Dialogues and HY-arguments*

Martin Caminada
Institute of Information and Computing Sciences
Utrecht University, PO Box 80.089
3508 TB Utrecht, The Netherlands
martinc@cs.uu.nl

Abstract

This paper introduces a new class called *hang yourself (HY)* arguments into the theory of defeasible argumentation. The novelty of such arguments is that they are inherently destructive: they cannot be used to support conclusions but only to attack other arguments. In this paper it is described what these arguments are, how they can be formalized, and what the formal consequences are of adding them to a logic for defeasible argumentation.

Introduction

A popular approach to the study of nonmonotonic reasoning is in terms of argumentation (Dung, 1995; Simari and Loui, 1992; Pollock, 1995; Vreeswijk, 1997; Bondarenko *et al.*, 1997). The main idea is that defeasible inference can be characterised in terms of the interaction between arguments for and against alternative conclusions. One of the advantages of this approach is that defeasible reasoning can be studied from the perspective of dialogues, in which two agents, a proponent and opponent, argue about the acceptablity of a certain statement (Simari and Loui, 1992; Brewka, 1994; Vreeswijk, 1997; Loui, 1998; Prakken and Sartor, 1997). Dialogues are very close to the human way of interacting when trying to convince each other of their respective points of view. It is a way of reasoning that people are relatively familiar with, so that the study of defeasible reasoning in terms of dialogues can help to decrease the gap between intuitive and formal accounts of defeasible reasoning.

In this paper, we examine one particular form of arguments that until now has received little attention in the field of defeasible logic and formal dialogues. In order to illustrate this form of arguments, a few informal examples are given in section . It is also shown that a formalization of the examples in existing systems can result in an outcome that is different than what one would expect based on intuitive grounds. In section , a new kind of formal argument (which we call HY-arguments) is introduced, and it is shown that the resulting formalism properly deals with the examples of section . In section , some of the formal properties of the resulting system are given.

*This research was supported by the Netherlands Organisation for Scientific Research (NWO) under project number 612.060.005.

The formal investigations of this paper will be carried out in terms of an example system for defeasible argumentation, viz. the one of (Prakken and Sartor, 1997). However, it is important to note that the problems can also arise in similar systems, such as default logic (Reiter, 1980), Pollock's OSCAR system (Pollock, 1995), defeasible logic (Nute and Erk, 1998; Nute *et al.*, 1998) or Simari and Loui's system (Simari and Loui, 1992).

The problem

In this section, the concept of HY-arguments is illustrated using a sample framework for defeasible argumentation, for which we have chosen the framework of Prakken and Sartor (Prakken and Sartor, 1997). One of the reasons to choose this system to illustrate the concept of HY-arguments is its ability to view defeasible reasoning as a dialogue, in which a proponent and an opponent discuss the validity of a certain statement. Based on the language of extended logic programming, the system is formulated as an instance of Dung's (Dung, 1995) grounded semantics, and is also given an equivalent dialogue-game formulation. In order to keep things concise, we have simplified Prakken and Sartor's system not to include priorities, weak negation or strict rules other than premises. Under these simplifications, roughly, arguments can be constructed by chaining rules, and arguments are in conflict when they use rules with conflicting heads. More precisely, our reference system (which is refered to as $DS_{classic}$) can be defined using the following definitions:

Definition 2.1.
A literal *is either an atomic proposition (P) or the negation of an atomic proposition ($\neg P$). A negation function ($- : literals \rightarrow literals$) is defined by $-P = \neg P$ and $-\neg P = P$.*

Rules come in two forms: defeasible rules and premises. Syntactically, the difference is indicated by the type of arrow. A short, single lined arrow ("→") indicates a premise, while a short double lined arrow ("⇒") indicates a defeasible rule. Another important difference is that premises always have an empty antecedent. The advantage of modeling premises as a kind of rules is that in this way arguments become more homogenous. In cases where the difference between premises and defeasible rules is not relevant, or where both kinds of rules are meant, a long single lined arrow is used (“$\longrightarrow$”).

Definition 2.2.
A rule *is an expression of the form* $L_0 \wedge \ldots \wedge L_{n-1} \longrightarrow L_n$*, where each* L_i *(*$0 \leq i \leq n$*) is a literal. The conjunction left of the arrow is the* antecedent *and the literal right of the arrow is the* consequent. *The following kinds of rules are distinguished:*

1. *premises:* $\rightarrow L$
2. *defeasible rules:* $L_0 \wedge \ldots \wedge L_{n-1} \Rightarrow L_n$

Definition 2.3.
An argument *is a finite sequence* $A = [r_0, \ldots, r_n]$ *of rules such that:*

1. *for every* i *(*$i \leq 0 \leq n$*), for every literal* L *in the antecedent of* r_i*, there is a* $h < i$ *such that* L *is the consequent of* r_h*, and*
2. *no two distinct rules in the sequence have the same consequent*

We say that L *is a* conclusion *of* A *iff* A *contains a rule with* L *as consequent.*

Definition 2.4.
A set S *of premises is* consistent *iff the set of all consequents of* S *is consistent.*

Definition 2.5.
A defeasible theory *is a pair* $(\mathcal{S}, \mathcal{D})$ *where* $\mathcal{S}$ *is a consistent set of premises and* $\mathcal{D}$ *is a set of defeasible rules. An argument* A *is* based on $(\mathcal{S}, \mathcal{D})$ *iff all rules in* A *are in* $\mathcal{S} \cup \mathcal{D}$.

Definition 2.6.
Let A_1 *and* A_2 *be two arguments.* A_2 attacks A_1 *on* L *iff* A_1 *has conclusion* L *and* A_2 *has conclusion* $-L$*. An argument* A *is* coherent *iff* $consequents(A) \cup consequents(\mathcal{S})$ *is consistent.*

The definition of defeat uses the definition of attack, with as additional condition that one cannot defeat any premises. Furthermore, an incoherent argument can *always* be defeated in a very trivial way.

Definition 2.7.
Let A_1 *and* A_2 *be two arguments. Then* A_2 defeats A_1 *iff:*

1. A_2 *attacks* A_1 *and on* L *and the rule in* A_1 *that has* L *as its consequent is not a premise.*
2. A_2 *is empty and* A_1 *is incoherent.*

We say that A_2 strictly defeats A_1 *iff* A_2 *defeats* A_1 *and* A_1 *does not defeat* A_2.

Definition 2.7 makes that incoherent arguments are basically discharged. This feature is not unique to the system of Prakken and Sartor. Pollock, for instance, argues that self-defeating arguments should themselves not be justified (“warranted”) and also not have any influence on whether other arguments are justified or not (Pollock, 1987; Pollock, 1992). The general idea is that an argument that contradicts itself should not be taken serious and should therefore not keep other arguments from becoming justified.[1]

[1]Apart from that, there also exist strong technical reasons for ruling out self-defeating arguments in advance; see for instance section 5.3 of (Caminada, 2004).

Based on the above described defeat relation between arguments, Prakken and Sartor use grounded semantics (Dung, 1995) to define a sceptical notion of defeasible consequence.

Definition 2.8.
A literal is justified *iff it is the conclusion of an argument in the grounded extension.*

An argumentation formalism based on grounded semantics can be given a dialectical proof theory.[2] The idea is that a proponent and an opponent are involved in a discussion about the validity of a main argument. They take turns, and in every turn they provide a counterargument against the other party's argument. The rules of the dialogue are such that each argument of the opponent should defeat the previous argument of the proponent, while each argument of the proponent should *strictly* defeat the previous argument of the opponent.[3] Furthermore, the proponent is not allowed to repeat any earlier moves, to prevent the dialogue from non-termination.

Based on the above formalism, the following are examples of dialogues between a proponent (P) and an opponent (O).

Example 2.1.
$\mathcal{S} = \{\rightarrow A,\ \rightarrow D\}$
$\mathcal{D} = \{A \Rightarrow B,\ B \Rightarrow C,\ D \Rightarrow E,\ E \Rightarrow \neg C\}$
P: $\rightarrow A,\ A \Rightarrow B,\ B \Rightarrow C \quad (A_1)$
O: $\rightarrow D,\ D \Rightarrow E,\ E \Rightarrow \neg C \quad (A_2)$

Here, A_2 attacks and defeats A_1.

Example 2.2.
$\mathcal{S} = \{\rightarrow A,\ \rightarrow D\}$
$\mathcal{D} = \{A \Rightarrow B,\ B \Rightarrow C,\ D \Rightarrow E,\ E \Rightarrow \neg B\}$
P : $\rightarrow A,\ A \Rightarrow B,\ B \Rightarrow C \quad (A_1)$
O : $\rightarrow D,\ D \Rightarrow E,\ E \Rightarrow \neg B \quad (A_2)$

Here, A_2 again attacks and defeats A_1.

Example 2.3.
$\mathcal{S} = \{\rightarrow A\}$
$\mathcal{D} = \{A \Rightarrow B,\ B \Rightarrow C,\ C \Rightarrow \neg B,\ \neg B \Rightarrow \neg D,\ A \Rightarrow D\}$
P: $\rightarrow A,\ A \Rightarrow D \quad (A_1)$
O: $\rightarrow A,\ A \Rightarrow B,\ B \Rightarrow C,\ C \Rightarrow \neg B,\ \neg B \Rightarrow \neg D \quad (A_2)$
P: $\emptyset \quad (A_3)$

Here, A_2 is an incoherent argument that is defeated by the empty argument.

Now that the formalities of $DS_{classic}$ are made clear, the next step is to provide a few natural language examples and to examine how these can be formalized. The first example is the well-known Nixon diamond.

[2]For the specific formalism of Prakken and Sartor, completeness and correctness of the dialectical proof theory is proved in (Prakken and Sartor, 1997). For a more general proof, for any grounded semantics argumentation system we refer to (Caminada, 2004).

[3] The idea is that there is a certain asymmetry in the dialogue. The opponent has the relatively easy task of casting doubt on a certain thesis (for which non-strict defeat is already enough). The proponent, however, should make sure that the thesis is casted away from all doubt (for which it needs to *strictly* defeat the possible counterarguments).

Example 2.4 (Nixon diamond).

P: *"Nixon is a pacifist, because he is a Quaker."*

O: *"Certainly not, because he is also a Republican."*

$\mathcal{S} \ = \ \{\rightarrow NQ, \ \rightarrow NR\}$

$\mathcal{D} \ = \ \{NQ \Rightarrow NP, \ NR \Rightarrow \neg NP\}$

P: $\rightarrow NQ, \ NQ \Rightarrow NP$

O: $\rightarrow NR, \ NR \Rightarrow \neg NP$

For the Nixon diamond, $DS_{classic}$ indeed derives the desirable outcome; in every possible dialogue meant to defend NP, the opponent will have the last word, thus NP is not justified (for similair reasons, $\neg NP$ is also not justified). Thus, the outcome of the Nixon diamond corresponds with what one may reasonably expect: proponent's argument is rejected because there exists a plausible counterargument that casts doubt on it. Other examples exists, however, where the formal outcome and the intuitive outcome are not necessarily the same.

Example 2.5 (shipment of goods).

P: *"The shipment of goods must have arrived in the Netherlands by now (a), because we placed an order three months ago (tma)"*

O: *"I don't think so. If the goods had really arrived in the Netherlands, then there would be a customs declaration (cd), and I can't see any such declaration in our information system ($\neg is$)."*

$\mathcal{S} \ = \ \{\rightarrow tma, \ \rightarrow \neg is\}$

$\mathcal{D} \ = \ \{tma \Rightarrow a, \ \neg is \Rightarrow \neg cd, \ a \Rightarrow cd\}$

P: $\rightarrow tma, \ tma \Rightarrow a$

O: $\rightarrow tma, \ tma \Rightarrow a, \ a \Rightarrow cd, \ \rightarrow \neg is, \ \neg is \Rightarrow \neg cd$

P: $\emptyset$

Example 2.6 (tax relief).

P: *"If all goes well, this administration will implement a huge tax relief (tr)."*

O: *"But in the current economical situation, you can only implement such a tax relief by accepting a significant budget deficit (bd), which means we will also get a huge fine from Brussels (fb).*[4]*There goes your tax relief."*

$\mathcal{S} \ = \ \emptyset$

$\mathcal{D} \ = \ \{\Rightarrow tr, \ tr \Rightarrow bd, \ bd \Rightarrow fb, \ fb \Rightarrow \neg tr\}$

P: $\Rightarrow tr$

O: $\Rightarrow tr, \ tr \Rightarrow bd, \ bd \Rightarrow fb, \ fb \Rightarrow \neg tr$

P: $\emptyset$

What the examples 2.5 and 2.6 have in common is that the opponent tries to attack the standpoint of the proponent not by providing an independent counterargument, (such as was seen in the examples 1, 2 and 3, as well as in the Nixon diamond), but by showing that the standpoint of the proponent is problematic as it can lead to conflicts. The opponent wants to indicate that the proponent drew its conclusions too fast and that if the proponent would have given the matter more thought, then after some additional reasoning steps he would have found out that his argument can lead to self-defeat. In the formalism of Prakken and Sartor, however, as well as in similar formalisms, it is the *opponent* that is blamed for making a self defeating argument (which, according to definition 2.7 is defeated by the empty argument), while, in fact, the opponent is only confronting the proponent with the consequences of its own reasoning.

Existing systems cannot account for this inherently destructive nature of HY-arguments; instead they treat each counterargument as a constructive argument for a conflicting conclusion. What is needed then, is a way to formalize reasoning with HY-arguments in such a way that they cannot be used to derive conclusions but only to attack other arguments.

HY-arguments

An HY-argument can be defined as an argument that shows the problematic nature of another argument by assuming the conclusions of this other argument and then entailing either a contradiction, or a conclusion that undercuts the other orgument[5]. In this section, a formal logic for defeasible reasoning is presented that can deal with this type of argument. The thus obtained logic is referred to as DS_{HY}.

Definition 3.1.
A rule is an expression of the form: $L_0 \wedge \ldots \wedge L_{n-1} \longrightarrow L_n$ *where each* L_i *(*$0 \leq i \leq n$*) is a literal. The conjunction at the left of the arrow is the antecedent and the literal at the right side of the arrow is the consequent of the rule. The following kinds of rules are distinguished:*

1. *premises:* $\rightarrow L$
2. *defeasible rules:* $L_0 \wedge \ldots \wedge L_{n-1} \Rightarrow L_n$
3. *foreign commitments:* $\rightsquigarrow L$

Foreign commitments are "imported" conclusions from the other agent's argument and are used in a way similair to reductio ad absurdum assumptions in classical logic. The requirement that for an argument to make sense, every foreign commitment should be based on an actual conclusion of another argument, will be formalized in the notion of attack.

Definition 3.2.
Let $(\mathcal{S}, \mathcal{D})$ *be a defeasible theory. An* argument *is a finite sequence* $A = [r_0, \ldots, r_n]$ *of rules such that:*

1. *for every* i *(*$0 \leq i \leq n$*), for every literal in the antecedent of* r_i*, there is a* $h < i$ *such that* L *is the consequent of* r_h
2. *no two distinct rules in the sequence have the same consequent*
3. *All premises of* A *are in* $\mathcal{S}$*, all defeasible rules of* A *are in* $\mathcal{D}$*, and for all foreign commitments* $\rightsquigarrow C$ *in* A *it holds that* C *is a conclusion of a rule in* $\mathcal{S} \cup \mathcal{D}$*.*

If A is an argument with conclusion L, then the set of rules *relevant* to L consist of all rules in A that have a role in deriving L, starting with the rule with L as its consequent and ending with one or more premises or foreign commitments.

[4]Countries within the euro-zone have to keep their budget deficit less than 3%, or face sanctions from the EU.

[5]In order to keep the discussion concise, the possibility of undercutting is not discussed in this paper

Definition 3.3.
Let A be an argument with conclusion L. The set of rules relevant *to L — written as $R_L(A)$ — is the smallest set such that:*

1. *$L_0 \wedge \ldots \wedge L_{n-1} \longrightarrow L \in R_L(A)$, where $L_0 \wedge \ldots \wedge L_{n-1} \longrightarrow L$ is a rule in A*
2. *if $L_0 \wedge \ldots \wedge L_{n-1} \longrightarrow L_n \in R_L(A)$ then also $R_{L_0}(A) \cup \ldots \cup R_{L_{n-1}}(A) \subseteq R_L(A)$*

Definition 3.4.
Let A be an argument with conclusion L. We say that L is fc-based iff $R_L(A)$ contains at least one foreign commitment.

Definition 3.5.
Let A_1 and A_2 be two arguments. A_2 attacks *A_1 on L iff:*

1. *A_2 has a conclusion L and a conclusion $-L$ where at least L is fc-based, and*
2. *for every foreign commitment $\rightsquigarrow C$ in A_2: C is a conclusion of A_1 that is not fc-based.*

An argument is coherent *iff it is not attacked by an argument without defeasible rules.*

For the definition of attack, it is important to notice that any "traditional" attacking argument can be converted into an attacking HY-argument. Suppose we have a defeasible theory $(\mathcal{S}, \mathcal{D}, <)$ with $\mathcal{S} = \{\rightarrow I,\ \rightarrow M\}$ and $\mathcal{D} = \{I \Rightarrow J,\ J \Rightarrow K,\ K \Rightarrow L,\ M \Rightarrow N,\ N \Rightarrow O,\ O \Rightarrow \neg L\}$. Then, a dialogue with "traditional" arguments would look as follows:

P: $\rightarrow I,\ I \Rightarrow J,\ J \Rightarrow K,\ K \Rightarrow L$
O: $\rightarrow M,\ M \Rightarrow N,\ N \Rightarrow O,\ O \Rightarrow \neg L$

A dialogue in which traditional defeat should be implemented by an HY-argument then looks as follows:

P: $\rightarrow I,\ I \Rightarrow J,\ J \Rightarrow K,\ K \Rightarrow L$
O: $\rightarrow M,\ M \Rightarrow N,\ N \Rightarrow O,\ O \Rightarrow \neg L,\ \rightsquigarrow L$

Definition 3.6.
Let A_1 and A_2 be two arguments. A_2 defeats *A_1 iff A_2 attacks A_1 on L and $\bigcup_{c_i \in \{c_i \mid \rightsquigarrow c_i \in R_L(A_2) \cup R_{-L}(A_2)\}} R_{c_i}(A_1)$ contains at least one defeasible rule.*

Note that, just as in Section , we have defined a binary defeat relation on a set of arguments, so the original definitions of grounded semantics and the corresponding dialogue-game version still apply. However, to capture the destructive nature of HY-arguments, the notion of a justified conclusion must be restricted to conclusions derived without foreign commitments.

Definition 3.7.
A literal is justified *iff it is the conclusion of a justified argument (grounded semantics) without foreign commitments.*

It is interesting to see how these definitions apply to the examples given earlier. The idea is simple: take one of the conclusions of the proponent's argument and then show that starting from this conclusion a contradiction can be derived.

Example 3.1 (Nixon diamond).
P: $\rightarrow NQ,\ NQ \Rightarrow NP$ *(A_1)*
O: $\rightarrow NR,\ NR \Rightarrow \neg NP,\ \rightsquigarrow NP$ *(A_2)*

In this example, a "traditional" attacking argument is written in the form of a HY-attack.

Example 3.2 (shipment of goods).
P: $\rightarrow tma,\ tma \Rightarrow a$ *(A_1)*
O: $\rightsquigarrow a,\ a \Rightarrow cd,\ \rightarrow \neg is,\ \neg is \Rightarrow \neg cd$ *(A_2)*

The proponent now does not have any argument (like $\emptyset$ in the case of $DS_{classic}$) that *strictly* defeats A_2.

Example 3.3 (tax relief).
P: $\Rightarrow tr$ *(A_1)*
O: $\rightsquigarrow tr,\ tr \Rightarrow bd,\ bd \Rightarrow fb,\ fb \Rightarrow \neg tr$ *(A_2)*

Here too, the proponent does not have any argument that strictly defeats A_2.

Properties

In this section, some of the formal effects of allowing HY-arguments are studied.

rule maximalization

The first thing to be noticed is that the addition of HY-arguments results in a logic that is based on a different principle than without HY-arguments. Take for instance the following example:
$\mathcal{S} = \{\rightarrow A, \rightarrow D\}$
$\mathcal{D} = \{A \Rightarrow B,\ B \Rightarrow C,\ C \Rightarrow \neg B,\ D \Rightarrow E\}$
In $DS_{classic}$, A, B, C, D and E are justified. In DS_{HY}, on the other hand, only A, D and E are justified. B, for instance, is not justified in DS_{HY}, since there is a counter-argument $\rightsquigarrow B,\ B \Rightarrow C,\ C \Rightarrow \neg B$ against $\rightarrow A,\ A \Rightarrow B$.

The different outcome of $DS_{classic}$ and DS_{HY} can be seen in the following perspective. Basically, there are two approaches to this example. The first approach is to try to find one or more maximal sets of conclusions, such that there exists a coherent argument for these conclusions. This is the usual approach in nonmonotonic logic; one tries to find extensions consisting of a maximal set of conclusions; a principle we call *conclusion maximization*.

Another approach would be to try to find maximal subsets of rules such that no incoherent argument can be constructed, a principle we call *rule maximization*. Formally, this principle can be stated as follows:

Definition 4.1.
Let $(\mathcal{S}, \mathcal{D})$ be a defeasible theory. A rule-maximal set of rules R_{rmax} is a maximal subset of rules from $\mathcal{D}$ such that no incoherent argument A from $DS_{classic}$ exists that is based on $(\mathcal{S}, \mathcal{D})$ with $defeasible(A) \subseteq R_{rmax}$.

Definition 4.2.
Let $(\mathcal{S}, \mathcal{D})$ be a defeasible theory. A literal L follows from *a set of defeasible rules R iff there is an argument A from $DS_{classic}$ based on $(\mathcal{S}, \mathcal{D})$ with conclusion L and $defeasible(A) \subseteq R$.*

In the case of the above example, this results in three rule maximal sets of rules, with associated conclusions that follow from it:
$\{B \Rightarrow C,\ C \Rightarrow \neg B,\ D \Rightarrow E\}$ (follows: A, D, E)
$\{A \Rightarrow B,\ C \Rightarrow \neg B,\ D \Rightarrow E\}$ (follows: A, B, D, E)
$\{A \Rightarrow B,\ B \Rightarrow C,\ D \Rightarrow E\}$ (follows: A, B, C, D, E)

If one takes for each rule-maximal set of rules the set of conclusions that can be derived (like is done above), and one takes the intersection of these sets (sceptical semantics), this results in a set containing only A, D and E. These are exactly the statements that are justified in DS_{HY}.

The fact that in the above example the effects of allowing HY-arguments corresponds with the results of applying rule-maximality is not a coincidence.

Theorem 4.1.
L is a justified conclusion in DS_{HY} iff L follows from every rule-maximal set of rules.

Proof. See (Caminada, 2004). □

cautious monotonicity

If one views nonmonotonic logic from the perspective of postulates, then a particularly interesting postulate is that of cautious monotonicity. This postulate can can expressed as follows:

If $(\mathcal{S}, \mathcal{D}) \vdash L$ and $(\mathcal{S}, \mathcal{D}) \vdash M$ then $(\mathcal{S} \cup \{\rightarrow L\}, \mathcal{D}) \vdash M$

In many systems for defeasible reasoning (such as (Prakken and Sartor, 1997) or (Reiter, 1980)) this property does not hold, as is illustrated by the following example:

$$\mathcal{S} = \{\rightarrow p\}$$
$$\mathcal{D} = \{p \Rightarrow q,\ q \Rightarrow r,\ r \Rightarrow \neg q,\ \neg q \Rightarrow s,\ \Rightarrow \neg s\}$$

It now holds that:

- $(\mathcal{S}, \mathcal{D}) \vdash r$ (there is a coherent classical argument $\rightarrow p,\ p \Rightarrow q,\ \Rightarrow r$ for r that has no classical coherent counterargument)
- $(\mathcal{S}, \mathcal{D}) \vdash \neg s$ (there is a coherent classical argument $\Rightarrow \neg s$ for $\neg s$ that has no classical coherent counterargument)
- $(\mathcal{S} \cup \{\rightarrow r\}, \mathcal{D}) \not\vdash \neg s$ (the argument $\Rightarrow \neg s$ now has a classical coherent counterargument $\rightarrow r,\ r \Rightarrow \neg q,\ \neg q \Rightarrow s$)

In DS_{HY}, however, the above counterexample against cautious monotonicity is no longer valid, since r is no longer a justified conclusion.

The fact that the above counterexample against cautious monotonicity no longer holds is not a coincidence.

Theorem 4.2.
Let $(\mathcal{S}, \mathcal{D}) \vdash_{DS_{HY}} L$ stand for "L is a justified conclusion in DS_{HY} under $(\mathcal{S}, \mathcal{D})$". Then it holds that:
If $(\mathcal{S}, \mathcal{D}) \vdash_{DS_{HY}} L$
and $(\mathcal{S}, \mathcal{D}) \vdash_{DS_{HY}} M$
then $(\mathcal{S} \cup \{\rightarrow L\}, \mathcal{D}) \vdash_{DS_{HY}} M$

Proof. See (Caminada, 2004). □

HY and contraposition

To some readers reasoning with HY-arguments may seem similar to contrapositive reasoning. An interesting question therefore is how the effect of adding HY-arguments compares to the effect of adding contraposition. We discuss three examples to illustrate the similarities and differences between these concepts. It will turn out that the main difference is that while contraposition can be used to derive new conclusions, HY arguments can only be used to attack other arguments. Given the controversial nature of the principle of default contraposition (cf. Example 4.2 below) this may be regarded as an advantage of the approach with HY arguments.

As the logic of Prakken and Sartor by itself does not validate the principle of default contraposition, we study the effects of contraposition by "manually" adding a contrapositive for each default.

Example 4.1.
$\mathcal{S} = \{\rightarrow A,\ \rightarrow \neg C\}$
$\mathcal{D} = \{A \Rightarrow B,\ B \Rightarrow C\}$

In $DS_{classic}$, there are justified arguments for A ($\rightarrow A$) and B ($\rightarrow A,\ A \Rightarrow B$). The (only) argument for C ($\rightarrow A,\ A \Rightarrow B,\ B \Rightarrow C$) is defeated by the strict argument $\rightarrow \neg C$ so C is not justified.

If we look at a system with HY-arguments, only A and $\neg C$ are justified, B is not. The reason is that the argument for B ($\rightarrow A,\ A \rightarrow B$) now has a HY-counterargument $\rightsquigarrow B,\ B \Rightarrow C,\ \rightarrow \neg C$.

Suppose we have a system without HY-arguments, but with contraposition. Contraposition essentially means that whenever we have a rule $A \Rightarrow B$ we may also use it in the contraposed way of $\neg B \Rightarrow \neg A$. This means that the number of usable rules can (at most) be doubled. If we allow contraposition in example 4.1, we thus get the following effective rule-bases:
$\mathcal{S}' = \{\rightarrow A,\ \rightarrow \neg C\}$
$\mathcal{D}' = \{A \Rightarrow B,\ B \Rightarrow C,\ \neg B \Rightarrow \neg A,\ \neg C \Rightarrow \neg B\}$
If we apply $DS_{classic}$ to these rule-bases, we obtain justified conclusions A and $\neg C$, and nothing else. B is not justified because there now is a (non-HY) argument ($\rightarrow \neg C,\ \neg C \Rightarrow \neg B$) against B.

The results, as far as justified conclusions are concerned, of example 4.1 can therefore be summarized as follows:

- $DS_{classic}$ $\{A, B, \neg C\}$
- DS_{HY}: $\{A, \neg C\}$
- $DS_{classic}$ + contrapos: $\{A, \neg C\}$

In example 4.1, we see that the effect of adding HY-arguments is the same as the effect of adding contraposition. The question is whether this is always the case.

Example 4.2.
$\mathcal{S} = \{\rightarrow \neg C\}$
$\mathcal{D} = \{A \Rightarrow B,\ B \Rightarrow C\}$

Here, $DS_{classic}$ allows us to derive nothing but the conclusion $\neg C$ (using argument $\rightarrow \neg C$), as there are simply no arguments for any other conclusion. If we allow HY-arguments, then still no other conclusions than $\neg C$ can be derived; as there are still no arguments for anything else. If, on the other hand, we allow contraposition, then we can also derive $\neg B$ ($\rightarrow \neg C,\ \neg C \Rightarrow \neg B$) and $\neg A$ ($\rightarrow \neg C,\ \neg C \Rightarrow \neg B,\ \neg B \Rightarrow \neg A$), and since these arguments do not have any counterarguments, both of them are justified.

The results of example 4.2 can therefore be summarized as follows:

- $DS_{classic}$: $\{\neg C\}$

- DS_{HY}: $\{\neg C\}$
- $DS_{classic}$ + contrapos: $\{\neg C, \neg B, \neg A\}$

An intuitive version of this example is: sailors are typically men, men typically have no beard, but captain Nemo has a beard. We don't want to conclude from this that captain Nemo is not a man and not a sailor.

Example 4.2 makes clear that the outcome of a system with HY arguments can be different from the outcome of a system with contraposition. This is not surprising, since HY-arguments are not able to generate new conclusions; instead, they merely cast doubt on other conclusions.

Example 4.3.
$\mathcal{S} = \{\rightarrow A\}$
$\mathcal{D} = \{A \Rightarrow B,\ B \Rightarrow C,\ C \Rightarrow D,\ D \Rightarrow E,\ E \Rightarrow \neg C\}$

Here, $DS_{classic}$ entails A, B, C, D and E. If we allow HY-arguments, however, then only A is remains justified. In order to see why this is, take for instance the argument for E: $\rightarrow A,\ A \Rightarrow B,\ B \Rightarrow C,\ C \Rightarrow D,\ D \Rightarrow E$. It now has a HY counterargument $\rightsquigarrow E,\ E \Rightarrow \neg C,\ \rightsquigarrow C$. HY-counterarguments against B, C and D are also available, so only A remains justified.

Contraposition allows us the justified conclusions of A, B and C (but not D or E). This can be seen as follows. Although there is an argument for D ($\rightarrow A,\ A \Rightarrow B,\ B \Rightarrow C,\ C \Rightarrow D$) there is also a counterargument ($\rightarrow A,\ A \Rightarrow B,\ B \Rightarrow C,\ C \Rightarrow \neg E,\ \neg E \Rightarrow \neg D$) so D is not justified. For a similar reason, E is also not justified. B and C, on the other hand *are* justified. C, for instance, has an argument $\rightarrow A,\ A \Rightarrow B,\ B \Rightarrow C$ that has no coherent counterargument since $\rightarrow A,\ A \Rightarrow B,\ B \Rightarrow C,\ C \Rightarrow \neg E,\ \neg E \Rightarrow \neg D,\ \neg D \Rightarrow \neg C$ is incoherent! For similar reasons, B is also justified.

The results of example 4.3 can therefore be summarized as follows:

- $DS_{classic}$: $\{A, B, C, D, E\}$
- DS_{HY}: $\{A\}$
- $DS_{classic}$ + contrapos: $\{A, B, C\}$

Summary and Conclusions

In this paper we have enriched the theory of defeasible reasoning with a notion of HY-arguments. The key novelty of this kind of argument is that it has an inherently destructive nature, which is the main reason why reasoning with such arguments does not simply boil down to contrapositive reasoning.

In this paper we formalised the notion of HY-arguments in the context of the example system of P&S. We showed that, unlike their original system, the adapted version with HY-arguments satisfies the postulate of cautious monotony. We also showed that, while the original system exhibits a principle of conclusion maximization, the adapted version with HY-arguments captures the alternative principle of rule maximization.

It is important to note that the formalism of Prakken and Sartor was chosen for illustrative purposes only. An analysis of how systems like default logic and Pollock's OSCAR can be enhanced with HY-arguments is provided in (Caminada, 2004), where it is also argued in which domains of reasoning HY-arguments are or are not applicable.

Acknowledgements

We would like to thank Henry Prakken for his useful comments and Leon van der Torre for helping to develop the initial idea.

References

A. Bondarenko, P.M. Dung, R.A. Kowalski and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence* 93:63-101 (1997)

G. Brewka. A logical reconstruction of Rescher's theory of formal disputation based on default logic, *Proceedings of the Eleventh European Conference on Artificial Intelligence* 366-370 (1994)

M.W.A. Caminada. *For the Sake of the Argument*. PhD thesis, Vrije Universiteit, Division of Mathematics and Computer Science, Amsterdam (2004)

P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. *Artificial Intelligence* 77:321-357 (1995)

R.P. Loui. Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence* 14: 1-38 (1998)

D. Nute and K. Erk. Defeasible logic graphs I. Theory. *Decision Support Systems* 22: 277-293 (1998)

D. Nute, Z. Hunter and Ch. Henderson. Defeasible logic graphs II. Implementation. *Decision Support Systems* 22: 295-306 (1998)

J.L. Pollock. Defeasible reasoning. *Cognitive Science* 11:481-518 (1987)

J.L. Pollock. How to reason defeasibly. *Artificial Intelligence* 57:1-42 (1992)

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

H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics* 7: 25-75 (1997) (special issue on 'Handling inconsistency in knowledge systems')

R. Reiter. A logic for Default Reasoning. *Artificial Intelligence* 13:81-132 (1980)

G.R. Simari and R.P. Loui. A mathematical treatment of defeasible argumentation and its implementation. *Artificial Intelligence* 53: 125-157 (1992)

G.A.W. Vreeswijk. Abstract Argumentation Systems. *Artificial Intelligence* 13: 81-132 (1980)