Formalization of the ad hominem argumentation scheme

Douglas Walton
University of Windsor

FORMALIZATION OF THE *AD HOMINEM* ARGUMENTATION SCHEME

Abstract

In this paper, several examples from the literature, and one central new one, are used as case studies of texts of discourse containing an argumentation scheme that has now been widely investigated in literature on argumentation. Argumentation schemes represent common patterns of reasoning used in everyday conversational discourse. The most typical ones represent defeasible arguments based on nonmonotonic reasoning. Each scheme has a matching set of critical questions used to evaluate a particular argument fitting that scheme. The project is to study how to build a formal computational model of this scheme for the circumstantial *ad hominem* argument using argumentation tools and systems developed in artificial intelligence. It is shown how the formalization built using these tools is applicable to the tasks of identification, analysis and evaluation of the central case studied. One important implication of the work is that it provides a foundational basis for showing how other argumentation schemes can be formalized.

The development of new computer technologies based on formal models of argumentation schemes is an exciting prospect for those of us working in the fields where we need to identify analyze and evaluate common arguments of the kind often used in conversational discourse (Reed and Norman, 2003). By taking a particular example of a text of discourse containing an argument of a kind fitting the form of one of the traditional fallacies, this paper investigates how useful such formalizations are, at their present state of development. The type of argument chosen as the focus of the investigation is the circumstantial *ad hominem* argument, an argumentation scheme that has now been widely investigated in the literature on argumentation (Walton, 1998; Verheij, 2003; Walton, 2006). It will be shown that the formalizations studied in the paper are applicable and most useful for the tasks of identification and analysis of the example argument chosen as the case to be studied. However, it will also be shown that the formalizations provide useful tools that are helpful for the project of evaluating the argument. Evaluating any actual *ad hominem* argument in a real case using an abstract formal model is tall order, but that is the task carried out.

First it is shown how to model the argumentation scheme for the circumstantial *ad hominem* argument, with its distinctive premises and conclusion, using either of two formalizations that have now been developed in artificial intelligence. The hardest problem is that of modeling the critical questions matching the scheme.[1] It is shown how one system of formalization of defeasible argumentation called DefLog (Verheij, 1999, 2001, 2003) can be applied to *ad hominem* arguments. It is also shown how another system called Carneades, designed to be integrated into the semantic web, has made considerable progress in not only formalizing the *ad hominem* scheme, but even incorporating in the model computational tools for showing how the critical questions should be used in an evaluation of any argument fitting the scheme (Gordon, 2005; Walton and Gordon, 2005). The limits around the edges of applying these formal models are also discussed, for, as is well known, natural language argumentation tends to resist any formalization into a precise format of the sort that can be recognized by artificial intelligence of the kind used by a computer (Gordon, 1995). Argumentation in natural language discourse contains value-laden language (Bench-Capon, 2003), implicit

---

[1] The usual device for analyzing and evaluating arguments that fit an argumentation scheme is the set of appropriate critical questions matching the scheme (Hastings, 1963; Perelman and Olbrechts-Tyteca, 1969; Kienpointner, 1992; Grennan, 1997).

premises, innuendo, and other forms of speech that are vague, ambiguous, and difficult to formulate as a set of propositions with precision and clarity (Perelman and Olbrechts-Tyteca, 1969; van Eemeren and Grootendorst, 2004). The problem posed is not one of getting a perfect match between the formal model and the example. It is more a question of whether the formal model is useful as a tool to help us analyze and evaluate the argumentation in the given example, once we have reformulated the argument using tools of argumentation theory clearly enough so that it can be brought into line with some known argumentation scheme (Prakken and Sartor, 1996; 1997). Only by such means can we develop objective methods of evaluating *ad hominem* arguments in real cases by collecting, analyzing and using textual evidence to determine whether an *ad hominem* fallacy has been committed or not, judging from what is known (or nor known) in the case.

1. Three Examples

In this section, three examples of the circumstantial *ad hominem* argument are presented. The first one is a summary of an article that appeared in the *National Post* (October 14, 2005, p. A10). It would be good for anyone reading this paper to read the whole article. It is only about one page, or one column in a newspaper format. Here, however, we will only present a brief summary of the main argument.

The Sealers Example

Rocco DiSpirito, a New York chef and best-selling food author, made famous as the star of the NBC reality show *The Restaurant*, wrote a public letter supporting a campaign by the U. S. Humane Society to end the Canadian seal hunt. The article quoted Mr. DiSpirito as saying, "Most of the seal clubbers [in Canada] are also snow crabbers. By refusing to use Canadian or Canadian-sourced snow crab in our restaurants, we can make a very vocal statement against the seal hunt." The Humane Society had been lobbying for an American boycott of Canadian seafood, especially snow crab from Atlantic Canada, advocating the boycott as an economic tactic to stop the seal hunt. Many American restaurants and seafood wholesalers had joined the boycott, pledging not to buy Canadian seafood. Newfoundland fishermen in the sealing industry replied by arguing that DiSpirito was a hypocrite for calling the seal hunt inhumane while serving *foie gras* made from the engorged livers of force-fed geese in his restaurant. This practice was officially banned in some European countries and California, where the humane society condemned it. Frank Pinhorn, managing director of the Newfoundland-based Canadian Sealers Association, was quoted as saying, "He's an absolute hypocrite, a man of double standards." Earl McCurdy, president of the Fish, Food and Allied Workers Union in St. John's was quoted as saying, "I think somebody who lives in a glass house shouldn't throw stones. It shows the hypocrisy of these celebrities, who know nothing about the seal hunt ... if he wants to serve *foie gras* in his restaurants, that's fine with me, but he shouldn't pass judgment on us." John Grandy, senior vice-president of the Humane Society, defended the chef. He was quoted as saying, "Absolutely the society is opposed to *foie gras*, but this issue is about seals, and a man of his distinction and abilities, who is simply appalled at the brutal destruction of these seals, well, *if* we can use him on the seals issue, we're happy to do so."

There is much going on in this particular text of discourse involving many different kinds of arguments. The sealers' argument is based on several argumentation schemes that have been widely studied in the literature. It is an instance of argument from a verbal classification, because it classifies certain actions or practices as inhumane. It is based on an implicit use of appeal to popular opinion, advocating the argument that since several states have banned the practice, that practice should be condemned. It is based on an argument from being in a position to know. The defenders of the sealers argue that New

Yorkers are not in a position to know about whether the practices of the sealers are humane or not, and the argument is arguably a straw man. The argument in the sealers example also fits the scheme of the type of *ad hominem* argument called guilt by association. The allegation is that snow crabbers are guilty by association with sealers. What the text of discourse says is that most of the seal clubbers are also snow crabbers. This statement implies that the association between the snow crabbers and the seal clubbers puts the snow crabbers in the same category as the seal clubbers, implying that they are guilty of these inhumane actions in the same way. The guilt by association argument in this case is at least questionable, and is well worth observing, but the central argument that is the unifying structure of the argumentation in the example as a whole is that of the circumstantial *ad hominem* argument.

Despite these other types of arguments being involved, we will focus on the central one in the example, that of the circumstantial *ad hominem* attack. Central to the case is the main thrust of the sealers' argument, which takes the form of a circumstantial *ad hominem* attack on what they take to be the position of the humane society, and the chef taken to share that position. The sealers argue that the chef is a hypocrite for condemning sealing as inhumane on the grounds that his personal activities include practices that should be considered equally inhumane. They allege that there is a practical inconsistency between the argument that the chef advocates and his own personal practice of using products like fattened goose liver in his restaurant. Although the chef alleged that the practices of the Canadians sealers can be described using the negative term 'inhumane', the sealers reply by arguing that his own personal practices merit the same description.

The sealers example needs to be compared to two other standard examples of the circumstantial *ad hominem* argument that have already been treated in the literature. One of these cases is the smoking example (Walton, 1998 p. 7).

The Smoking Example

A parent argues to her child that smoking is associated with chronic disorders and that smoking is unhealthy, therefore the child should not smoke. The child replies "You smoke yourself. So much for your argument against smoking!"

In this case, the child observes the action of the parent, and compares his observation of the parent's action of smoking against her argument for not smoking. The parent advocates the policy of not smoking, but as the child observes, she herself smokes. It is a classic case of arguing against somebody's argument using the counter-argument that the original arguer does not practice what she preaches. In this case, it appears that the child commits a circumstantial *ad hominem* fallacy, because it appears that the child absolutely rejects the parent's argument as worthless. However, the argument that the parent gives against smoking could be a good one. For example, the parent might show evidence of diseased lungs full of chronic obstructive lung disease caused by smoking. If the child is absolutely rejecting the parent's argument that smoking is unhealthy, he could be committing a fallacy of jumping too quickly to the conclusion that the parent's argument is worthless. On the other hand, if the child is merely questioning the sincerity of the parent's argument by pointing out a pragmatic inconsistency between her words and

actions, expressing doubt about the argument by questioning the parent's credibility, the child would appear to have a good point.

The additional point can be made (Walton 1998, p. 31) that the parent could reply to the child's allegation of practical inconsistency by stating that she has tried very hard to give up smoking, but that nicotine is addictive. She might counsel the child not to start smoking because it is an addictive habit and because it is unhealthy. In such a case, the parent has explained her way out of the practical inconsistency used by the child to attack her advice. Even though she admits that she smokes, she argues that her action of smoking should not be held against her advice on what the child should do. In this case then, whether or not the child commits a circumstantial *ad hominem* fallacy depends on how the child is taking the conclusion of his own argument against the parent's argument. If the child is absolutely rejecting the parents argument as worthless, that would be committing an *ad hominem* fallacy, whereas if the child is only raising critical questions about the parent's argument, based on his observations of the pragmatic inconsistency, his *ad hominem* argument could be reasonable.

The sealers example also needs to be compared with another standard example that it is similar to in other respects. In this case (Walton 1998 p. 32), a critic attacks a hunter for killing animals for sport, whereupon the hunter attacks back by questioning the personal circumstances of the critic.

The Hunter Example

A hunter is accused of barbarity for his sacrifice of innocent animals to his own amusement or sport hunting. His reply to his critic: "why do you feed on the flesh of harmless cattle?"

The problem posed by the hunter example is similar to the one in the smoking example. The hunter accuses the critic of a circumstantial inconsistency and uses this accusation to suggest that the critic is not really a sincere advocate of his own argument. By throwing doubt on the sincerity of the critic, the hunter shifts a burden of proof to the critic's side, implying that the critic is a hypocrite, or at any rate is not a sincere person who should be thought to have much credibility in the debate.

The difference between the hunter example and smoker example is that, in the latter case, a direct inconsistency is involved. The parent argues for a policy of not smoking, but admits that she personally smokes. In this case, the policy advocated is the opposite, or negation, of the description of the action that the parent is actually carrying out in her personal circumstances. The relationship between the two propositions in the hunter case is more indirect. It may be that the critic is not a vegetarian, and eats the occasional steak or pork chop. But is this personal circumstance enough to defeat the accusation of the hunter? It may be, because there is a connection (in the form of an implicit generalization that is a matter of common knowledge), between eating meat and hunting animals: killing animals is the necessary means for obtaining meat used for human consumption. Thus the hunter has some support of logic on his side. On the other hand, we have to be very careful in evaluating the argument in this case. The critic is presumably not herself a hunter, and she is criticizing the hunter for engaging in the sacrifice of animals for his own amusement or sport. She classifies this activity using the word 'barbarity'. Thus the critic is not being inconsistent in the way the hunter argument may appear to suggest. De Morgan (1847, 65) put this point very well in commenting on the parallel between the

situation of the hunter and the critic: "the parallel will not exist until, for the person who eats meat, we substitute one who turns butcher for amusement."

2. *Ad Hominem* Argumentation Schemes

The most basic form of the *ad hominem* is the direct or personal type. The scheme for the direct *ad hominem* argument is given in *Fundamentals of Critical Argumentation* (Walton, 2006, p. 123) as follows.

Argumentation Scheme for the Direct Ad Hominem Argument

Character Attack Premise: *a* is a person of bad character.

Conclusion: *a*'s argument should not be accepted.

In this scheme, *a* stands for an agent that is proponent of an argument that has been previously put forward. The *ad hominem* argument then brought forward attacks this prior argument by arguing that *a* is a person of bad character. For example, the attacker may say that *a* has often lied in the past. What is the basis of such an attack? The basis is that, in many cases of argumentation, a person's argument depends on his presumed honesty, ethical character, and trustworthiness. For example in a case of witness testimony, it is an important underlying assumption that the witness can be counted on to tell the truth. Thus an *ad hominem* argument is most effective when it raises doubts about an arguer's credibility, so that his argument is discounted. This type of argument is typically called the abusive *ad hominem* argument in logic textbooks. But this label is misleading because it suggests that direct *ad hominem* arguments are fallacious. In fact, they are often reasonable, as in a case where the credibility of a witness is attacked during cross-examination in court (Walton, 1998).

The following three critical questions are offered (Walton, 2006, p. 123) as appropriate for raising doubts about the direct *ad hominem* argument.

1. How well supported by evidence is the allegation made in the character attack premise?
2. Is the issue of character relevant in the type of dialogue in which the argument was used?
3. Is the conclusion of the argument that *A* should be (absolutely) rejected, even if other evidence to support *A* has been presented, or is the conclusion merely (the relative claim) that *a* should be assigned a reduced weight of credibility as a supporter of *A*, relative to the total body of evidence available?

The scheme, along with the critical questions, is meant to be used by a critic to evaluate any given instance of a direct *ad hominem* argument found in a text of discourse. First, the critic has to verify that the argument in question fits the scheme, and thus is a genuine instance of the direct *ad hominem* type of argument. Then the critic can raise doubts about whether the argument holds by asking any one of the critical questions matching the scheme. The evaluation is a dialog process. If one of the questions matching the

scheme cannot be answered, revealing a critical gap in the argument by pinpointing a required assumption that is not justified, the original argument defaults (Verheij, 2005).

The circumstantial *ad hominem* argument combines the scheme for the direct *ad hominem* argument with the scheme for argument from inconsistent commitment (Walton, 2006, p. 123). In *Fundamentals of Critical Argumentation* (2006, p. 125), the following argumentation scheme for the circumstantial *ad hominem* argument is given. The inconsistent commitment premise represents argument from inconsistent commitment as one component. By questioning the arguer's character as a sincere person who can be trusted, the credibility questioning premise represents the direct *ad hominem* argument as a component. It is important here to distinguish between an argument from inconsistent commitment that is not an *ad hominem* argument and one that is. Only the latter should properly be classified as an *ad hominem* argument, according to the analysis defended at some length in (Walton, 1998).

Argumentation Scheme for the Circumstantial Ad Hominem Argument (Walton, 2006, p. 125)

Argument Premise: *a* advocates argument α, which has proposition *A* as its conclusion.

Inconsistent Commitment Premise: *a* is personally committed to the opposite (negation) of *A*, as shown by commitments expressed in her/his personal actions or personal circumstances expressing such commitments.

Credibility Questioning Premise: *a*'s credibility as a sincere person who believes in his own argument has been put into question (by the two premises above).

Conclusion: The plausibility *a*'s argument α is decreased or destroyed.

The following four critical questions are cited (p. 126) as appropriate for the circumstantial *ad hominem* argument.

1. Is there a pair of commitments that can be identified, as shown by evidence, to be commitments of *a*, and taken to show that *a* is practically inconsistent?
2. Once the practical inconsistency is identified that is the focus of the attack, could it be resolved or explained by further dialogue, thus preserving the consistency of the arguer's commitments in the dialogue, or showing that *a*'s inconsistent commitment does not support the claim that *a* lacks credibility?
3. Is character an issue in the dialogue, and more specifically, does *a*'s argument depend on his/her credibility?
4. Is the conclusion the weaker claim that *a*'s credibility is open to question or the stronger claim that the conclusion of α is false?

One of the most important things to realize is that the circumstantial *ad hominem* argument, as defined by the scheme above, is not the same as an argument from inconsistent commitment. The latter, a species of argument from commitment has the following general form (Walton, 2006, p. 120).

Argumentation Scheme for Argument from Inconsistent Commitment

Initial Commitment Premise: *a* has claimed or indicated that he is committed to proposition *A* (generally, or in virtue of what he said in the past).

Opposed Commitment Premise: other evidence in this particular case shows that *a* is not really committed to *A*.

Conclusion: *a*'s commitments are inconsistent.

Argument from inconsistent commitment is also a defeasible form of argument that can be cast into doubt by asking one or more of a list of matching critical questions. To resolve the issues raised by the use of an argument from inconsistent commitments, it has to be judged how well the questions can be answered. To do this, a critic will need to go further into the textual details of the given case and judge matters like whether the alleged inconsistency is apparent or can be proved to be real, based on the arguer's commitments. Even if the arguer being attacked admits there really is an inconsistency of the alleged sort in his commitment set, he might still be able to explain how the conflict can be dealt with and resolved.

The important thing to recognize is that the difference between argument from inconsistent commitment and the circumstantial *ad hominem* argument is that the latter, but not the former, contains within it a direct *ad hominem* argument that is the basis of it. The problem is that throughout the long history of the subject (Walton, 1998, pp. 106-111) *ad hominem* has often been characterized as identical to argument from inconsistent commitment. It was argued in (Walton, 1998) that this view should be rejected, because all *ad hominem* arguments as a class, including the circumstantial type, should be defined as personal attack arguments. In other words, the thesis is that all genuine *ad hominem* arguments should contain an attack on the arguer's ethical character as an essential requirement. In the scheme for the circumstantial *ad hominem* argument above, that requirement is taken care of by the credibility questioning premise.

3. Applying the Scheme to the Case

There are two sides to the case. Let's call them the sealers group (S) and the humane society group, including the chef (C). We can attribute the following argument to C.

Argument Alpha

The seal hunt is inhumane.

Therefore it should be stopped.

We can help to stop it by boycotting Canadian seafood.

Therefore, we should boycott Canadian seafood.

This argument is an instance of practical reasoning, the first part of itself being an argument. It is implied that the first argument rests on the reason (implicit premise) that inhumane treatment of animals should be stopped. This yields argument beta.
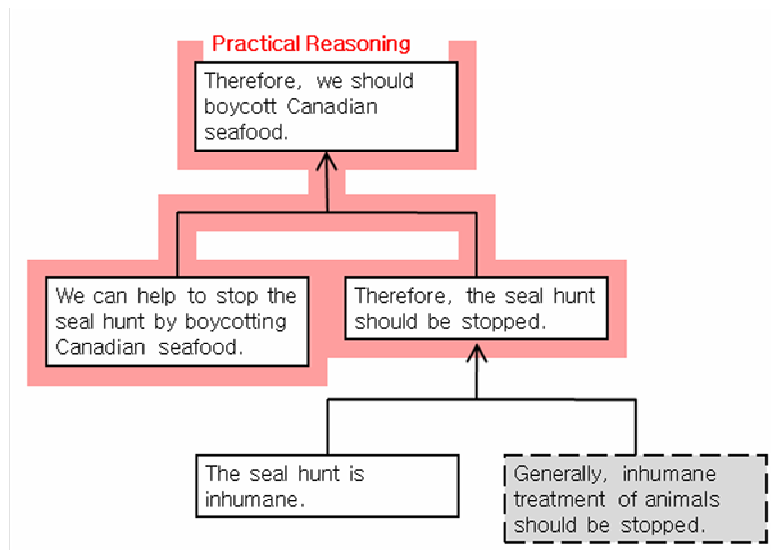
Argument Beta

The seal hunt is inhumane.

Implicit Reason: Generally, inhumane treatment of animals should be stopped

Therefore the seal hunt should be stopped.

Argument alpha is a typical chain of argumentation in which the conclusion of one argument reappears as a premise in an adjoining argument. Argument beta is an enthymeme, an argument in which an implicit premise is needed to combine with an explicit premise in order to support the conclusion. The two arguments are combined, as shown in the argument diagram in figure 1 below, constructed by using the free software tool for argument diagramming called *Araucaria* (Reed and Rowe, 2005). It aids a user when constructing a diagram of the structure of an argument by inserting the text to be analyzed as a text document into *Araucaria*. Each statement is represented in a text box that appears on the screen. Next a user can then draw in arrows from each premise to each conclusion it supports, producing an argument diagram connecting all the premises and conclusions in one comprehensive diagram.

Figure 1: Argument Diagram of Arguments Alpha and Beta

Practical Reasoning

Therefore, we should boycott Canadian seafood.

We can help to stop the seal hunt by boycotting Canadian seafood.

Therefore, the seal hunt should be stopped.

The seal hunt is inhumane.

Generally, inhumane treatment of animals should be stopped.

This argument diagram shows how the two premises displayed in the middle are linked together by an argumentation scheme called practical reasoning. At the bottom level, a

linked argument is displayed in which one premise is shown in a darkened box with dotted lines around the edges. This argument is an enthymeme, in which the premise on the right is an implicit assumption, not explicitly stated in the text of discourse of the sealers example. This premise is combined with the other one, as shown in the diagram, based on the scheme for argument from a verbal classification. The rationale is that the premise on the left at the bottom classifies the seal hunt as something that is inhumane.

S attacks argument beta, and thereby argument alpha, by using the following circumstantial *ad hominem* argument.

Circumstantial *Ad Hominem* Identified in the Sealers Example

Argument Premise: C advocates the proposition P that inhumane treatment of animals should be stopped.

Inconsistent Commitment Premise: C is personally committed to the opposite (negation) of P, as shown by commitments expressed in his personal actions or circumstances of serving *foie gras* made from the livers of force-fed geese (an inhumane practice, as C agrees).

Credibility Questioning Premise: C's credibility as a sincere person who believes in his own argument has been put into question (by the two premises above).

Conclusion: The plausibility of C's argument beta is decreased or destroyed.

The structure of this argument can be represented on an Araucaria diagram in a way that is highly revealing. The screen shot in figure 2 below shows how the circumstantial *ad hominem* is selected form the list of schemes called "Walton".

Figure 2: Screen Shot of the Selection of the Scheme in Araucaria
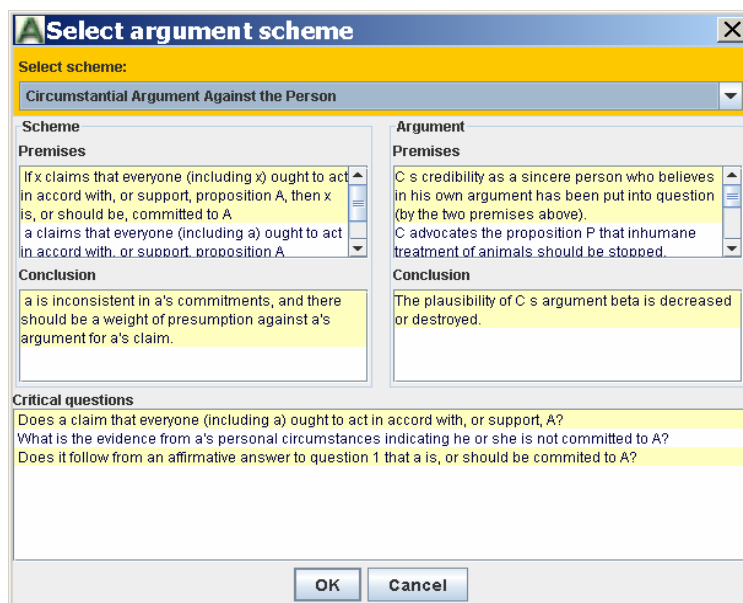
In figure 2 it is shown how the scheme is called circumstantial argument against the person (circumstantial *ad hominem*) in the Walton scheme list.

    Now it can be shown how the argument diagram produced by selecting this scheme looks in *Araucaria*. Figure 3 shows the structure of the circumstantial *ad hominem* argument in the sealers case below. It shows the premises and the conclusion of the argument, displays the argumentation scheme for the circumstantial *ad hominem* as applied to the argument, and shows how the scheme is fitted to the text of discourse of the sealers example, identifying the circumstantial *ad hominem* as a type of argument that can be shown to be present in the case.

Figure 3: *Araucaria* Diagram of the Circumstantial *Ad Hominem* Argument in the Sealers Example
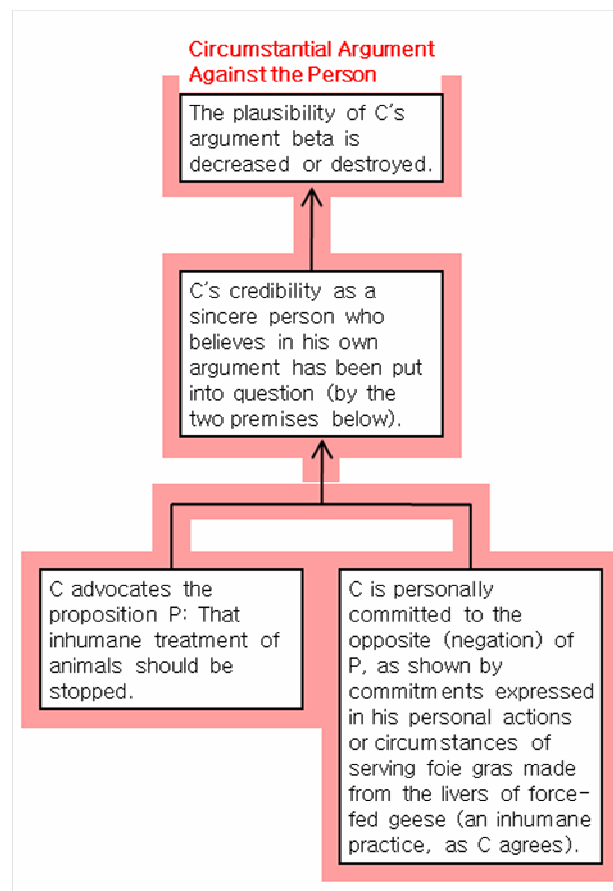
The interesting thing about the diagram in figure 3 is that it shows something unexpected. You would normally expect all the premises of the circumstantial *ad hominem* to go together as a linked argument supporting the conclusion. What the diagram reveals is that the two premises at the bottom go together in a linked argument configuration that supports the premise represented in the box in the middle of the diagram. What the diagram tells us is that the pragmatic inconsistency displayed in the bottom two premises functions as a reason for attacking the arguer's credibility, as displayed in the middle box. What is shown is that the top step in the argument shown in figure 3 can itself be

regarded as a direct *ad hominem* argument. The diagram reveals how the circumstantial *ad hominem* argument is connected to the direct *ad hominem* argument. It shows how the circumstantial *ad hominem* argument is a specials subspecies of the direct *ad hominem* argument that attacks the arguer's character for sincerity based on a perceived practical inconsistency in that person's argumentation and circumstances.

4. Verheij's Formalization

Verheij (2003) proposed a method for formalizing argumentation schemes, and used the schemes for *argumentum ad hominem* (Walton, 1998) as his primary body of material for illustrating how the proposed method will work. Verheij's system (p. 170) attempts to show how pragmatic argumentation schemes can be systematically analyzed using formal methods. His method is based on the key observation that there is a structural resemblance between logical rules of inference like *modus ponens* and pragmatic argumentation schemes like those for the *ad hominem* argument. Based on this resemblance, his method is to treat argumentation schemes as inferences having a premise-conclusion form (p. 170). He postulates his analysis (p. 176) on the assumption that any argumentation scheme is taken to have the following general form.

Premise 1. Premise 2. . . . Premise n. Therefore *Conclusion.*

Verheij (p. 177) uses an argument diagramming method called *ArguMed* to represent argumentation schemes and show how they apply to arguments in a given case. How the basic structure of any scheme is represented in *ArguMed* is shown by the argument diagram in figure 4, redrawn from the diagram in (Verheij, 2003, p.177).

Figure 4: Diagram of the Structure of an Argumentation Scheme in *ArguMed*
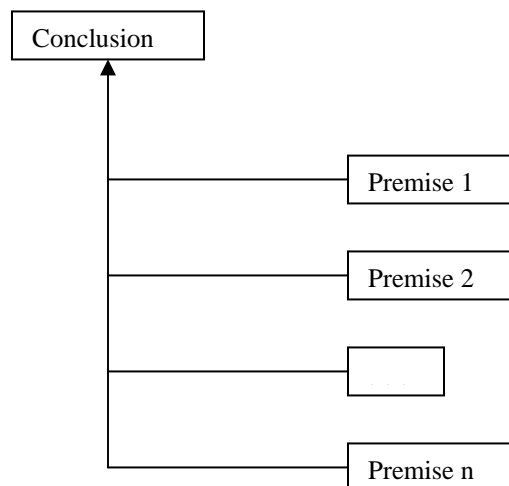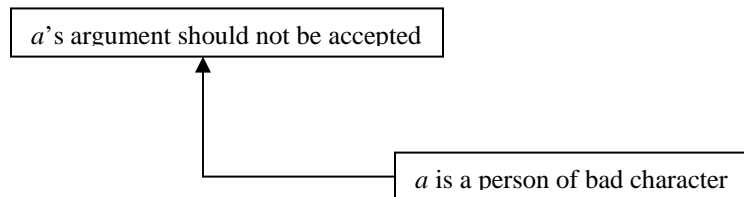

This argument structure can fit any of the argumentation schemes, including, as will be

shown below, the various types of schemes for *ad hominem* argumentation modeled in (Walton, 1998).

Verheij (2003 p. 177) offered a graphical representation of the argumentation scheme for the direct type of *ad hominem* argument found in (Walton 1998), similar to the one drawn in figure 5 below.
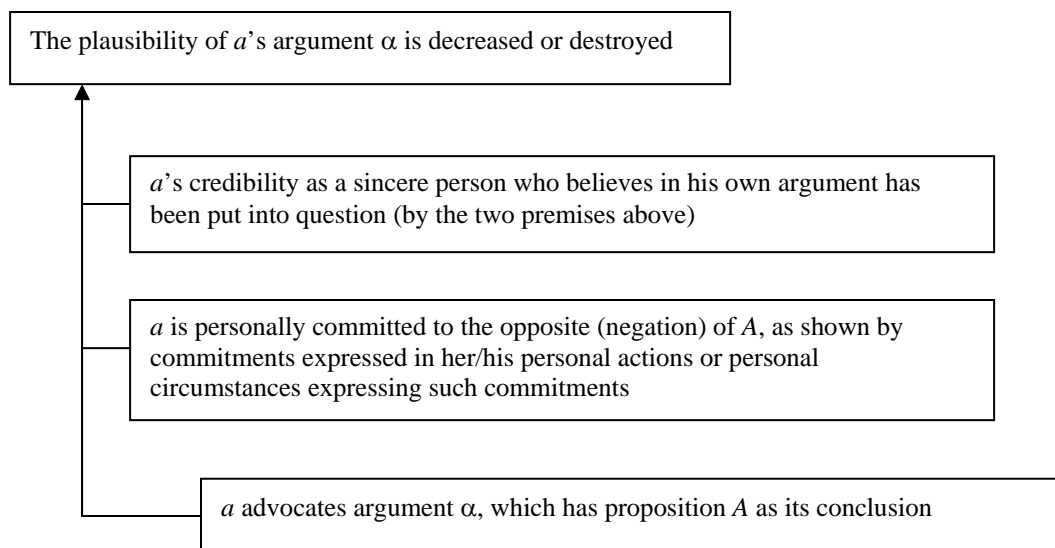
Figure 5: *ArguMed* Diagram for the Direct *Ad Hominem* Argument

```
┌─────────────────────────────────────┐
│ a's argument should not be accepted  │
└─────────────────────────────────────┘
                 ↑
                 │
         ┌───────────────────────────────┐
         │ a is a person of bad character │
         └───────────────────────────────┘
```

Another illustration can be given to show how Verheij represents a more complex type of *ad hominem* argument in his system of argument diagramming. He offers (Verheij, 2003, p. 178) a different graphical representation of the argumentation scheme for the guilt by association type of *ad hominem* argument found in (Walton 1998). It should be noted that these are two distinct argumentation schemes representing two different types of *ad hominem* arguments. The guilt by association type of *ad hominem* argument incorporates the direct type of *ad hominem* argument as an essential part.

Extrapolating from Verheij's presentation of these two types of *ad hominem* arguments in his system of argument diagramming, we can show how he might represent the circumstantial type of *ad hominem* argument of the kind represented by the scheme from (Walton, 2006) presented above.

Figure 6: *ArguMed* Diagram for the Circumstantial *Ad Hominem* Argument

```
┌────────────────────────────────────────────────────────────┐
│ The plausibility of a's argument α is decreased or destroyed │
└────────────────────────────────────────────────────────────┘
     ↑
     │  ┌──────────────────────────────────────────────────────────┐
     ├──│ a's credibility as a sincere person who believes in his own │
     │  │ argument has been put into question (by the two premises above) │
     │  └──────────────────────────────────────────────────────────┘
     │
     │  ┌──────────────────────────────────────────────────────────┐
     ├──│ a is personally committed to the opposite (negation) of A, as │
     │  │ shown by commitments expressed in her/his personal actions or │
     │  │ personal circumstances expressing such commitments            │
     │  └──────────────────────────────────────────────────────────┘
     │
     │  ┌──────────────────────────────────────────────────────────┐
     └──│ a advocates argument α, which has proposition A as its conclusion │
        └──────────────────────────────────────────────────────────┘
```

13

A clear difference between this argument diagram and the previous *Araucaria* diagram representing the circumstantial *ad hominem* argument in figure 3 is that the *Araucaria* diagram explicitly represented the distinction between linked and convergent arguments. The argument in the sealers case in figure 3 showed how the circumstantial argument against the chef is a linked argument in which two of the premises are bound together in the argumentation scheme for the circumstantial *ad hominem* argument, and then combined with the remaining premise.[2]

Verheij sees pragmatic argumentation schemes as being inherently defeasible. In a typical pragmatic argumentation scheme, the premises do not justify the conclusion exclusively on the basis of the premises. Schemes are subject to exceptions. Thus one of the important steps in the methodology of his investigation of formalizing argumentation schemes is to determine the exceptions blocking the use of a given argumentation scheme (p. 174). On his approach it is important not only to determine the types of premises characteristic of an argumentation scheme, and to analyze the logical structure of the scheme as an inference, but also to determine what kinds of exceptions might make an argument fitting an argumentation scheme default. In *ArguMed*, blocking moves that make an argument default are drawn by a device called entanglement. Entanglement is represented as a line that meets another line at a junction marked by an X, indicating that new evidence attacks the inferential link between the premises and conclusion of the original argument, making the original argument default.

The idea of entanglement is that the connection between a reason and its conclusion, represented by the arrow between them, can be the subject of argumentation, just like other claims. Such argumentation can be either supporting, by giving reasons for the connection between the reason and its conclusion, or attacking, by giving reasons against the connection between the reason and its conclusion. The positive version of entanglement is represented graphically by an ordinary arrow pointing to another arrow. The negative version of entanglement is represented graphically by an arrow ending in a cross (X) pointing to another arrow.

The argument diagram in figure 7 illustrates how new evidence of this kind can make an instance of the circumstantial *ad hominem* argument default. It illustrates the negative version of entanglement.

Figure 7: Default of a Circumstantial *Ad Hominem* Argument in *ArguMed*

---

[2] Note also that *ArguMed* version shown in figure 6 is a single step graphical representation of the argumentation scheme presented for the circumstantial *ad hominem* argument, whereas the graphical representation shown in figure 3, the *Araucaria* version, is a different analysis. The latter version has two premises, one intermediate conclusion and one final conclusion. The difference between the former and the latter versions is that in the latter, the intermediate conclusion is a premise is a next step in the argument. This difference suggests the hypothesis that the analysis shown in figure 3 may be better because it displays a dependency between the three premises in a way that is made explicit. However, this difference between the two versions does not imply a difference between *Araucaria* and *ArguMed* as visualization tools for analysis. In both it is possible to chain argument steps.

14
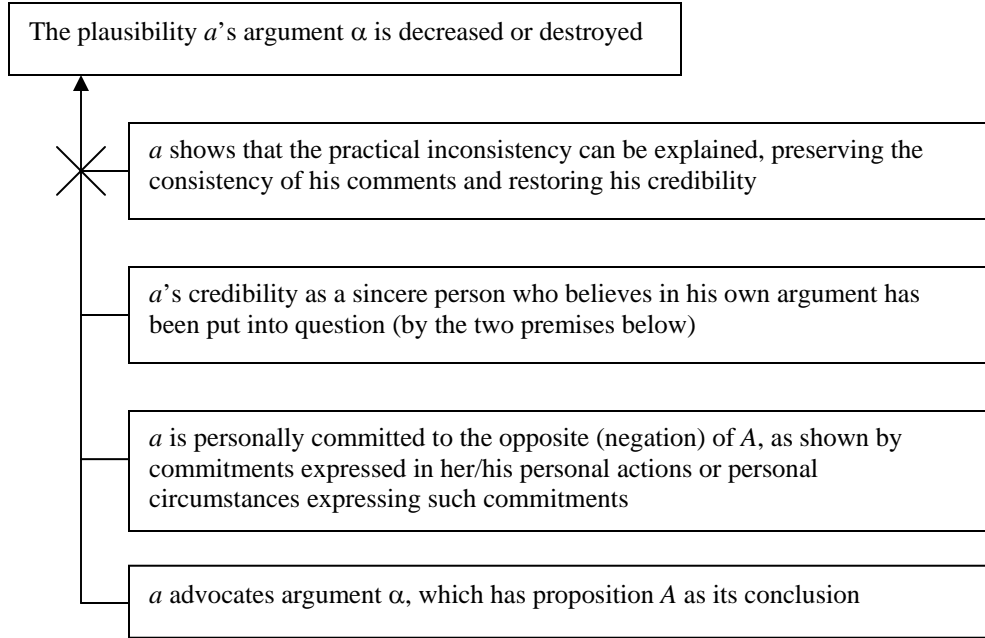
As shown in figure 7, the argument fits the form of the circumstantial *ad hominem* type of argument, as indicated by the conclusion and the three bottom premises, but the fourth premise, when added, makes the *ad hominem* argument default.

This last step takes us into the question of how the critical questions matching a given scheme should be used to raise doubts about the applicability of the scheme and make an *argumentum ad hominem* default. Can the critical questions be modeled as implicit premises that, when added to an argument, can either support it or make it default? This possibility is extremely exciting, from a point of view of formalizing argumentation schemes, because the job would be made much easier if we could represent the critical questions as propositions that function as additional premises of a given argument, thus relieving us of the problem of how to deal with the logical form of questions.

5. Defeasible Generalizations and Argument Defeat

In this section it is shown that there are three kinds of generalizations that need to be distinguished: (1) the universal (absolute) generalization, (2) the inductive (statistical) generalization, and (3) the presumptive (defeasible) generalization. The statement 'All birds fly' can be interpreted as a universal generalization, if taken as a statement that is falsified by a single counter-example. Taking the predicate letter F to stand for the property of being a bird, and the predicate letter G to stand for the property of being something that flies, the universal generalization 'All birds fly' may be represented has the standard logical form $(\forall x)(Fx \supset Gx)$ in classical deductive logic. As an example of a type of argument based on this kind of generalization, consider the following deductively valid inference.

$(\forall x)(Fx \supset Gx)$

Fa

_____

Ga

Consider an inductive argument, in contrast, based on the inductive generalization '98.6 per cent of birds fly', with the additional premise that Tweety is a bird. These premises support the conclusion that Tweety flies by an inductively strong argument. It is logically possible for both premises to be true and the conclusion false, but it is improbable for both premises to be true while the conclusion is false. We are familiar with recognizing the distinction between deductively valid and inductively strong inferences. While some will argue about the details of how the distinction should be made, there is wide general acceptance that such a distinction is necessary and important.

The classic example of a third kind of generalization we might call the defeasible or presumptive type (terminology does not appear to be settled) is the statement 'Birds fly' taken to mean that birds generally fly, subject to exceptions that cannot all be anticipated in advance. This type of generalization says, 'Given something is a bird, we expect it to fly, in a normal type of case, but in special circumstances, it might fail to fly'. In the special circumstances that Tweety is a penguin, or a bird with a broken wing, Tweety cannot fly, and in such a special case, the generalization is subject to an exception, making any argument based on it default. As Reiter (1987, p. 149) showed, this kind of argument is tenable in cases of uncertainty and absence of knowledge, and is based on a qualified generalization with a clause saying "absence of information to the contrary". The argument works on a principle of a shifting burden of proof, implying that it holds tentatively during an investigation, as long as it has not been disproved, but it can fail if new evidence is collected that refutes it, as applied to a special case. If the proponent fails to give the evidence required to prove the claim, once it has been shown to default, he must retract the claim. But in this case, the claim is presumptive in nature. The defeasible generalization occupies a ground tentatively. As long there is nothing special about Tweety indicating that he cannot fly, the argument for the conclusion that Tweety can fly holds. But if an opponent cites new evidence that revealing special features of the case that present an exception to the generalization, it fails to hold, and must be given up.

Presumptive generalization is different from inductive reasoning of the Bayesian kind because it is based on what one can presume to be true in a normal type of case, based on what is known so far in an investigation, even though it is known that not everything is known yet. Special circumstances can make the generalization default in way that cannot be known or predicted in advance by probabilities (Rescher, 1976). What really matters is not statistical regularity, but whether the generalization is subject to exceptions of a kind that are not all known in advance at the stage the investigation has now reached (Prakken and Sartor, 2003). This pattern can be illustrated by expressing the classic Tweety argument in the following form.

Given that Tweety is a bird, generally, but subject to qualifications in special circumstances, Tweety can fly.

It can be established that Tweety is a bird.

So far, there are no special circumstances of the given case that would make the generalization in the first premise default.

Therefore, as far as we know at this point in the investigation, Tweety can fly.

This argument holds, as things stand, because generally it is true that birds fly. However, it can also fail in specific instances. In a case where it is discovered that Tweety is a penguin, the first two premises of the argument above hold as true but the third one fails, making the argument default.

Verheij (2001, p. 232) proposed the hypothesis that argumentation schemes can be modeled using defeasible generalizations (1999, p. 113) based on a distinction between two rule-based forms of inference. The first one is the familiar *modus form* of argument from classical deductive logic.

Modus Ponens

As a rule, if *P* then *Q*

 P

Therefore *Q*

Modus Non Excipiens

As a rule, if *P* then *Q*

P

It is not the case that there is an exception to the rule that if *P* then *Q*

Therefore *Q*

On his model, (2000, p. 5), in an instance in which only strict rules are involved, *modus ponens* can be applied, but *modus non excipiens* needs to be applied in an instance where both strict rules and rules not admitting of exceptions might possibly come into play. For example if the given argument is based on a universal generalizations about all triangles, without exception *modus ponens* can be applied. But *modus non excipiens* always needs to be applied in a case where the generalization is subject to qualifications. Rather than using cumbersome Latin expressions like *modus non excipiens*, these two forms of argument might better be called strict *modus ponens* (SMP) and defeasible *modus ponens* (DMP). The terminology remains unsettled. Some would claim that DMP is not really a *modus ponens* type of argument, and should not even be called a *modus ponens* argument at all. Others are happy to use the label DMP, as long as the distinction between it and the deductive form SMP is made clear.

It is already clear from the literature on argumentation schemes that some schemes have a DMP structure. The scheme for argument from commitment is a case in point. In the following representation of this scheme (Walton, 1995, p. 144), *P* is a participant in a discussion and *A* and *B* are propositions.

Scheme for Argument from Commitment

Generally, if *P* is committed to *A* than *A* is also committed to *B*.

 P is committed to *A*.

Therefore *P* is committed to *B*.

Many of the common schemes studied in the literature are defeasible, and it is not very plausible that they are reducible to a deductive form or to some known inductive form like the statistical syllogism (Reed and Walton, 2005).
    The first step in revising the scheme for the direct *ad hominem* to fit the requirements of a system of defeasible argumentation schemes is to add the generalization that functions as a warrant to license the inference from the other premise to the conclusion.

Revised Argumentation Scheme for the Direct Ad Hominem Argument

Character Attack Premise: *a* is a person of bad character.

Generalization: given that *a* is a person of bad character, then generally, but subject to qualifications in special circumstances, *a*'s argument should not be accepted.

Qualification: so far, there are no special circumstances of the given case that would make the generalization premise default.

Conclusion: as far as we know at this point, *a*'s argument should not be accepted.

Now we come to the critical questions. As pointed out by Verheij, the first critical question merely repeats the first premise. And as noted above, the second question is a matter of relevance. Assuming that both are redundant in a formalized treatment, the third critical question is the only one that needs to be considered.
    The third critical question draws a distinction in how to interpret the conclusion. This question functions like a warning not to uncritically take the conclusion in a stronger way, when usually *ad hominem* argumentation is better taken in the weaker way.

Revised Argumentation Scheme for the Circumstantial Ad Hominem Argument

Argument Premise: *a* advocates argument α, which has proposition *A* as its conclusion.

Inconsistent Commitment Premise: *a* is personally committed to the opposite (negation) of *A*, as shown by commitments expressed in her/his personal actions or personal circumstances expressing such commitments.

Credibility Questioning Premise: *a*'s credibility as a sincere person who believes in his own argument has been put into question (by the two premises above).

Generalization: given that *a* advocates argument α, which has proposition *A* as its conclusion, and *a* is personally committed to the opposite (negation) of *A*, as shown by commitments expressed in her/his personal actions or personal circumstances expressing such commitments, and *a*'s credibility as a sincere person who believes in his own argument has been put into question, then generally, but subject to qualifications in special circumstances, the plausibility *a*'s argument α is decreased or destroyed.

Qualification: so far, there are no special circumstances of the given case that would make the generalization premise default.

Conclusion: as far as we know at this point, the plausibility *a*'s argument α is decreased or destroyed.

But now, having revised the structure of the scheme for the circumstantial *ad hominem* argument, we have the problem of how such a scheme could be formalized.
This choice raises the problem of how such defeasible arguments are to be evaluated, which depends on the controversial notion of argument defeat. A first pass through the literature on argument defeat must begin with the distinction between undercutters and defeaters (Pollock, 1995, pp. 40-41) A defeater, sometime also called a rebuttal or refutation of an argument, is defined by Pollock as another argument that has the opposite (negation) of the original conclusion as its conclusion.[3] A Pollockian undercutter is a counter-argument that attacks the inferential link between the premises and the conclusion in the original argument. The undercutter seems to be the weaker form of attack, one that only raises questions on whether the original argument supports it conclusion, leaving room for doubt. The defeater appears to be a stronger form of attack.

6. Two Formal Systems: DefLog and Carneades

Another most difficult, central theoretical problem in current argumentation theory is the analysis of the concept of argument defeat. Argument defeat is obviously something different from classical negation, but exactly what it is has never been made clear. Two types of argument rebuttals called defeaters and undercutters have been recognized in the literature (Pollock, 1995), but the precise nature of the distinction, and how it is to be applied to legal argumentation, has proved to be a source of some uncertainty.
Verheij (2003a) developed a formal system of defeasible reasoning called DefLog that, as shown above, can be used to model the argumentation schemes for *argumentum ad hominem*, including both the direct and circumstantial types. His model represents

---

[3] Pollock (1995, p. 40) calls a "rebutting defeater" a reason for denying the conclusion of an argument where the premises offer a *prima facie* (defeasible) reason for the conclusion.

some of the critical questions as undercutters of a scheme and others as defeaters. Critical questions that function as undercutters or rebutters of an argumentation scheme are modeled in DefLog using three propositional operators. The arrow ~> , called the defeasible conditional (Verheij, 2003, p. 184) is a form of implication that conforms to only one logical rule of inference, *modus ponens*. The x is the unary operator for dialectical negation, used to express the claim that a sentence is defeated (Verheij, 2003, p. 184). We have to be careful to observe that if a proposition's dialectical negation is justified, that shows the argument is defeated, but the converse does not obtain. If a proposition is justified, that does not necessarily show that its dialectical negation is defeated (Verheij, 2003a, p. 327).[4] The & (ampersand) is the operator for conjunction.

An undercutter is directed to attacking the inferential link between the premises and the conclusion in the original argument. In its logical form, the dialectical negation applies to the whole defeasible conditional of the argument attacked.

x (premise 1, premise 2, . . . , premise n ~> conclusion)

Arguments attacking the conclusion of the original argument, are modeled (below) as defeaters. Here the dialectical negation applies only to the consequent.

premise 1, premise 2, . . . , premise n ~> x conclusion

Undercutters have a different kind of underlying logic than defeaters in Verheij's system. This distinction will come to be important when we come to see how Verheij approaches the problem of modeling the critical questions matching the two *ad hominem* argumentation schemes, the problem to be taken up in the next section.

A new computational system of reasoning of with argumentation schemes called Carneades provides a model that enables argumentation schemes to be integrated into the semantic web. It defines structures for integrating basic elements of argumentation, including atomic propositions, arguments, cases, issues, argumentation schemes and proof standards. In one respect, Carneades is an extension of Verheij's proposed system for formalization of argumentation schemes, because it portrays every scheme as based on a set of premises, one of which is a generalization that can be expressed in the form of a defeasible conditional statement. Thus, like Verheij's system, it casts the various presumptive schemes of the kind commonly analyzed in (Walton, 1995) as having the DMP form. A useful feature of Carneades for formalizing argumentation schemes is that it provides a way of formalizing the critical questions as components of any given defeasible argumentation scheme.[5]

Carneades is not a formalization of argument in the manner of a deductive formal system of logic. It is a computational model that builds on ontologies from the semantic web to provide a platform for employing argumentation schemes in legal reasoning. In effect, the model is an abstract functional specification of a computer program that can be

---

[4] According to Verheij (2003a, p. 327), neither of the double negation rules of classical logic holds for dialectical negation

[5] Note however that Verheij's format for argumentation schemes has four kinds of elements: premises, exceptions, conditions of use and conclusions. Each of these elements corresponds to a different kind of critical question.

implemented in any programming language. Arguments modeled in the Carneades system can be visualized using an argument diagram because the basic structure Carneades uses, following the model of the semantic web, is that of the directed labeled graph. The nodes represent objects and the arrows represent binary relations.[6] It is especially important for our purpose to look carefully at how Carneades models the notion of argument defeat. What are called defeaters or rebuttals in Pollock's language are modeled as arguments that are opposed to another argument at issue. For example if one argument at issue is *pro*, its rebuttal would be another argument *con* that argument. Premise defeat is modeled by an argument *con* an antecedent or assumption, or *pro* an exception (Gordon, 2005, p. 56). In Carneades, a Pollock-style undercutter of an argument *n* is modeled as an argument against the implicit applicability presumption of a scheme, i.e. an argument *con* the atom (applies arg-*n* true).

The Carneades system defines three kinds of defeat relations or argument rebuttals. The first type consists of arguments con the consequent of a pro argument. The second type consists of arguments pro the consequent of a con argument. The third type consists of arguments pro alternative positions on the same issue. Undercutters correspond to arguments con applicability presumptions. An example is the assumption that the generalization 'Birds fly' is applicable to Tweety.

An output representing an argument in Carneades can be visualized using an *Araucaria* argument diagram. The basic units of the system, called atoms, as in the example, (asserts Gloria (killed joe sam)), are made up of subjects, objects and binary predicates. Atoms are defined as RDF triples of the following kind (Gordon, 2005, 54).

```
 type atom =
{ predicate: symbol'
 subject: symbol,
 object: datum }
```

Argumentation is viewed as a model construction process that tracks issues about which atoms should be included in a domain model. An *issue* is defined as a record for keeping track of the arguments *pro* and *con* each position (Gordon, 2005, p. 55). A *position*, which can be accepted, rejected or undecided, is a proposed or claimed value of an atom. An argument is defined as a single atom called a consequent and another set of atoms called premises. The distinctive feature of how Carneades models arguments is that is recognizes three different types of premises called antecedents, assumptions and exceptions. Antecedents are normal premises that are assumed to be acceptable, and must be justifiable to make an argument acceptable. Assumptions are assumed to be acceptable unless called into question. Exceptions are premises that are not assumed to be acceptable, but are taken for granted for the sake of argument unless they are challenged. This distinction can be clarified by comparing the three kinds of premises to critical questions matching an argumentation scheme. They can be classified into three categories, depending on whether they are treated as antecedents, presumptions or exceptions. Antecedents are like premises that are already present as required premises in

---

[6] The resource description framework (RDF) of the semantic web provides an XML syntax representing a graph.

a scheme, and so critical questions questioning them can be seen as redundant. Assumptions are premises that are assumed to be true, while exceptions are premises that assumed to be false, even though they may later be shown to be true.

It is important to note that an exception is different from the negation of a presumption. An assumption is an additional premise of the generalization (conditional) that represents the structure of the argument. Consider the conditional in the Tweety example, shown as an argument in Carneades.

```
(if and (isa tweety bird) (applies arg-1 true)) (canfly tweety true))
```

If we tried to model an exception as the negation of presumption, we get the conditional:

```
if (and (isa tweety bird) (not (isa tweety penguin)) (applies arg-1
true)) (canfly tweety true))
```

The proper way to model an exception is by a conditional of this form:

```
if (and (isa tweety penguin) (applies arg-1 true))bottom)
```

An exception, as modeled in Carneades, expresses a constraint on an argument. It cannot be treated in the same way as an application of an argument.

Each argument is provided with an identifier called an *id*. The following definition (Gordon, 2005, p. 56) shows how all the various elements are combined in the definition of the data type for an argument. Arguments are provided with an identifier allowing propositions to be about arguments, because the identifier of an argument can be used as the subject or object of an atom. Using this feature, an applicability presumption of the form (applies <argument-id> true) is added to every argument.

```
type argument =
{ id: id,
  direction: {pro, con},
  consequent: atom,
  antecedent: atom list,
  assumptions: atom list,
  exceptions: atom list}
```

As noted above, premise defeat is modeled by an argument *con* an antecedent or assumption, or *pro* an exception (Gordon, 2005, p. 56). Also as noted above, Carneades follows Verheij's lead in modeling the most familiar sorts of argumentation schemes as having the DMP form.

There are many interpretations of the Tweety argument, but let's consider one that allows for different ways for Tweety to be abnormal, and to be an exception to the generalization that birds normally fly. The term 'abnormal' should be taken to mean abnormal because of being a nonflying bird, and not abnormal in some other way. This version of the Tweety argument can be visualized in Carneades as shown in figure 8. Assumptions are shown as statements leading to a conclusion by a closed dot arrowhead, while exceptions are shown with an open dot arrowhead.

Figure 8: Diagram of the Stage 1 of the Tweety Argument
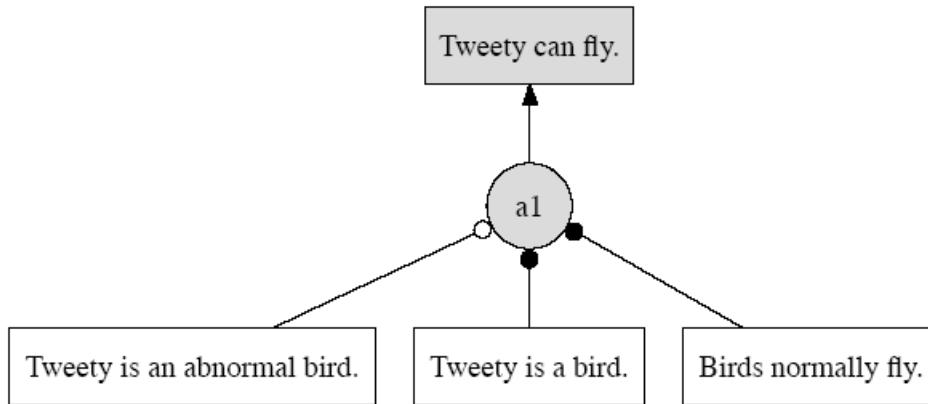
In the diagram in figure 8, the statement 'Tweety is abnormal' could still be supported with further argument, but is not, and thus it does not make the argument default. Both premises 'Tweety is a bird' and 'Birds normally fly' are assumptions in this argument, each shown to lead to the conclusion by a closed dot arrowhead. Thus they support the argument a1, leading to the conclusion that Tweety can fly. The statement 'Tweety is an abnormal bird' is shown as an exception (indicated by the open dot arrowhead). It is assumed to be false, unless further evidence indicates that it applies. The argument a1 is shown in a darkened node, indicating that it is acceptable, and thus the conclusion drawn from it is also shown as acceptable.

So far, we have seen only the first stage of the Tweety argument, where there is no evidence so far indicating that Tweety is other than a normal bird. If we add new evidence that would support the statement that Tweey is an abnormal bird, it would defeat the previous argument shown in figure 8. This development is shown in figure 9, introducing the additional statement that Tweety is a penguin.

Figure 9: Stage 2 of the Tweety Argument
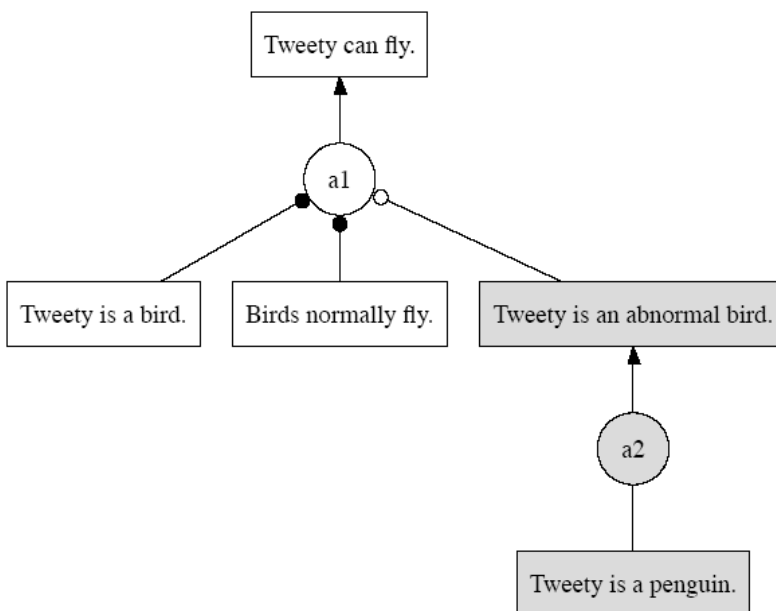
In figure 9, the statement 'Tweety is a penguin' is darkened, showing that it is accepted. Thus argument a2 is acceptable, leading to the accepted conclusion that Tweety is an abnormal bird. Adding this evidence shows that a1 is not longer acceptable, and that the conclusion 'Tweety can fly' must now be rejected. Thus the original argument shown in figure 8 has defaulted, because new evidence shows that it is no longer acceptable.

7. Formally Modeling the Critical Questions

The general problem is how to represent the critical questions as ways of challenging or refuting arguments that fit one of these forms of argument. Verheij suggested that it may be useful to treat some of the questions in a different way from others. Critical questions that point to exceptions to a general rule only undercut an argument, while others could be seen refuting the argument by denying implicit assumptions on which it rests, or by pointing to counter-arguments. He began by showing that critical questions can have four different kinds of roles.
1. They can be used to question whether a premise of a scheme holds.
2. They can point to exceptional situations in which a scheme should not be used.
3. They can set conditions for the proper use of a scheme.
4. They can point to other arguments that might be used to attack the scheme.
Verheij argued that critical questions that criticize the premises of a scheme are redundant because they merely ask whether the premise is true. It is a condition of the use of any argument that the premises are true, or at least are acceptable. Thus he argued that critical questions that merely restate a premise of an argumentation scheme are redundant, and can be ignored. For example, the field question, in the list of critical questions matching argument from expert opinion above, could be said to be redundant, because the major premise already says that E is an expert in field F containing proposition A.

We now come to the difficult problem of modeling the critical questions by attempting to integrate them as components of the argumentation scheme for the *argumentum ad hominem*. Verheij (2003, p. 82) begins by citing the first critical question matching the direct *ad hominem* argument. It asks whether the premise is true, or well supported, that the arguer is a bad person. He notes (p. 182) that it is a precondition of the use of any scheme that its premises are true or well supported. He concludes that critical questions of this kind are already implicitly given in the argumentation scheme itself, and that therefore it is not necessary to have them as separate critical questions. From a formal viewpoint, this analysis is reasonable. Such critical questions are very useful for practical purposes of teaching students in an argumentation course to think more critically by looking at standard ways and argument should be questioned. However when it comes to the project of formalizing argumentation schemes, it is not necessary to have specific questions merely asking whether a premise is true. The reason is that we can automatically assume that a critic can question whether any premise of the scheme is true.

Critical questions of the third kind, that set conditions on the applicability of the scheme, are modeled in DefLog using a different kind of rule. In this rule, C represents a condition on the applicability of a scheme.

(premise 1, premise 2, . . . , premise n ~> conclusion) only if *C*

According to this rule, if condition *C* does not apply, the scheme is undercut. As an examples of this kind of critical question Verheij (2003, p. 183) cites the critical question matching the *ad hominem* type of argument that asks whether character is relevant. Thus it would seem that critical questions asking about relevance would come under this classification. Relevance represents a fourth way of attacking an argument, apart from the usual three ways commonly recognized – rebutting, undercutting and attacking a premise.

When Verheij (2003, p. 183) examines the second critical question, which asks whether the allegation that the arguer is a bad person is relevant to judging his argument, he suggests that this type of critical question has a role that corresponds to what he calls "conditions for a schemes use". This means that if the allegation that the arguer is a bad person is irrelevant to judging his argument, a kind of problem that often occurs, we can say that the argument used in a given case does not meet the required conditions for the correct use of a scheme. In other words, the criticism here is that the scheme has been misapplied by attempting to use it in a context where it should not properly be used. A problem with this approach is that much the same thing could be said about the redundancy of the critical question that was said about the first critical question matching the direct *ad hominem* argumentation scheme. With respect to any given argument matching any one of the schemes, the question can always be asked, and perhaps should always be asked, of whether the argument is relevant. Thus it would seem that neither of the critical questions is necessary in the case of the direct *ad hominem* type of argument. Both are redundant, when it comes to formalization, because both are routine questions that would apply to any scheme.

The practical problem remains however that when giving students advice on how to analyze and evaluate an *ad hominem* argument in a given case, there are two criticisms that are especially important and fundamental. Many direct *ad hominem* arguments are based on innuendo, where very little evidence is brought forward to support the allegation that the arguer has bad character. It is most important to warn students about this, because the *ad hominem* argument is such a powerful attack that it can often unseat an opponent or critic who fails to think twice that the evidence lacking to back up the allegation is one source of questioning or attacking the argument.

DefLog provides a formal model of argumentation that allows us to represent critical questions as different kinds of undercutting or defeating counter-arguments that can be used to attack an argument fitting one of the *ad hominem* schemes. Thus it, for the first time, provides a method that can be used not only to identify and analyze *ad hominem* arguments, but also to evaluate them. This step forward is a big one. Carneades goes even further by showing how the critical questions can be incorporated into the scheme itself.

Carneades is a computational system that builds on technologies from the semantic web and provides a platform for using argumentation schemes for argument analysis and evaluation (Gordon and Walton, 2006). An argument is defined as a triple, made up of a statement designated as the conclusion, a direction, pro or con, and a set of premises (statements). Arguments in Carneades are visualized with argument diagrams that are compatible with the semantic web, using an XML syntax. The Carneades diagram visualizes an argument as a directed graph in which the nodes are statements or

arguments, and the lines joining the nodes represent inferences from a set of premises to a conclusion, or from an argument to a conclusion. An argument is identified by recognizing its scheme (*id*), and its direction, *pro* or *con* the statement at issue. Arguments are judged as acceptable or not in relation to an issue being discussed. An issue functions as a record for keeping track of the arguments *pro* and *con* each position as the argument progresses through the dialogue (Gordon and Walton, 2006). A statement can be accepted, rejected or at issue. Whether a premise holds in any given argument depends on its dialectical status and the type of premises it contains.

The key motivation of the Carneades system is its capability for dealing with two kinds of critical question (Walton and Gordon, 2005). Carneades recognizes three types of premises, called ordinary premises[7], assumptions and exceptions (Gordon and Walton, 2006). Assumptions are assumed to be acceptable unless called into question. Ordinary premises are automatically classified as presumptions. Both of these kinds of premises are taken to hold unless they are at issue. Exceptions, on the other hand, are taken not to hold. An exception can block or undercut the acceptability of an argument as a dialogue proceeds if evidence comes in supporting the statement classified as an exception in the argument. It does this by revealing assumptions and exceptions as implicit premises in a given argument as the argument is critically questioned or attacked.

Carneades models the argumentation scheme for the direct *ad hominem* argument by including the factors corresponding to the critical questions as premises of the scheme. It accomplishes this task by distinguishing between two kinds of premises called assumptions and exceptions. The scheme for the direct type of *ad hominem* argument is formalized as follows.

Formalization of the Direct *Ad Hominem* Scheme in Carneades

```
id: arg-i
direction: con,
scheme: direct-ad-hominem-argument,
conclusion: (a's argument should not be accepted),
ordinary premises:
   (a isa person of bad character),
presumptions:
   (assertion (a isa person of bad character (based-on-evidence))
   (character is at issue),
exceptions:
   (a's argument should be accepted true)absolutely rejected)
```

The argumentation scheme for the circumstantial *ad hominem* argument can be formalized in the Carneades system as follows.

Formalization of the Circumstantial *Ad Hominem* Scheme in Carneades

```
id: arg-i
direction: con,
scheme:circumstantial-ad-hominem argument,
conclusion: (a's-argument-plausible-true),
ordinary premises:
   (personal-circumstances a show a con a's argument),
assumptions:
```

---

[7] Ordinary premises are the premises explicitly stated in a scheme. Assumptions and exceptions correspond to critical questions matching a scheme.

```
   (pair-of-commitments shown to be practically inconsistent)
   (based-on-evidence e true)
   (character is at-issue),
exceptions:
  (practical inconsistency can be resolved by further dialogue)
  ((a's argument should be accepted true)absolutely rejected)
```

Carneades offers a formalization of these schemes that removes the need to deal with the critical questions in a purely dialogue format, by incorporating them into the premises of the scheme. Using this method of formalization of argumentation schemes, all of the schemes for the various types of *ad hominem* arguments analyzed in (Walton, 1998) could be formalized in the same manner. This work can be suggested as a project for further research on the *ad hominem* argument.

8. Evaluation of the *Ad Hominem* Argument in the Sealers Example

Now we come back to the original question: how could the circumstantial *ad hominem* argument identified in the sealers example the evaluated? First, we have to comment on how the critical questions were analyzed. What was suggested by the remarks in sections 6 and 7 is the importance of carefully distinguishing between the project of formalizing the critical questions matching an argumentation scheme and the project of teaching critical argumentation skills by formulating common failings of a type of argument that a critic needs to be aware of. Critical questions in this more practical sense function as pointers to gaps in the logical structure of an argument that a critic needs to be aware of as requiring justification.

Now let's turn to the four critical questions cited as appropriate for the circumstantial type of *ad hominem* argument. The first question is important because the pair of commitments said to be inconsistent need to be identified. The first step, in order to evaluate the charge of inconsistency, is to identify the pair of propositions that are supposed to be inconsistent. In many cases one or both of these propositions are implicit premises of a chain of argumentation in the discourse. In such cases much of the argument analysis that is required prior to evaluating the argument is the task of digging out this pair of propositions in such a way that both can be explicitly stated in a way that is fair to the text of discourse. Unfortunately however, from a theoretical point of view, this question has been formulated as a yes-no question, merely asking the critic whether such a pair of propositions can be said to exist, yes or no. Although this form of question works well enough as a practical tool to aid critical argumentation, in order to formalize this argumentation scheme it might be better to ask the critic to specifically identify the pair of propositions.

The second critical question asks whether the inconsistency that has now been identified could be resolved or explained by further dialogue. The third critical question in effect asks the question of relevance, by asking whether the arguer's original argument should depend on his or her credibility. If credibility is an issue then an *ad hominem* argument is relevant whereas otherwise it tends not to be. Therefore this critical question comes very close to asking whether the *ad hominem* argument is relevant in the given case. The fourth critical question concerns the conclusion of the *ad hominem* argument attacking the original argument. Essentially, this question asks whether the *ad hominem* argument should be interpreted as a defeater or as an undercutter of the original

argument. This question is of practical importance because in many instances, an *ad hominem argument* can be a legitimate way of asking questions about the arguer's credibility, but it is such a powerful form of argument that in many cases it is taken to resoundingly defeat the original argument. It is this observation that is the basis of the *ad hominem* fallacy (Walton, 1998). Such an interpretation suggests that the argument cannot be rehabilitated, and that the original argument has been so strongly refuted that no further attempts to defend it could ever possibly succeed. In other words, this interpretation of how to evaluate such an argument is one that suggests that the investigation should be closed off. The problem is that in many cases such a strong interpretation of the argument is inappropriate.

Now let us return to the sealers example, and discuss the problem of how to evaluate the circumstantial *ad hominem* identified in it as reasonable or fallacious. S has a good argument, assuming the premises of the circumstantial *ad hominem* argument against C are justified as supported by the evidence in the case. How good is C's reply? C's reply could perhaps be justified by claiming it fits critical question 2 or critical question 3. Let's consider them one at a time. It could be argued that even though the practical inconsistency has been identified, and a plausible case for it has been made, it could be resolved or explained by further dialogue showing that C's inconsistent commitment does not support the claim that he lacks credibility. It might be argued that the "brutal destruction" of the seals is so inhumane that stopping it is the most important thing, and that being consistent does not matter, compared to that. The argument is that C has credibility, because his worthy ethical motive is to save the seals.

Similarly it might be argued, by citing critical question 3, that character is not the issue in the dialogue, and that C's argument does not depend on his credibility. The reason offered is that the "brutal destruction" of the seals is so inhumane that stopping it is the most important thing, and that one person's character does not matter, compared to that. Both attempted rebuttals are based on relevance, the argument essentially being that saving the seals is so important that it is the only issue to be considered, and that matters of consistency and character are irrelevant. S puts forward this argument saying that the *foie gras* issue is not relevant, because the issue is about seals. One way in which an *ad hominem* argument is typically fallacious concerns its relevance (Walton, 1998). *Ad hominem* arguments can often be relevant for example in political discourse and election campaigns, or in questioning the character of a witness in a trial. But there are also plenty of cases where they are not relevant, but they're simply used to distract an audience by bringing up colorful and interesting allegations of character, or bad ethical practices, that can quickly and effectively prejudice an audience against an arguer.

In some cases where a pragmatic inconsistency is alleged in an *ad hominem* argument, it is possible to resolve the conflict by a further dialogue. In some cases, for example, the person accused can explain the contradiction by showing it does not really show that he or she is being hypocritical or insincere, as in the smoking example. In other cases however, there is a sharp contradiction between what an arguer claims and what she strongly professes to be her own basic principles, in previous actions, or in a previous argument that she has put forward. In the sealers case the chef has argued against inhumane treatment of animals, but his actions suggest otherwise. In the sealers case, such a resolution of the conflict does not seem to be possible, because the circumstances are strongly opposed, and neither side is willing to make concessions. However, one

possible way of continuing the dialogue would be for C to claim that he uses only "decent" foie gras, that does not conform to the French legal definition of foie gras, and does not have the same quality.

It has been shown that the evaluation of the *ad hominem* in the sealers example depends most centrally on the factors specified in the antecedent, assumption and exceptions of the scheme for the circumstantial *ad hominem* presented above. The antecedent is the statement that C's personal circumstances are inconsistent with the argument that C put forward. Whether or not this statement is justified by the evidence in the example is something that cannot be proved or disproved using the scheme only. As we saw, this example is comparable to the hunters example, where the claim of inconsistency depends on how the involvement of C is described. Still, the antecedent cites the key statement assumed to be true in the *ad hominem* argument put forward by S. The first assumption is that the pair of commitments in the example can be shown to be practically inconsistent. The second assumption is the inconsistency that is taken to show that C is a person of bad character, in this instance, a hypocrite. It shows that the parties putting forward the argument against S are not sincere in following their own policies they advocate for others. The third assumption is that character is at issue. This presumption is the statement that character is a relevant issue in the case, and it is around this assumption that much of the discussion of the evaluation of the argument outlined above has to be based on.

The evaluation of the sealers argumentation is shown in figure 10, where the premises and conclusion of the argument are represented as nodes containing statements, and an argument is a node represented as a circle containing the identifier of the argument and its argumentation scheme.

Figure 10: Evaluation of the Sealers Argument in Carneades

29
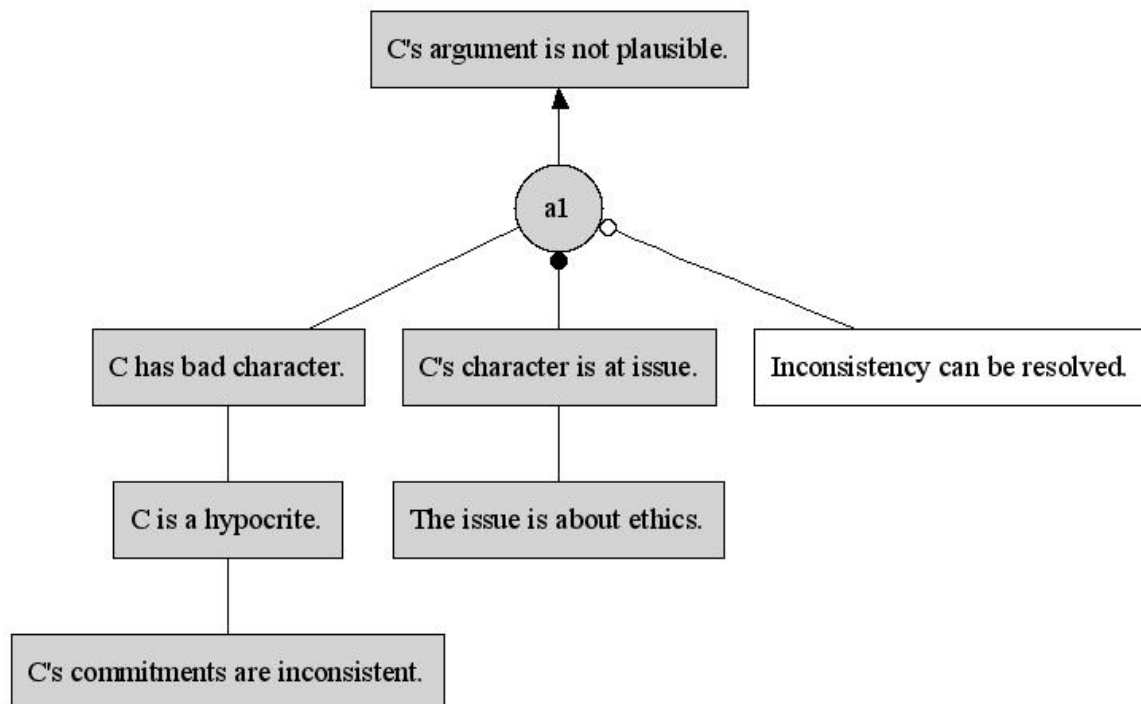
In figure 10, argument a1 is the sealers argument and its argumentation scheme is that for the circumstantial *ad hominem*. Directly under a1, three premises are represented. The first one, the statement that C has bad character, is an ordinary premise. The second one, the statement that C's character is at issue, is an assumption. The third one, the statement that the inconsistency can be resolved, is an exception. Recall that exceptions have to be proved in order to defeat an argument. In other words, the respondent has to show that the inconsistency can be resolved by further dialogue, or otherwise the proponent's argument stands. On the other hand, assumptions function like ordinary premises. They have to be proved, or supported by sufficient evidence, in order for the argument to hold. Such supporting evidence is presented in figure 10 under the first two premises. The ordinary premise that C has bad character is supported by the statement that C is a hypocrite, which is in turn supported by the evidence in the case that sees commitments are inconsistent. The assumption that C's character is at issue is supported by the statement that the issue is about ethics. Normally in Carneades, what is at issue is defined at a different level (an earlier stage of the dialogue), but to simplify the representation of this case, we will insert the evidence in the case that the issue is about ethics as a statement.

Note that all the nodes in figure 10 are darkened except for the one on the right containing the statement that the inconsistency can be resolved by further dialogue. This premise of a1 is an exception. In the sealers case it was not proved by any further dialogue that ensued in the case. But as we look over the details of the sealers case and the analyses of the above, it is clearly justified to claim that the issue is about ethics, and that C's commitments have been shown to be inconsistent. To evaluate the argument, we have to start from the bottom left. That C's commitments were shown to be inconsistent shows that C is a hypocrite, which in turn shows that C has bad character. This premise, along with the premise that C's character is at issue, and the failure of the exception (that the inconsistency can be resolved by further dialogue), shows that argument a1 is acceptable. Since argument a1is a circumstantial *ad hominem* argument against C's prior argument, the argumentation as a whole, as displayed in figure ten, shows that C's argument is not plausible. That statement is displayed at the top of figure 10 representing the conclusion of argument a1.

The Carneades diagram in figure 10 thus represents the analysis that the premises of the circumstantial *ad hominem* argument against C are justified, as supported by the evidence in the case. It is assumed that the practical inconsistency has been identified, that it shows that C is a hypocrite (at least defeasibly), and that the issue is about ethics. All of these assumptions are justified by the data presented in the original sealers case, or so it has been argued by the analysis of that case presented above in arguments alpha and beta. There is no evidence that C has resolved or explained the practical inconsistency.[8] The issue is about ethics, and therefore sees character is an issue. It follows that C's character does matter, and that the attempted counterargument that a person's character does not matter is defeated. To sum up, what has been shown is that the circumstantial argument against C shifts the burden of proof back on to C to show that has original

---

[8] This evaluation of the dialogue shown in the text of the example does not imply that the dialogue could not be continued by expanding the example. For example, it might be found that C believes that foie gras is not really very bad, for some reason, or that the badness is outweighed by its gastronomical excellence. This extension of the dialogue could bring in value-based practical reasoning, re-opening the dialogue.

argument is plausible, by asking critical questions that are appropriate for the argumentation scheme for the circumstantial *ad hominem* argument. Thus it can be said that the circumstantial *ad hominem* argument in the sealers example is a reasonable one.

9. Conclusions

As also shown by the analysis and evaluation in section 8, the two exceptions in the scheme are the key issues in the evaluation of the *ad hominem* in the case. These are the two exceptions to the effect that (1) the practical inconsistency can be resolved by further dialogue, and (2) the argument for C is being absolutely rejected by S. These statements are assumed to be false, but if evidence in the example shows that either is true, the generalization in the scheme would default, meaning that the circumstantial *ad hominem* argument against C would be evaluated as not acceptable. As shown in the Carneades evaluation of the argument in figure 10, the three worrisome premises are the statement that C has bad character, the statement that C's character is at issue, and the statement that the inconsistency can be resolved. As shown in figure 10, the first two can be proved by supportive evidence, while the third is an exception that does not need to be proved, in order for the argument to be accepted. This analysis shows that the circumstantial *ad hominem* argument against C is justified. It is concluded that the text of discourse in the sealers case provides sufficient evidence to show that the circumstantial *ad hominem* argument in the case is a reasonable one, when all the evidence is fairly evaluated.

In this paper we began by showing that there are many different kinds of arguments used in the sealers case, and that as with any attempt to interpret a natural language text of discourse, the argumentation may need to be diagrammed in several different ways. We showed that it needs to be recognized that several argumentation schemes are centrally or tangentially involved. The process of carrying out an analysis of an argument must begin by making several kinds of careful judgments about how to interpret the argumentation in the given case (van Eemeren and Grootendorst, 2004). New material will have to be added, by assumption, because many of the arguments will depend on implicit premises. Some of the given material will have to be deleted as unimportant. But such a judgment of importance depends on the purpose of the analysis, and what is hoped to be achieved. In this case the main purpose of our analysis was to illustrate an interesting example of the circumstantial *ad hominem* argument, and use that to pose various theoretical problems about how to formalize the argumentation scheme for this type of argument. As shown by comparing the sealers example to other previously studied examples, namely the smoking example and the hunter example, it was shown how subtleties are involved in attempting to analyze and evaluate this kind of argumentation. It is not for nothing that this type of argument has traditionally been classified as an informal fallacy (Walton, 1998). We have shown that it well deserves its reputation as a tricky and deceptive tactic with many subtle features that need to be carefully analyzed.

It is concluded that the foregoing effort to identify and analyze the argument in the sealers case has been successful. This task was carried out by identifying the argumentation scheme for the circumstantial *ad hominem* type of argument, and showing how that scheme is applicable to the data given in the text of discourse for the sealers example. It was shown how the scheme applied to the case, first of all by identifying

argument alpha and argument beta in section 3, and then by showing how the one party used the circumstantial *ad hominem* argument identified in the sealers example to attack argument beta. It was then shown how this particular instance of the circumstantial *ad hominem* argument can be represented by identifying its premises and conclusion in the argument diagramming system *Araucaria*. It was shown how each of the premises in the scheme was there to be found in the argument against argument alpha. Each premise is essential to the scheme and to identifying its existence in a given case. The argument premise is essential because it is sometimes overlooked that an *ad hominem* argument is not merely a character attack or a personal smear against a person, but is the use of this specific kind of attack to attack another party's argument. The inconsistent commitment premise is essential because the defining characteristic of the circumstantial type of *ad hominem argument*, as opposed to the direct type or other types, is the presence of an alleged inconsistency. Essentially, the critic argues that the original arguer does not practice what he preaches, implying that his personal actions and circumstances express commitments that are inconsistent with his argument. The credibility questioning premise is essential in order for the argument to be a genuine *ad hominem* argument, as noted above. We showed how all the premises and the conclusion can be identified in the sealers case. We made no comment on how well each of them is justified by the evidence given in the case, but noted that C did not dispute them

Next we looked at how this scheme could be formalized, first examining the system of Verheij. Then we examined some problems of modeling generalizations on which defeasible reasoning is based. We saw that a revised argumentation scheme for the circumstantial *ad hominem* argument can be developed by showing, following Verheij (2003), that this form of argument has a DMP structure. On Verheij's formal analysis, an *ad hominem* scheme is based on a rule or generalization that is subject to exceptions. We then confronted the problem of modeling the critical questions, we moved on to the computational system for reasoning with argumentation schemes called Carneades. We showed how Carneades could also be used to solve the evaluation problem for the circumstantial *ad hominem* argument by modeling the critical questions as components of the scheme itself. Finally, we showed how this scheme can be used to structure the evaluation of the argument in the sealers example by specifying all the elements that need to be proved to make the argument acceptable in the case.

References

Trevor Bench-Capon, 'Persuasion in Practical Argument Using Value-Based Argumentation Frameworks', *Journal of Logic and Computation*, 13, 2003, 429-448.

Floris Bex, Henry Prakken, Chris Reed and Douglas Walton, 'Towards a Formal Account of Reasoning about Evidence: Argumentation Schemes and Generalizations', *Artificial Intelligence and Law*, 12, 2003, 125-165.

Augustus DeMorgan, *Formal Logic*, London, Taylor and Walton, 1847.

Richard Foot, 'Celebrity Chef Brings Sealers to the Boil', *National Post*, October 14, 2005, p. A10.

Thomas F. Gordon, *The Pleadings Game: An Artificial Intelligence Model of Procedural Justice*, Dordrecht, Kluwer, 1995.

Thomas F. Gordon, 'A Computational Model of Argument for Legal Reasoning Support Systems', *Argumentation in Artificial Intelligence and Law*, ed. Paul E. Dunne and Trevor Bench-Capon, IAAIL Workshop Series, Nijmegen, Wolf Legal Publishers, 2005, 53-64.

Thomas F. Gordon and Douglas Walton, 'The Carneades Argumentation Framework: Using Presumptions and Exceptions to Model Critical Questions', *Computational Models of Argument: Proceedings of COMMA 2006*, ed. P. E. Dunne and T. J. M. Bench-Capon, Amsterdam, IOS Press, 2006, 195-207.

Wayne Grennan, *Informal Logic*, Kingston, McGill-Queen's University Press, 1997.

Arthur C. Hastings, *A Reformulation of the Modes of Reasoning in Argumentation*, Evanston, Illinois, Ph.D. Dissertation, 1963.

Manfred Kienpointner, *Alltagslogik : Struktur und Funktion von Argumentationsmustern*, Stuttgart, Fromman-Holzboog, 1992.

Chaim Perelman and Lucie Olbrechts-Tyteca, *The New Rhetoric*, Notre Dame, University of Notre Dame Press, 1969.

John Pollock, *Cognitive Carpentry*, Cambridge, Mass., MIT Press, 1995.

Henry Prakken, 'AI & Law, and Argument Schemes', *Argumentation*, 19, 2005, 303-320. Available at: http://www.cs.uu.nl/people/henry/

Henry Prakken and Giovanni Sartor, 'A Dialectical Model of Assessing Conflicting Arguments in Legal Reasoning', *Artificial Intelligence and Law*, 4, 1996, 331-368.

Henry Prakken and Giovanni Sartor (1997), 'Argument-based Extended Logic Programming with Defeasible Priorities', *Journal of Applied Non-classical Logics*, 7, 1997, 25-75.

Henry Prakken and Giovanni Sartor, 'The Three Faces of Defeasibility in the Law', *Ratio Juris,* 16, 2003, 495-516

Chris Reed and Timothy J. Norman, *Argumentation Machines: New Frontiers in Argument and Computation*, Dordrecht, Kluwer, 2003.

Chris Reed and Glenn Rowe, *Araucaria, Version 3,* 2005. Available free at http://www.computing.dundee.ac.uk/staff/creed/araucaria/.

Chris Reed and Douglas Walton, 'Towards a Formal and Implemented Model of Argumentation Schemes in Agent Communication', *Autonomous Agents and Multi-Agent Systems*, 11, 2005, 173-188.

Ray Reiter, 'Nonmonotonic Reasoning', *Annual Review of Computer Science*, 2, 1987, 147-186.

Nicholas Rescher, *Plausible Reasoning*, Assen, Van Gorcum, 1976.

Frans H. van Eemeren and Rob Grootendorst, *A Systematic Theory of Argumentation*, Cambridge, Cambridge University Press, 2004.

Bart Verheij, (1999), 'Logic, Context and Valid Inference', *Legal Knowledge Based Systems. JURIX 1999: The Twelfth Conference*, eds. H.J. van den Herik, M.-F. Moens, J. Bing, B. van Buggenhout, J. Zeleznikow, and C.A.F.M. Grütters, Nijmegen, Gerard Noodt Instituut, 1999, 109-121.

Bart Verheij, 'Legal Decision Making as Dialectical Theory Construction with Argumentation Schemes', *The 8th International Conference on Artificial Intelligence and Law: Proceedings of the Conference*, New York Association for Computing Machinery, 2001, 225-236. Available at http://www.ai.rug.nl/~verheij/publications.htm.

Bart Verheij, 'Dialectical Argumentation with Argumentation Schemes: An Approach to Legal Logic', *Artificial Intelligence and Law*, 11, 2003, 167-195.

Bart Verheij, 'DefLog: on the Logical Interpretation of Prima Facie Justified Assumptions', *Journal of Logic and Computation*, 13, 2003a, 319-346. Available at http://www.ai.rug.nl/~verheij/publications.htm.

Bart Verheij, *Virtual Arguments*: *On the Design of Argument Assistants for Lawyers and Other Arguers*, The Hague, Asser Press, 2005.

Douglas Walton, *A Pragmatic Theory of Fallacy*, Tuscaloosa, University of Alabama Press, 1995.

Douglas Walton, *Ad Hominem Arguments*, Tuscaloosa, University of Alabama Press, 1998.

Douglas Walton, 'Are Some *Modus Ponens* Arguments Deductively Invalid?', *Informal Logic,* 22, 2002, 19-46.

Douglas Walton, 'Justification of Argumentation Schemes', *The Australasian Journal of Logic*, 3, 2005, 1-13.

Douglas Walton, *Fundamentals of Critical Argumentation*, Cambridge, Cambridge University Press, 2006.

Douglas Walton and Thomas F. Gordon, 'Critical Questions in Computational Models of Legal Argument', *Argumentation in Artificial Intelligence and Law*, IAAIL Workshop Series, ed. Paul E. Dunne and Trevor Bench-Capon, Nijmegen, Wolf Legal Publishers, 2005, 103-111.