the way in which it is actually played, since the status of a claim as justified or overruled is defined in terms of all dialogues involving all possible defeating arguments of the proponent. In accordance with this, their model does not attempt to formulate strategic or predictive criteria for which arguments are actually going to be made, in contrast to the present approach.

The Pleading's Game [11, 12] is a normative formalisation and computational model of civil pleading. Its purpose is to regulate pleading between parties engaged in a legal conflict, which is meant to identify the issues at dispute before the case is tried in court. After pleading the court is then only confronted with issues on which no agreement could be reached. Gordon's model is normative in the sense that it prescribes the procedural rules of pleading as they should be according to standards like efficiency and fairness.

There is no judge or referee and there are only two players involved in Gordon's Game, namely the proponent of the main claim and the opponent. Before the actual pleading starts the main claim is known, as well as a set of statements for each player. Statements can be seen as utterances concerning a formula of some logic (conditional entailment, a kind of default logic, see [10]), for instance the statement that some formula is claimed or that a set of formulas is a argument for some formula (p.126). Pleading comes down to making assertions about statements, formula sets or rules, for instance conceding a statement or declaring a rule (p.127). Such assertions can be considered the moves of Gordon's Game.

Each player has a set of open statements, a set of denied statements and one of conceded statements. The open statements are the ones that have not yet been responded to by the other party, while the denied and conceded statements are the ones that have been denied and conceded, respectively. The rules of the game (pp. 131–133 and pp. 137–140) say under which conditions the different moves are allowed and which effects each move has on the players' sets of open, denied or conceded statements.

Pleading ends if all relevant statements have been answered by an appropriate assertion, such as denying or conceding a claim. Briefly, a statement is relevant if its subject formula is an issue with respect to the main claim, or if it is the denial of a relevant statement (p. 141). Stated very roughly and briefly, an issue is a claimed formula that is not known to be derivable given the conceded statements, that is, neither the subject of a conceded statement nor entailed by formulas known to be derivable (cf. p. 130 for the technical meaning of 'known' in this context). Furthermore, the issue must be relevant to the claim in the sense that it plays a role in an argument for or against the claim, or for or against another issue with respect to the claim (see p. 164 for an exact definition).

The conceded statements of both parties together establish a set of claims conceded as facts, and rules whose declaration is conceded (pp. 128–129). The proponent of the main claim has now won the game in the end if there are no issues left and the claim is entailed by the agreed facts and rules. The legal reward of winning the pleading is that the proponent is entitled to a summary judgment, that is, a judgment that is made routinely in favour of the

proponent. The opponent wins if there are no issues left and the main claim is not entailed. If neither player wins the game ends in a draw and the case has to be tried by the court.

How does Gordon's model compare to the present approach? The model does not allow for treating premises as having a probability that does not equal one, unlike the present approach. The purpose of the game is to identify the issues that have to be tried by a judge. In a sense, then, both players are winners since both benefit from clarifying the issues that divide them. Gordon focuses on defining dialogue rules that guarantee a fair and just procedure, rather than on prescribing or even predicting the actual course of pleading. This is in contrast to the present approach where such an attempt is made.

Lodder's DiaLaw [19] is a (Prolog implemented) dialogue model of legal justification, incorporating the idea of the pure procedure approach [26] in which there is nothing but the procedure itself to arrive at justification. The model takes the form of a dialogue game between two players. It is the purpose of the game to convince one's opponent of that some claim is justified. The actions that can be taken are illocutionary acts with a propositional content that is expressed by some sentence of a logical language (extended somewhat, pp. 47f.; [18]). There are four types of speech act that can be made, namely claiming a sentence, questioning a sentence, accepting a sentence, or withdrawing a sentence claimed earlier.

A dialogue is a sequence of moves, where each move also includes the player that makes it. There are dialogue rules that determine the moves that can be made at each stage of the dialogue.

The dialogue ends as soon as the termination criterion is fulfilled, which informally says that there are no disputed (or 'open', p. 36) sentences left (recall that the purpose is persuasion). To define the set of disputed sentences a commitment store is used, telling which players are committed to which sentences. The commitment store starts empty (p. 36). There are rules that govern the way commitments can change after each move (p. 40 and pp. 50f.), such as the rule that a player becomes committed to a sentence he claimed himself. Special commitment rules concern forced changes in commitment, based on a logic of legal reasoning with rules [17, 18]. An example is that by being committed to a legal rule's validity and to that rule's condition one also becomes committed to the sentence that the rule is applicable (p. 53).

Given the commitment store the set of disputed or sentences is defined as the set of sentences occurring in the commitment store, with the property that only one of the players is committed to it (p. 36). The dialogue ends, then, if the set of disputed sentences is empty. This can happen only if the original claim is accepted by the other player, or if the claim is withdrawn.

A player can be said to win the dialogue if the initial claim is accepted by the other player, and to lose the game if forced to withdraw the initial claim. In the former case the initial claim is justified, in the latter it is not.

How does Lodder's DiaLaw relate to the present work? As was the case for Prakken and Sartor's and Gordon's approach, Lodder does not assign probabilities to the factual premises involved in the debate. His purpose is to define the rules of legal debate in such a way that it can serve as a model of legal justification through persuasion. Again, in a sense both players involved can be considered winners, since both are guaranteed a fair procedure. Accordingly, no attempt is made to prescribe or even predict the actual course of play, contrary to the present approach.

10. CONCLUSIONS

We have treated legal argumentation as a game in its game the-oretical sense, allowing for an account that enables us to prescribe and to some extent even predict the arguments that are actually going to be played out by the parties. A primitive notion thereby is probability, which intuitively is a measure of strength of the defeasible proof for a conclusion. We used Defeasible Logic in combination with standard probability calculus that a defeasible proof holds, on the basis of the probabilities assigned to factual premises. This probability of a claim was then interpreted in the game theoretical sense as the payoff for the proponent of the claim. This allowed us to analyse an example and prescribe the strategies that players should adopt in the example.

A brief survey of existing argument games in AI and Law showed that the approach with probabilities has been almost lacking from these approaches so far, and that the same holds for the prescriptive and predictive character of the present account, which comes to light in the fact that the strategies the parties adopt are found by a game theoretical analysis.

A number of ways of generalising the present approach have suggested themselves along the way. First, one could allow the probabilities to be subjective or party-dependent in the sense that different players may assess the probability of the factual premises differently. The only requirement then is that each player's assignment of probability be common knowledge among both players. A second generalization would be to allow for different claims being raised at the same time, each of them having a different subjective preference value for the players. Then the final payoff for players could be obtained as an expected value of a variable that ranges over the set of valued claims raised. A third generalisation would be to incorporate also strict derivability next to its defeasible counterpart, into the definition of payoffs. These extensions are left for future research.

11. ACKNOWLEDGMENTS

12. REFERENCES

[1] G. Antoniou, D. Billington, G. Governatori, and M. J. Maher. Representation results for defeasible logic. *ACM Transactions on Computational Logic*, 2(2):255–287, 2001.

[2] G. Antoniou, D. Billington, G. Governatori and M. J. Maher. On the Modelling and Analysis of Regulations. In *Proceedings of the Australian Conference Information Systems*, pages 20–29, 1999.

[3] G. Antoniou, D. Billington, and M. J. Maher. On the analysis of regulations using defeasible rules. In *HICSS*, 1999.

[4] D.G. Baird, R.H. Gertner. R.C. Picker. *Game Theory and the Law*. Harvard University Press, Cambridge, 1994.

[5] T. Bench-Capon. Post congress tristesse. In *Specification and Implementation of Toulmin Dialogue Game*, pages pp 5–20. JURIX, 1998.

[6] D. Billington. Defeasible logic is stable. *Journal of Logic and Computation*, 3(4):379–400, 1993.

[7] M. Dastani, G. Governatori, A. Rotolo, and L. van der Torre. Programming cognitive agents in defeasible logic. In G. Sutcliffe and A. Voronkov, editors, *Logic for Programming, Artificial Intelligence, and Reasoning, 12th International Conference*, volume 3835 of *LNAI*, pages 621–636. Springer, 2005.

[8] A.P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B*, 30:205–247,1968.

[9] D. Dubois and H. Prade. Possibility theory, probability theory and multiple-valued logics: A clarification. *Annals of Mathematics and Artificial Intelligence*, 32(1-4):35–66, 2001.

[10] H. Geffner and J. Pearl. Conditional Entailment: Bridging Two Approaches to Default Reasoning. *Artificial Intelligence*, 53(2-3): pp. 209-244, 1992.

[11] T.F. Gordon. The pleadings game: formalizing procedural justice. In *ICAIL '93: Proceedings of the 4th international conference on Artificial intelligence and law*, pages 10–19, New York, NY, USA, 1993. ACM Press.

[12] T.F. Gordon. The Pleadings Game. An Artificial Intelligence Model of Procedural Justice. Dordrecht: Kluwer Academic Publishers, 1995.

[13] G. Governatori. Representing business contracts in RuleML. *International Journal of Cooperative Information Systems*, 14(2-3):181–216, 2005.

[14] G. Governatori, M. Dumas, A. H. ter Hofstede, and P. Oaks. A formal approach to protocols and strategies for (legal) negotiation. In H. Prakken, editor, *Procedings of the 8th International Conference on Artificial Intelligence and Law*, pages 168–177. ACM Press, 2001.

[15] G. Governatori, M.J. Maher, G. Antoniou, and D. Billington. Argumentation semantics for defeasible logic. *Journal of Logic and Computation*, 14(5):675–702, 2004.

[16] G. Governatori, A. Rotolo, and V. Padmanabhan. The cost of social agents. In P. Stone and G. Weiss, editors, *5th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 513–520, New York, 10–12 May 2006. ACM Press.

[17] J.C. Hage. *Reasoning with Rules. An essay on Legal Reasoning and Its Underlying Logic*. Dordrecht: Kluwer Law and Philosophy Library, 1997.

[18] J.C. Hage and B. Verheij. Reason-Based Logic: a logic for reasoning with rules and reasons. *Law, Computers and Artificial Intelligence*, 3(2/3):130–155, 1994.

[19] A.R. Lodder. *DiaLaw. On legal justification and dialog games*. Dissertation Universiteit Maastricht, 1998.

[20] M. J. Maher. Propositional defeasible logic has linear complexity. *Theory Practice Logic Programming*, 1(6):691–711, 2001.

[21] D. Nute. Defeasible Reasoning. In *Handbook of Logic in Artificial Intelligence and Logic Programming* (eds. Gabbay, D., Hogger, C. and Robinson, J.), Vol 3. Oxford: Oxford University Press, pp. 353-395, 1994.

[22] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1999.

[23] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal Logic and Computation*, 15(6):1009–1040, 2005.

[24] H. Prakken and G. Sartor. A dialectical model of assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law*, 4(3-4):331–368, 1996.

[25] H. Prakken and G. Sartor. Modelling reasoning with precedents in a formal dialogue game. *Artificial Intelligence and Law*, 6(2-4):231–287, 1998.

[26] J. Rawls. *A Theory of Justice*. Oxford: Oxford University Press, 1972.

[27] A.C. Roth. *Case-based reasoning in the law. A formal theory of reasoning by case comparison*. Dissertation Universiteit Maastricht, 2003.

[28] G. Shafer. *A Mathematical Theory of Evidence*. Princeton, Princeton University Press, 1976.

[29] L. Zadeh. Fuzzy Sets as the Basis for a Theory of Possibility. *Fuzzy Sets and Systems*, 1:3-28, 1978.