ON THE ACCEPTABILITY OF ARGUMENTS
AND ITS FUNDAMENTAL ROLE IN
NONMONOTONIC REASONING, LOGIC PROGRAMMING
AND N-PERSONS GAMES

Phan Minh Dung
Division of Computer Science
Asian Institute of Technology
GPO Box 2754, Bangkok 10501, Thailand
E-mail:dung@cs.ait.ac.th

"The true basis of the logic of existence and universality lies in the human activities of seeking and finding"
Jaakko Hintikka [26,pp33]

Abstract

The purpose of this paper is to study the fundamental mechanism, humans use in argumentation, and to explore ways to implement this mechanism on computers.

We do so by first developing a theory for argumentation whose central notion is the acceptability of arguments. Then we argue for the "correctness" or "appropriateness" of our theory with two strong arguments. The first one shows that most of the major approaches to nonmonotonic reasoning in AI and logic programming are special forms of our theory of argumentation. The second argument illustrates how our theory can be used to investigate the logical structure of many practical problems. This argument is based on a result showing that our theory captures naturally the solutions of the theory of n-person games and of the well-known stable marriage problem.

By showing that argumentation can be viewed as a special form of logic programming with negation as failure, we introduce a general logic programming-based method for generating metainterpreters for argumentation systems, a method very much similar to the compiler-compiler idea in conventional programming.

The results in this paper (except those of part 2, and 3.3.2) have been published in condensed form in [10].

Keyword: Argumentation, nonmonotonic reasoning, logic programming, n-person games, the stable marriage problem.

1. Introduction

Argumentation constitutes a major component of human's intelligence. The ability to engage in arguments is essential for humans to understand new problems, to perform scientific reasoning, to express, clarify and defend their opinions in their daily lives. The way humans argue is based on a very simple principle which is summarized succinctly by an old saying "*The one who has the last word laughs best*". To illustrate this principle, let us take a look at an example, a mock argument between two persons I and A, whose countries are at war, about who is responsible for blocking negotiation in their region.

Example 1[1]

I: My government can not negotiate with your government because your government doesn't even recognize my government

A: Your government doesn't recognize my government either

The explicit content of the I's utterance is that the failure of A's government to recognize I's government blocks the negotiation. This establishes the responsibility of A's government for blocking the negotiation by an implicit appeal to the following commonsense interpretation rule:

> <u>Responsibility attribution</u>: If some actor performs some action which causes some state of affairs then that actor is responsible for that state of affairs unless its action was justified.

A uses the same kind of reasoning to counterargue that I's government is also responsible for blocking the negotiation as I's government doesn't recognize A's government either.

At this point, neither arguer can claim "victory" without hurting his own position. Consider the following continuation of the above arguments:

I: But your government is a terrorist government

This utterance justifies the failure of I's government to recognize A's government. Thus the responsibility attribution rule can not be applied to make I's government responsible for blocking the negotiation. So this represents an attack on the A's argument. If the exchange stops here, then I clearly has the "last word", which means that he has successfully argued that A's government is responsible for blocking the negotiation. ■

The goal of this paper is to give a scientific account of the basic principle "*The one who has the last word laughs best*" of argumentation, and to explore possible ways for implementing this

---

[1] This example is inspired by a similar example in [5].

principle on computers.

The problems of understanding argumentation and its role in human's reasoning have been addressed by many researchers in different fields including philosophy, logic and AI. Toulmin [58] has given an excellent philosophical account of the general structure of arguments. The relation between argumentation (in the form of a dialogue-game) and classical (monotonic) logic has been studied by Lorenz and Lorenzen [4] who have showed that classical first order logic can be viewed as dialogue-game logic where propositions are entities which can be either won or lost.

In AI, much work has been done to analyze the structure of arguments and to build computer systems which can engage in exchange of arguments. Argument systems which can understand editorials or engage in political dialogues have been built by Alvarado [1] and Birnbaum et all [5,6,22]. An in depth analysis of argument structure has been provided by Cohen [9]. These works can be considered as forming an heuristic approach to argument-based commonsense reasoning.

Roughly, the idea of argumentational reasoning is that a statement is believable if it can be argued successfully against attacking arguments. In other words, whether or not a rational agent believes in a statement depends on whether or not the argument supporting this statement can be successfully defended against the counterarguments. Thus, the beliefs of a rational agent are characterized by the relations between the "internal" arguments supporting his beliefs and the "external" arguments supporting contradictory beliefs. So, in certain sense, argumentational reasoning is based on the "external stability" of the accepted arguments. This is quite different and at the same time inherently related to the mainstream approaches to nonmonotonic reasoning in AI and logic programming [2,24,23,39,40,41,50,52] which are based on a kind of "internal stability" of beliefs[2]. These two kinds of "stability" are like the two sides of the same coin. Their relationship is very much similar to the relationship between Hintikka's game-theoretic semantics and Tarskian semantics of logic and natural language [4,26,53].

The understanding of the structure and acceptability of arguments is essential for a computer system to be able to engage in exchanges of arguments. Much work has been done to analyze the structure of arguments. Significant progress has been achieved here [58,9,1,5,6,22,37,45,46,60]. In contrast, it is still not clear how to understand the acceptability of arguments. The lack of progress here leaves the question about the semantical relations between argumentation and nonmonotonic reasoning remaining open until today. One of the goals of this paper is to provide an answer to these problems.

Moore distinguished between default reasoning and autoepistemic reasoning [40]. According to

[2]. A set of beliefs is "internally stable" if it can "reproduces" itself in a way which is solely determined by the set itself. In other words, its stability is totally determined by the "internal" relations between its elements.

him, default reasoning is drawing plausible inferences in the absence of information to the contrary while autoepistemic reasoning is like reasoning about one's own knowledge or beliefs. Thus default reasoning is like arguing with the Nature, where a conclusion, supported by some argument, can be drawn in the absence of any counterargument. On the other hand side, reasoning about one's own knowledge or beliefs is much like arguing with oneself. So both autoepistemic reasoning and default reasoning are two forms of argumentation. In fact, we will demonstrate in this paper that many of the major approaches to nonmonotonic reasoning in AI and logic programming are different forms of argumentation. This result should not be very surprising as it may seem since all forms of reasoning with incomplete information rest on the simple intuitive idea that a defeasible statement can be believed only in the absence of any evidence to the contrary which is very much like the principle of argumentation. In [11], this idea has been applied to develop a simple and intuitive framework for semantics of logic programming unifying many other previously proposed approaches [2,24,23,50]. Later, Kakas, Kowalski and Toni [28] have pointed out that the framework given in [11] is in fact an argumentational approach to logic programming. This important insight constitutes a major source of inspiration and motivation for this paper[3].

Argumentation is a major method humans use to justify their solutions to their social and economic problems. We demonstrate this by pointing out that many solutions to the n-person games modelling meaningful economic systems [42,15,56], are based on our theory of argumentation. Further, using the stable marriage problem as the benchmark, we show that our theory captures naturally the way humans argue to justify their solutions to many social problems. The results we gain here provides also a strong argument to defeat an often held opinion in AI and logic programming community that if a knowledge base has no stable semantics then there must be some "bug" in it.

Though argumentation is a powerful method for problem solving, it turns out that it can be "implemented" easily in logic programming. We demonstrate this by showing that argumentation can be viewed as logic programming with negation as failure. This result shows that logic programming is the perfect tool for implementing argumentation systems.

It seems necessary to point out again that our primary intention in this paper is Not to study the relationship between logic programming and nonmonotonic reasoning though much light is shed on this relationship from our result that both of them are forms of argumentation. Our main goal is to give an analysis of the nature of human's argumentation in its full generality. This is done in two steps. In the first step, a formal, abstract but simple theory of argumentation is developed to capture the notion of acceptability of arguments. In the next step, we demonstrate the "correctness" (or "appropriateness") of our theory. It is clear that the "correctness" of our theory can not be "proved" formally. The only way to accomplish this task is to provide relevant and

[3]Recently inspired by this paper, Bondarenko, Toni and Kowalski [7] have developed an argumentational assumption-based framework to nonmonotonic reasoning unifying many other approaches in a very interesting way.

convincing examples. Two "examples" are provided. The first one shows how our theory can be used to investigate the logical structure of many human's economic and social problems. The second one shows that many major approaches to nonmonotonic reasoning in AI and logic programming [52,40,41,45,46,11,24,50,23,57] are in fact different forms of our theory of argumentation.

This paper provides four novel results. The first one is a theory of acceptability of arguments which, in fact, is a formal account of the principle of argumentation. The second result shows the fundamental role our theory of argumentation can play in investigating the logical structure of many social and economic problems. The third result shows that logic programming as well as many major formalisms to nonmonotonic and defeasible reasoning in AI are argumentation systems. That means that all these systems are based on the same principle. They differ only by the structure of their arguments. The fourth result introduces a general method for implementing argumentation systems by showing that argumentation can be viewed as logic programming with negation as failure. This method is very much similar to the compiler-compiler idea in conventional programming.

List of Contents

The prerequisite-relation between the chapters is illustrated in the following tree:

part 1

part 2 part 3 part 4

Part 1: Acceptability of Arguments

1.1. Argumentation Frameworks

Our theory of argumentation is based on a notion of argumentation framework defined as a pair of a set of arguments, and a binary relation representing the attack-relationship between arguments. Here, an argument is an abstract entity whose role is solely determined by its relations to other arguments. No special attention is paid to the internal structure of the arguments.

Definition 1 An *argumentation framework* is a pair

$$AF = \langle AR, attacks \rangle$$

where AR is a set of arguments, and attacks is a binary relation on AR, i.e. attacks $\subseteq$ AR $\times$ AR. ■

For two arguments A,B, the meaning of attacks(A,B) is that A represents an attack against B.

Example 2 (continuation of example 1). The exchange between I and A can be represented by an argumentation framework <AR,attacks> as follows: AR = $\{i_1, i_2, a\}$ and attacks = $\{(i_1,a),(a,i_1),(i_2,a)\}$ with i_1,i_2 denoting the first and the second argument of I, respectively, and a denoting the argument of A. ■

Remark 1 From now on, if not explicitly mentioned otherwise, we always refer to an arbitrary but fixed argumentation framework AF = <AR,attacks>. Further, we say that A attacks B (or B is attacked by A) if attacks(A,B) holds. Similarly, we say that a set S of arguments attacks B (or B is attacked by S) if B is attacked by an argument in S.

Definition 2 A set S of arguments is said to be *conflict-free* if there are no arguments A,B in S such that A attacks B. ■

For a rational agent G, an argument A is acceptable if G can defend A (from within his world) against all attacks on A. Further, it is reasonable to assume that a rational agent accepts an argument only if it is acceptable. That means that the set of all arguments accepted by a rational agent is a set of arguments which can defend itself against all attacks on it. This leads to the following definition of an admissible (for a rational agent) set of arguments.

Definition 3 (1) An argument A $\in$ AR is *acceptable* with respect to a set S of arguments iff for each argument B $\in$ AR: if B attacks A then B is attacked by S.[4]

(2) A conflict-free set of arguments S is *admissible* iff each argument in S is acceptable wrt S. ■

The (credulous) semantics of an argumentation framework is defined by the notion of preferred

[4]See remark 1.

extension.

Definition 4 A *preferred extension* of an argumentation framework AF is a maximal (wrt set inclusion) admissible set of AF. ■

Example 3 (Continuation of example 2). It is not difficult to see that AF has exactly one preferred extension E = $\{i_1,i_2\}$. ■

Example 4 (Nixon Diamond). The well-known Nixon diamond example can be represented as an argumentation framework AF = <AR,attacks> with AR = {A,B}, and attacks = {(A,B),(B,A)} where A represents the argument "Nixon is anti-pacifist since he is a republican", and B represents the argument "Nixon is a pacifist since he is a quaker". This argumentation framework has two preferred extensions, one in which Nixon is a pacifist and one in which Nixon is a quaker. ■

Lemma 1 (Fundamental Lemma)
Let S be an admissible set of arguments and A,A' be arguments which are acceptable wrt S. Then

1) S' = S $\cup$ {A} is admissible, and
2) A' is acceptable wrt S'.

Proof 1) We need only to show that S' is conflict-free. Assume the contrary. Therefore, there exists an argument B $\in$ S s.t. either A attacks B or B attacks A. From the admissibility of S and the acceptability of A, there is an argument B' in S such that B' attacks B or B' attacks A. Since S is conflict-free, it follows that B' attacks A. But then there is an argument B" in S s.t. B" attacks B'. Contradiction !!
2) Obvious. ■

The following theorem follows directly from the fundamental lemma.

Theorem 1 Let AF be an argumentation framework.

(1) The set of all admissible sets of AF form a complete partial order wrt set inclusion.
(2) For each admissible set S of AF, there exists an preferred extension E of AF such that S $\subseteq$ E. ■

Theorem 1 together with the fact that the empty set is always admissible, implies the following corollary:

Corollary 2 Every argumentation framework possesses at least one preferred extension. ■

Hence, preferred extension semantics is always defined for any argumentation framework.

Stable Semantics for Argumentation

To compare our approach with other approaches, we introduce in the following the notion of stable extension.

Definition 5 A conflict-free set of arguments S is called a *stable extension* iff S attacks each argument which does not belong to S. ■

In part 2, we will show that in the context of game theory, our notion of stable extension coincides with the notion of stable solutions of n-person games introduced by Von Neuman and Morgenstern fifty years ago [42].

It is easy to see that

Lemma 3 S is a stable extension iff S = { A | A is not attacked by S}. ■

It will turns out later (part 3) that this proposition underlines exactly the way the notions of stable models in logic programming, extensions in Reiter's default logic, and stable expansion in Moore's autoepistemic logic are defined.

The relations between stable extension and preferred extension are clarified in the following lemma.

Lemma 4 Every stable extension is a preferred extension, but not vice versa.

Proof It is clear that each stable extension is a preferred extension. To show that the reverse does not hold, we construct the following argumentation framework: Let AF = (AR,attacks) with AR = {A} and attacks = {(A,A)}. It is clear that the empty set is a preferred extension of AF which is clearly not stable. ■

It is not difficult to see that in the above examples, preferred extension and stable extension semantics coincide.

Though stable semantics is not defined for every argumentation system, an often asked question is *whether or not argumentation systems with no stable extensions represent meaningful systems ???* In part (2) of this paper, we will provide meaningful argumentation systems without stable semantics, and thus provide an definite answer to this question.

1.2. Fixpoint Semantics and Grounded (Skeptical) Semantics

We show in this chapter that argumentation can be characterized by a fixpoint theory providing an elegant way to introduce grounded (skeptical) semantics.

Definition 6 The *characteristic function*, denoted by F_{AF}, of an argumentation framework AF = <AR,attacks> is defined as follows:

$$F_{AF}: 2^{AR} \rightarrow 2^{AR}$$

$$F_{AF}(S) = \{ A \mid A \text{ is acceptable wrt } S \} \quad \blacksquare$$

Remark 2 As we always refer to an arbitrary but fixed argumentation framework AF, we often write F instead of F_{AF} for short.

Lemma 5 A conflict-free set S of arguments is admissible iff $S \subseteq F(S)$.

Proof The lemma follows immediately from the property "If S is conflict-free then F(S) is also conflict-free". So we need only to prove this property. Assume that there are A,A' in F(S) such that A attacks A'. Thus, there exists B is S such that B attacks A. Hence there is B' in S such that B' attacks B. Contradiction !! So F(S) is conflict free. ■

It is easy to see that if an argument A is acceptable wrt S then A is also acceptable wrt any superset of S. Thus, it follows immediately that

Lemma 6 F_{AF} is monotonic (wrt set inclusion). ■

The skeptical semantic of argumentation frameworks is defined by the notion of grounded extension introduced in the following.

Definition 7 The *grounded extension* of an argumentation framework AF, denoted by GE_{AF}, is the least fixed point of F_{AF}. ■

Example 5 (Continuation of example 3). It is easy to see: $F_{AF}(\phi) = \{i_2\}$, $F_{AF}{}^2(\phi) = \{i_1,i_2\}$, $F_{AF}{}^3(\phi) = F_{AF}{}^2(\phi)$. Thus $GE_{AF} = \{i_1,i_2\}$. Note that GE_{AF} is also the only preferred extension of AF. ■

Example 6 (Continuation of the Nixon-example). From AF = <AR,attacks> with AR = {A,B}, and attacks = {(A,B),(B,A)}, it follows immediately that the grounded extension is empty, i.e. a skeptical reasoner will not conclude anything. ■

The following notion of complete extension provides the link between preferred extensions (credulous semantics), and grounded extension (skeptical semantics).

Definition 8 An admissible set S of arguments is called a *complete extension* iff each argument which is acceptable wrt S, belongs to S. ■

Intuitively, the notion of complete extensions captures the kind of confident rational agents who believes in every thing he can defend.

Lemma 7 A conflict-free set of arguments E is a complete extension iff E = $F_{AF}(E)$. ■

The relations between preferred extensions, grounded extensions and complete extensions is given in the following theorem.

Theorem 2 (1) Each preferred extension is a complete extension, but not vice versa.
(2) The grounded extension is the least (wrt set inclusion) complete extension.
(3) The complete extensions form a complete semilattice[5] wrt set inclusion.

Proof (1) It is obvious from the fixpoint definition of complete extensions that every preferred extension is a complete extension. The Nixon diamond example provides a counter example that the reverse does not hold since the empty set is a complete extension but not a preferred one.
(2) Obvious
(3) Let SE be a nonempty set of complete extensions. Let LB = { E | E is admissible and E $\subseteq$ E' for each E' in SE }. It is clear that GE $\in$ LB. So LB is not empty. Let S = $\cup$\{ E | E $\in$ LB\}. It is clear that S is admissible, i.e. S $\subseteq$ F(S). Let E = lub(F^i(S)) for ordinals i. Then it is clear that E is a complete extension and E $\in$ LB. Thus E = S. So E is the glb of SE. ■

Remark In general, the intersection of all preferred extensions does not coincide with the grounded extension.

In general, F_{AF} is not continuous, but if the argumentation framework is finitary then it is.

Definition 9 An argumentation framework AF = <AR,attacks> is *finitary* iff for each argument A, there are only finitely many arguments in AR which attack A. ■

Lemma 8 If AF is finitary then F_{AF} is ω-continuous.

Proof Let $S_0 \subseteq \ldots \subseteq S_n \subseteq ..$ be an increasing sequence of sets of arguments, and let $S = S_0 \cup \ldots \cup S_n \cup \ldots$ Let $A \in F_{AF}(S)$. Since there are only finitely many arguments which attack A, there exists a number m s.t. $A \in F_{AF}(S_m)$. Therefore, $F_{AF}(S) = F_{AF}(S_0) \cup \ldots \cup F_{AF}(S_n) \cup \ldots$. ■

An example of a non-finitary argumentation framework is given in appendix A.

1.3. Sufficient Conditions for Coincidence between Different Semantics

Well-Founded Argumentation Frameworks

We want to give in this paragraph a sufficient condition for the coincidence between the grounded semantics and preferred extension semantics as well as stable semantics.

---

[5] A partial order (S,≤) is a complete semilattice iff each nonempty subset of S has a glb and each increasing sequence of S has a lub.

Definition 10 An argumentation framework is *well-founded* iff there exists no infinite sequence $A_0,A_1,\ldots,A_n,\ldots$ such that for each i, A_{i+1} attacks A_i. ■

The following theorem shows that well-founded argumentation frameworks have exactly one extension.

Theorem 3 Every well-founded argumentation framework has exactly one complete extension which is grounded, preferred and stable.

Proof Assume the contrary, i.e. there exist a well-founded argumentation framework whose grounded extension is not a stable extension. Let AF=(AR,attacks) be such a argumentation framework such that $S = \{ A \mid A \in AR \backslash GE_{AF}$ and A is not attacked by $GE_{AF} \}$ is nonempty. Now we want to show that each argument A in S is attacked by S itself. Let $A \in S$. Since A is not acceptable wrt GE_{AF}, there is an attack B against A s.t. B is not attacked by GE_{AF}. From the definition of S, it is clear that B does not belong to GE_{AF}. Hence, B belongs to S. Thus there exists an infinite sequence $A_1,A_2,\ldots$ such that for each i, A_{i+1} attacks A_i. Contradiction !! ■

Coherent Argumentation Frameworks

Now, we want to give a condition for the coincidence between stable extensions and preferred extensions. In general, the existence of a preferred extension which is not stable indicates the existence of some "anomalies" in the corresponding argumentation framework[6]. For example, the argumentation framework $<\{A\},\{(A,A)\}>$[7] has an empty preferred extension which is not stable. So it is interesting to find sufficient conditions to avoid such anomalies.

Definition 11 1) An argumentation framework AF is said to be *coherent* if each preferred extension of AF is stable.

2) We say that an argumentation framework AF is *relatively grounded* if its grounded extension coincides with the intersection of all preferred extensions. ■

It follows directly from the definition that there exists at least one stable extension in a coherent argumentation framework.

Imagine an exchange of arguments between you and me about some proposition C. You start by putting forward an argument A_0 supporting C. I don't agree with C, and so I present an argument A_1 attacking your argument A_0. To defend A_0 and so C, you put forward another argument A_2 attacking my argument A_1. Now I present A_3 attacking A_2. If we stop at this point, A_0 is defeated. It is clear that A_3 plays a decisive role in the defeat of A_0 though A_3 does not directly attack A_0.

---

[6]The existence of "anomalies" does not mean that something is wrong in the concerned argumentation frameworks.

[7]The argumentation framework corresponding to the logic program $p \leftarrow not\ p$ is of this kind.

A_3 is said to represent an indirect attack against A_0. In general, we say that an argument B *indirectly attacks* A if there exists a finite sequence $A_0,..,A_{2n+1}$ such that 1) A = A_0 and B = A_{2n+1}, and 2) for each i, $0 \le i \le 2n$, A_{i+1} attacks A_i. We say that an argument B *indirectly defends* A if there exists a finite sequence $A_0,..,A_{2n}$ such that 1) A = A_0 and B = A_{2n}, and 2) for each i, $0 \le i < 2n$, A_{i+1} attacks A_i. An argument B is said to be *controversial* wrt A if B indirectly attacks A and indirectly defends A. An argument is *controversial* if it is controversial wrt some argument A.

Definition 12 1) An argumentation framework is *uncontroversial* if none of its arguments is controversial.

2) An argumentation framework is *limited controversial* if there exists no infinite sequence of arguments $A_0,...,A_n,...$ such that A_{i+1} is controversial wrt A_i. ■

It is clear that every uncontroversial argumentation framework is limited controversial but not vice versa.

Theorem 4 1) Every limited controversial argumentation framework is coherent.

2) Every uncontroversial argumentation framework is coherent and relatively grounded.

Proof 1) Assume that there exists an limited controversial argumentation framework AF which is not coherent. Let E a preferred extension of AF which is not stable. Let define: AR' = { A' | A' ∈ AR\E and A' is not attacked by E}. It is clear that AR' is nonempty. Let attacks' be the restriction of attacks on AR'. Let AF' = (AR',attacks'). From lemma 9, there exists a nonempty complete extension E' of AF'. It is easy to see that E ∪ E' is again an extension of AF. Contradiction !!

2) follows immediately from the following lemmas 9,10. ■

For the proof of lemmas 9,10, we need a couple of new notations. An argument A is said to be a *threat* to a set of argument S if A attacks S and A is not attacked by S. A set of arguments D is called a *defense* of a set of argument S if D attacks each threat to S.

Lemma 9 Let AF be a limited controversial argumentation framework. Then there exists at least a nonempty complete extension E of AF.

Proof If the grounded extension of AF is not empty, then the lemma is proved. Suppose now that the grounded extension of AF is empty. Therefore, it follows immediately that each argument A in AF is attacked by some other argument (otherwise, the grounded extension would not be empty). Let A be an argument s.t. there exists no B s.t. B is controversial wrt A. The existence of such an argument is clearly guaranteed by the limited controversity of AF. Define E_0 = {A}. For each natural number i>0, define the set E_i as follows: $E_i = E_{i-1} \cup D_{i-1}$ where D_{i-1} is a minimal (wrt set inclusion) defense of E_{i-1}. Now we prove by induction that for each i:

(*) E_i is conflict-free, and each argument B ∈ E_i indirectly defends A.

It is clear that this holds for i=0. Let i>0, and assume that (*) holds for each i-1. From the fact that each argument in AF is attacked by some other argument, it is clear that there exists a minimal defense D_{i-1} of E_{i-1}. From the induction hypothesis that each argument in E_{i-1} indirectly defends A, it is not difficult to see that all arguments in D_{i-1} indirectly defend A, too. Thus from the induction hypothesis, each argument in E_i indirectly defends A. Assume now that E_i is not conflict-free. Thus there exist two arguments B,B' in E_i s.t. B attacks B'. Since each argument in E_i indirectly defends A, B is clearly controversial wrt A. Contradiction !! So E_i is conflict-free.

Let $F = \cup_i E_i$. It is clear that F is admissible. Let define E to be the least complete extension containing F. Hence E is the desired extension. ■

Lemma 10 Let AF be an uncontroversial argumentation framework, and A be an argument such that A is not attacked by the grounded extension GE of AF and $A \notin GE$. Then

1) There exists a complete extension E_1 such that $A \in E_1$, and
2) There exists a complete extension E_2 such that E_2 attacks A.

Proof Let AR'= {A' | A'∈ AR\GE and A' is not attacked by GE}. Hence A ∈ AR'. Thus AR' is not empty. Let attacks' be the restriction of attacks on AR'. Let AF'= (AR',attacks').

1) Similar to the proof of lemma 9, we can show that there exists a complete extension E0 of AF' s.t. A ∈ E0. Let $E_1 = GE \cup E0$. It is clear that E_1 is the desired extension.

2) Since A is attacked by some argument in AR', there exists B ∈ AR' s.t. B attacks A. So there exists a complete extension E1 of AF' s.t. B ∈ E1. Hence $E_2 = GE \cup E1$ is the desired extension. ■

Corollary 11 Every limited controversial argumentation framework possesses at least one stable extension. ■

This corollary in fact gives the answer to an often asked question about the existence of stable semantics of knowledge representation formalism like Reiter's default logic, logic programming or autoepistemic logic. Much works have been done to study this kind of questions [12,35,25,18,54,16]. The uncontroversity of argumentation frameworks is a generalization of the results given in these works.

Part II: Argumentation, N-Persons Games and Stable Marriage Problem

In the next two chapters, we will demonstrate the "correctness" of our theory of argumentation through two examples in which we show how our theory can be used to investigate the logical structure of the solutions to many practical problems.

2.1. Argumentation in N-Persons Games

In the theory of n-person game developed by Von Neuman and Morgenstern [42], a social economy is viewed as a game whose players are the major forces of the economy. Like a program, a game has two aspects: the operational aspects concerning the question: *How to play ??*, and the specification aspects concerning the question: *What is the payoff ??*

Classical game theory as presented in [42,56,15] is mostly concerned with the specification aspects of the games. In other words, the theory of n-person games is a theory about the possible payoffs to the players of the game. The central notion of the theory of n-person game is the notion of solution of a game which is a set of payoff vectors called imputations, to its participants. Formally, an imputation of a n-person game is defined as a vector $(p_1,...,p_n)$ of numbers giving the utilities each player gets after the game. Hence in considering a social economy as a n-person game, imputations model the ways the wealth is distributed in an economy. The distribution of wealth in a stable economy does not consist of a rigid system of apportionment, i.e. of imputation, but a variety of alternatives that though following certain commonsense principles, nevertheless differ among themself in many particular aspects. Such a system of imputation describes the "established order of the society" or the "accepted standard of behavior" [42].

Formally, a cooperative n-person game (in normal form) is defined by a characteristic function V which associated with each coalition a number determining the minimum amount that coalition can obtain if all its members join together and play as a team. The only condition imposed on the characteristic function is the superaddivity which says that for each two disjoint coalitions A,B, if A and B join together, they will get more than staying independent, i.e. $V(A \cup B) \geq V(A) + V(B)$. The stability of a coalition is determined fully by the amount each of its member can get. So if any of the member of a coalition can get more in another coalition then he will defect thus causing a new imputation of the game. This is modelled by the notion of *domination* between imputations. An imputation $(p_1,...,p_n)$ is said to *dominate* another imputation $(q_1,...,q_n)$ if there is a (nonempty) coalition $K \subseteq \{1,...,n\}$ such that for each $i \in K$, $p_i > q_i$ and $p_{i1} + ... + p_{ik} \leq V(K)$ where $K = \{i1,..,ik\}$. Von Neuman and Morgenstein [42] define a *solution* of a cooperative n-person game, referred to as NM-solution, as a set of imputations satisfying the following two postulates (NM1),(NM2).

(NM1) No s in S is dominated by an s' in S

(NM2) Every s not contained in S is dominated by some s' contained in S

The first postulate expresses the condition that the "established order of the society" represented by S is free from inner contradiction. The second postulate expresses the fact that any attempt to build a coalition to impose a new imputation $s \notin S$ will be blocked by some imputation $s' \in S$ which dominates s. In other words, it is not possible to deviate from the "established order of the society". That means that every thing has to conform to this "established order". It turns out that this "extremist standpoint" of NM-solution is the cause for the nonexistence of NM-solution to many meaningful economic systems.

To illustrate the intuition behind NM-solutions, let us consider the following example taken from [15]. Suppose that P_1,P_2,P_3 are players in a three-person game in which any coalition with either two or three players can get 2 units of wealth and a player alone get nothing. This game has infinitely many solutions. We will look at two of them. The first solution S_1 consists of three imputations $s_1 = \{P_1{:}1,P_2{:}1,P_3{:}0\}$, $s_2 = \{P_1{:}1,P_2{:}0,P_3{:}1\}$, $s_3 = \{P_1{:}0,P_2{:}1,P_3{:}1\}$. Let us check that this solution really satisfies the postulates NM1,NM2. It is clear that S_1 satisfies the first postulate. Let s now be an arbitrary imputation $\{P_1{:}v_1,P_2{:}v_2,P_3{:}v_3\} \notin S_1$. Then it is easy to see that there exists $\{P_i,P_j\}$ with $i \neq j$ such that $v_i + v_j < 2$ and $\max(v_i,v_j) < 1$ (otherwise s would belong to S_1). Without loss of generality, we can assume that i=1 and j=2. It is easy to see that s is dominated by s_1. The "established order" characterized by this solution dictates that a bigger coalition is not tolerated if the same result can be achieved with a smaller one, and the participants in a coalition are treated equally. In the second solution - a discriminatory solution - two players join, give the third player something less than his "fair share" of 2/3 and take the rest for themself. This is similar to what happens in an apartheid society.

It is not difficult to see that the argument for the building a coalition K is the payoff for each of its participants. Thus each imputation represents an argument for building some coalition. So the set of imputations together with the domination relation between them forms an argumentation framework. It is obvious that the following theorem holds:

Theorem 5 Let IMP be the set of imputations of some cooperative n-person game G and dom = {(s,s') | s dominates s' }. Then each NM-solution of the game G is a stable extension of (IMP,dom) interpreted as an argumentation framework, and vice versa. ■

Von Neuman and Morgenstein believed that each cooperative n-person game possesses at least one NM-solution. "There can be, of course, no concession as regards existence. If it should turn out that our requirements concerning a solution S are, in any special case, unfulfillable - this would necessitate a fundamental change in the theory. Thus a general proof of the existence of solutions for all particular cases is most desirable. It will appear from our subsequent investigations that this proof has not yet been carried out in full generality but that in all cases considered so far solutions were found" [42]. Twenty years later, F.W. Lucas constructed a ten-person game which has no NM-solution [56]. Later, Shubik [56] has pointed out that despite having no NM-solution, Lucas's games model meaningful economic systems. The conclusion here is this:

Stable extensions do not capture the intuitive semantics of every meaningful argumentation system.

We will come back to this point again in the next paragraph.

As preferred extensions exist for every argumentation framework, we can introduce the *preferred solutions* to n-person games by defining them as the preferred extensions of the corresponding argumentation system (IMP,dom). The new solutions satisfy both conditions of a rational standard behavior: freeness from inner contradiction and the ability to withstand any attack from outside. This is clearly a contribution to the theory of n-person games.

Another notion of solution of an n-person game is *the core* defined as the set of imputations whose members are not dominated by any other imputation [56]. It is not difficult to see that

Theorem 6 Let IMP be the set of imputations of some n-person game G and dom be the corresponding domination relation between them. Then the core of G coincides with $F(\phi)$ where F is the characteristic function of (IMP,dom) interpreted as an argumentation framework. ■

2.2. Argumentation and The Stable Marriage Problem[8]

Given two sets M,W of n men and n women respectively. The stable marriage problem (SMP) is the problem of finding a way to arrange the marriage for the men and women in M,W where it is assumed that all the men and women in M,W have expressed mutual preference (each man must say how he feels about each woman and vice versa)[9]. The marriages have to be stable in the sense that if for example A is married to B then all those whom A prefers to B must be married to someone whom they prefer to A. Formally, a solution to the SMP is a one-one correspondence $S: M \rightarrow W$ such that there exists no pair $(m,w) \in M \times W$ such that m prefers w to S(m) and w prefers m to $S^{-1}(w)$.

The SMP can be formalized as the task of finding a stable extension of an argumentation framework AF = (AR,attacks) as follows:

It is clear that D represents a threat to a marriage (A,B) if A prefers D to B. In other words, a hypothetical marriage of A to D poses an attack to (A,B). But this attack is eliminated if D is married to some one whom D prefers to A. Let

$AR = M \times W$

attacks $\subseteq AR \times AR$: (C,D) attacks (A,B) iff 1) A = C and A prefers D to B, or 2) D = B and B prefers C to A

Theorem 7 A set $S \subseteq AR$ constitutes a solution to the SMP iff S is a stable extension of the corresponding argumentation framework.

Proof "=>" Let S be a solution of the SMP. Since S is a one-one correspondence between M,W, it is clear that S is conflict-free. Let $(m,w) \notin S$. Then from the definition of S, either m prefers

---

[8]Mathematically, the stable marriage problem is a special case of the graph matching problem which has been studied extensively in the literature due to its wide applicability. For example, in the USA, a quite complicated system has been set up to place graduating medical students into hospital residency positions. Each student lists several hospital in order of preference and each hospital lists several students in order of preference. The problem is to assign the students to positions in a fair way respecting all the stated preferences [55].

[9]That means that associated with each person is a strictly ordered preference list containing all members of the opposite sex.

S(m) to w or w prefers $S^{-1}(w)$ to m. Hence, (m,w) is attacked by at least one element from $\{(m,S(m)),(S^{-1}(w),w)\} \subseteq S$.
"<=" Let S be a stable extension of AF. From the definition that S is conflict-free, it is clear that S,S^{-1} are partial functions from M into W and from W into M respectively. Assume now that S is not a total function from M into W. Then it is clear that there exists $(m,w) \in AR\backslash S$ such that both S(m) and $S^{-1}(w)$ are undefined. Therefore, (m,w) is not attacked by S which is a contradiction. Hence both S,S^{-1} are total functions, i.e. S is a one-one correspondence between M,W. Further it is also easy to see that it follows directly from the stability of S that there exists no pair $(m,w) \in M \times W$ such that m prefers w to S(m) and w prefers m to $S^{-1}(w)$.
■

To demonstrate once more that there are practically relevant argumentation systems which have no stable semantics, in the following we introduce the Stable Marriage Problem with Gays (SMPG) which is a modification of the SMP in which individuals of the same sex can be married to each other. The condition for the stability of a marriage is defined as in the SMP. The problem now is finding a way to arrange the marriage for a maximal numbers of persons. In contrast to the SMP, the SMPG corresponds to the problem of finding a preferred extension in an argumentation framework AF = (AR,attacks) with $AR = P \times P$ where P is the set of persons involved and attacks is defined as in the SMP. The following example shows that in general, the argumentation framework corresponding to a SMPG has no stable semantics.

Let P = {m,w,p1,p2,p3} where m is a man, w is a woman. For short we say that x loves y if x prefers y to all others. Suppose that m and w are in love with each other. Further suppose that there is a love triangle between p1,p2,p3 as follows: p1 loves p2, p2 loves p3 and p3 loves p1. So it is not difficult to see that there is no way to arrange a stable marriage for any among p1,p2,p3. The only stable marriage is between m and w. Indeed, the corresponding argumentation framework has exactly one preferred extension containing only the pair (m,w). It is clear that there is nothing wrong in the above argumentation framework. If something is "wrong" then it is the problem, i.e. the world we are trying to model is somehow "wrong". But it is "normal" that there are lots of things which are "wrong" in some ways in the world around us. So it is natural to expect that any knowledge system representing this world may not have a stable semantics. Further, due to the result that nonmonotonic reasoning and logic programming are different forms of argumentation, and an argumentation system itself can be transformed into an equivalent logic program (see coming parts), the conclusion we draw here can be formulated as follows:

Let P be a knowledge base represented either as a logic program, or as a nonmonotonic theory or as an argumentation framework. Then there is not necessarily a "bug" in P if P has no stable semantics.

This theorem defeats an often held opinion in logic programming and nonmonotonic reasoning community that if a logic program or a nonmonotonic theory has no stable semantics then there is something "wrong" in it.

Though it has been recognized earlier in [41a] that the stable marriage problem can be viewed as a nonmonotonic reasoning problem, argumentation presents a direct and more natural representation and analysis of this problem.

Part 3: Nonmonotonic Reasoning and Logic Programming as Argumentation

A number of different approaches to nonmonotonic reasoning has been proposed in AI [52,40,41,45,46,47,48,57] which are very different at the first look. But it turns out that all of them are different forms of argumentation. Due to the lack of space, we only show in this chapter, that two of them, the Reiter's default logic, as representative of the extension-based approach [52,40,41,47,48], and the Pollock's inductive defeasible logic, as representative of the argument-based approach [45,46,57], are different forms of argumentation. This clarifies the relationship between these two approaches to nonmonotonic reasoning, a problem which has been open until today. Further we also show that logic programming is a form of argumentation, too. Readers who are interested in more details about the relations between argumentation and nonmonotonic reasoning are referred to a recent paper of Bondarenko, Toni and Kowalski [7] who, generalizing the results given in this chapter, have developed an interesting assumption-based framework to nonmonotonic reasoning unifying many other previously proposed formalism.

3.1. Reiter's Default Logic As Argumentation

A default is an expression of the form $(p:j_1,...,j_k/w)$ where $p,j_1,..,j_k,w$ are closed first order sentences with p being called the pre-requisite, $j_1,...,j_k$ the justifications and w the conclusion of the default.

A default theory is a pair (D,W) where D is a set of defaults and W is a set of closed first order sentences. A default theory is said to be consistent if W is consistent [52]. A Reiter's extension (or R-extension for short) [52] of a default theory (D,W) is a closed first order theory E satisfying the following conditions:

$$E = \cup\{ W_i \mid i \text{ is a natural number } \}$$

where

$$W_0 = W$$

$$W_{i+1} = Th(W_i) \cup \{ w \mid \exists(p:j_1,..,j_k/w) \text{ in } D \text{ s.t. } \{j_n\} \cup E \text{ is consistent for } k \geq n \geq 1, \text{ and } p \in W_i \}$$

with $Th(W_i)$ denoting the first order closure of the theory W_i.

Let S be a set of defaults. The set of all justifications of defaults in S is denoted by Jus(S).

Let T = (D,W) and $K = \{j_1,..,j_m\} \subseteq Jus(D)$. A closed wff k is said to be a *defeasible consequence* *of T,K* if there is a sequence $(e_0,e_1,...,e_n)$ with $e_n = k$ such that for each e_i, either $e_i \in W$ or e_i is a logical consequence of the preceding members in the sequence or e_i is the conclusion w of a default $(p:j_1',..,j_r'/w)$ whose pre-requisite p is a preceding member in the sequence and whose justifications $j_1',..,j_r'$ belong to K. K is said to be a *support for k wrt T*.

A default theory T = (D,W) can be interpreted as an argumentation framework AF(T) = $\langle AR_T, attacks_T \rangle$ as follows:

$AR_T = \{ (K,k) \mid K \subseteq jus(D)$: K is a support for k wrt T$\}$

$(K,k)\ attacks_T\ (K',k')$ iff $\neg k \in K'$

The following lemma shows that the argumentation framework AF(T) is a "meaningful" one.

Lemma 12 Let S be an admissible set of arguments in AF(T). Let $H = \cup\{ K \mid (K,k) \in S \}$. Then $T,H \nvdash$ false iff T is consistent.

Proof "=>" Obvious
"<=" Assume the contrary. Thus there is a finite nonempty subset K of H s.t. $T,K \vdash$ false. Thus for each closed wff k, $(K,k) \in AR_T$. Let $(K',k') \in S$ such that K' is not empty. So K represents an attack against (K',k'). From the admissibility of S, there is A = (H',h') in S s.t. $\neg h' \in K$. That means that A attacks some argument B in S. Hence S is not conflict-free. Contradiction !! ■

The correspondence between the R-extensions of a default theory T and the stable extensions of the corresponding argumentation framework AF(T) is captured by the following mapping:

Definition 13 Let S be a first order theory and S' be a set of arguments of AF(T). Define

$arg(S) = \{ (K,k) \in AR_T \mid \forall j \in K: \{j\} \cup S$ is consistent $\}$

$flat(S') = \{ k \mid \exists (K,k) \in S' \}$. ■

From the definition of R-extension, it is not difficult to see that the following lemma holds:

Lemma 13 Let T be a default theory, and E be a first order theory. Then E is a R-extension of T iff E = flat(arg(E)). ■

It follows directly

Theorem 8 Let T = (D,W) be a default theory. Let E be a R-extension of T and E' be a stable extension of AF(T). Then

(1) arg(E) is a stable extension of AF(T)

(2) flat(E') is a R-extension of T

Proof 1) Let A = (H,h) be an argument in AF(T). Then it is easy to see that A is not attacked by arg(E) iff for each $(K,k) \in$ arg(E): $\neg k \notin H$ iff $\forall k \in E$: $\neg k \notin H$ (from lemma 13) iff $\forall j \in H$: $E \cup \{j\}$ is consistent iff $A \in$ arg(E). So from the lemma 3, it follows immediately that arg(E) is a stable extension of AF(T).

2) It is easy to see that for each argument (K,k): $(K,k) \in$ arg(flat(E')) iff $\forall j \in K$: $\{j\} \cup$ flat(E') is consistent iff (K,k) is not attacked by E' iff $(K,k) \in$ E' (from the fact that E' is a stable extension of AF(T) and lemma 3). Hence E'= arg(flat(E')). So flat(E') = flat(arg(flat(E'))). From lemma 13, it follows that flat(E') is an R-extension of T. ■

It is obvious that the preferred extension semantics of AF(T) generalizes the R-extension semantics of T. Moreover, we argue that preferred extension semantics of AF(T) captures in a more natural way the intuition of a default $(p:j_1,..,j_k/w)$ which says that in the absence of any evidence to the contrary of the justifications $j_1,..,j_k$, concludes w if p holds. It is clear that this intuitive understanding of defaults does not say that the existence of such a "paradox" default like $(:\neg p/p)$ can prevent us from concluding q in the default theory $T=(\{(:\neg p/p)\},\{q\})$. But R-extension semantics does exactly that while preferred extension semantics not. Supporters of R-extension may argue that T in this case has a bug and we have to fix it before conclude something from T. How can we fix it ? How can we know that the bug is in D and not in W ? We know it thanks to the preferred extension of AF(T) !! Further, interpreting a default theory T as a shorthand of its corresponding argumentation framework AF(T) makes it also possible to introduce a skeptical semantics to Reiter's default logic, thus building a bridge to other skeptical approaches to nonmonotonic reasoning.

3.2. Pollock's Inductive Defeasible Logic as Grounded Argumentation

Starting from the ideas of prima facie reasons and defeaters in philosophy, Pollock [45,46] has constructed a theory for defeasible reasoning that is based on the relations between arguments supporting contradictory conclusions. Pollock's work is one of the most general and influential approaches to defeasible reasoning which deviate from the mainstream approaches to nonmonotonic reasoning in AI. In this paragraph, we will show that Pollock's inductive theory of defeasible reasoning is based on our notion of grounded extension. A byproduct of this result is the illumination of the inherent relations between argument-based [45,46,57] and extension-based [41,40,52,24,23] nonmonotonic reasoning in AI.

Given an argumentation framework <AR,attacks>, Pollock's theory of defeasible reasoning is based on a hierarchy of arguments defined as follows:

All arguments are level 0 arguments

An argument is a level n+1 argument iff it is not attacked by any level n argument

Definition 14 An argument is *indefeasible* iff there is an m such that for each n>m, the argument is a level n argument. ■

Let AR_i denote the set of level i arguments. It is clear that for each i, $AR_i = Pl_{AF}(AR_{i-1})$ where

$Pl_{AF}: 2^{AR} \rightarrow 2^{AR}$

$Pl_{AF}(S) = \{ A \mid \text{no argument in } S \text{ attacks } A \}$.

The operator Pl_{AF} is very closely related to F_{AF} as the following lemma shows.

Lemma 14 $F_{AF} = Pl_{AF} \circ Pl_{AF}$

Proof Let S be a set of arguments in AF and A be an arbitrary argument in AF. Then
$A \in F_{AF}(S)$
iff each attack against A is attacked by an argument in S
iff each attack against A belongs to $AR \backslash Pl_{AF}(S)$
iff no attack against A belongs to $Pl_{AF}(S)$
iff no argument in $Pl_{AF}(S)$ attacks A
iff A belongs to $Pl_{AF}(Pl_{AF}(S))$. ■

The relations between Pollock's indefeasible arguments and our grounded extension semantics is illuminated in the following lemma and theorem.

Lemma 15 Let GE_{AF} be the grounded extension of AF. Then

$$\phi \subseteq AR_1 \subseteq ... \subseteq AR_{2i-1} \subseteq AR_{2i+1} \subseteq ... \subseteq GE_{AF} \subseteq ... \subseteq AR_{2i+2} \subseteq AR_{2i} \subseteq ... \subseteq AR_0 = AR$$

Proof It is easy to see that $AR_1 = F_{AF}(\phi)$. Further, from the fact that for each $n \geq 0$: $AR_{n+2} = F_{AF}(AR_n)$, it follows immediately that for each $i \geq 0$: $AR_{2i} = F_{AF}{}^i(AR_0) = F_{AF}{}^i(AR)$, and for each $i \geq 1$: $AR_{2i-1} = F_{AF}{}^{i-1}(AR_1) = F_{AF}{}^i(\phi)$. The lemma follows then directly from the monotonicity of F_{AF}, and the fact that GE_{AF} is a fixpoint of F_{AF}. ■

Let $AR_{inf} = \cup\{ AR_{2i-1} \mid i \geq 1 \}$. It follows immediately that

Theorem 9 1) An argument A is indefeasible iff $A \in AR_{inf}$.
2) $AR_{inf} \subseteq GE_{AF}$
3) If AF is finitary then $AR_{inf} = GE_{AF}$. ■

3.3. Logic Programming as Argumentation

It is widely accepted today that logic programming provides an ideal environment for the implementation of knowledge bases. So, it is not surprising that much work has been done to study the semantics of logic programming. The semantics of logic programming depends on whether we view negation as finite failure or as possibly infinite failure. The first view can provide computable semantics [8,38,36] but fails to capture the intended semantics in many cases. The second view captures better the intended semantics of a logic program [17,11,24,23,29,50] but is incomputable in general. In [11,28], an argument-based framework for logic programming with negation as possibly infinite failure has been given unifying many previously proposed approaches. Continuing this line of research, we will show in this chapter that a logic program can be considered as a schema for generating arguments. Different semantics will result from the difference in the structure of the arguments. The computability of a semantics is determined by the computability of the arguments involved.

A logic program is a finite set of clauses of the form $b_0 \leftarrow b_1,..,b_m,\neg b_{m+1},..,\neg b_{m+n}$ where b_i's are atoms. For a logic program P, G_P denotes the set of all ground instances of clauses in P. For each literal h, the complement of h is denoted by h*. Further, for each set of ground atoms M, let $\neg.M = \{\neg b \mid b \in M\}$.

3.3.1. Negation As Possibly Infinite Failure

Let $K = \{\neg b_1,...,\neg b_m\}$ be a set of ground negative literal. A ground atom k is said to be a defeasible consequence of P,K, denoted by $P,K \vdash k$, if there is a sequence of ground atoms $(e_0, e_1,...,e_n)$ with $e_n = k$ such that for each e_i, either $e_i \leftarrow \in G_P$ or e_i is the head a clause $e_i \leftarrow a_1, ..,a_t,\neg a_{t+1},..,\neg a_{t+r}$ in G_P such that the positive literal $a_1,..,a_t$ belong to the preceding members in the sequence and the negative literal $\neg a_{t+1},..,\neg a_{t+r}$ belong to K. K is said to be a *support for k wrt P*.

A logic program P is transformed into an argumentation framework $AF_{napif}(P) = \langle AR,attacks\rangle$ as follows:

$$AR = \{(K,k) \mid K \text{ is a support for } k \text{ wrt } P\} \cup \{ (\{\neg k\},\neg k) \mid k \text{ is a ground atom } \}$$

$$(K,h) \text{ attacks } (K',h') \text{ iff } h^* \in K'$$

Remark An argument of the form $(\{\neg k\},\neg k)$ captures the idea that k would be concluded false if there is no acceptable argument supporting k.

The semantics of P defined by the preferred extensions of $AF_{napif}(P)$ is called preferred extension semantics. It is not difficult to see that this semantics coincides with the preferential semantics defined in [11].

Correspondence between Stable Models of P and Stable Extensions of $AF_{napif}(P)$

Let M be a Herbrand interpretation (a set of ground positive literal) of P. M is said to be a *stable model* of P iff M is the least Herbrand model of the program obtained from G_P by (1) deleting every clause in G_P whose body contains a negative literal $\neg b$ with $b \in M$, and (2) deleting all negative literal from the remaining clauses [24].

For each interpretation M, define CM = { a | a is a ground atom and a $\notin$ M }. Let $AR_{napif}(P)$ = (AR,attacks), and for each stable model M, let E_M = {(K,k) $\in$ AR | K $\subseteq$ ¬.CM }. It is easy to see that k $\in$ M $\cup$ ¬.CM iff $\exists$(K,k) $\in$ E_M. Hence, for each argument A=(K,k) $\in$ AR, A $\in$ E_M iff $\forall$¬b $\in$ K: b $\notin$ M iff A is not attacked by E_M. From lemma 3, it follows:

Theorem 10 Let P be a logic program. Then a Herbrand interpretation M is a stable model of P iff there is a stable extension E of $AF_{napif}(P)$ s.t. M $\cup$ ¬.CM = { k | $\exists$(K,k) $\in$ E }. ■

Correspondence between Well-Founded Model of P and Grounded Extension of $AF_{napif}(P)$

A consistent set of ground literal is called a partial interpretation of P. The definition of well-founded model [23] is based on the following notion of unfounded sets: A set S of positive ground atoms is an unfounded set of a logic program P wrt a partial interpretation I iff for each clause C in G_P whose head belongs to S, either the body of C is false wrt I or it contains a positive literal l such that l $\in$ S. The *well-founded model* of a logic program P is defined as the least fixed point of the following monotonic operator

$$V_P(I) = \neg.GUS(I) \cup T_P(I)$$

where $T_P(I)$ = { b | $\exists$C $\in$ G_P s.t. head(C) = b and body(C) is true wrt I}, and GUS(I) is the greatest unfounded set of P wrt I.

The following theorem shows the equivalence between well-founded model of P and grounded extension of AF(P).

Theorem 11 Let P be a logic program, and WFM be the well-founded model of P. Let GE be the grounded extension of $AF_{napif}(P)$. Then

$$WFM = \{ h \mid \exists(K,h) \in GE \}$$

Proof See appendix B.

Though considering a logic program P as an argumentation framework $AF_{napif}(P)$ provides an appropriate platform for capturing the intended semantics of P, any semantics based on $AF_{napif}(P)$ is incomputable due to the result that in general, the set of arguments of $AF_{napif}(P)$ is incomputable.

Lemma 16 Let P be an arbitrary logic program and $AF_{napif}(P)$ = (AR,attacks). Then, in general, AR is not recursively decidable, i.e. there is no algorithm which always terminates and decides for each pair (K,k) whether or not (K,k) $\in$ AR.

Proof Let f be an n-ary partial recursive function which is not totally recursive. Then according to theorem 9.6 in [38], there is a definite program P and an n+1-ary predicate symbol p_f such that all computed answers for P $\cup$ {$\leftarrow p_f(s^{k1}(0),..,s^{kn}(0),x)$} have the form {$x/s^k(0)$} and for all nonnegative integers k1,..,kn, we have f(k1,..,kn)=k iff $\{x/s^k(0)\}$ is the computed answer for P $\cup$ $\{\leftarrow p_f(s^{k1}(0),..,s^{kn}(0),x)\}$.

For any sequence of nonnegative integers k1,..,kn, B = $(\emptyset, p(s^{k1}(0),..,s^{kn}(0),k))$ is an argument in $AF_{napif}(P)$ iff f(k1,..,kn) is defined and f(k1,..,kn) = k. Since f is partially recursive but not totally recursive, there exists no algorithm which always terminates and can decide whether B is an argument. ■

It follows immediately that

Theorem 12 Let P be an arbitrary logic program. Then the stable, well-founded and preferred extension semantics of P are in general incomputable. ■

3.3.2. Negations As Finite Failure

To capture the semantics of negation as finite failure, a logic program P is transformed into an argumentation framework $AF_{naff}(P)$ = <AR,attacks> as follows:

$$AR = \{(K,k) \mid \exists C \in G_P: head(C)=k \text{ and } body(C)=K\} \cup \{ (\{\neg k\},\neg k) \mid k \text{ is a ground atom } \}$$

$$(K,h) \text{ attacks } (K',h') \text{ iff } h^* \in K'$$

Remark The definition of $AF_{naff}(P)$ means that each ground instance of a clause of P constitutes an argument for its head.

It follows immediately

Lemma 17 The set of arguments in $AF_{naff}(P)$ for each logic program P is computable. ■

The relationship between the Clark's completion semantics and the $AF_{naff}(P)$-based semantics is clarified by the following theorems.

Theorem 13 Let P be an arbitrary logic program. Then a Herbrand interpretation M is a model of the Clark's completion of P, comp(P)[10], iff there is a stable extension E of $AF_{naff}(P)$ s.t. M $\cup$ $\neg$.CM = { k | $\exists$(K,k) $\in$ E }.

Proof "=>" Let $E_M = \{(K,k) \in AR \mid k \in M \text{ and } K \text{ is true wrt } M\} \cup \{(\{\neg k\},\neg k) \mid k \notin M\}$. Let (H,h) be a n arbitrary argument in AR. Then (H,h) $\in$ E_M iff $\forall l \in H$: l is true in M iff $\forall l \in H$: l* is not true in M iff (H,h) is not attacked by E_M. Hence from lemma 3, it follows that E_M is a stable extension.
"<=" Let M = { k | $\exists$(K,k) $\in$ E and k is an atom}. Since E is stable, it is clear that for each b $\in$ M, there is a C $\in$ G_P s.t. head(C) = b and body(C) is true in M and for each b $\notin$ M, for each

[10]The formal definition of comp(P) is given in the appendix C.

$C \in G_P$ if head(C) = b then body(C) is false in M. So it is clear that M is a model of comp(P). ■

For each logic program P, each partial interpretation I, the operator $F_P(I)$ is defined as follows:

$$F_P(I) = \{ k \mid \exists C \in G_P: \text{head}(C) = k \text{ and body}(C) \text{ is true wrt } I\}$$

$$\cup \{\neg k \mid \forall C \in G_P: \text{head}(C) = k \text{ implies: body}(C) \text{ is false wrt } I\}$$

The *Fitting's model* of a logic program P is defined as the least fixpoint of F_P [19].

Theorem 14 Let P be a logic program, and FM be the Fitting's model of P. Let GE be the grounded extension of $AF_{naff}(P)$. Then

$$FM = \{ h \mid \exists(K,h) \in GE \}$$

Proof Let F be the characteristic function of $AF_{naff}(P)$. We prove by induction that for each ordinal i, $F_P^i(\emptyset) = \{h \mid \exists(K,h) \in F^i(\emptyset)\}$.

It is clear that $F_P^i(\emptyset) = \{h \mid \exists(K,h) \in F^i(\emptyset)\}$ for i=0 and for any limit ordinal i if $F_P^j(\emptyset) = \{h \mid \exists(K,h) \in F^j(\emptyset)\}$ holds for any j<i. Further, it is also not difficult to see that the equation $F_P^i(\emptyset) = \{h \mid \exists(K,h) \in F^i(\emptyset)\}$ holds for n+1 if it holds for n. ■

3.3.3. Coincidence Between Different Semantics

Due to the fact that a logic program can have different semantics, it is often important for practical purposes to find sufficient syntactical conditions guaranteeing the existence and the equivalence of these semantics. In this chapter, we want to show that all the conditions which have been given in the logic programming literature to guarantee the equivalence of different semantics can be captured by our newly introduced notions of well-foundedness and uncontroversity of argumentation frameworks

Let Pred be the set of all predicate symbols occurring in a logic program P. The predicate dependency graph [2] of P is a directed graph with signed edges. The nodes are the elements of Pred. An edge from p to q is positive (resp. negative) iff p occurs in the head of a clause C of P and q occurs in a positive (resp. negative) literal in the body of C. Define $\geq_{+1}$ and $\geq_{-1}$ by: $p \geq_{+1} q$ (resp. $p \geq_{-1} q$) iff there is a (nonempty) path from p to q containing an even (resp. odd) number of negative edges in the predicate dependency graph. Further, let us define $p \geq q$ iff $p \geq_{+1} q$ or $p \geq_{-1} q$, and $p \equiv q$ iff $p \geq q$ and $q \geq p$.

A program is said to be *hierarchical* [38] if there is no p such that $p \geq p$. A program is said to be *stratified* [2] if we never have both $p \equiv q$ and $p \geq_{-1} q$. A program is *strict* [35,54] if there are no p,q such that $p \geq_{+1} q$ and $p \geq_{-1} q$. A program is *call-consistent* [35,54,12,18] if there is no predicate symbol p such that $p \geq_{-1} p$.

It is not difficult to see that the following theorems hold:

Theorem 15 1) If P is stratified then $AF_{napif}(P)$ is well-founded.
2) If P is hierarchical then $AF_{naff}(P)$ is well-founded. ■

Theorem 16 1) If P is strict then both $AF_{napif}(P)$ and $AF_{naff}(P)$ are uncontroversial.
2) If P is call-consistent then both $AF_{napif}(P)$ and $AF_{naff}(P)$ are limited controversial. ■

Therefore, it follows immediately the following results.

Corollary 17 1) The stable and well-founded semantics of stratified logic programs coincide.
2) The Clark's completion of a hierarchical program P has exactly one Herbrand model which coincides with the Fitting's model of P. ■

Corollary 18 1) The well-founded semantics, stable semantics and preferred extension semantics of any strict logic program P coincide in the sense that for each ground literal k, $k \in WFM_P$ iff k is true in each stable model of P.
2) The stable semantics and preferred extension semantics of call-consistent logic programs coincide.
3) Each maximal three-valued Herbrand model of comp(P) is two-valued if P is call-consistent. ■

Corollary 19 1) There exists at least one stable model for each call-consistent logic program.
2) The Clark's completion of call-consistent P, comp(P), is consistent. ■

Though corollaries 17,19 are not new, theorems 15,16 give a much deeper insight into the nature of strictness, stratification and call-consistency. Further, they give also a much simpler proof for these results.

Part 4: Argumentation As Logic Programming: A Generator of Metainterpreters for Argumentation Systems

There are extremely interesting relations between argumentation and logic programming. In the previous chapter, we have seen that logic programming can be shown to be a form of argumentation. In this chapter, we will show that argumentation itself can be "viewed" as logic programming. This result is of great importance. It introduces in fact a general method for generating metainterpreters for argumentation systems. This method is very much similar to the compiler-compiler idea in conventional programming.

Any argumentation system is composed from two essential components: One for generating the arguments together with the attack-relationship between them. The other is for determining the acceptability of arguments. So we can think of an argumentation system as consisting of two units, an argument generation unit, AGU, and an argument processing unit, APU. The argument processing unit APU is in fact a very simple logic program consisting of the following two clauses:

(C1) $acc(X) \leftarrow \neg defeat(X)$

(C2) $defeat(X) \leftarrow attack(Y,X), acc(Y)$

where acc(X) stands for "argument X is acceptable" and defeat(X) for "argument X is defeated".[11]

The above described architecture of an argumentation system is illustrated by the following picture

AGU

$attacks(A_1,A_2)$
.....
$attacks(B_1,B_2)$

APU

acceptable arguments

Let AF=(AR,attacks) be an argumentation framework. Let P_{AF} denote the logic program defined by

$$P_{AF} = APU + AGU$$

with APU = {C1,C2} and $AGU = \{attacks(A,B) \leftarrow \mid (A,B) \in attacks\}$ [12].

[11]Clause C2 means that an argument is defeated if it is attacked by an acceptable argument. C1 means that X is acceptable if it is not defeated (or equivalently, each clause which attacks X is not acceptable (i.e. defeated)).

[12]Each argument is considered as a distinct element in the Herbrand Universe of P_{AF}.

Further, for each extension E of AF, denote

$$m(E) = AGU \cup \{acc(A) \mid A \in E\} \cup \{defeat(B) \mid B \text{ is attacked by some } A \in E\}$$

The following theorem shows the correctness of the above architecture.

Theorem 17 Let AF be an argumentation framework and E be an extension of AF. Then

1) E is a stable extension of AF iff m(E) is a stable model of P_{AF}.

2) E is a grounded extension of AF iff $m(E) \cup \{\neg defeat(A) \mid A \in E\}$ is the well-founded model of P_{AF}.

3) The well-founded model and Fitting's model of P_{AF} coincide.

Proof 1) Obvious from the definition of stable model [24], and from lemma 3.

2) Let $AF0 = AF_{napif}(P_{AF})$. Let X,Y be two arguments in AF0. Hence X attacks Y iff there is an argument A in AF such that X = (K,defeat(A)), and Y = (K',k') s.t. $\neg defeat(A) \in K'$. Let F, F_0 be the characteristic functions of AF,AF0, respectively. It is not difficult to prove by induction on i that for each ordinal i, $m(F^i(\phi)) \cup \{\neg defeat(A) \mid A \in F^i(\phi)\} = \{h \mid (K,h) \in F_0^i(\phi)\}$.

3) Obvious. ■

Kowalski [32] has pointed out that logic-based knowledge bases can be described by the equation "Knowledge Base = Knowledge + Logic". Further logic-based knowledge bases can be viewed as argumentation systems where the knowledge is coded in the structure of arguments and the logic is used to determine the acceptability of arguments. In that sense, the above architecture of argumentation systems can be considered as a schema for generating metainterpreters for knowledge bases. To increase the efficiency of this metainterpreter, techniques of partial evaluation and program transformation [3,59] can be applied.

5. Conclusions and Discussions

In this paper, we have developed a highly abstract but simple theory of argumentation where the central notion of acceptability of arguments is captured in a general way. Then we proceed to argue for the appropriateness of our theory first by demonstrating how our theory can be used to investigate the logical structure of many problems in human's social and economic affairs, and second by showing that nonmonotonic reasoning in AI and logic programming is just a form of argumentation.

Our work can have many practical consequences. First, the theory of argumentation frameworks proposed in this paper provides an unified foundation for the different approaches to knowledge representation and reasoning in AI, philosophy and logic programming. Therefore, our results can serve as the foundation for the development of knowledge representation formalism capable of communicating knowledge among different knowledge representation systems. This is especially important in constructing large knowledge bases as such systems will require a sustained effort over a large geography by many teams which will be forced to use different knowledge representation languages in developing their subsystems since no single formalism to knowledge representation can satisfy all the "basic properties" of a knowledge base system[13][49,32,27,44].

The uncover of the relationship between argumentation and n-person games point out the relationship between argumentation and negotiation. While negotiation can be viewed as the (operational) process to find a solution, argumentation is needed to justify a proposed solution. Hence, it is clear that there is no negotiation without argumentation. In other words, argumentation is an integral part of negotiation. So we expect that our theory will have some impact to the study of DAI.

This paper is the first in a series of works devoted to study the theory, architecture and development of argumentation systems. In [13], we study the relations between argumentation, gametheoretical semantics and logic programming. In general, we expect that the attacks against some arguments may have different strengths, one may be more "deadly" than the others. So a study of how to differentiate the strength of arguments is necessary. A first step has been done in [14] where we have identify two kind of attacks, the reductio ad absurdum attacks and the specificity attacks. Bondarenko, Toni and Kowalski [7] have also studied this problem in a more general context to provide an unified argumentational assumption-based approach to nonmonotonic reasoning. Still, more works need to be done here. An interesting topic of research is the problem of selfdefeating arguments as illustrated in the following example. Consider the argumentation framework <{A,B},{(A,A),(A,B)}>. The only preferred extension here is empty though one can argue that since A defeats itself, B should be acceptable. Pollock in [51] gives a convincing analysis of the importance and the nature of this problem. This problem has also been studied by Kakas, Mancarella and Dung in [29,29a] in the context of logic programming. We plan to look at this problem in our framework of argumentation in the future.

Many other argument systems have also been proposed in the literature [37,45,46,57,60]. The focus of most of these works is on the structure of arguments. Vreeswijk classifies arguments into deductive arguments, statistically based inductive arguments and generic inductive arguments. According to Vreeswijk's classification, the systems in [37,57] are deductive and generic inductive argument systems. The attacks-relation between arguments in Simari and Loui's system [57] is based on Poole's formalization [48] of the principle of specific information overriding more general one. Simari and Loui [57] adopt Pollock's criterium (see chapter 3.2) for determining the acceptability of arguments. In [37], the attack-relationship between arguments and their acceptability are not discussed at all. But by pointing out that different approaches to nonmonotonic reasoning in AI can be viewed as argument systems satisfying certain completeness conditions, Lin and Shoham [37] implicitly recognize the need for a mechanism

[13] Poole [49] shows that no default reasoning system can have all of the following basic properties: conditioning, finite conjunctive closure,Horn representativity,consistency,arbitrary defaults.

for determining the acceptability of arguments in any argument system.

Acknowledgement

I am indebted to Bob Kowalski for his support, encouragement, and especially for the many spiritful discussions with him which have been the major source of motivation and inspiration for me to carry on with this work. So many many thanks to him. Thanks also to Franchesca Toni and Kostas Stathis for their vital help with the literature and also for some strong arguments. I am also grateful to Luis Monis Pereira and David Perkins for the constructive comments on an earlier version of this paper. John Pollock has been also very helpful and supportive. So lots of thanks to him. I also would like to express my sincere thanks to an anonymous referee for his constructive and critical comments.

This paper has been partially supported by the Abduction Group at the Imperial College under a grant from Fujitsu. Support from the EEC activity KIT011 - LPKRR is also acknowledged.

References

[1] Alvarado S.J. ’Argument Comprehension’, Encyclopedia of AI, Stuart C. Shapiro (ed.), pp 30-52

[2] Apt K., Blair H., Walker A. ’Towards a theory of declarative knowledge’ in Minker J (ed.) ’Foundation of deductive databases and logic programming, 1988, Morgan Kaufman

[3] Aravindan C.,Dung P.M. ’partial deduction of logic programs wrt well-founded semantics’, To appear in New Generation Computing (also in Proc. of Third International conference on Algebraic and Logic programming, LNCS 632, Springer Verlag, Volterra, Italy, Sep. 1992)

[4] Barth E.M., Martens J.L. (eds) ’Argumentation: Approaches to Theory Formation’, CLCS Series, 1982, Amsterdam/John Benjamins B.V.

[5] Birnbaum L., ’Argument Molecules: A Functional Representation of Argument Structure’, in Proc. of AAAI’82 pp 63-65

[6] Birnbaum L., Flowers M., McGuire R. ’Towards an AI Model of Argumentation’, In Proc. of AAAI’80, 313-315

[7] Bondarenko A., Toni F., Kowalski B. ’An assumption-based framework to nonmonotonic reasoning’, Invited paper in Proc. of LPNMR, 1993, MIT press

[8] Clark,K.L. ’Negation as Failure’ in Logic and Database, Gallaire H., Minker J. (eds), Plenum, New York,1978

[9] Cohen R. ’Analyzing the Structure of Argumentative Discourse’, Computational Linguistics, Vol 13, No 1-2, pp 11-24, 1987

[10] Dung P.M. ’On the acceptability of arguments and its fundamental role in nonmonotonic reasoning and logic programming’, In Proc. of IJCAI’93, France.

[11] Dung P.M. ’Negations as Hypotheses: an Abductive Foundation for Logic Programming’ In Proc. of ICLP’91, MIT Press

[12] Dung P.M. ’On the relations between stable and well-founded semantics of logic programs’, Theoretical Computer science 105, 1992, 7-25

[13] Dung P.M. ’Logic programming as dialogue games’, Technical report, Dec 1992, Division of Computer Science, AIT

[14] Dung P.M. ’An argumentational semantics to logic programming with explicit negation’ In Proc. of ICLP’93, MIT press

[15] Davis M. ’Game theory: A nontechnical introduction’ Basic Books, Inc., Pub. New York, London

[16] Etherington D.W. ’Reasoning with incomplete information: Investigation of nonmonotonic reasoning’ Research notes in AI, Pitman, London, 1987

[17] Eshghi K., Kowalski R.A. ’Abduction Compared with Negation by Failure’, Proc. of 6th ICLP, 1989

[18] Fages F. ’Consistency of Clarks’ completion and existence of stable models’ J. of Methods of Logic in Computer Science, 1992, To appear

[19] Fitting M.
’A Kripke-Kleen semantics for logic programs’
J. logic programming, 1985,2, 295-312

[20] Gabbay D ’Labelled Deductive Systems, Part 1’, CIS Bericht 90-22

[21] Gefner H. ’Beyond Negation as Failure’ Proc. 2nd Int Conf on Principles of Knowledge Representation and Reasoning, Cambridge, Mass., 1991

[22] McGuire R., Birnbaum L., Flowers M. ’Opportunistic Processing in Arguments’, in Proc. of Seventh IJCAI, 1981, pp. 58-60

[23] Van Gelder A., Ross K., Schlipf J.S. ’Unfounded sets and Well-Founded Semantics for General Logic Programs’ in PODS 1988

[24] Gelfond M.,Lifschitz V. ’The stable model semantics for logic programs’, Proc. of the 5th Int Conf/Sym on Logic Programming, MIT Press, 1988

[25] Gelfond M.,Lifschitz V. ’Representing Actions in Extended Logic Programming ’, Proc. of the JICSLP’92, MIT Press, 1992

[26] Hintikka J. ’The Game of Language’, D Reidel Publishing Company, Dordrecht Holland, 1983

[27] Israel D.J ’What is Wrong with Nonmonotonic logic?’ Readings in Nonmonotonic Reasoning, M.L. Ginsberg (ed.), Morgan Kaufman, 1987

[28] Kakas T., Kowalski R., Toni F. ’Abductive Logic Programming’, J. of Logic and Computations, Vol 2, No 6, pp 719-770, 1992

[29] Kakas T., Mancarella P. ’Stable Theories for Logic Programs’, Proc. ISLP’91, MIT Press

[29a] Kakas T., Mancarella P., Dung P.M. ’The Acceptability Semantics for Logic programs’ In Proc. of the 11th ICLP, 1994, Italy, MIT Press

[30] Konolige K. ’On the Relation between Default and Autoepistemic Logic’ Readings in Nonmonotonic Reasoning, M.L. Ginsberg (ed.), Morgan Kaufman, 1987

[31] Kowalski R.A. ’Logic for problem solving’, Elsvier North Holland, New York, 1979

[32] Kowalski R.A ’The Limitations of Logic and its role in AI’ in ’Foundation of knowledge Base Management: Contribution from Logic, Databases and AI’, J.W. Schmidt, C. Thanos (eds.), Springer verlag, 1989

[33] Kowalski R.A ’Logic Programming in AI’ invited lecture at IJCAI’91

[34] Kowalski R.A ’Logic-based Open Systems’ Proc. of the Stuttgart Conference on Discourse Representation, Dialogue tableaux and logic programming, J. Ph. Hoepelman (ed) Tuebingen, 1988

[35] Kunen K. ’Signed data dependencies in logic programming’ J. of LP, 7, 1989, 231-245

[36] Kunen K. ’Negation in Logic Programming’, in JLP 1987,4,289-308

[37] Lin F., Shoham Y ’Argument Systems: an uniform basis for nonmonotonic reasoning’, KR’89

[38] Lloyd J.W. ’Foundations of Logic Programming’, Springer Verlag, 1987

[39] McCarthy J ’Circumscription - A Form of Nonmonotonic Reasoning’ in Readings in Nonmonotonic Reasoning, M.L. Ginsberg (ed.), Morgan Kaufman, 1987

[40] Moore R. ’Semantical Considerations on Nonmonotonic Logic’ Readings in Nonmonotonic Reasoning, M.L. Ginsberg (ed.), Morgan Kaufman, 1987

[41] McDermott, Doyle J. ’Nonmonotonic Logic I’, Readings in Nonmonotonic Reasoning, M.L. Ginsberg (ed.), Morgan Kaufman, 1987

[41a] Marek W., Nerode A., Remmel J. ’A Theory of Nonmonotonic Rule Systems I’ Annals of Mathematics and Artificial Intelligence, 1 (1990) 241-273

[42] Von Neuman J., Morgenstern O. ’Theory of games and economic behavior’ John Wiley & Sons, Inc.,

[43] Pereira L.M., Aparicio J.N., Alferes J.J. ’Nonmonotonic reasoning with well-founded semantics’, ICLP 1991, MIT press

[44] Perlis D ’On the Consistency of Commonsense Reasoning’ in Readings in Nonmonotonic Reasoning, M.L. Ginsberg (ed.), Morgan Kaufman, 1987

[45] Pollock J.L ’Defeasible reasoning’ Cogn. Sci. 11 (1987) 481-518

[46] Pollock J.L. ’A Theory of Defeasible Reasoning’, International J. of Intelligent Systems, vol 6, no 1, Jan 1991

[47] Poole D., ’A Logical Framework for Default Reasoning’ Artificial Intelligence 1988

[48] Poole D., ’On the Comparison of Theories: Preferring the most Specific Explanation’, IJCAI 1985

[49] Poole D ’The Effect of knowledge on belief: conditioning, specificity and the lottery paradox in default reasoning’, J of AI, vol 49, pp 281-307, 1991

[50] Przymusinski T.C., ’On the Declarative Semantics of Deductive Databases and Logic Programs’ in Foundations of Deductive Databases & Logic Programming, J. Minker (ed.) 1988

[51] Pollock J.L. "Justification and Defeat", To appear in J. of AI

[52] Reiter R. ’A Logic for Default Reasoning’, Readings in Nonmonotonic Reasoning,

M.L. Ginsberg (ed.), Morgan Kaufman, 1987

[53] Richard B. 'Game-theoretical semantics and logical form' in Jaakko Hintikka, ed. Radu J. Bogdan, D Reidel Publishing Company, 1987

[54] Sato T. 'Completed Logic programs and their consistency' J. of LP 1990, vol 9, 33-44

[55] Sedgewick R. 'Algorithms in C' Addison-Wesley Publishing Company, 1990

[56] Shubik M. 'Game theory in the social sciences' MIT press, 1985

[57] Simari G.R., Loui R.P. 'A Mathematical Treatment of Defeasible Reasoning and Its Implementation', Artificial Intelligence 53 (1992), 125-157

[58] Toulmin S. 'The Uses of Arguments', Cambridge University Press, Cambridge, Mass., 1958

[59] Tamaki H., Sato T. 'Unfold/Fold Transformation of logic programs'. In Proc. of second ICLP, 1984

[60] Vreeswijk G. 'The Feasibility of Defeat in Defeasible Reasoning', in KR'91

Appendix A

Example The argumentation framework AF(P) of the following logic program P is not finitary.

P:
$r \leftarrow \neg p$
$p \leftarrow \neg q(x)$
$q(x) \leftarrow even(x)$
$q(x) \leftarrow \neg even(x)$
$even(s(x)) \leftarrow \neg even(x)$
$even(0) \leftarrow$

■

Appendix B

Let I be an partial interpretation. Let $I_0 = I$ and $I_{i+1} = T_P(I_i) \cup I_i$. Define $T_P{}^*(I) = \cup\{ I_i \mid i$ is a natural number$\}$. It is easy to see that WFM is the least fixed point of the following operator

$$W_P(I) = \neg.GUS(I) \cup T_P{}^*(I \cup \neg.GUS(I))$$

Let F be the characteristic function of $AF_{napif}(P)$. To prove the theorem, it is enough to show that for each ordinal i

$$W_P^i(\emptyset) = \{ h \mid \exists(K,h) \in F^i(\emptyset) \}$$

We show this by induction. It is obvious that the above equation holds for i=0 and also for limit ordinal i if it holds for all ordinals less than i. Suppose now that the above equation holds for i. We want to show now that it holds also for i+1. Let $I = W_P^i(\emptyset)$ and $S = F^i(\emptyset)$. From the definitions of W_P and F, it is clear that to show the above equation, it is enough to show:

$$GUS(I) = \{ k \mid (\{\neg k\},\neg k) \in F^{i+1}(\emptyset) \}$$

which itself follows directly from the following lemma:

Lemma B1 Let P be an logic program and let E be an admissible set of arguments from $AF_{napif}(P)$. Further let I be a partial interpretation defined by $I = \{ h \mid \exists(K,h) \in E\}$. Then for each ground atom k, $(\{\neg k\},\neg k) \in F(E)$ iff $k \in GUS(I)$.

Proof First, we need the following notion of proof trees.

If $a \leftarrow$ is a clause in G_P then the tree

$$\begin{array}{c} a \\ \\ \square \end{array}$$

is a proof tree of a.

If $a \leftarrow \neg a_1,\ldots,\neg a_n$ is a clause in G_P, then the tree

$$\begin{array}{c} a \\ \\ \neg a_1 \quad \ldots\ldots \quad \neg a_n \end{array}$$

is a proof tree of a.

If $a \leftarrow a_1,\ldots,a_n,\neg a_{n+1},\ldots,\neg a_{n+m}$ is a clause in G_P, and $T_1,\ldots,T_n$ are proof trees of $a_1,\ldots,a_n$ respectively, then the tree

$$\begin{array}{c} a \\ \\ T_1 \ \ldots \ T_n \ \ \neg a_{n+1} \ \ldots \ \neg a_{n+m} \end{array}$$

is a proof tree of a.

"<=" Let $k \in GUS(I)$. Assume that $(\{\neg k\},\neg k) \notin F(E)$, i.e. there is an argument A = (K,k) such that A is not attacked by E. There exists a proof tree Tr of k wrt P such that all of its terminals nodes are either $\Box$ or an element from K. We first prove the following proposition.

Proposition There is a path from the root k of Tr to a terminal node $\neg h$ in Tr such that all the positive literal on this path belong to GUS(I) and $h \in I$.

Proof By induction on the height (the length of the longest path from the root to a terminal node) of Tr.

Base case: The height of Tr is 1. The proposition follows directly from the fact that GUS(I) is an unfounded set.

Induction case: Let the height of Tr be n. Let C be the clause such that body(C) is the set of all children of k in Tr. Since GUS(I) is an unfounded set and $k \in GUS(I)$, it follows that either $I \cup body(C)$ is inconsistent or there is $b \in body(C) \cap GUS(I)$.

Case 1: $I \cup body(C)$ is inconsistent. Case 1.1: There is an atom $b \in body(C)$ such that $\neg b \in I$. Hence, $(\{\neg b\},\neg b) \in E$. As $(K,b) \in AR$ and E is admissible and (K,b) attacks $(\{\neg b\},\neg b)$, there is $\neg a \in K$ such that $a \in I$. Contradiction to the fact that A is not attacked by E. So case 1.1 does not occur. Case 1.2: There is $\neg b$ in body(C) such that $b \in I$. This leads to a contradiction since $\neg b \in K$. So case 1.2 can not occur either.

Case 2: There is $b \in body(C) \cap GUS(I)$. The subtree Tr' with root b of Tr is again a proof tree of b wrt KB. As height of Tr' is n-1 and $b \in GUS(I)$, it follows that there is a path from the root b to a terminal node $\neg a$ in Tr' such that all the positive literal on this path belong to GUS(I) and $a \in I$. The proposition follows then immediately. ■

From the proposition, it follows immediately that A is attacked by E. Contradiction. So $(\{\neg k\},\neg k) \in F(E)$.

"=>" Let $X = \{ b \mid (\{\neg b\},\neg b) \in F(E)\}$. We want to prove that X is an unfounded set of P wrt I. Assume that X is not an unfounded set wrt I. Then there is an atom $a \in X$ and a ground instance $a \leftarrow Bd$ of a clause in P such that $I \cup Bd$ is consistent and no positive subgoal in Bd belongs to X. Thus for each positive subgoal b in Bd there exists an argument (K_b,b) such that $b' \notin I$ for each $\neg b' \in K_b$ (otherwise $(\{\neg b\},\neg b)$ would be acceptable wrt E, a contradiction). Let $K = \cup\{ K_b \mid b \text{ is a positive subgoal in } Bd\} \cup \{ \neg b \mid \neg b \in Bd\}$. Then it is clear that $(K,a) \in AR$. Since $(\{\neg a\},\neg a) \in E$, there is $\neg b' \in K$ such that $b' \in I$. Thus $\neg b' \in Bd$. Contradiction to the fact that $I \cup Bd$ is consistent !! So X is unfounded wrt I. ■

Appendix C

The following definition of comp(P) is taken from [38].

The definition of a predicate p in a logic program P is the set of all clauses in P which have p in their head.

To define comp(P), each clause $p(t_1,..,t_n) \leftarrow b_1,...,b_m$ is transformed into $p(x_1,..,x_n) \leftarrow \exists y_1,..,\exists y_r (x_1=t_1),..,(x_n= t_n),b_1,..,b_n$ where x_i's are variables not appearing in the original clause, and y_j's are the variables of the original clause.

Let $p(x_1,..,x_n) \leftarrow E_1$

.....

$p(x_1,..,x_n) \leftarrow E_k$

be the transformed clauses of the definition of p.

Then the completed definition of p is defined as

$$\forall x_1,..,\forall x_n\ (p(x_1,..,x_n) \leftrightarrow E_1 \text{ v..v } E_k$$

The completed definition of a predicate p whose definition in P is empty is

$$\forall x_1,..,\forall x_n\ \neg p(x_1,..,x_n)$$

comp(P) is defined as the collection of all predicates in P together with the Clark's equality theory.