

DEFEASIBLE REASONING AND INFORMAL FALLACIES

The purpose of this paper is to advance fallacy theory beyond its current state of development by linking it to the notion of defeasible reasoning. Defeasible reasoning has turned out to be very important for computing, especially in view of the attention paid to modeling argumentation (Bench-Capon and Dunne, 2007) and the use of argumentation schemes (Verheij, 2003) and dialogs (Prakken, 2000, 2006) to study problems of nonmonotonic reasoning. The advent of argumentation frameworks (Dung, 1995) can be shown to provide an elegant way of subsuming much previous work on defeasible reasoning (Bondarenko, Dung, Kowlaski and Toni, 1997). Although much has been written on individual fallacies, there is comparatively little on the general theory of fallacy, except for the pragma-dialectical theory (van Eemeren and Grootendorst, 1992) and the pragmatic theory (Walton, 1995). Defeasible reasoning is uncontestedly important for helping us to better grasp the notion of fallacy and rethink it as a concept useful for modern logic, but so far the link between the two notions has not been studied.

Many of the most common forms of argument associated with major fallacies, like argument from expert opinion, *ad hominem* argument, argument from analogy and argument from correlation to cause, have now been analyzed using the device of argumentation schemes (Walton, Reed and Macagno, 2008). Recent research in computing has embraced the use of argumentation schemes, and linked them with key logical notions like burden of proof that are also related to the study of fallacies (Gordon, Prakken and Walton, 2007). Argumentation schemes have been put forward as a helpful way of characterizing structures of human reasoning, like argument from expert opinion, that have proved troublesome to view deductively. Attempting to deductivize the reasonable examples, by viewing the major premise as a conditional not subject to exceptions (e.g. if X says Y then Y is true) does not work at all well, as this type of argument is typically defeasible. The traditional logic textbooks treated these forms of argumentation only under the heading of informal fallacies, but more and more research in argumentation/informal logic is supporting the view that they are not always fallacious, that the nonfallacious instances are common, and that they are very useful arguments in many cases (Hansen and Pinto, 1995). In this paper evidence is given to suggest that argumentation schemes and formal dialog models are the tools needed for studying the properties of defeasible reasoning and informal fallacies.

Section 1 is a brief survey that outlines the main approaches taken by the two leading theories of fallacy and discusses how dialog rules can be applied to the study of informal fallacies. In section 2, the reader is introduced to argumentation schemes, and it is shown that twelve of the traditional major informal fallacies correspond to argumentation schemes, while seven of them do not. Section 3 gives a quick summary of the main features of defeasible reasoning that are especially important for the study of fallacies. Section 4 introduces the reader to formal dialog systems, citing the main properties of such systems that are important for analyzing defeasible reasoning and informal fallacies. In section 5, the profile of dialog tool is applied to the fallacy of many questions to show how dialog models can be applied in an easy way to real instances of argumentation treated under the category of a traditional informal fallacy. Section 6 formulates dialog conditions that connect defeasibility to fallacies and that reveal underlying features that can be used to analyze sequences of reasoning associated with fallacies. Section 7 takes

as a case in point the fallacy most closely related to defeasible reasoning, *secundum quid* (the fallacy of neglect of qualifications). Section 8 features another case in point, the defeasible lack of knowledge type of inference associated with the traditional fallacy of *argumentum ad ignorantiam*. Section 9 draws conclusions from what has been shown.

1. Theories of Fallacy

The most fully developed theories of fallacy so far are the pragmatic theory and the pragma-dialectical theory of the Amsterdam School (Tindale, 1997). The two theories have much in common, but define the notion of fallacy in different ways. According to the pragmatic theory (Walton 1995, 237-238), there are six basic characteristics of fallacy. These characteristics state conditions on what reasonably should be taken to be a fallacy according to the criteria discussed in (Walton 1995).

1. A fallacy is a failure, lapse, or error, subject to criticism, correction, or rebuttal.
2. A fallacy is a failure that occurs in what is supposed to be an argument.
3. A fallacy is associated with a deception or illusion.
4. A fallacy is a violation of one or more of the maxims of reasonable dialog or a departure from acceptable procedures in that type of dialog.
5. A fallacy is an instance of an underlying, systematic kind of wrongly applied technique of reasonable argumentation.
6. A fallacy is a serious violation, as opposed to an incidental blunder, error, or weakness of execution.

These six basic characteristics set in place requirements that any theory should meet, or at least should be prepared to cope with, in order to be a satisfactory theory of fallacy.

The fallacy of many questions arguably fails to meet condition 2, because the speech act of asking a question is different from the speech act of putting forward an argument. For example, a complex and loaded question like, ‘Are you confused when you are not on your medications?’ has traditionally been taken to fall under the category of the fallacy of many questions even though, at least apparently, it does not have the form of an argument. What needs to be said here are two things. First, this question implicitly contains an argument, and indeed that is partly what is fallacious about it. The argument is concealed. Second, it uses an argumentation strategy of attempting to take three turns at once in a dialog, instead of following the proper rule of only taking one turn at a time.¹

The leading competitor to the pragmatic theory of fallacy (Walton, 1995) has so far been the evolving doctrines of the Amsterdam School. According to their earliest version of their theory, a fallacy is a violation of a rule of a critical discussion (van Eemeren and Grootendorst, 1992). This theory was a good advance at the time because it went beyond the older treatments of the logicians who so often tended to analyze a fallacy by only seeing it as a failed inference of some sort. It was a breath of fresh air. However, it was argued in (Walton, 1995) that this theory is still too narrow for several reasons, and these reasons will be illustrated and amplified below using a few examples. First, as shown in section 2 below, many if not most of the traditional fallacies are associated with argumentation schemes that represent forms of inference of various kinds. Both the inferential aspect and the dialectical aspect of the notion of a fallacy need to be taken into account. But the Amsterdam theory is also too narrow for other reasons cited in (Walton,

¹ This issue will be discussed further in section 5.

1995, 298). Violating a rule of a critical discussion should not be itself equated with the committing of a fallacy, for some such violations are merely blunders and not fallacies. Another problem with their theory is that many different fallacies can be analyzed as different ways of violating the same rule. Still another problem, as will be illustrated below, is that in order to properly analyze many of the fallacies one has to examine the argumentation strategy the fallacy was based on. The theory has been more recently been strengthened by the work of van Eemeren and Houtlosser (2006) on strategic maneuvering.

The following short definition can be given, based on the pragmatic (1995) theory of fallacy: a fallacy is an argument, a pattern of argumentation, or something that purports to be an argument, that falls short of some standard of correctness as used in a conversational context but that, for various reasons, has a semblance of correctness about it in context, and poses a serious obstacle to the realization of the goal of the dialog. A pattern of argumentation is an ordered sequence of moves by two parties in a dialog. This theory is inherently dialectical in that not only is the structure of the inference from the set of premises to the conclusion taken into account in evaluating argumentation in a particular case, but whether or not an argument is fallacious in that case also needs to depend on the context of how it was used in a sequence of moves in a dialog.

Many of the fallacies are fallacious moves in a dialog not because of the inherent unreasonableness of the argument, but of the way it is used in a sequence of moves to try to prevent the respondent from questioning it, or even continuing the dialog at all. The classic case of this type of fallacy is the fallacy of poisoning the well. For example in the Cardinal Newman case, the attack alleged that as a Catholic, Newman could have no regard for the truth of the matter in any political discussion (Walton, 2006). Or to cite another excellent example, an appeal to expert opinion might cite a scientific authority, and dismiss any reasonable attempt to ask critical questions about the argument from expert opinion by declaring that any evidence appearing to go against it must be dismissed as anecdotal. The fallacy in both instances is not found in the argumentation scheme, as applied to a single argument, but in a pattern that can only be found by examining a connected sequence of moves by both parties.

On this dynamic approach, a distinction has to be drawn between two kinds of fallacies. In some cases, a fallacy is merely a blunder or an error, while in other cases, it is a sophistical tactic used to try to get the best of a speech partner in dialog unfairly, typically by using verbal deception or trickery. The evidence of the use of such a tactic is found in the pattern of moves made by both sides in the dialog. It is important for fallacy theory to avoid being impaled on the horns of a dilemma between these two traditional types of problematic argumentation moves. To confront the dilemma, the pragmatic theory (1995) distinguished between two kinds of fallacies. The paralogism is the type of fallacy in which an error of reasoning is committed typically by making a blunder by failing to meet some necessary requirement of an argumentation scheme. The sophism is a sophistical tactic used to try to unfairly try to get the best of a speech partner.

Some would say that a fallacy is an intentional deception, distinguishing between the two kinds of fallacy by saying that one is an intentional deception while the other is a mere mistake that is not intentional. This way of drawing the line misses the point however. Many fallacies are committed because the proponent has such strong interests at stake in putting forward a particular argument, or is so fanatically committed to the

position advocated by the argument, that she is blind to weaknesses in it that would be apparent to others not so committed. In this kind of case the deception may not quite be intentional, because the proponent doesn't see the argument as faulty, but it is close to being intentional because she is so committed to this kind of argumentation that she pushes ahead with it blind to errors that others might find in it. In this kind of case one could say that the committing of the fallacy is on the borderline between being intentional and unintentional. But whether it is intentional or unintentional doesn't really matter from a point of view of analyzing the argument and deciding whether it should be considered fallacious. What is important is the logical weakness in the argument, or some fault in the pattern of argumentation.

The sophisticated tactics type of fallacy tends to be a more serious kind of problem than the error of reasoning one. It is based on the idea that an organized rule-governed dialog in which arguments are exchanged, like a critical discussion, is partly adversarial but also partly cooperative. Collaborative procedural rules are very important in such a dialog, but there is also an adversarial element. A participant in a critical discussion is an advocate of his own viewpoint and a critic of the opposing viewpoint. Thus a critical discussion is like a free market economy in which each side tries to win by having the strongest argument that will triumph over those of its opponents. The problem is that bad things can happen, including the committing of fallacies by shifting to a quarrel that is purely adversarial. Any conversation in which reasonable argumentation is to be used for some constructive purpose must strike the right balance between this adversarial aspect and the need for all parties to follow the Gricean maxims of polite conversation that are required to make the contributions of each participant useful to move the dialog toward its goal.

Another problem is that to analyze fallacies properly, we have to explain how each of them is used as an effective deceptive tactic that does work to fool people. The theory of strategic maneuvering now claims to accomplish this task by taking the rhetorical dimension of fallacies into account. These considerations introduced the factor that van Eemeren and Houtlosser (2006) call strategic maneuvering. Strategic maneuvering refers to the efforts of arguers in a discussion to reconcile their twin aims of rhetorical effectiveness and maintaining dialectical standards of reasonableness. These twin aims arise from the fact that the critical discussion has both an adversarial and a collaborative component. Participants have to collaborate by following the rules for the critical discussion, for example by taking turns putting forward arguments. At the same time however, each side is trying to convince the other side to accept its point of view by using the strongest arguments possible to support its thesis and to refute the thesis of the opposed side. This dual aspect of the critical discussion means there is always a tension between rhetorical effectiveness and using arguments that fit dialectical standards of reasonableness. The critical discussion is like a free market economy in which people have to follow laws and social rules, but within the framework of these rules they are free to try to maximize their own profits. In such a competitive situation, an arguer is free to use strategic maneuvering to try and put forward a winning argument. It is up to the other side to ask critical questions, and generally to probe into an argument to try to find the weak points in it, and if possible to refute it by counter-arguments.

Even over and above strategic maneuvering, however, participants in a critical discussion tend to have interests at stake. In a philosophy discussion about some theoretical subject in a classroom, participants who are students might want to show that

they are better than the other students by using strong but fallacious argument to try to refute the arguments of their opponents. Moreover, they might want to impress their professor in the hope of showing how clever they are, thinking that this might help them get a better grade in the course. What these observations show is that interests are involved outside the framework of the critical discussion itself. These interests relate to the goals of the participants outside the critical discussion in relation to how they might use their arguments in the critical discussion for some external purpose. What this may suggest is that the explanation for committing fallacies may be found in the observation that the participants involved in a critical discussion are also involved in other goal-directed types of activities at the same time.

2. Argumentation Schemes and Fallacies

Many of the most common forms of argument associated with major fallacies, like argument from expert opinion, *ad hominem* argument, argument from analogy and argument from correlation to cause, have now been analyzed using the device of argumentation schemes (Walton, Reed and Macagno, 2008). The traditional logic textbooks mainly treated these forms of argumentation under the heading of informal fallacies, but more and more research in argumentation/informal logic is supporting the theory that they are not always fallacious (Hansen and Pinto, 1995), and the view that in many instances, they are reasonable but defeasible arguments. The formal and inductive fallacies, like affirming the consequent and arguing from too small a sample to a generalization, can be analyzed with deductive and inductive forms of reasoning familiar to formal logicians. However, of the major informal fallacies, the following twelve need to be analyzed with defeasible argumentation schemes of the sort that can be found in (Walton, Reed and Macagno, 2008, chapter 9).

1. *Ad Misericordiam* (Scheme for Argument from Distress, 334)
2. *Ad Populum* (Scheme for Argument from Popular Opinion and its subtypes, 311)
3. *Ad Hominem* (*Ad Hominem* Schemes; direct, circumstantial, bias, 336-338)
4. *Ad Baculum* (Scheme for Argument from Threat, 333; Fear Appeal, 333)
5. Straw Man (Scheme for Argument from Commitment, 335)
6. Slippery Slope (Slippery Slope Schemes; four types, 339-41)
7. *Ad Consequentiam* (Scheme for Argument from Consequences, 332)
8. *Ad Ignorantiam* (Scheme for Argument from Ignorance, 327)
9. *Ad Verecundiam* (Scheme for Argument from Expert Opinion, 310)
10. *Post Hoc* (Scheme for Argument from Correlation to Cause, 328)
11. Composition and Division (Argument from Composition, 316; Division, 317)
12. False Analogy (Scheme for Argument from Analogy, 315)

Of course, these are not the only fallacies that can be analyzed with the help of argumentation schemes. We include here for discussion only the major fallacies that are most commonly treated in the most widely used logic textbooks.

An example is the scheme for argument from expert opinion with its matching set of critical questions (Walton, Reed and Macagno, 2008, 310).

MAJOR PREMISE: Source *E* is an expert in subject domain *S* containing proposition *A*.

MINOR PREMISE: *E* asserts that proposition *A* is true (false).

CONCLUSION: *A* is true (false).

This scheme represents a defeasible form of argument that is not well modeled as deductive or inductive. Instead, it is viewed as an argument that can hold tentatively under conditions of lack of knowledge of the full facts of a case, but that can be defeated or cast into doubt by the asking of appropriate critical questions. The six basic critical questions useful for guidance on how to evaluate the argument are these (310).

CQ₁: *Expertise Question*. How credible is *E* as an expert source?

CQ₂: *Field Question*. Is *E* an expert in the field that *A* is in?

CQ₃: *Opinion Question*. What did *E* assert that implies *A*?

CQ₄: *Trustworthiness Question*. Is *E* personally reliable as a source?

CQ₅: *Consistency Question*. Is *A* consistent with what other experts assert?

CQ₆: *Backup Evidence Question*. Is *E*'s assertion based on evidence?

On this view, argument from expert opinion can be a reasonable argument in some instances of its use, provided it is realized that it is a defeasible heuristic that can be given up if more facts come to be known about the case so that we no longer need to rely on expert opinion. In other instances of its use however, it can be fallacious. For example, if the proponent of the argument treats it as infallible, and refuses to concede that it is even open to critical questioning, that would be a fallacious misuse of the argument.

Arguments like appeal to expert opinion are tricky because they are often very useful, and in many instances they are the best resources we have for navigating through a world where time and resources are limited, but where a decision to accept a hypothesis or course of action needs to be taken. Argument from expert opinion needs to be seen as open to critical questioning in a dialog. You can treat it as a deductively valid argument or as an inductively strong argument in some instances, but in the vast majority of cases, this way of treating it would lead to serious problems (even fallacies). Generally, it is not a good policy when examining expert witness testimony to assume that the expert is omniscient. Indeed, such a policy would be highly counterproductive when evaluating this kind of evidence in a common law trial. However, there is a natural tendency in everyday reasoning to respect expert opinions and even to defer to them, or hesitate to question them. Questioning the opinion of an expert can seem impolite, unless done in a circumspect way. However, as a matter of fact, experts are often wrong, or what they say can be highly misleading, leading to the drawing of a conclusion that is wrong or not supported by the evidence. As a practical matter, one often needs to be prepared to critically question the opinion of an expert by asking the right questions. Such an argument needs to be seen as defeasible, or subject to default.

Tindale (2007) agrees that many fallacies are misuses of argumentation schemes that are legitimate but defeasible argument strategies citing *ad hominem* and *ad verecundiam* as leading cases in point. However, he cites the straw man as an example of a fallacy that does not fit into this category. He writes (2007, 12) that there is no way we can make it fit “unless we conjure up something trivial such as Real Man.” But as Krabbe, 2007, 129) noted, “Real Man is not so trivial, considering how hard it is to correctly explicitize the implicit elements in one’s opponent’s position [their caps in quotes]”. Straw man, Krabbe adds, can be seen as a derailment of such explicitization strategies.

As well, the straw man fallacy can be seen as a derailment of argument from commitment. The scheme for argument from commitment (version 1) from (Walton, Reed and Macagno, 2008, 335) is given below.

COMMITMENT EVIDENCE PREMISE: In this case it was shown that *a* is committed to proposition *A*, according to the evidence of what he said or did.

LINKAGE OF COMMITMENTS PREMISE: Generally when an arguer is committed to *A*, it can be inferred that he is also committed to *B*.

CONCLUSION: In this case, *a* is committed to *B*.

The following are the two critical questions matching argument from commitment (Walton, Reed and Macagno, 2008, 335).

CQ₁: What evidence in the case supports the claim that *a* is committed to *A*, and does it include contrary evidence, indicating that *a* might not be committed to *A*?

CQ₂: Is there room for questioning whether there is an exception, in this case, to the general rule that commitment to *A* implies commitment to *B*?

The fallacy of straw man occurs where the first party in a dialog has distorted what is taken to the second party's implied commitment to *B*, making the attributed commitment appear more implausible or more extreme, and thereby more open to attack. In any given case however, this fallacy cannot be (completely) pinned down by using the argumentation scheme for argument from commitment. Matters of the context of dialog need to be taken into account. We need to examine carefully what each party said in relation to what the other party claimed he said, and how she used that in the argument.

Note however that version 2 of the scheme for argument from commitment given in (Walton, Reed and Macagno, 2008, 335) does take dialog factors into account.

MAJOR PREMISE: If arguer *a* has committed herself to proposition *A*, at some point in a dialog, then it may be inferred that she is also committed to proposition *B*, should the question of whether *B* is true become an issue later in the dialog.

MINOR PREMISE: Arguer *a* has committed herself to proposition *A* at some point in a dialog.

CONCLUSION: At some later point in the dialog, where the issue of *B* arises, arguer *a* may be said to be committed to proposition *B*.

This version of the scheme, however, is not sustainable, for the following reasons. If *A* entails *B* and a participant is committed to *A*, then it may be inferred that he is also committed to *B*, whether or not *B* is an issue. In other words, whether or not *B* has become an issue at some later point in the dialog should be irrelevant for the purpose of an argumentation scheme. Evidence for this is found in the need for consistency with the formulation of other argumentation schemes. No other argumentation scheme restricts the conclusions that may be drawn specifically to conclusions that are relevant for issues that have been raised in the dialog. Such a relevancy condition is not suitable to be placed within an argumentation scheme. Its proper setting is that of dialog rules.² The lessons of these points are significant, for it has been shown that while the scheme for argument from commitment is necessary for the determination of instances of the straw man fallacy, it is not sufficient. Dialog factors also need to be taken into account.

Next we need to see that there are at least seven major fallacies remaining that do not fit any of the argumentation schemes.

1. Equivocation
2. Amphiboly
3. Accent
4. *Petitio Principii* or Begging the Question

² I would like to thank an anonymous referee for bringing this important point to my attention.

5. *Ignoratio Elenchi* (Irrelevance: species are red herring and wrong conclusion)

6. *Secundum Quid* (Neglecting Qualifications)

7. Many Questions

The three linguistic fallacies, equivocation, amphiboly and accent, are not based on specific argument types. They have to do with ambiguous communications (speech acts) in dialogs. Begging the question occurs in circular chains of reasoning where the links in the chain can consist of many different kinds of arguments. Irrelevance is failure to prove a specified conclusion that is supposed to be proved. Many kinds of arguments can commit this error. The error is not specific to type of argument used. Irrelevance and begging the question are fallacies that have to do with sequences of extended argumentation where the links of inference making up the reasoning can be of different kinds, including any types represented by the various argumentation schemes. The fallacy of *secundum quid* is not specific to a type of argument, unless it is perhaps arguing from a generalization to an instance or arguing from an instance to a generalization. But such inferences can occur in any kind of defeasible argumentation. The fallacy of neglecting qualifications occurs in any such instances where proper qualifications are ignored or suppressed, as will be shown in section 5. However, it should be noted that there are a number of argumentation schemes in which this fallacy plays an important role. As noted above, the fallacy of many questions is not identified with a specific type of argument, but rather a strategy of questioning in a dialog format. Hence none of these seven fallacies is a misuse of any particular argumentation scheme.

3. Defeasible Reasoning

The etymology of the term ‘defeasible’ comes from medieval English contract law, referring to a contract that has a clause in it that could defeat the contract in a case where the circumstances fit the clause. However, the origin of the term in modern philosophy and law is a paper called ‘The Ascription of Responsibility and Rights’ by H. L. A. Hart (1949; 1951). Hart’s work was attacked in subsequent years by philosophers who criticized it heavily (Loui, 1995, 21), even though a few, like Toulmin, accepted and used it. But his view turned out to be prescient, in light of the importance defeasible reasoning turned out to have in computing. As noted in the introductory section above, many formal systems of defeasible reasoning were produced in the fields of artificial intelligence and logic. There are several different approaches. Some place defeasible reasoning in a context of new information coming in that annuls a previous conclusion drawn by inference. Others see defeasibility as operating in a framework of belief revision, as an agent updates his/her beliefs. Hart saw it as circumstances fitting an exception to a general rule.

The originating idea behind Hart’s way of defining the term can be appreciated from the following quotation from the 1951 version of his paper, taken from (Loui, 1995, 22).

Claims can usually be challenged or opposed in two ways. First, by a denial of the facts upon which they are based and secondly by something quite different, namely a plea that although all the circumstances on which a claim could succeed are present, yet in the particular case, the claim . . . should not succeed because other circumstances are present which brings the case under some recognized head of exception, the effect of which is either to defeat the claim . . . altogether, or to “reduce” it . . . (1951, 147-148).

Judging from this quotation, it would appear that Hart had the idea of a claim being at first acceptable because it is supported by reasoning, but that is later defeated because circumstances are present that bring the case under an exception. Thus we recognize the idea of a defeasible argument, of a kind so common in law.

However, that is not the only way Hart saw defeasibility. He also discussed defeasible concepts. His most famous example is from *The Concept of Law* (1961). Consider the rule that no vehicles are allowed in the park. This rule could be defeated by special circumstances, for example during a parade, but it could also be defeated because of the open texture of the concept of a vehicle. For example, a car would definitely be classified as a vehicle, and be excluded from the park, but what about a bicycle? Is it a vehicle? Both sides could be argued, unless the law makes a specific ruling on bicycles. The literature on computing has concentrated on defeasibility of arguments rather than on defeasibility of concepts³, and these two notions seem to be quite different. However, the precise distinction between these two kinds of defeasibility needs to be clarified. The notion of a defeasible concept presupposes that the concept already has some definition, but the existing definition turns out to be inadequate in some new case where it is unclear whether an entity in the case fits the definition or not. The most obvious way to handle this problem is to redefine the concept so that it is made clear whether the entity in the case fits the definition or not. But there is another way to deal with the problem. This second way is to bring forward a new set of inference rules that supplement the existing definition of the concept. Two examples of the second way of dealing with the problem of open-textured contexts in law are useful to consider.

The first is the case of the drug-sniffing dog (Weinreb, 2005). If a trained dog sniffs luggage left in a public place and signals to the police that it contains drugs, should this event be classified as a search according to the Fourth Amendment? If it can be classified as a search, information obtained as a result of the dog sniffing the luggage is not admissible as evidence. The problem is that although the concept of a search is partially defined in law, it may be open to contention whether this case fits the existing rules that provide the partial definition.⁴ Weinreb (2005, p. 24) cited two rules established by prior court decisions that can helpfully be applied to the argumentation in the problem. One is the rule that if a police officer opens luggage and then observes something inside the luggage, the information collected is classified as a search. Another is the rule that if a police officer obtains information about a person or thing in a public place without intrusion on the person or taking possession of or interfering with the use of the thing, it is not a search for purposes of the Fourth Amendment. In the case of the drug-sniffing dog, the police officer did not open the luggage, so it can be argued on the basis of the first rule that what he did was not a search. Since there was no intrusion or interference, it can also be argued on the basis of the second rule that what he did should not be classified as a search.

³ This claim needs to be qualified. The HYPO line of work, initiated by (Rissland and Ashley, 1987), modeled defeasible reasoning with cases about the meaning of the open-textured term 'trade secret'. It should also be noted that logic programs can often be regarded as offering a set of definitions to work with, and these definitions are often defeasible. This is especially true in legal experts systems based on formalization of legislation, a well known example being (Sergot et al., 1986).

⁴ Here we need to be careful to note that many legal theorists do not consider precedent cases as defining legal rules. Reasoning with precedent cases is a theory construction process in which rules are hypothesized and then critically evaluated, and the *ratio decidendi* of a case is not legally binding.

In the case of *Popov v. Hayashi*, Barry Bonds made his record-breaking 73rd home run in 2001 by hitting a ball into the stands where it was stopped by a fan named Popov in the upper webbing of his baseball mitt.⁵ Before he was given the chance to complete the catch, he was thrown to the ground by a mob of fans trying to grab it, and when the melee was sorted out, another fan named Hayashi had secured possession of the ball. The case went to trial in the Supreme Court to decide which of these two fans can be said to have secured possession of the ball. Part of the problem in the case is the issue of whether Popov may properly be said to have caught the ball. Gray (2009), in his brief on the case, formulated some rules that are helpful in defining the notion of a catch. The first is the rule that a catch does not occur simply because the ball hits the fan on the hands or enters the pocket or webbing of the fan's baseball glove. The second is the rule that a catch does occur when the fan has the ball in his hand or glove, the ball remains there after its momentum has ceased, and even remains there after the fan makes incidental contact with a railing, wall, the ground or other fans who are attempting to catch the baseball or get out of the way. Both rules provide foundations for arguments that Popov did not, properly speaking, catch the ball.

Cases like these can be used to show that there is a very close connection between the defeasibility of concepts as studied by Hart and the defeasibility of rule-based arguments of the kind studied in AI. Indeed, it may even be suggested by a consideration of such cases that defeasible concepts can be reduced to defeasible arguments. Work on case-based reasoning in the field of AI and law, for example (Rissland and Ashley, 1987) and (Costantini and Lanzarone, 1995), has even attempted to model reasoning with open textured concepts to reasoning with defeasible rules of the kind studied in mainstream AI.

Default logic (Reiter, 1980) and circumscription (McCarthy, 1986) were developed around the same time to deal with nonmonotonic inference, and both formalisms have been extensively researched in computer science since that time. Both are designed for reasoning in the absence of complete information, where a tentative conclusion is drawn based on plausible assumptions needed to fill in missing details necessary to carry out an action or solve a problem. New incoming information may require the retraction of the conclusion so arrived at if it turns out that the assumption fails to hold once this new information comes in. One application of these formalisms is to cases of communication conventions. The following example was presented by McCarthy (1986, 3-4). Suppose I hire you to build me a birdcage and you fail to put a top on the cage. It would be ruled by a judge that I do not have to pay for the cage even though I had never explicitly said to you that my bird can fly. On the other hand, if I were to complain that you wasted money by putting a top on the cage that I intended for a penguin, the judge would rule that if the bird was of a kind that could not fly, I should have told you this before you commenced work on the cage.

In default logics of the kind used in AI, first-order logic is extended with domain specific rules called defaults (Reiter, 1980). A default rule has the following form, where P is a set of statements that act as given premises and D is another statement that could be called a default blocker. The form of such a default rule is: $P: D; \text{ therefore } C$. A rule of this form tells you that if you know A , and you have no evidence of not D , then you may infer C . Another way to formulate a default rule is as a knowledge-based conditional of

⁵ It may be interesting to note that the decision in the case of *Popov v. Hayashi* has been modeled by Wyner, Bench-Capon and Atkinson (2007) using a set of argumentation schemes.

the following form: if you know that *P* is true, and you have no evidence that *D* applies, then you may infer *C*. In the case of the Tweety argument, *A* is ‘Tweety is a bird’, *D* is ‘Tweety is not an exceptional bird’ and *C* is ‘Tweety flies’. As long as the default blocker applies, then the default rule works and the defeasible argument can be treated like any other deductively valid argument. Essentially then, on this theory, a defeasible argument is analyzed as a default inference in which the warrant is a default rule. Immediately the reader will recognize that both forms of defeasible reasoning bear a strong resemblance to the argument from ignorance in the list of fallacies in section 2. This resemblance will be studied in greater detail in section 8 on the argument from ignorance.

The problem with formal default logics is how you know in a given case whether the default blocker applies. We may not know, for example, that Tweety is an exceptional bird, but then later we may find out that he is. It is in the nature of many defeasible arguments that we don’t know what lies in the future as new knowledge comes in. Indeed, according to the theory expressed by default logics, all defeasible arguments are arguments from ignorance. Thus in evaluating any given instance of such an argument, it depends on how far along an investigation has gone. If there is no evidence that *D* applies, an arguer can put forward a default argument and the respondent of the argument has to accept the conclusion, at least provisionally. But matters of burden of proof complicate such cases. We may think that Tweety is not an exceptional bird, for example, but if we are very worried that he might be, we might draw a different conclusion. Suppose Tweety has to carry an important message to military forces that depend on the information in the message. In such a case we might have doubts about how much weight we can put on the assumption that Tweety is not an exceptional bird, and look to also using other methods of sending the message. Such matters of burden of proof are very important for evaluating defeasible argumentation of the kind associated with fallacies. Prakken and Sartor (1996, p. 194) have modeled defeasible legal argumentation by using the notion of reversal of burden of proof. Defeasible arguments often have to do with presumptions that involve a reversal of the normal burden of proof.

Thus there are various distinctive aspects of defeasible arguments that suggest they are more complex than they seem. The default rule does indicate how they work, but is limited in certain respects in explaining how they should be evaluated. The same default argument may be evaluated quite differently at a different stage in the procedure whereby new information is collected and arguments evaluated. At different stages of the process, defeasible arguments may vary in how they should be evaluated. The burden of proof may shift back and forth, depending on how far along an argument has proceeded.

In law, there are three kinds of burden of proof (Williams, 1977; Prakken and Sartor, 2009). The burden of persuasion is set before the beginning the trial, and never (or only rarely) shifts throughout the whole trial⁶. How persuasive such a winning argument needs to be depends on the standard of proof for that type of trial. In a criminal trial, the prosecution has to prove all the elements of the offense beyond a reasonable doubt, whereas in a civil trial the winning side must merely have a stronger argument than the losing side. In contrast, the evidential burden (also often called the burden of producing evidence or the burden of production), can shift from one side to the other (Fleming, 1961; Williams, 1977). The evidential burden refers to “the burden of the producing

⁶ It seems to a bone of contention whether the burden of persuasion never shifts or whether it can sometimes shift but rarely does (Fleming, 1961, 62).

evidence on an issue on pain of having the trial judge determine that issue in favor of the opponent” (Williams, 2003, 166). The burden of production refers to the quantity of evidence that the judge is satisfied with to be considered by the jury as a reasonable basis for making the verdict in favor of one side (Wigmore, 1940, 279). According to Williams (2003, 166) and (Prakken and Sartor, 2009, 228), there is also a third meaning of burden of proof. In this sense of the term, ‘burden of proof’ means that if the party “does not produce evidence or further evidence he or she runs the risk of ultimately losing on that issue.” This third type of burden of proof involves a tactical evaluation of who is winning or losing at a particular point during the sequence of argumentation in the trial, and so Williams calls it the tactical burden, as opposed to the evidential burden. Gordon and Walton (2009) clarified the distinction between the burden of persuasion and the tactical and evidential burdens by showing that a trial has three stages, an opening stage an argumentation stage and a closing stage. The argumentation stage can then be broken down into a sequence of smaller stages, where each smaller stage consists of all the arguments which have been put forward by both parties so far in the proceeding. The parties take turns in a dialog putting forward arguments. The burden of persuasion is set at the opening stage, and is used at the closing stage to determine which side won the trial. The two other burdens apply only during the argumentation stage. In some instances there can be meta-dialogs to argue about whether burden of proof should be changed or which side should have burden of proof (Walton, 2007).

According to (Prakken and Sartor, 2009) and (Gordon and Walton, 2009), there are also two other types of burden of proof that can be distinguished in law. A person who feels he has a right to some legal remedy has the burden of claiming, that is, the burden of initiating the proceeding by filing a complaint alleging facts entitling him to some remedy. The burden of questioning requires that during pleading, an allegation of fact by either party is to be implicitly conceded unless it is denied.

The notion of burden of proof would surely seem to be best analyzed in a dialog argumentation framework rather than in a more restricted framework in which an argument is merely seen as a set of reasons (premises) supporting a claim (conclusion). There were attempts to adopt a dialogical view in AI in the early 1990’s, e.g. (Bench-Capon et al, 1991) and Gordon’s influential pleadings game (1995). Now the dialogical approach has become fairly common in AI and law. One serious motivation that brought argumentation theory into use in AI was the need to conduct reasoning under conditions where the circumstances are changing and there is a lack of complete knowledge (Bench-Capon and Dunne, 2007, 621). The linkage between argumentation and defeasibility was seen in the adoption of the view in which an argument was taken to be a device for presenting a justification for a claim made, including the notion that an argument can be defeated as new information comes in. It seemed at this stage, and no doubt still does to many, that an argument should be viewed as a one-sided process in which a single party merely presents the reason that might be defeated by later evidence. However, in the more recent work on argumentation in AI, there has been increasing recognition of argumentation as a dialogical process.

One of the driving forces behind the dialogical view of argumentation has been work in multi-agent systems, where dialog was seen as a natural model of interaction between agents (McBurney and Parsons, 2002). Another is the need to study informal fallacies. The insight here (Bench-Capon and Dunne, 2007, 623-624) is that particular fallacies can

only be analyzed by studying the given argument in relation to potential attacks on it by an opposed party. The most important general conclusion to be drawn from these observations is that many of the most important fallacies, and particularly those closely related to the notion of burden of proof, are only amenable to precise and useful analysis if argumentation is viewed as part of a goal-directed dialogical process for arriving at a reasoned conclusion by weighing the relevant arguments for and against it. These considerations take us to the conclusion that the evaluation of an argument in a given case as fallacious or not depends on the assumption that the argument was being used for some conversational purpose in a dialog between two parties.

4. Dialog Systems

Such a dialog has a goal, and it starts from an opening stage, moves through an argumentation stage, and concludes in the closing stage. A framework of evaluation is presupposed (Walton, 1995, 261) in which the actual text of discourse surrounding the argument needs to be taken into account and modeled as a dialog that is moving forward from an opening stage to the supposed fulfillment of a goal at the closing stage. Not only the text of discourse of the actual discussion needs to be taken into account, but also the projected completion of the dialog as it moves towards its goal.

In formal dialog systems the two participants, called the proponent and the respondent, take turns making moves. Each has what is called a commitment set (Hamblin, 1970, 1971). A commitment set, in the simplest case, is just a set of statements. They could be written on a blackboard for example, so that both commitment sets are visible to both participants. Various rules in a dialog govern what kinds of moves can be made, how the other party must respond to a given move at the next move, and what happens to this commitment set at each move. Commitment rules determine what statements go into or are taken out of each commitment set at each move (Prakken, 2000). For example, if the proponent asserts statement *A* at some move, then *A* is added to her commitment set. If a participant retracts commitment to statement *B* at some point in a dialog, and he or she was committed to *B* previously, then *B* is now removed from his or her commitment set. Among the most difficult problems is the formulation of rules of retraction of commitments for various types of dialog (Krabbe, 2001).

Commitment rules for several types of dialog are set out in (Walton and Krabbe, 1995). One of the most common types of dialog is called persuasion dialogue. 'Persuasion' in formal dialog theory refers not to psychological persuasion but to rational persuasion (Prakken, 2006). A proponent persuades a respondent to commit to a statement *A* in this sense by presenting a structurally correct argument with *A* as the conclusion containing only premises that are already commitments of the respondent. The goal of persuasion dialog has been accomplished in such a case because the respondent was not committed to this particular statement, but now he is. Several different formal models of persuasion dialog have been constructed in (Walton and Krabbe, 1995). Some, called rigorous persuasion dialogs or RPDs, have rules that do not allow the participants much flexibility, but the advantage of an RPD is that it is fairly simple while at the same time being formally rigorous. But RPDs model argumentation only in a simplistic way that does not express many of the important features of empirical cases of natural

language argumentation. Another kind, called PPD or permissive persuasion dialog, is more flexible and is closer to empirical argumentation.

No matter what type of dialog is concerned, and no matter which rules are selected, and there can be many variations, arguments are always evaluated in light of three factors. These are how the argument was put forward in a dialog, how that move affects the commitment sets of both parties, and in some cases how the respondent replied to the argument. The device of commitment is useful, and does not get into all the problems that have been encountered with the BDI (belief-desire-intention) model of defeasible reasoning as belief revision. A belief is an internal entity, and using the BDI model can involve an argument evaluator in the mysterious metaphysics of iterated beliefs. Commitment is a less opaque concept. You are committed only to statements you have gone on record as accepting in a dialog. The idea is that a public record is kept of a participant's set, as each move is made, so that if an arguer claims that he never said something, the other party can go back over what he said or didn't say, and use this as evidence in determining commitment. Thus commitment represents acceptance of a specific statement, judged by the evidence available from the prior text of dialog in a given case. Dialog models can represent argumentation in a dynamic way that represents an argument as not just a static set of premises and conclusion, but as a speech act put forward by one party and replied to by the other party.

There can be different types of dialog. In a persuasion dialog, the goal is to resolve or throw light on some conflict of opinions or unsettled issue (Prakken, 2006). The proponent of an argument tries to get the respondent to commit to the conclusion by using the premises as reasons. The proponent uses the commitments of the other party as these premises. If she puts forward a structurally correct argument that has premises that are all commitments of the respondent, then the respondent is rationally obliged to commit to the conclusion. That is the process called rational persuasion. But there are other types of dialog as well (Walton and Krabbe, 1995). Some dialogs take the form of an investigation that collects facts and then tries to prove or disprove some statement using these supposed facts as premises.

A dialog, in the simplest case, has two participants, called the proponent and the respondent. The two participants take turns making moves. The moves are essentially speech acts of various kinds. For example, asking a question is a kind of move. Making an assertion is a kind of move. Putting forward an argument is another common kind of move. A type of dialog is defined formally as a set of participants, a set of rules defining permitted or required moves, a set of rules for determining how one participant must reply to the prior move of the other participant, and a set for determining when a completed sequence of moves fulfills the goal of the dialog (so-called "win-loss" rules). The general idea is that a dialog is a sequence of moves, starting at a first move and ending at a last move. In the dialog theory of Hamblin (1970; 1971), the proponent makes the first move, the respondent makes the next move, and then they take turns, producing an orderly sequence of moves. Each member in the sequence is defined by Hamblin (1971, p. 130) as a triple, $\langle n, p, l \rangle$. n represents the length of the dialog (the number of moves so far). p is a participant. And l is what Hamblin calls a locution, or what we now call a speech act. Such systems of dialog have been proposed by Mackenzie (1981, 1990) and Walton and Krabbe (1995).

A dialog should be seen as having three characteristic stages (Gordon and Walton, 2009).⁷ The sequence of argumentation in the argumentation stage should be seen as having started at the opening stage and moving towards the closing stage. The three stages are shown in figure 1.

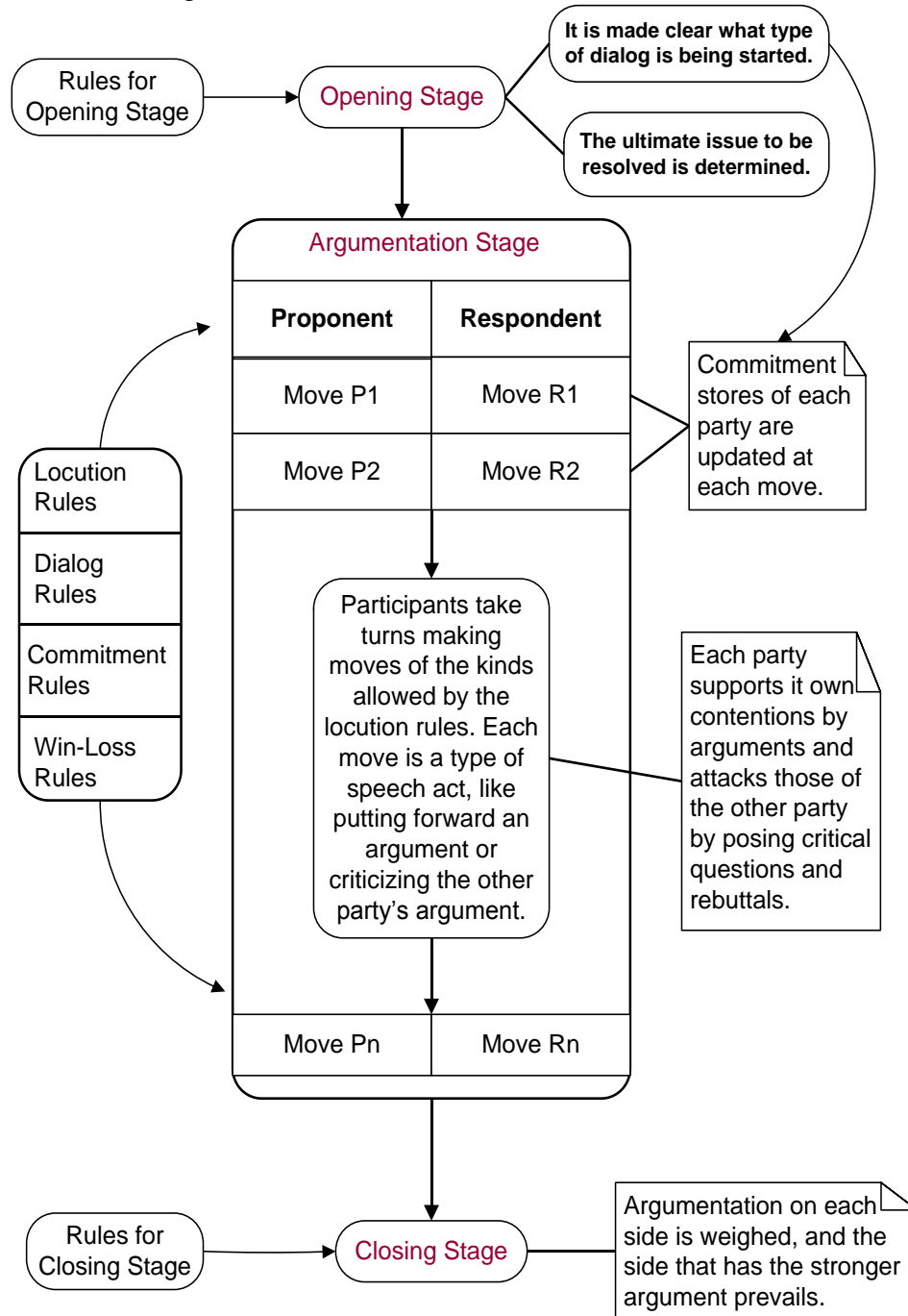


Figure 1: the Three Stages of a Dialog

⁷ The confrontation stage of van Eemeren and Grootendorst (1992, pp. 34-37) can be classified as part of the opening stage.

During the opening stage, the rules for allocating the requirements for what constitutes a winning argument that will apply over the following two stages are set, and both parties become committed to following these rules in order to settle the issue by rational argumentation. Next there is an argumentation stage in which the arguments and rebuttals on both sides are brought forward and replied to. Each side takes its turn to make moves. Finally there is a closing stage in which it is judged whether the issue has been settled.

Each type of dialog has its own special rules of procedure that need to be followed by both sides during the argumentation stage. The question is whether fallacies can be identified simply with breaches of the rules. In some cases they can, and the rules provide valuable normative guidelines that give insight into what is wrong about a fallacious argument or move. However, there is no one-to-one correspondence between a particular fallacy and the violation of some particular procedural rule of a dialog. As we look over the various fallacies, the problem is that a single fallacy can be committed in a number of ways. One such rule is that a participant in a critical discussion must not prevent the other party from putting forward arguments or asking questions that are legitimate contributions to the dialog. However, several of the traditional major informal fallacies appear to violate this rule. Appeal to force is one, but many of the other major informal fallacies also appear to be fallacies because they commit such an infraction. As shown above, *argumentum ad verecundiam* often fits this diagnosis. And, as will be shown below in section 7, the *secundum quid* fallacy, the fallacy of ignoring exceptions to a generalization, can be held to be fallacious because the arguer ignores critical questions that should be raised about exceptions to a general rule.

In some cases, fallacious arguments are too weak because a required premise has not been supported adequately by bringing forward enough evidence, but despite this weakness, because the argument is so powerfully impressive to the given audience at a particular moment, it has a devastating impact and carries the day. A good example of this phenomenon is the *ad hominem* type of argument, like the use of a negative campaign tactics at the right moment in a political campaign. Such an argument may be so powerful and convincing to an audience that it is taken for a much stronger argument than it really is. Perhaps the argument is based on very little or no evidence that can be verified, but is merely innuendo based on suspicion, or perhaps some rumor from an unspecified source. Despite these deficiencies, the argument may still be very effective, and swing voters one way or the other, especially when the time for refuting it, or even seriously examining it, is not available because the time between the negative attack and the election is very short. In other cases, the problem is not so much the weakness of the argument, its lack of support, but the way it is used as a powerful tactic to close off further discussion.

5. Profiles of Dialog

Using the whole apparatus of a formal dialog structure with all its stages and rules (see section 4 above) may not be necessary to help analyze a text of argumentation in some examples. Often the most useful tool is the profile of dialog (Krabbe, 1999). A profile of dialog is a relatively short table of moves with the proponent's moves listed sequentially in the left column and the respondent's matching moves (MV) in the right column. An

example is the small profile of dialog shown in table 1. In this example, the proponent began at move 1 by asking a why-question. The respondent replied to the question by putting forward an argument, giving a reason why the proponent should accept the statement *A* that she questioned.

MV	Proponent	Respondent
1.	Why should I accept <i>A</i> ?	Because <i>B</i> .
2.	Why should I accept <i>B</i> ?	Because <i>C</i> .
3.	I do not accept <i>C</i> .	Do you accept 'If <i>D</i> then <i>C</i> '?
4.	Yes.	Do you accept <i>D</i> ?
5.	Yes.	Well then you must accept <i>C</i> .

Table 1: Example of a Profile of Dialog

As the dialog proceeds, the respondent keeps trying to persuade the proponent to accept *A*. He uses a *modus ponens* form of argumentation at moves 3-5 to try to get the proponent to accept *C*. By this means he hopes to get her to accept *B*, and ultimately *A*.

How the profile of dialog works as applied to an example of the fallacy of many questions above is shown in the profile presented in figure 1 (Walton, 1995, 203), representing the proper sequence of turn-taking by both parties. The question asked is, 'Have you stopped cheating on your income taxes?' The reason that asking this question is taken to be fallacious is that the respondent may never have cheated on his income taxes, or may never have even made income tax returns in the past. But if he has to answer the question directly, yes or no, he can't deny these assumptions. This sequence is shown in table 2.

MV	Questioner	Respondent
1.	Have you made income tax returns in the past?	Yes.
2.	Have you cheated on those income tax returns in the past?	Yes.
3.	Have you stopped cheating on your income tax?	

Table 2: Profile of Dialog for the Fallacy of Many Questions

Once the questioner has made moves 1 and 2, and the respondent has answered as shown in table 2, then finally the questioner can properly ask the complex question.

A profile of dialog can also be represented as a graph structure, as shown in the example of the question, 'Are you confused when you are not on your medications?'. Asking this question could be considered to be fallacious if it is used as an attempt to suggest that the respondent may not be mentally stable. In turn, that conclusion could be used as the basis of an *ad hominem* attack alleging that the person may not be mentally stable, and that therefore no serious attention should be paid to his argument. A profile of dialog for this example is shown in figure 2.

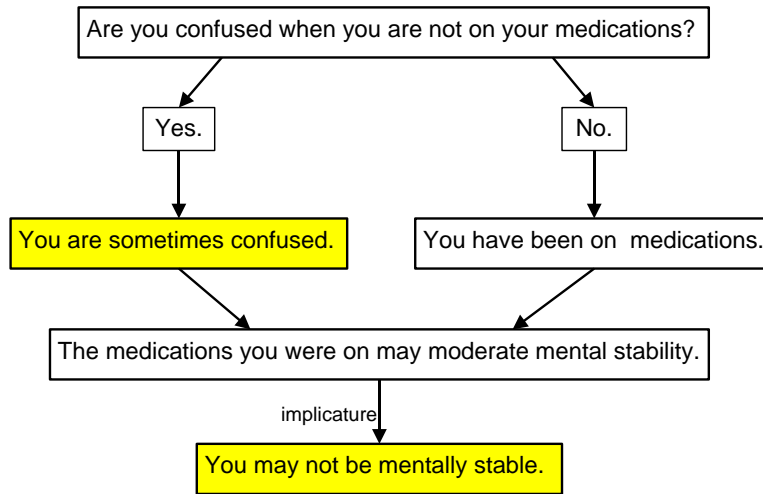


Figure 2: Profile 1 for the Fallacy of Many Questions

Tracking the sequence of questioning shown in the profile in figure 2, we can see that no matter which way the respondent answers, he is inevitably led to one or more of the dangerous admissions shown in the darkened boxes that represent dangerous admissions.

The profile in figure 3 shows the proper sequence of questions and answers for a non-fallacious instance of the same complex question.

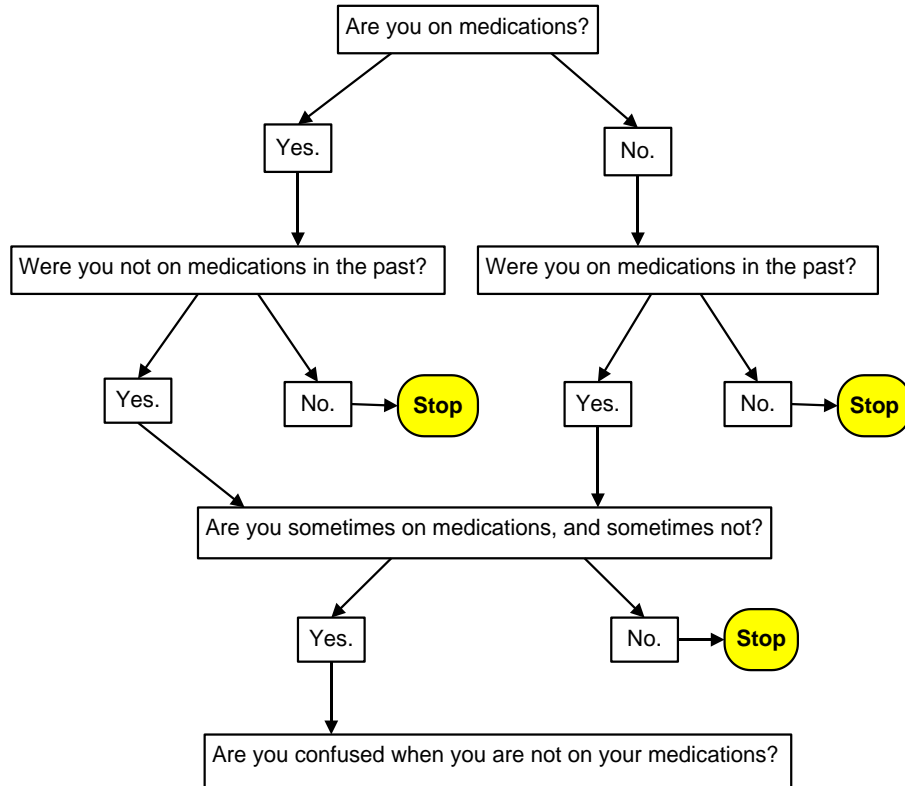


Figure 3: Profile 2 for the Fallacy of Many Questions

In figure 3, the following correct sequence of questioning is represented. Before asking the final complex question shown at the bottom of figure 4, the questioner must first ask the two prior sequences of questions shown above it. If all the questions are asked in the right order as shown in figure 3, asking the complex question in that context of dialog could be reasonable. This profile illustrates the point that the same complex question can be fallacious or not, depending on the context of dialog, referring to the sequence of questions and replies and the order in which they were put forward.

This discussion suggests that requirement 2 in the definition of ‘fallacy’ proposed in section 1 needs to be modified as follows: a fallacy is a failure that occurs in what is supposed to be an argument or an argumentation strategy used as a move in a dialog. Two other potential exceptions to requirement 2 also need to be considered. Common instances of the fallacy of begging the question and the fallacies of relevance, coming under the traditional heading of *ignoratio elenchi*, involve chains of argumentation rather than single arguments. The modified version of requirement 2 makes room for these kinds of fallacies as well. The profile of dialog represents the ordered moves in a dialog before and after the move where the argument in question was put forward. The profile of dialog is a useful tool for the analysis of argumentation associated with fallacies (Krabbe, 1999). A fallacy can also be a sequence of argumentation moves in a dialog.

6. Dialog Conditions for Defeasible Arguments

To illustrate how defeasible reasoning can be modeled in a dialog format, consider an instance of the Tweety argument displayed in the profile in table 3.

MV	Proponent	Respondent
1.	Tweety flies.	How can you prove that?
2.	Tweety is a bird.	So what?
3.	Birds fly.	Yes, they generally do.
4.	Therefore Tweety flies.	But Tweety is a penguin.
5.	Oh, really?	Therefore Tweety does not fly.

Table 3: The Tweety Profile of Dialog

In the Tweety profile, the proponent put forward a defeasible argument. At move 3, the respondent conceded the major premise of the argument. At move 4, the proponent drew the conclusion. But then, at his turn in move 4, the respondent defeated the argument by bringing in new information about Tweety. This profile shows how things should go ideally in a dialog in which one party has brought forward a defeasible argument. She properly used a defeasible generalization as a premise. This premise, along with the other one, formed an argument having the form of DMP (defeasible *modus ponens*). The respondent indicated at move 2 that he accepts the premise that Tweety is a bird. At move 3, he indicates acceptance of the generic premise that birds fly. Thus at the next move, the respondent must commit to the conclusion that Tweety flies, in order to follow the commitment rules of a persuasion dialogue. The big problem is whether the respondent can now retract commitment from one of the premises. At move 4, he doesn’t retract

directly. Instead he offers a new bit of evidence in the form of the statement ‘Tweety is a penguin’.

The Tweety profile of dialog illustrates a key feature of defeasible argumentation. The proponent conceded that Tweety does not fly, and so his prior argument is now defeated. An important feature of a defeasible argument in a dialog is the *Openness to Defeat (OTD) Condition*: When the proponent has put forward a defeasible argument during the argumentation stage in a dialog, s/he must be open to giving it up and admitting its defeat at any future move by the respondent that defeats the argument at any point before the closing stage of the dialog. Principles somewhat like the OTD condition have been recognized in the argumentation literature. Van Eemeren and Grootendorst (1992, 108) state that parties in a critical discussion cannot declare their viewpoint as sacrosanct, so that they are rendered immune to criticism. Johnson (2000, 224) formulated a principle of vulnerability of arguments comparable to Popper’s notion of falsifiability: “to be a legitimate argument, an argument must be vulnerable to criticism”.

What sort of move defeats a defeasible argument like the Tweety argument? The defeater has to be an exception to the rule postulated by the generic premise. Such a defeasible rule is subject to exceptions. So the respondent, to defeat the argument, has to come up with information in the given case that presents an exception to the rule. The OTD condition requires that if the respondent comes up with such a case, and the proponent admits that it constitutes an exception to the rule, then the proponent has to give up his argument. Thus the respondent no longer has to accept the conclusion. Indeed, both parties must now retract commitment to the conclusion.

The OTD condition says something about all future moves of a dialog between the move where the defeasible argument was put forward and the closing move of the dialogue. It says that at all such moves, the proponent must retract the conclusion, if the respondent makes the sort of move that defeats the argument used to prove it. There is no way to predict when a respondent may come up with such a defeater before closure of a dialogue. Defeasible arguments are characteristically open-ended⁸. They have to do not just with the argument as presented, but with the future stretch of dialog beyond the argument in a continuing dialog that provides the framework of argument use.

A second feature of defeasible argumentation illustrated by the Tweety dialog is the phenomenon often called the shifting of the burden of proof. At move 1, the proponent has made a claim, and so has incurred a burden of proof. The respondent asks, “How can you prove that?”, referring to the proponent’s claim that Tweety flies. The proponent then fulfills this burden by putting forward a defeasible argument to support it. That should be the end of it, if the argument was not defeasible. But then at move 4, the respondent asserts that Tweety is a penguin. The respondent made an allowable move by bringing forward an exception to the rule posited by the proponent’s major premise in his prior argument. This pattern is characteristics of a defeasible argumentation.

The pattern of argumentation characteristic of defeasible argumentation has the following characteristic dialog sequence for defeasible argumentation (DSD).

1. The proponent puts the defeasible argument forward.
2. One premise is a generalization or conditional (rule) that admits of exceptions.

⁸ This observation presents a computational problem. For this reason, much modeling of dialogs in computing uses the restriction that an argument must be attacked immediately or conceded. See for example (Vreeswijk and Prakken, 2000).

3. If the respondent is committed to the premises, he must commit to the conclusion.
 4. But his commitment to the conclusion can be retracted.
 5. The dialog must remain open to the respondent's finding an exception to the rule.
 6. As soon as the respondent cites such an exception, the proponent's argument defaults.
 7. The proponent must now retract commitment to the conclusion.
- The argumentation in the profile of dialog shown in table 4 fits the DSD pattern.

MV	Proponent	Respondent
1.	$P: D$; therefore C .	I accept that argument.
2.	Do you accept premise P ?	Yes.
3.	So you must accept C .	Yes.
4.	C has been proved then?	I do not accept C .
5.	Why not?	Here is evidence for <i>not-D</i> .
6.	OK, I accept that evidence.	Will you now retract C ?

Table 4: An Example of a Profile of Dialog Fitting the DSD

Let's say that the dialog is a critical discussion and that C is a statement that the proponent needs to prove, in order to prove her ultimate thesis in the dialog. Let's also assume, however, that the respondent can still make a reply at move 4. The closing stage has not been reached yet. If he can find an exception to the default rule in the proponent's defeasible argument put forward at move 1, she must retract C . Normatively speaking, she must retract C because the DSD applies. Hence she must answer 'yes' at the next move. If she answers 'no', she commits the *secundum quid* fallacy.

This sequence of argumentation has the reversal of burden of proof characteristic of the typical presumptive argument of the kind often used in law. A presumption is accepted tentatively in a dialog as a way of moving the dialog forward, but it can be defeated if new evidence comes into the dialog. For example, in law, a person who has disappeared without a trace for a determined number of years may be presumed dead, for purposes of settling an estate. The basis of the reasoning is an argument from ignorance. But in this instance, the argument from ignorance is reasonable, not fallacious, as long as it is treated as a defeasible argument that can default in the face of new evidence. If there is no evidence that the person is alive, then the conclusion can be drawn by inference (after the stated period) that he is dead. But if such a person turns up, the presumption that he is dead is defeated. This type of argumentation will be treated at more length in section 8, concerning arguments from ignorance. If it turns out later that the person who turned up is not the missing person, then the new conclusion will have to be retracted and we may revert to the old one. The theory is that we track defeasible chains of argumentation through a dialog.

Retraction of the conclusion of the argument is shown as a key dialectical feature of defeasible argumentation at stage 7 of the DSS. In a persuasion dialog, retraction is allowed, although not in all circumstances. Retraction of one statement may require retraction of other statements that it is closely related to by inference. For example, if the proposition that a person is dead is retracted then the proposition that someone else may inherit his estate may also need to be retracted, even though it was accepted previously. However, the problem of retraction has not been solved for persuasion dialog, even

though proposed solutions to it have been put forward (Krabbe, 2001). Thus in the dialogical theory of defeasible reasoning, the problem of defeasibility is recast as a problem of determining conditions for retraction in the various types of dialog.

7. The Fallacy of *Secundum Quid*

The single fallacy most closely related to defeasible reasoning in general is the one called *secundum quid*. The *secundum quid* fallacy was very clearly explained in one logic textbook, (Joseph, 1916, p. 589), in a way that makes it appear similar to the way Aristotle described it. However, as shown in detail in (Walton, 1999a), its treatment in nearly all the other logic textbooks is mixed in with other fallacies and defined using ancient and unhelpful terminology, like “accident”, and “converse accident”. The textbook treatments of this fallacy and other related errors of reasoning are more than just inconsistent. They are highly confusing. *Secundum quid* is mixed up with related fallacies like hasty generalization, glittering generality, converse accident, over-simplification, *de dicto simpliciter*, and so forth. These supposed fallacies are conflictingly defined and mixed in together in a way that shows an urgent need for clarification and systematization. The best place to begin is with some examples.

Secundum quid (in Greek, *para to pe*), means ‘in a certain respect’, and refers to qualifications attaching to a term or generalization (Hamblin, 1970, 28). The *secundum quid* fallacy is the error of neglecting qualifications when drawing a conclusion by inference. The clearest account of what this fallacy is supposed to consist in was given by Joseph (1916, 589) using this example: “Water boils at a temperature of 212 degrees Fahrenheit; therefore boiling water will be hot enough to cook an egg hard in five minutes: but if we argue thus at an altitude of 5,000 feet we shall be disappointed; for the height, through the difference in the pressure of the air, qualifies the truth of our general principle”. The fallacy of *secundum quid* cited here can be specified more precisely by examining the structure of the following argument

Explicit General Premise: Water boils at a temperature of 212 degrees.

Implicit General Premise: Immersing an egg in water at a temperature of 212 degrees Fahrenheit will be hot enough to cook an egg hard in five minutes.

Conclusion: Immersing an egg in boiling water will be hot enough to cook an egg hard in five minutes.

This argument works fine, assuming that standard conditions, like those of altitude, hold. But let’s add a new premise, the statement that we are at 5,000 feet, a higher altitude than we might be assuming to be standard. What happens is that the explicit general premise fails to hold, not generally, but in these specific conditions. In these conditions water boils at a different temperature, and the inference defaults. The error here can be diagnosed as one of neglecting qualifications to the explicit general premise of the argument, and thereby drawing a wrong conclusion.

The fallacy here seems to be precisely an error of defeasible reasoning. The initial argument was fine, as applied to what is taken to be a normal or standard situation, but defaults when new knowledge comes in about the particulars of the case. Once the new data comes in, the old conclusion must be given up. It no longer holds in the special circumstances of the case. It’s not that the old argument was bad, generally speaking. It’s just that it doesn’t work, once the new data about altitude is added in as a premise.

According to Aristotle’s description of *secundum quid* (*On Sophistical Refutations* 166b38- 167a21), this fallacy arises from two ways an expression can be used, absolutely, or in a certain respect. “If, for example, after securing an admission that the Ethiopian is black, one were to ask whether he is white in respect of his teeth, and then, if he be white in this respect, were to think that he had finished the interrogation and had proved dialectically that he was both black and not black”. The premise that the Ethiopian is black generally (meaning, in respect of skin color), is true. But from this premise it would be erroneous to draw the conclusion that he has black teeth. For when we say ‘The Ethiopian is black’, we (normally) mean that he is black in a certain general respect, one that does not refer specifically to the color of his teeth.

The Ethiopian example is not quite the same as the egg example, because the former refers to a generality implicit in standard word usage while the latter refers to an empirical generalization. Also, the analysis of the egg example based on defeasible reasoning does not seem to apply to the Ethiopian example. We now need to recall that in the section on defeasible reasoning above (section 4), following Hart a distinction was drawn between defeasible concepts and defeasible reasoning in arguments. It is this distinction that brings out the difference between the defeasibility in the two examples. The egg example is a standard case of defeasible reasoning of the kind usually taken to be central in computing, corresponding to the description of defeasible reasoning cited in Hart’s paper in section 4. The Ethiopian example is an instance of a defeasible concept. When we say the Ethiopian is black, arguing that he can’t be black because his teeth are white is fallacious because that is not an exception. As we saw, Aristotle offered a dialectical analysis of the Ethiopian example, showing how failing to draw the distinction between ‘absolutely’ and ‘in a certain respect’ leads to a contradiction.

This analysis can be displayed using the profile in table 5 (Walton, 1999, 164).

MV	Proponent	Respondent
1.	Is the Ethiopian black?	Yes.
2.	Is he white in respect to his teeth?	Yes.
3.	If something is white, must it follow that it is not black?	Well, yes, of course.
4.	So by your admission the Ethiopian is black and not black.	Respondent loses.

Table 5: Profile of Dialog for the Ethiopian Example

The profile in table 5 shows that while each of the respondent’s single replies answers the proponent’s question reasonably, the ordered dialectical sequence of them leads to inconsistent commitments. In Aristotelian dialectic, a fallacy can be analyzed as a set of opinions that seem individually plausible (endoxic, generally accepted) but that are collectively shown to lead to a contradiction (or to a comparable logical difficulty). The analysis reveals that one should be careful in accepting a statement to be clear whether the statement is meant to be true absolutely, in an unrestricted sense, or whether it is meant to be true only in a certain respect.

This is not the end of the story on the fallacy of *secundum quid*. How general rules should be applied to particular cases is centrally important in scientific reasoning as well as legal reasoning of the most common sort. There are special argumentation schemes

(Walton, Reed and Macagno, 2008, 343-345) representing arguing from rules and arguing from exceptions to rules. Also, more needs to be done to map the relationship of *secundum quid* to related fallacies like hasty generalization.

Both the egg and the Ethiopian example show how fallacies can arise when an arguer is too rigid, and views an inference as based on a strict universal generalization modeled by the universal quantifier in deductive logic instead of a defeasible generalization. The fallacy can be seen as a species of violation of the OTD condition. It can be argued that many of the fallacies, especially the twelve based on argumentation schemes, are failures of this sort (even though other faults can be involved as well). Let us next consider another fallacy that is also closely associated with defeasible reasoning.

8. Lack of Knowledge Inferences

The simplest formulation of the scheme for the *argumentum ad ignorantiam* of the logic textbooks is this: statement A is not known to be true (false), therefore A is false (true). As noted above, this form of argument is often called the lack of evidence argument in the social sciences or an *ex silentio* argument in history, where it is presumed not to be fallacious. In both fields it is commonly regarded as a reasonable but inconclusive form of argument. To cite an example (Walton, 1999), there is no evidence that Roman soldiers received posthumous decorations, or medals for distinguished service, as we would call them. We only have evidence of living soldiers receiving such awards. From this lack of evidence, it has been considered reasonable by historians to put forward the hypothesis that Roman soldiers did not receive posthumous decorations. Of course, such a conjecture is not based on positive evidence, but only on a failure to find evidence that would refute it. Such arguments from ignorance are common in many fields, not least in law, as will be shown below.

As shown in (Walton, 1999), argument from ignorance needs to be seen as an inherently dialectical form of argumentation. The context most often helpful to grasping the structure of this form of argumentation is that of an ongoing investigation in which facts are being collected and inserted into a knowledge base. In such a context, the argument from ignorance can be represented using the following argumentation scheme for epistemic argument from ignorance given in (Walton, Reed and Macagno, 2008, 328). In this knowledge-based scheme, D is a domain of knowledge and K is a knowledge base in a given domain, or field of knowledge.

All the true propositions in D are contained in K .

A is in D .

A is not in K .

For all A in D , A is either true or false.

Therefore, A is false.

This form of argument can be deductively valid in a domain D where K is closed, meaning that it contains all the statements that can ever be known in that domain. But in a vast majority of cases, argument from ignorance is a defeasible inference that may default as an investigation proceeds and new knowledge come in. Thus one of the most important critical questions in evaluating any given instance of an argument from ignorance is whether the knowledge base is open or closed.

An example that can be used to illustrate how the nonfallacious *ad ignorantiam* works as an argument is the foreign spy argument: Mr. X has never been found guilty of breaches of security, or of any connection with agents of the foreign country he is supposedly spying for, even though the Security Service has checked his record; therefore, Mr. X is not a foreign spy. This argument from ignorance is defeasible, because it is not possible to be absolutely certain that Mr. X is not a foreign spy. Mr. X could have avoided detection through many security searches, as Kim Philby did. Hence arguments from ignorance tend to be defeasible arguments, even though they can be conclusive in some cases. The argument from ignorance can be seen to be a very common kind of defeasible argumentation, once you learn to recognize it.

There is a very common principle often appealed to in knowledge-based systems in AI called the closed world assumption (Clark, 1978). Essentially, the closed world assumption means that all the information that there is to know or find is listed in the collection of information one already has, but there are different ways of representing information. According to Reiter (1980, p. 69), the closed world assumption is met if all the positive information in a database is listed, and therefore negative information is represented by default. Reiter (1987, 150) offers the example of a database for an airline flight schedule to show why negative information is useful. It would be too much information to include in such a database all flights and the city pairs they do not connect. This amount of information would be overwhelming. Instead, the closed world assumption is invoked. If a positive flight connection between a pair of cities is not asserted on the screen representing the database, the conclusion is drawn that there is no flight connecting these two cities. If the system searches for a flight of the designated type and does not find one in the data base, it will reply “no”. Reiter (1980, p. 69) described the form of argument used in this sequence of knowledge-based reasoning as: “Failure to find a proof has sanctioned an inference.” As noted above, this kind of inference by default from lack-of-knowledge has traditionally been called the *argumentum ad ignorantiam* in logic. The argument from negative evidence may be merely a defeasible inference that leads to a provisional commitment to a course of action, but should be seen as open to new evidence that might come into an investigation and needs to be added to the knowledge base.

In legal argumentation, argument from ignorance is closely associated with what is often called the presumption of innocence in a criminal trial.⁹ The prosecution has the burden of proof and must bring forward enough evidence to satisfy the proof standard of beyond reasonable doubt. The defendant need only bring forward enough evidence to prevent the prosecution from meeting its burden of proof, by casting enough doubt on the prosecution’s attempt to prove its claim. This asymmetry involves an argument from ignorance. If the defense can show that there is a lack of evidence to support the prosecution’s claim (ultimate thesis to be proved in the trial), then the defense has shown that this claim does not hold up and must be rejected. This form of argumentation meets the requirements for the argumentation scheme of the argument from ignorance. Thus argument from ignorance is fundamental to the argumentation structure of the trial in the adversary system.

Argument from ignorance is also common in more special forms in legal argumentation. For example, as shown by Park, Leonard and Goldberg (1998, p. 103),

⁹ This terminology is misleading. It should be said that the burden of persuasion is on the prosecution side.

there is a presumption that some writing has been accurately dated: “unless the presumption is rebutted, the writing in question will be deemed accurately dated”. Another example (p. 153) concerns character evidence. Suppose a first person was in a position to hear derogatory statements about a second person if any were made. And suppose the first person testifies that he heard no such comments. This testimony counts as evidence of the first person’s good character. The form of argument in such a case is that of argument from ignorance. If no evidence of bad character was found or reported by the witness, this lack of such a finding may be taken as evidence of good character.

Arguments from ignorance having the form of the argumentation scheme set out above are best analyzed as defeasible arguments at some stage of a dialog or investigation in which evidence is being collected and assessed. The typical pattern of reversal of burden of proof characteristic of the *argumentum ad ignorantiam* can be modeled by the following dialog sequence discussed by Krabbe (1995, p. 256).

Proponent: Why A?

Respondent: Why not-A?

This pattern of shifting of the burden of proof in dialog is characteristic of the argument from ignorance. In a case study in (Walton, 1996, 118-122), during a political debate the opposition party asked the government minister to prove with absolute certainty that Canadian uranium was not being used for military purposes. The minister replied that the opposition should give evidence to support their allegation that Canadian uranium was being used for military purposes.

Argument from ignorance is not always a defeasible argument. In a case where a knowledge base is closed, the argument from ignorance can be conclusive, assuming that the closed world assumption holds, for example if a dialog has reached the closing stage. Thus generally speaking, arguments from ignorance need to be analyzed and evaluated using two tools. One is the argumentation scheme. The other is the placement of that scheme in a context of dialog representing an investigation that has some standard and burden of proof set at the opening stage, so that it is known whether the closed world assumption applies or not.

Argument from ignorance displays a pattern common to many if not all of the twelve fallacies based on schemes. That pattern is for the proponent to press ahead too aggressively to jump to a conclusion uncritically by overlooking the defeasibility of the argumentation scheme in question. The proponent may have collected some evidence, even in the form of negative evidence, but has not collected enough to satisfy the standard of proof that should be required. When confronted by critical questions, or by evidence that suggests an exception to the defeasible rule on which the inference is based, the proponent violates the OTD condition. There are various ways to try to carry out such a strategy by trying to escape the need for retraction. Each way is associated with a single fallacy, or a group of fallacies. One way is to try to browbeat the respondent into accepting the conclusion (for example, by arguing that nothing he says is of any value, because he is not an expert). Another is to try to shift the burden of proof to the other side to avoid proof (as in argument from ignorance). Still another is to mount a personal attack (*ad hominem* argument) that attempts to discredit the respondent, so that any objections or counter-arguments he raises are discounted as worthless.

9. Conclusions

The dialog-based theory applied in this paper has identified three characteristics that are essential to the defeasible argumentation on which the twelve scheme-based fallacies are built: openness to defeat, reversal of burden of proof, and retraction of commitment. It was shown how these three characteristics can best be modeled using a dialog model in which an argument is seen as a sequence of moves made by a proponent and a respondent who take turns. With respect to defeasibility, there are two ways to view a dialog in a given case. It can be open or closed. In typical cases, defeasible arguments are viewed as having been put forward, and as not yet being defeated, assuming that the dialog is still open. At the opening or argumentation stages, commitment to the conclusion of the argument is still tentative and subject to possible retraction as the dialog proceeds. This tentativeness has often been the basis of a traditional feeling of distrust about defeasible arguments. They have often been seen as unreliable or even fallacious. The argument from ignorance is an excellent case in point. We need to get over this wholesale rejection of defeasible arguments and see them as arguments that can go wrong, but are often quite reasonable. They are typically reasonable in cases of uncertainty and lack of knowledge where some decision for action needs to be made or a presumption adopted. They are dangerous but necessary.

One of the most important things about defeasible arguments is that they are often used during the argumentation stage of a dialog, before it has reached the closing stage. One needs to take a stance of being open-minded about such arguments, and be ready to give them up if new evidence comes in. It is perhaps because of the human tendency to be reluctant to admit defeat, and thus to close off argumentation too soon, that the problems with defeasible arguments often arise as fallacies. These matters have to do with dialectical factors like presumption, default, burden of proof, and openness and closure of argumentation in a dialog. As shown in this paper, such arguments depend on a default blocker that makes acceptance of an argument conditional on lack of evidence to support a key assumption. Such an argument is based on a mixture of knowledge and lack of knowledge. It is very often an argument from ignorance.

In an argument from ignorance, it is argued that a statement *A* is not yet known to be true (false), and argued on this basis that *A* is false (true). The premise may be taken to mean that *A* is not yet known to be true (false) at some point in an investigation or collection of data that is underway. At least, that is what the premise means if the case is a typical sort where the dialog is not closed, and the argumentation is still open to possible defeat in the future as more information is collected. The proponent urges the respondent to accept the conclusion that *A* is false (true) based on the findings to that point in the dialog. If the respondent does accept the conclusion of the argument from ignorance it is a tentative commitment that is open to retraction as the dialog proceeds. Only when the dialog reaches the closing stage does the argument from ignorance become non-defeasible. The conclusion is then based on the closed world assumption.

The notion of fallacy should be defined by a more concise but amplified definition that improves on the pragmatic definition (1995) proposed in section 1.

- An argument
- that is often an instance of a defeasible argumentation scheme
- that is reasonable, but is somehow used wrongly, and

- that falls short of the standard of proof set for it in the dialog the arguer is supposed to be taking part in
- but that plausibly seems correct [in its given context of dialog]
- and committing it poses a serious obstacle to reaching the goal of the dialog.

The tools shown to be most useful for the study of fallacies in this paper are defeasible reasoning, defeasible argumentation schemes with matching sets of critical questions, formal models of dialog, the OTD condition, the DSD sequence and profiles of dialog.

References

- Aristotle, *On Sophistical Refutations*, trans. E. S. Forster, Loeb Classical Library, Cambridge, Mass., Harvard University Press, 1928.
- Bench-Capon, T. J. M., D. Lowes and A. M. McEnery, 'Argument-Based Explanation Logic Programs', *Knowledge-Based Systems*, 4 (3), 1991, 177-183.
- Bench-Capon, T. J. M. and P. E. Dunne, 'Argumentation in Artificial Intelligence', *Artificial Intelligence*, 171, 2007, 619-641.
- Bondarenko, A., P. M. Dung, R. A. Kowlski and F. Toni, 'An Abstract Argumentation-Theoretic Approach to Default Reasoning', *Artificial Intelligence*, 93, 1997, 63-101.
- Clark, K. L., 'Negation as Failure', *Logic and Data Bases*, ed. H. Gallaire and J. Minker, New York, Plenum Press, 1978, 293-322.
- Costantini, S. and A. Lazarone, 'Explanation-Based Interpretation of Open-Textured Concepts in Logical Models of legislation', *Artificial Intelligence and Law*, 3, 1995, 191-208.
- Dung, P. M., 'On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-person Games', *Artificial Intelligence*, 77, 1995, 321-358.
- Fleming, J., 'Burdens of Proof', *Virginia Law Review*, 47, 1961, 51-70.
- Gordon, T. F., *The Pleadings Game: An Artificial Intelligence Model of Procedural Justice*, Dordrecht, Kluwer, 1995.
- Gordon, T. F., H. Prakken and D. Walton, 'The Carneades Model of Argument and Burden of Proof', *Artificial Intelligence*, 171, 2007, 875-896.
- Gordon, T. F. and Walton, D., 'Proof Burdens and Standards', *Argumentation and Artificial Intelligence*, ed. Iyad Rahwan and Guillermo Simari, Berlin, Springer, 2009, 239-260.
- Gray, B. E., 'Reported and Recommendations on the Law of Capture and First Possession: Popov v. Hayashi', *Superior of the State of California for the City and County of San Francisco*, Case no. 400545, November 6, 2002. Available May 24, 2009 at: http://web.mac.com/graybe/Site/Writings_files/Hayashi%20Brief.pdf
- Hamblin, C. L., *Fallacies*, London, Methuen, 1970.
- Hamblin, C. L., 'Mathematical Models of Dialogue', *Theoria*, 37, 1971, 130-155.
- Hansen, H. V. and R. C. Pinto (eds.), *Fallacies: Classical and Contemporary Readings*, University Park, Pennsylvania State University Press, 1995, 251-264.
- Hart, H. L. A., 'The Ascription of Responsibility and Rights', *Proceedings of the Aristotelian Society*, 49, 1949, 171-194. Reprinted in *Logic and Language*, ed. A. Flew, Oxford, Blackwell, 1951, 145-166.
- Hart, H. L. A., *The Concept of Law*, Oxford, Oxford University Press, 1961.

- Johson, R. H., *Manifest Rationality*, Mahwah, N. J., Erlbaum, 2000.
- Joseph, H. W. B., *An Introduction to Logic*, Oxford, Clarendon Press, 1916.
- Krabbe, E. C. W., 'Appeal to Ignorance', *Fallacies: Classical and Contemporary Readings*, Hans V. Hansen and Robert C. Pinto (eds.), University Park, Pennsylvania State University Press, 1995, 251-264.
- Krabbe, E. C. W., 'Profiles of Dialogue', *JFAK: Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday*, ed. J. Gerbrandy, M. Marx, M. de Rijke and Y. Venema, Amsterdam, Amsterdam U Press, 1999, 25-36.
- Krabbe, E. C. W., 'The Problem of Retraction in Critical Discussion', *Synthese*, 127, 2001, 141-159.
- Krabbe, E. C. W., 'Review of Tindale (2007)', *Argumentation*, 23, 2009, 127-131.
- Ronald P. Loui, 'Hart's Critics on Defeasible Concepts and Ascriptivism', *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, New York, ACM Press, 1995, 21-30. Available at: <http://portal.acm.org/citation.cfm?id=222099>
- Mackenzie, J. D., 'The Dialectics of Logic,' *Logique et Analyse*, 94, 1981, 159-177.
- Mackenzie, J. D., 'Four Dialogue Systems,' *Studia Logica*, 49, 1990, 567-583.
- McBurney, P. and S. Parsons, 'Games That Agents Play: A Formal Framework for Dialogues Between Autonomous Agents', *Journal of Logic, Language and Information*, 11, 2002, 315-334.
- McCarthy, J. 'Applications of Circumscription to Formalizing Common Sense Knowledge', *Artificial Intelligence*, 28, 1986, 89-116.
- Park, R. C., D. P. Leonard and S. H. Goldberg, *Evidence Law*, West Group, St. Paul, Minnesota, 1998.
- Prakken, H., 'On Dialogue Systems with Speech Acts, Arguments and Counterarguments', *Proceedings of JELIA 2000, the European Workshop on Logic for Artificial Intelligence*, ed. Manuel Ojeda-Aciego, Imma P. de Guzman, Gerhard Brewka and Luis M. Pereira, Berlin, Springer, 2000, 224-238.
- Prakken, H., 'Formal Systems for Persuasion Dialogue', *The Knowledge Engineering Review*, 21, 2006, 163-188.
- Prakken H., and G. Sartor, 'A Dialectical Model of Assessing Conflicting Arguments in Legal Reasoning', *Artificial Intelligence and Law*, 4, 1996, 331-368.
- Prakken H., and G. Sartor, 'A Logical Analysis of Burdens of Proof', *Legal Evidence and Proof: Statistics, Stories, Logic*, ed. Hendrik Kaptein, Henry Prakken and Bart Verhiej, Farnham, Ashgate, 2009, 223-282.
- Reiter, R., 'A Logic for Default Reasoning', *Artificial Intelligence*, 13, 1980, 81-132.
- Reiter, R., 'Nonmonotonic Reasoning', *Annual Review of Computer Science*, 2, 1987, 147-186.
- Rissland, E. and K. Ashley, 'A Case-Based System for Trade Secrets law', *Proceedings of the First International Conference on Evidence and Law*, Boston, 1987, 60-66.
- Sergot, M., F. Sadri, R. A. Kowalski, F. Kriwaczek, P. Hammond and H. T. Cory, 'The British Nationality Act as a Logic Program', *Proceedings ZNP-83 Congress*, ed. G. Van Nevel and F. Balfroid, New York, Elsevier North-Holland, 29 (5), 1986, 370-386.
- Tindale, C. W., 'Fallacies, Blunders and Dialogue Shifts: Walton's Contributions to the Fallacy Debate', *Argumentation*, 11, 1997, 341-354.
- Van Eemeren, F. H. and R. Grootendorst, *Argumentation, Communication and Fallacies*, Hillsdale, N. J., Erlbaum, 1992.

- Van Eemeren, F. H. and P. Houtlosser, 'Strategic Maneuvering: A Synthetic Recapitulation', *Argumentation*, 20, 2006, 381-392.5
- Verheij, B., 'Dialectical Argumentation with Argumentation Schemes: An Approach to Legal Logic', *Artificial Intelligence and Law*, 11, 2003, 167-195.
- Vreeswijk, G. and H. Prakken, 'Credulous and Sceptical Argument Games for Preferred Semantics', *Proceedings of the 7th European Workshop on Logics in Artificial Intelligence*, Springer Lecture Notes in AI 1919, Springer Verlag, Berlin, 2000, 239-253.
- G. Vreeswijk and H. Prakken, 'Credulous and sceptical argument games for preferred semantics', *Proceedings of JELIA'2000, The 7th European Workshop on Logics in Artificial Intelligence*. Springer Lecture Notes in AI 1919, Springer Verlag, Berlin, 2000, 239-253.
- Walton, D., *A Pragmatic Theory of Fallacy*, Tuscaloosa, U of Alabama Press, 1995.
- Walton, D., *Arguments from Ignorance*, University Park, Pa., Penn State Press, 1996.
- Walton, D., 'Profiles of Dialogue for Arguments from Ignorance', *Argumentation*, 13, 1999, 53-71.
- Walton, D., 'Rethinking the Fallacy of Hasty Generalization', *Argumentation*, 13, 1999, 161-182.
- Walton, D., 'Poisoning the Well', *Argumentation*, 20, 2006, 273-307.
- Walton, D., 'Metadialogues for Resolving Burden of Proof Disputes', *Argumentation*, 21, 2007, 291-316.
- Walton D. and E. C. W. Krabbe, *Commitment in Dialogue*, Albany, State University of New York Press, 1995.
- Walton, D., C. Reed and F. Macagno, *Argumentation Schemes*, Cambridge, Cambridge University Press, 2008.
- Weinreb, L. L., *Legal Reason: The Use of Analogy in Legal Argument*, Cambridge, Cambridge University Press, 2005.
- Wigmore, J. H., *A Treatise on the Anglo-American System of Evidence in Trials at Common Law*, 3rd ed., vol. 1, Boston, Little, Brown and Company, 1940.
- Williams, G., 'The Evidential Burden: Some Common Misapprehensions', *New Law Journal*, Feb. 17, 1977, 156-158.
- Williams, A. R., 'Burdens and Standards in Civil Litigation', *Sydney Law Review*, 25, 2003, 165-188.
- Wyner, A., T. J. M. Bench-Capon and K. Atkinson, 'Arguments, Values and Baseballs: Representation of Popov v. Hayashi', *Legal Knowledge and Information Systems: JURIX 2007, The Twentieth International Conference*, ed. A. R. Lodder and L. Mommers, Amsterdam, IOS Press, 2007, 151-160.

Acknowledgments

Work on this paper was supported by a research grant from the Social Sciences and Humanities Research Council of Canada. The author would also like to thank the members of CRRAR for comments when an earlier version of the paper was read, and the two anonymous referees who provided many valuable suggestions for improvements.

Abstract

This paper argues that some traditional fallacies should be considered as reasonable arguments when used as part of a properly conducted dialog. It is shown that argumentation schemes, formal dialog models, and profiles of dialog are useful tools for studying properties of defeasible reasoning and fallacies. It is explained how defeasible reasoning of the most common sort can deteriorate into fallacious argumentation in some instances. Conditions are formulated that can be used as normative tools to judge whether a given defeasible argument is fallacious or not. It is shown that three leading violations of proper dialog standards for defeasible reasoning necessary to see how fallacies work are: (a) improper failure to retract a commitment, (b) failure of openness to defeat, and (c) illicit reversal of burden of proof.