UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E INFORMÁTICA INDUSTRIAL

MIRIAM MARIELA MERCEDES MORVELI ESPINOZA

CÁLCULO DA FORÇA DE ARGUMENTOS RETÓRICOS E SUA UTILIZAÇÃO EM DIÁLOGOS DE NEGOCIAÇÃO PERSUASIVA EM SISTEMAS MULTIAGENTE

TESE

CURITIBA

2018

MIRIAM MARIELA MERCEDES MORVELI ESPINOZA

CÁLCULO DA FORÇA DE ARGUMENTOS RETÓRICOS E SUA UTILIZAÇÃO EM DIÁLOGOS DE NEGOCIAÇÃO PERSUASIVA EM SISTEMAS MULTIAGENTE

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná como requisito prévio para obtenção do grau de "Doutor em Ciências" – Área de Concentração: Informática Industrial.

Orientador: Prof. Dr. Cesar Augusto Tacla

CURITIBA

2018

Dados Internacionais de Catalogação na Publicação

M892c
2018

Morveli Espinoza, Miriam Mariela Mercedes
Cálculo da força de argumentos retóricos e sua utilização em diálogos de negociação persuasiva em sistemas multiagente / Miriam Mariela Mercedes Morveli Espinoza.-- 2018.
158 f. : il. ; 2018

Disponível também via World Wide Web
Texto em inglês com resumo em português
Tese (Doutorado) - Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Curitiba, 2018
Bibliografia: f. 149-154

1. Sistemas multiagentes. 2. Agentes inteligentes (Software). 3. Retórica - Processamento de dados. 4. Engenharia elétrica - Teses. I. Tacla, Cesar Augusto. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial. III. Título.

CDD: Ed. 22 -- 621.3

Biblioteca Central da UTFPR, Câmpus Curitiba
Bibliotecário: Adriano Lopes CRB-9/1429

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Diretoria de Pesquisa e Pós-Graduação

TERMO DE APROVAÇÃO DE TESE Nº 174

A Tese de Doutorado intitulada **"Cálculo da Força de Argumentos Retóricos e sua Utilização em Diálogos de Negociação Persuasiva em Sistemas Multi-Agente",** defendida em sessão pública pelo(a) candidato(a) **Miriam Mariela Mercedes Morveli Espinoza**, no dia 10 de agosto de 2018, foi julgada para a obtenção do título de Doutor em Ciências, área de concentração Engenharia da Computação, e aprovada em sua forma final, pelo Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial.

BANCA EXAMINADORA:

Prof(a). Dr(a). Cesar Augusto Tacla - Presidente – (UTFPR)
Prof(a). Dr(a). Rafael Heitor Bordini – (PUC-RS)
Prof(a). Dr(a). Edson Emílio Scalabrin – (PUC-PR)
Prof(a). Dr(a). Adolfo Gustavo Serra Seca Neto – (UTFPR)
Prof(a). Dr(a). Gustavo Alberto Giménez Lugo - (UTFPR)

A via original deste documento encontra-se arquivada na Secretaria do Programa, contendo a assinatura da Coordenação após a entrega da versão corrigida do trabalho.

Curitiba, 10 de agosto de 2018.

This thesis is dedicated to my family
Melquiades, Victoria, and Karl.

ACKNOWLEDGMENTS

First I want to thank God for giving me the opportunity to undertake this research study and to persevere and complete it satisfactorily. Thanks to him I found the right place and the right people to work with. Thus, I would like to express my sincere gratitude and appreciation to my supervisor, Professor Cesar Tacla, who has always been supportive and patient throughout the whole period of this study. In the same way, I want to thank my "battle mate" Ayslan Possebom whose determination and positivity inspired me in many occasions. I am very grateful for the many enriching discussions that allowed me to shape my research.

I would like to thank to Professors Rafael Bordini, Edson Scalabrin, Adolfo Neto, and Gustavo Gimenez-Lugo for taking the time to read this thesis and for examining me. I would also like to thank to Professors Myriam Delgado, Josep Puyol-Gruart, and Juan Carlos Nieves, with whom I worked in some articles. Thanks to all of you for your comments and advices because all these helped to improve my work and my research skills.

Finally, this thesis would have not been possible without a scholarship, which allowed me to stay completely focused on my research. For that, thanks CAPES.

RESUMO

MORVELI-ESPINOZA, Mariela. CÁLCULO DA FORÇA DE ARGUMENTOS RETÓRICOS E SUA UTILIZAÇÃO EM DIÁLOGOS DE NEGOCIAÇÃO PERSUASIVA EM SISTEMAS MULTIAGENTE. 158 f. Tese – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2018.

A negociação entre agentes inteligentes é chamada de persuasiva quando as propostas são apoiadas por argumentos retóricos (ameaças, recompensas ou apelações). Esta tese aborda o problema de cálculo da força destes tipos de argumentos, cujo papel é tentar persuadir o agente oponente a aceitar as propostas enviadas mais rapidamente. Nos trabalhos relacionados, o valor da força de um argumento retórico é representado por um vetor de dois elementos: o valor da incerteza das crenças que constituem o argumento e o valor da importância do objetivo do oponente. No entanto, existe uma necessidade de uma análise mais profunda destes componentes e das características dos participantes que podem influenciar no valor da força. Portanto, o objetivo deste trabalho é estudar estes tipos de argumentos e fornecer um modelo de cálculo da força que seja mais expressivo.

Esta tese contém três partes principais. A primeira foca-se na delineação de uma arquitetura de agentes que é baseada no modelo de processamento de objetivos definido por Castelfranchi e Paglieri, o qual pode ser considerado uma extensão do modelo Crenças-Desejos-Intenções. Neste modelo, os objetivos passam por quatro etapas de filtragem, nas quais um objetivo começa em um estado adormecido e vira uma intenção ao passar a última etapa. O trabalho apresenta uma formalização computacional deste modelo baseada em argumentação, onde o avanço dos objetivos de uma etapa para outra é suportada por argumentos.

A segunda parte desta tese foca-se no modelo de cálculo da força de argumentos retóricos. Primeiramente, apresenta-se a arquitetura dos agentes negociadores e as definições lógicas de ameaça, recompensa, e apelação. Em seguida, são apresentados os critérios que são utilizados no modelo de cálculo, tais como a importância do objetivo do oponente, a efetividade de dito objetivo e a credibilidade dos agentes participativos. A efetividade do objetivo é calculada tomando como base o estado deste –segundo o modelo de Castelfranchi e Paglieri– e o tipo de argumento retórico.

A última parte apresenta um conjunto de experimentos que visam avaliar empiricamente o modelo de cálculo proposto. Com este propósito, primeiramente apresenta-se um modelo de negociação que rege o comportamento dos agentes participantes durante o diálogo. Os experimentos avaliam a eficiência da proposta, comparando-a com a proposta mais próxima encontrada na literatura. Os resultados demonstram que o modelo proposto é mais eficiente em termos de número de ciclos de negociação, número de argumentos trocados pelos agentes e número de acordos alcançados.

Palavras-chave: negociação persuasiva, força dos argumentos retóricos, agentes inteligentes

ABSTRACT

MORVELI-ESPINOZA, Mariela. CALCULATING RHETORICAL ARGUMENTS STRENGTH AND THEIR UTILIZATION IN DIALOGUES OF PERSUASIVE NEGOTIATION IN MULTIAGENT SYSTEMS. 158 f. Tese – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2018.

A negotiation between agents is called persuasive when the proposals are backed by rhetorical arguments (threats, rewards, or appeals), whose role is to try to persuade the opponent agent to accept the proposal more readily. This thesis tackles the problem of calculating the strength value of these kinds of arguments. In the related work, the strength value of a rhetorical argument is represented by a vector of two elements: the value of the uncertainty of the beliefs that make up the argument and the value of the importance of the opponent's goal. Nevertheless, there is a need of a further analysis of these components and of the characteristics of the participant agents that may impact on the strength value. Therefore, the objective of this work is to study these kinds of arguments and to propose a more expressive model for calculating their strength values.

This thesis contains three main parts. The first one concerns the design of an agent architecture that is based on the goal processing model proposed by Castelfranchi and Paglieri, which can be considered an extension of the Beliefs-Desires-Intentions (BDI) model. In this model, the goals go through four stages of filtering from being mere desires until they become an intention. We propose an argumentation-based formalization of this model, which means that the passage of the goals from one stage to the next is supported by arguments.

The second part of this thesis concerns the strength value calculation model. First of all, the architecture of the negotiation agents and the logical definitions of each kind of rhetorical argument are presented. After that, the criteria that are taken into account for the strength calculation are presented. Thus, besides considering the importance of the opponent's goal, we also consider the effectiveness of this goal and the credibility of the participant agents. The effectiveness of the opponent's goal is calculated based on its status – according to the model of Castelfranchi and Paglieri – and the kind of rhetorical argument it makes up.

The last part presents a set of experiments that aims to empirically evaluate the proposed model. With this purpose, firstly, a negotiation model that rules the behavior of the participant agents during the dialogue is presented. The experiments evaluate the efficiency of the proposal by comparing it with the closest proposal found in literature. The results demonstrate that the proposed model is more efficient in terms of number of negotiation cycles, number of exchanged arguments during the negotiation, and the number of achieved agreements.

Keywords: persuasive negotiation, rhetorical arguments strength, intelligent agents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BDI	Beliefs-Desires-Intentions
BBGP	Belief-based Goal Processing
AF	Argumentation Framework

CONTENTS

1 INTRODUCTION

Argumentation is an usual and important activity in people's daily lives. We use argumentation when we are faced with conflicting situations or information and we have to select a way to follow or what to belief or not. This process may occur only inside our minds or it may occur during a conversation with another person. For example, during a medical appointment, a doctor has to give a treatment to a patient based on his/her symptoms; he could have a list of pros and cons for a particular drug medicine. These pros and cons can be seen as arguments and counterarguments that will help the doctor to make a decision. Consider another situation where two people need to decide about a travel destination, each of them can formulate arguments in favor and against a given city. The final decision of the dialogue (if it exists) will be based on such arguments. Thus, argumentation can be seen as a process by which arguments and counterarguments are constructed, compared and evaluated in order to determine whether any of them are warranted according to some criterion (BESNARD; HUNTER, 2008).

Over the last twenty years, argumentation has became increasingly important within Artificial Intelligence (AI) (BENCH-CAPON; DUNNE, 2007). This is due to an increasing mutual influence between the two areas[1]. Thus, the study of AI connects in many ways with the study of argumentation in areas like non-monotonic reasoning (e.g., (TONI, 2017; KAKAS et al., 2014; ČYRAS; TONI, 2015)), defeasible reasoning (e.g., (MARTINEZ et al., 2012; BIKAKIS; ANTONIOU, 2010; FERRETTI et al., 2014), argumentation and dialogues (e.g., (FAN; TONI, 2011; HUNTER, 2015; KOK et al., 2010)).

In the context of intelligent agents[2], argumentation can be used for their internal tasks (e.g., revising their beliefs, preferences and/or goals) or during a dialogue with other agents

[1] We refer the reader to (EEMEREN et al., 2014) for a more detailed exposition of the influence between AI and argumentation and for a survey of the literature up until approximately 2002, we refer the reader to (REED; NORMAN, 2003). For an expansive presentation of the achievements that connect AI and argumentation, the reader is referred to (RAHWAN; SIMARI, 2009).

[2] An agent is a computer system. According to Wooldridge e Jennings (1995), two usages of the term 'agent' can be distinguished. On one hand, from a weak notion of agency, an agent is a software-based computer system that has autonomy, social ability, pro-activeness, and reactivity. On the other hand, from a stronger notion of agency, an agent can be characterized by using mentalistic notions, such as knowledge, belief, intention, and obligation.

(e.g., for sending additional information that support their opinions). In the former case, the argumentation is called monological and in the latter one, it is called dialogical (BESNARD; HUNTER, 2008). Whether it is a monological or a dialogical argumentation, the arguments involved during the process can be equipped with a weight that allows the agent to assess them. This value is called the strength of the argument. The calculation and the assessment of the strength of arguments is relevant to (i) compare arguments, (ii) modify the status of the arguments by removing some counterarguments[3], or (iii) define decision principles[4].

In the literature, we can find two types of argument. On one hand, the arguments that are made up only of beliefs, which are called epistemic arguments in the decision making context (AMGOUD; PRADE, 2009) and explanatory arguments in the negotiation dialogues (AMGOUD; PRADE, 2005). On the other hand, the arguments that have in their structure beliefs and goals of the agent and/or goals of his opponent (when the argumentation is dialogical). These arguments are called practical in the decision making context (AMGOUD; PRADE, 2009) and rhetorical arguments in the persuasive negotiation dialogues (KRAUS et al., 1998; RAMCHURN et al., 2003; AMGOUD; PRADE, 2005). For the epistemic or explanatory arguments, the strength is calculated based on the certainty level of the beliefs that make up the argument and for the practical (or rhetorical) arguments, it is also considered the importance of the goal that is part of the argument (AMGOUD; HAMEURLAIN, 2006).

This work focuses on rhetorical arguments. Thus, we will study and characterize (i) how to measure the strength of rhetorical arguments and (ii) the impact of strength calculation on persuasive negotiation dialogues, in which agents exchange rhetorical arguments. This impact will be studied in terms of the *number of negotiation cycles*, the *number of exchanged arguments during the negotiation*, and the *number of achieved agreements*. An agreement is achieved when one of the agents accept the proposal of the other agent.

In what follows, two scenarios of persuasive negotiation are presented (Section1.1). These scenarios are then employed to show the importance of studying this topic (Section 1.2). Next, it is presented a more detailed explanation of the problem that is studied in this thesis, including the research questions and the goals of this work (Section 1.3). The main contributions of this thesis are enumerated in Section 1.4 and the publications that are the consequence of the development of this work are presented in Section 1.5. Finally, the structure of this thesis is presented in Section 1.6.

[3] After the argumentation process ends, the arguments adopt a status. They can be acceptable or not. If a given argument is attacked by a counterargument with less strength, such counterargument is not considered anymore.

[4] During a decision process, there is a list of arguments an counterarguments in favor or against a given option. The decision principles state which of such arguments will be taken into account for making the final decision (AMGOUD; PRADE, 2009)

1.1 SCENARIOS

In this section, two scenarios of persuasive negotiation are presented. In both scenarios, the participants try to resolve their conflicts by exchanging arguments. The first one is composed of robot agents and the second one is composed of software agents that act in behalf of their human users.

1. **Rescue robots scenario**: This is a scenario of a natural disaster, where a set of robot agents have a set of tasks such as: (i) looking through rubble to find survivors, (ii) wandering the area in search of people needing help, (iii) helping disabled people do tasks, and (iv) bringing supplies for survivors. When they find a person who is seriously injured, the robots take him/her to the hospital, otherwise he/she is sent to a shelter. The robots can communicate with each other in order to ask for/send information or to ask for help.

   The disaster area is divided into numbered zones, which are named by using ordered pairs. In the disaster area, there is also a robot workshop, where they can supply of power to keep working and be fixed, in case of a damage or failure.

   Each agent is in charge of a certain zone and must achieve its own goals with respect to that zone, which are closely related to its tasks. However, robots can help each other in certain situations, for example, to remove heavy debris. It is under these conditions where a persuasive negotiation dialogue may arise, because robots have to decide whether to continue with their tasks and accomplish their own goals or stop to help another robot.

2. **Software agents scenario**: This is a scenario about a Consumer Complaint Website, whose goal is to try to resolve a conflict between consumers and companies. In this scenario, a software agent (hereafter, it is referred as `CONSUMER`) makes a complaint about a service or product in behalf of a human user and another software agent representing a company (hereafter, it is referred as `COMPANY`), offers possible solutions.

   A persuasive negotiation arises when `CONSUMER` is not pleased with the initial solution offered by `COMPANY`. From that moment, both agents can use rhetorical arguments to try to resolve the conflict and reach an agreement.

1.2 MOTIVATION

During a dialogue of persuasive negotiation between two agents namely the proponent and the opponent, the proponent agent may utter rhetorical arguments to try to force or convince

the opponent to accept a given proposal (RAMCHURN et al., 2003). These arguments can be divided in:

- **Threats:** the proponent tries to persuade the opponent agent by using the argument that something negative will happen to him if he does not accept to do the requirement sent by the proponent;
- **Rewards:** the proponent tries to persuade the opponent by using the argument that something positive will happen to him if he accept to do the requirement sent by the proponent; and
- **Appeals:** the proponent tries to persuade the opponent by using also a positive argument, as in the rewards, but this positive event will depend on the opponent; hence, appeals can be seen as self-rewards (AMGOUD; PRADE, 2006).

In order to better understand the relevance of calculating the strength of these kinds of arguments, three situations, each for one kind of argument, are presented below.

SITUATION 1: Threats

Let `CONSUMER` be a proponent agent that represents a passenger of an airline and `COMPANY` be his opponent agent, which represents the airline. The user of `CONSUMER` missed an international flight due to a schedule change and he wants the airline company to reimburses him the total price of the ticket. The following is the beginning of the conversation between both agents:

`CONSUMER`: Since I was not properly informed about the schedule change of my flight I ask for the reimbursement of the total cost of the ticket.

`COMPANY`: We are sorry, but we sent an e-mail about the schedule change to every passenger. According to our policies we only can refund the 20% of the total price of the ticket, without including taxes.

At this point, the strategy of `CONSUMER` is to try to force `COMPANY` to accept his proposal and decides to send a threat. The following are three threats that `CONSUMER` generates:

- th_1 : *You should refund the total price of the ticket, otherwise I will never buy a ticket in your company anymore.*
- th_2 : *You should refund the total price of the ticket, otherwise I will destroy your reputation in social networks.*

- th_3 : *You should refund the total price of the ticket, otherwise I will take legal actions against your company.*

The question is: which of these threats will CONSUMER send to try to achieve his goal?

SITUATION 2: Rewards

Suppose that CONSUMER sent one of his threats, the strategy of COMPANY is to try to offer something positive and he generates the three next rewards:

- rw_1 : *If you agree with the 20% refund, we will give you 10000 miles.*
- rw_2 : *If you agree with the 20% refund, we will sell you an executive ticket for the price of an economic one for any national destination.*
- rw_3 : *If you agree with the 20% refund, we will give you our service of assistance for the elderly for free for any destination.*

Like in the previous case, COMPANY has to choose one of the rewards to send to CONSUMER. How will he decide which reward to send?

SITUATION 3: Appeals

For this situation, consider the scenario of the rescuer robots. Let BOB and TOM be two robot agents that are wandering the disaster area. At a given moment, BOB needs to remove heavy debris and sends a message to TOM asking for help. However, TOM is still performing its own tasks and is trying to achieve its own goals; hence, his answer is negative. The strategy of BOB is to try to persuade TOM by appealing to any of TOM's goals, and BOB generates the following appeals:

- ap_1 : *If you help me, you can win utility points.*
- ap_2 : *If you help me, you can recharge your battery since the workshop is next to this zone.*
- ap_3 : *If you help me, you can fix your sensor since the workshop is next to this zone.*

Once again, the question is: which appeal may work out better than the others?

Thus, given an agent that is able to generate more than one threat, reward, or appeal. The question is the same for the three situations: which of these threats (rewards or appeals) will

the agent choose to try to persuade his opponent to accept his proposal? According to Guerini e Castelfranchi (2006), a rhetorical argument has to meet some pre-conditions in order for the proponent to reach a negotiation favorable to him; therefore, the chosen argument has to be in the set of arguments that meet such pre-conditions. However, before the proponent decides what argument to send, he needs to have a way of differentiating the arguments of that set. A way of doing it is by measuring their strengths (RAMCHURN et al., 2003).

Thus, studying how to calculate the strength of rhetorical arguments is useful because it provides a measure that allows to compare these arguments, and during a persuasive negotiation dialogue, the involved agents may use it as a reference to choose the argument that will be sent to their opponents.

1.3 PROBLEM STATEMENT

Threats, rewards, and appeals are rhetorical argument whose components are mainly related to the opponent agent, namely the action the proponent wants the opponent to perform (e.g., refunding the total price of the ticket, agreeing with the 20% refund, or helping to remove heavy debris) and the opponent's goal (e.g., having always a good reputation, getting a discount in the price of a executive ticket, or winning utility points). This last component is especially important because the rhetorical arguments are generated with the aim of convincing the opponent and during a persuasive negotiation dialogue, the proponent uses the goals he knows about his opponent to construct such rhetorical arguments.

Against this background, we state below the primary research question of this thesis:

What criteria should an intelligent agent take into account in order to measure the strength of a rhetorical argument and how should this measurement be done?

Literature related to the measurement of the strength of rhetorical arguments uses the value of the importance of the opponent's goal to make the calculation, and besides they consider the certainty level of the beliefs that made up the rhetorical argument (AMGOUD; PRADE, 2004, 2005, 2006). However, considering only these criteria can be limiting, since there exist situations in which other criteria are needed in order to perform a more exact calculation. To make this discussion more concrete, consider the following circumstances:

1. Agent `BOB` knows that "fixing a sensor" (denoted by go_2) is a more important goal – for agent `TOM` – than "winning utility points" (denoted by go_1) or "recharging the battery"

(denoted by go_3). Taking into account only the importance, the appeal generated using go_2 would be the strongest one. However, what happens if agent BOB also knows that go_2 is not an achievable goal since the spare part is not available yet? (and TOM also knows it). In this case, the importance is no longer the best or the unique criterion, related to the goal of the opponent, for calculating the strength of an appeal.

2. Agent COMPANY has already offered rewards before and rarely she has fulfilled it, and agent CONSUMER knows about it. In this case, the strength of a reward offered by COMPANY is also influenced by his credibility, i.e. the credibility of the proponent agent.

In the first case, notice that besides importance, there exists another criterion to evaluate the quality of the goal of an opponent, because it does not matter how important a goal is if it is far from being achieved. Considering the circumstance in item 2 above, the credibility of the proponent should also be considered, since even when the goal of an opponent is very important and/or achievable, a low level of credibility could diminish the value of the strength of a rhetorical argument.

1.3.1 OBJECTIVE OF THIS THESIS

The main objective of this thesis is:

- To propose and evaluate a model for the calculation of the strength of rhetorical arguments, which, besides contemplating the importance of the opponent's goal, takes into account the status of this goal in the opponent's mind and the credibility of the proponent. The use of these criteria impacts on the results returned by the model. Thus, the results can be more efficient and effective than the results of the approach that is based only on the importance of the opponent's goal. The efficiency is measured in terms of two variables: (i) *number of exchanged arguments* and (ii) *number of negotiation cycles*, and the efficacy is measured in terms of the variable *number of reached agreements*. In order to attain this goal, the following tasks will be performed:
  - A computational formalization of the Belief-Based Goal Processing (BBGP) model proposed in (CASTELFRANCHI; PAGLIERI, 2007). This formalization is important because it will be used as the basis for determining the status of a goal, which is used in the calculation of the arguments strength.
  - An analysis of the elements that make up rhetorical arguments with the aim of distinguishing the particular characteristics of each type of rhetorical argument. The

result of this analysis will allow us to propose a customized strength calculation, i.e. a different way for calculating the strength of each type of rhetorical argument.

- A set of simulations that evaluate the proposed model in order to compare its performance with the approach based only on the importance of the opponent's goal.

1.4 MAIN CONTRIBUTIONS

The main contributions of this thesis are:

1. (Chapter 3) An argumentation-based formalization of the BBGP model proposed in (CASTELFRANCHI; PAGLIERI, 2007). From the point of view of the BDI model (BRATMAN, 1987), it can be said that this formalization supports the process of formation of intentions in practical agents. However, the BBGP model can be considered a more expressive an refined model than the BDI model since it contemplates further statuses for a goal. Thus, this formalization focuses on the progress of goals since they are active goals (these have the same meaning of desires) until they become executive goals (these have the same meaning of intentions), including the intermediary statuses (pursuable goals and chosen goals) and the conditions under which a goal can be cancelled. Structured argumentation (BESNARD et al., 2014) is used to support the passage of the goals from their initial state until their final state.

   As formal results, the properties of the formalization are studied, especially the properties indicated by the authors in (CASTELFRANCHI; PAGLIERI, 2007), which are diachrony and synchrony. The first means that the support is given since the goal is a desire until it becomes an intention, and the second means that the support can be tracked, i.e. there is a memory of the cognitive path from the beginning of the process until the end.

2. (Chapter 4) A model for the calculation of the strength of rhetorical arguments (i.e., threats, rewards, and appeals). This model is based on three suggested criteria (i) the credibility of the proponent agent, (ii) the effectiveness of the opponent's goal, and (iii) the importance of the opponent's goal.

   As formal results, the properties of the model are studied and a set of propositions about the behavior of the calculation model is delineated. Besides, the model is applied on two case studies, which are based on the scenarios presented previously. As empirical results, the efficiency of the proposed model is evaluated by means of simulations (Chapter 5).

1.5 PUBLICATIONS

The following publications are a direct consequence of the development of this work.

- (Chapter 3) Mariela Morveli-Espinoza, Ayslan Trevizan Possebom, and César Tacla: *Resolving Resource Incompatibilities in Intelligent Agents*. 6th Brazilian Conference on Intelligent Systems, BRACIS 2017, Uberlândia, Brazil.
- (Chapter 3) Mariela Morveli-Espinoza, Ayslan Trevizan Possebom, Josep Puyol-Gruart, and César Tacla: *Argumentation-based Intention Formation Process*. DYNA, Revista de la Facultad de Minas de la Universidad Nacional de Colombia (Medellin). *In Press*.
- (Chapter 4) Mariela Morveli-Espinoza, Ayslan Trevizan Possebom, and César Tacla: *Construction and Strength Calculation of Threats*. 6th International Conference on Computational Models of Argument, COMMA 2016, Potsdam, Germany.
- (Chapter 4) Mariela Morveli-Espinoza, Ayslan Trevizan Possebom, and César Tacla: *Constructing and calculating the strength of rewards*. XIII Encontro Nacional de Inteligência Artificial e Computacional, ENIAC 2016, Recife, Brazil.
- (Chapter 4) Mariela Morveli-Espinoza, Ayslan Trevizan Possebom, and César Tacla: *Strength calculation of rewards*. 16th Workshop on Computational Models of Natural Argument, CMNA 2016, New York, USA.
- (Chapter 4) Mariela Morveli-Espinoza: *Calculating rhetorical arguments strength and its application in dialogues of persuasive negotiation*. Second Summer School on Argumentation: Computational and Linguistic Perspectives, SSA 2016, Potsdam, Germany.

The work published in the following papers has also contributed to some ideas that are related to this thesis, even though not explicitly included.

- Mariela Morveli-Espinoza and César Tacla: *Generating arguments based on Data-oriented Belief Revision Model*. VII Workshop-Escola de Sistemas de Agentes, seus Ambientes e Aplicações, WESAAC 2014, Porto Alegre, Brazil.
- Mariela Morveli-Espinoza, Myriam Delgado, and César Tacla: *Agente negociador baseado em técnicas fuzzy*. XI Encontro Nacional de Inteligência Artificial e Computacional, ENIAC 2014, São Carlos, Brazil.

- Mariela Morveli-Espinoza, Ayslan Trevizan Possebom, Guilherme F. Mendes, and César Tacla: *Using argumentation for cooperative decision making process*. 19th IEEE International Conference on Computer Supported Cooperative Work in Design, CSCWD 2015, Calabria, Italy.

- Mariela Morveli-Espinoza, Ayslan Trevizan Possebom, and César Tacla: *An abstract-argumentation-based approach for the portfolio selection problem*. XI Workshop-Escola de Sistemas de Agentes, seus Ambientes e Aplicações, WESAAC 2017, São Paulo, Brazil.

- Mariela Morveli-Espinoza, Ayslan Trevizan Possebom, Josep Puyol-Gruart, and César Tacla: *Dealing with Incompatibilities among Goals*. 16th Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2017, São Paulo, Brazil.

- Mariela Morveli-Espinoza: *Persuasive Negotiation Dialogues using Rhetorical Arguments*. 16th Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2017, São Paulo, Brazil.

- Mariela Morveli-Espinoza, Juan Carlos Nieves, Ayslan Trevizan Possebom, Josep Puyol-Gruart, and César Tacla: *Resolving Incompatibilities between Procedural Goals: An Argumentation-based Approach*. Workshop on Formal Methods & Logical Aspects of Multiagent Systems, FMLAMAS 2018, Stockholm, Sweden.

- Mariela Morveli-Espinoza, Juan Carlos Nieves, Ayslan Trevizan Possebom, and César Tacla: *Resolving Incompatibilities among Procedural Goals under Uncertainty*. 6th International Workshop on Engineering MultiAgent Systems, EMAS 2018, Stockholm, Sweden.

1.6 STRUCTURE OF THE WORK

This thesis is structured as follows:

Chapter 2. Background: The goal of this chapter is to provide the theoretical framework that supports this research. This chapter presents the basic notions of abstract and logic-based argumentation. Besides, the processing model of Castelfranchi and Paglieri is also introduced.

Chapter 3. Formalization of the belief-based goal processing model: This chapter proposes an architecture for a BBGP-based agent, including the main building blocks and the operative part. This operative part is based on logic-based argumentation, more specifically, on rule-based argumentation systems. Thus, this chapter presents the kinds of arguments that are employed, the kinds of attacks between them, and defines how these arguments are evaluated. Finally, an agent of the rescue robots scenario is modeled according to this proposal.

Chapter 4. Calculation of the strength of rhetorical arguments: This chapter presents the core part of this thesis. Thus, the model for the strength calculation is introduced. We extend the BBGP-based agents by endowing them with the ability of generating rhetorical arguments and of calculating the strength value of these arguments. We discuss the new criteria that are taken into account in the calculation model. We make a theoretical analysis of the proposed model and show the application of it in both scenarios, the software agents scenario and the rescue robots scenario.

Chapter 5. Persuasive negotiation dialogue: simulations and results: This chapter presents the evaluation of the model proposed in Chapter 4. We start by defining a negotiation model that includes a communication language, a protocol, a decision function that rules the answers of the agents during the dialogue, and a strategy that determines which argument the agent will send to this opponent. We present and discuss four experiments, which aim to evaluate the performance of the proposed model.

Chapter 6. Final remarks: This chapter concludes this thesis with a summary and a list of possible future works.

2 BACKGROUND

In this chapter we study the following topics. Section 2.1 covers the main concepts about abstract and logic-based argumentation. Although the approach proposed in this thesis fits into the definition of logic-based argumentation, it is important to study the abstract theory because it provides necessary and important definitions that are the base to evaluate the arguments in Chapter 3. In Section 2.2, we present the conceptual definitions of the BBGP model, which serve as the underlying theory of our proposal of Chapter 3. Finally, Section 2.3 concludes.

2.1 ARGUMENTATION

In argumentation theory, a widely accepted definition of argumentation was given in (EEMEREN et al., 1996). The authors define it as follows:

"Argumentation is a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener or reader, by putting forward a constellation of propositions intended to justify (or refute) the standpoint before a rational judge."

The above definition starts by asserting that argumentation is an activity that involves reasoning. This means that the construction of arguments is based on an inference process since according to Walton (2005), the building blocks of reasoning are inferences. The authors also state that it is a verbal and social activity. This means that it involves the interaction of people through natural language. In the case of software entities, argumentation is being studied for dialogues among intelligent agents and among intelligent agents and humans. Notice also that argumentation is a goal-directed activity. Thus, the participants engage in argumentation when they have the goal of decreasing or increasing the acceptability of a certain standpoint or solution. At last, the participants use reasons in support of or against that standpoint to try to reach their goal.

In artificial intelligence, the argumentation research is focused on building computational models of arguments. We can categorize this research in two: an abstract perspective of the arguments and a structured perspective. In both perspectives, the argumentation process starts with a set of arguments, which may interact. Finally, these arguments are evaluated to determine the set of arguments that are acceptable together.

2.1.1 ABSTRACT ARGUMENTATION

An abstract argument system or argumentation framework (AF) was introduced in the seminal paper of Dung (1995) and it is one of the most significant developments in the computational modelling of argumentation in recent years. The framework is made up of a set of arguments and a binary relation encoding attacks between arguments. The framework is abstract because neither the structure of the arguments nor the structure of the attacks is specified.

Definition 2.1. (Abstract argumentation framework) An abstract AF is a pair $\mathcal{AF} = \langle \texttt{ARG}, \texttt{att} \rangle$, where $\texttt{ARG}$ is a set of arguments and $\texttt{att} \subseteq \texttt{ARG} \times \texttt{ARG}$ is a binary relation representing attacks between arguments. For two arguments $A, B \in \texttt{ARG}$, the notation A $\texttt{att}$ B or $(A,B) \in \texttt{att}$ means that A attacks B.

An AF can be represented as a directed graph whose nodes represent the arguments of the framework and the edges stand for attacks between them.

Example 2.1. Let $\mathcal{AF} = \langle \texttt{ARG}, \texttt{att} \rangle$ be an abstract AF where $\texttt{ARG} = \{A,B,C,D,E\}$ and $\texttt{att} = \{(A,B),(B,A),(C,A),(D,B),(D,D),(C,E),(E,C)\}$. Figure 1 shows the graphical representation of the framework.

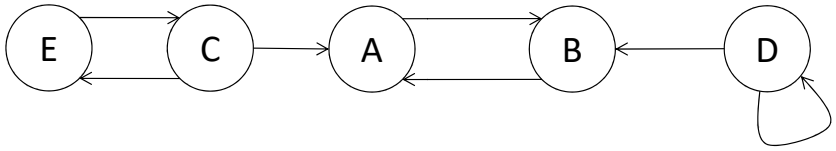


Figure 1: Graphical representation of the abstract argumentation framework of Example 2.1.

An important step in the argumentation process is the one in which arguments are evaluated using an acceptability semantics. In argumentation theory, an acceptability semantics is a function in charge of returning sets of arguments called extensions, which represent a coherent point of view or a coherent position. Thus, each extension should be internally consistent, that is, it must not contain arguments which attack each other.

Various argument-based semantics have been proposed in argumentation literature. We can distinguish two main families of approaches for defining a semantics: declarative approaches and labeling-based ones. The declarative approaches specify which sets of arguments are acceptable. In this family, we have the semantics proposed by Dung (1995) (i.e. admissible, complete, preferred, stable, and grounded) and others based on them: prudent semantics (COSTE-MARQUIS et al., 2005), semi-stable (CAMINADA, 2006), and ideal (DUNG et al., 2007). The labeling-based approaches follows two steps: i) to assign a label to each argument using a particular labeling function, and ii) to compute the extensions (CAMINADA, 2008). Generally three labels are assumed: in, stating that the argument is acceptable; out, meaning that the argument is rejected; and undec, meaning that the status of the argument is unknown. In this family, we have robust semantics (JAKOBOVITS; VERMEIR, 1999) and stage semantics (VERHEIJ, 1996). Additionally, it was also shown that Dung's semantics can be redefined using labeling functions (CAMINADA, 2006).

For the purpose of this thesis, we will recall Dung's semantics (DUNG, 1995). We begin by introducing the notions of conflict-freeness and defense, on which are defined most of the semantics.

Definition 2.2. (Conflict-freeness, Defense) Let $\mathcal{AF} = \langle \mathtt{ARG}, \mathtt{att} \rangle$ be an AF and $\mathcal{E} \subseteq \mathtt{ARG}$.

- $\mathcal{E}$ is conflict-free iff $\nexists A, B \in \mathcal{E}$ such that $(A,B) \in \mathtt{att}$.
- $\mathcal{E}$ defends an argument A iff $\forall B \in \mathtt{ARG}$ if $(B,A) \in \mathtt{att}$ then $\exists C \in \mathcal{E}$ such that $(C,B) \in \mathtt{att}$.

Dung's semantics are based on a notion of admissibility. The idea behind admissibility is that a set of arguments is acceptable if for any argument – that belongs to an extension – that is attacked from outside of the extension, there is an argument that defends it.

Definition 2.3. (Admissibility) Let $\mathcal{AF} = \langle \mathtt{ARG}, \mathtt{att} \rangle$ be an AF and $\mathcal{E} \subseteq \mathtt{ARG}$. $\mathcal{E}$ is an admissible set of $\mathcal{AF}$ iff $\mathcal{E}$ is conflict-free and defends all its arguments.

A property of admissible sets is that every AF has at least one admissible set and the empty set is admissible in every AF. Based on admissibility, we can define the argumentation semantics.

Definition 2.4. (Semantics) Let $\mathcal{AF} = \langle \mathtt{ARG}, \mathtt{att} \rangle$ be an AF and $\mathcal{E} \subseteq \mathtt{ARG}$.

- $\mathcal{E}$ is a *complete extension* of $\mathcal{AF}$ iff $\mathcal{E}$ is conflict-free and $\mathcal{E} = \{A \in \mathtt{ARG} \mid \mathcal{E}$ defends $A\}$.

- $\mathcal{E}$ is a *preferred extension* of $\mathcal{AF}$ iff $\mathcal{E}$ is maximal (with respect to set inclusion) complete extension.
- $\mathcal{E}$ is a *stable extension* of $\mathcal{AF}$ iff $\mathcal{E}$ is conflict-free and $\forall A \in \texttt{ARG}$, $\exists B \in \mathcal{E}$ such that $(B,A) \in \texttt{att}$.
- $\mathcal{E}$ is a *semi-stable extension* iff $\mathcal{E}$ is a complete extension where $\mathcal{E} \cup \mathcal{E}^+$ is maximal where $\mathcal{E}^+ = \{A \mid A \in \texttt{ARG}$ and $(B,A) \in \texttt{att}$ and $B \in \mathcal{E}\}$. If $\mathcal{E}$ is a complete extension, then $\mathcal{E} \cup \mathcal{E}^+$ is called its range. This notion was introduced in (VERHEIJ, 1996).
- $\mathcal{E}$ is a *grounded extension* of $\mathcal{AF}$ iff $\mathcal{E}$ is minimal (with respect to set inclusion) complete extension.

There exists a partial ordering between the Dung's admissibility-based semantics. Every stable extension is a semi-stable extension, every semi-stable extension is a preferred extension, every preferred extension is a complete extension, and every grounded extension is a complete extension (CAMINADA, 2008). We can observe that the converse is not true, for instance, the empty set is a complete extension but not a preferred one. This ordering is represented in Figure 2[1].

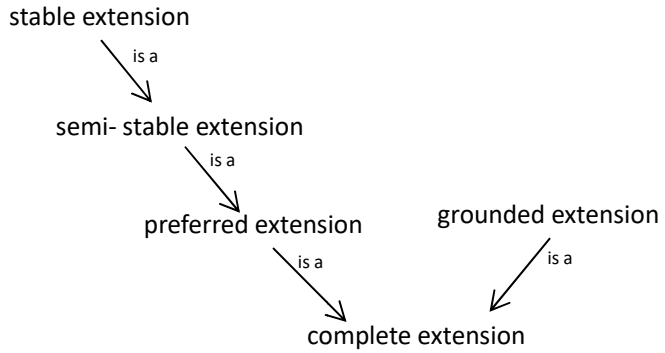


Figure 2: Partial ordering between the admissibility-based semantics

Some other characteristics of Dung's semantics are:

- An $\mathcal{AF}$ has at least one preferred extension.
- A drawback of stable semantics is that its existence is not guaranteed.
- All the semantics, except for the grounded one, may return more than one extension. This means that arguments may be accepted in some extensions and rejected in others.
- An $\mathcal{AF}$ has exactly one grounded extension (which may be empty).
- The grounded extension of an $\mathcal{AF}$ is exactly the set-theoretic intersection of all complete extensions of $\mathcal{AF}$.
- The grounded extension of an $\mathcal{AF}$ is a subset of any preferred extension of $\mathcal{AF}$.

[1]This figure was extracted from (CAMINADA, 2008).

Example 2.2. From the framework of Example 2.1, the following extensions can be obtained:

- There are eight conflict-free sets: $\emptyset, \{E\}, \{B\}, \{B,E\}, \{C\}, \{B,C\}, \{A\}, \{A,E\}$.
- There are four admissible sets: $\emptyset, \{E\}, \{A,E\}, \{C\}$.
- There are four complete extensions: $\emptyset, \{E\}, \{A,E\}, \{C\}$.
- There is no stable extension.
- There is one semi-stable extension: $\{A,E\}$.
- There are two preferred extensions $\{C\}, \{A,E\}$.
- There is one grounded extension: $\emptyset$.

Figure 3 depicts the graphical representation of the relationship among the extension sets presented above.

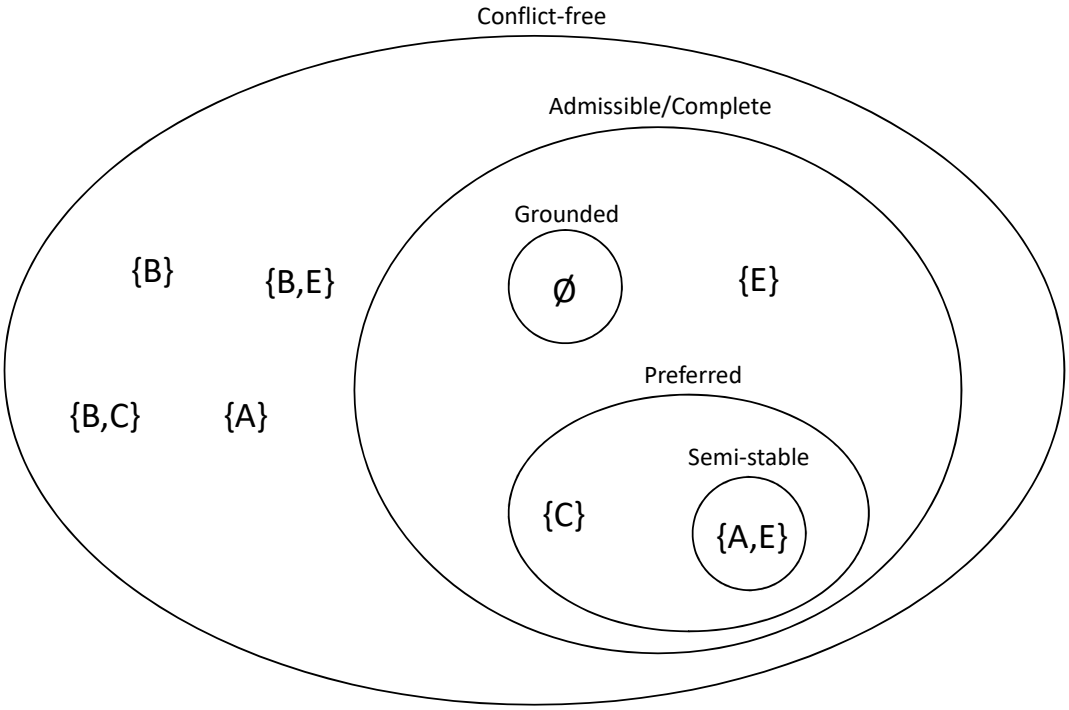


Figure 3: Graphical representation of the relationship among the extension sets.

Besides returning extensions under a given semantics, abstract AFs also return a status for each argument.

Definition 2.5. (Statuses of arguments) Let $\mathcal{AF} = \langle \texttt{ARG}, \texttt{att} \rangle$ be an AF, $\texttt{EXT}(\mathcal{AF})$ be the set of n extensions under a given semantics, and $A \in \texttt{ARG}$ an argument.

- A is *sceptically* accepted iff $A \in \bigcap \mathcal{E}_i$, where $\mathcal{E}_i \in \texttt{EXT}(\mathcal{AF})$ with $1 \leq i \leq n$
- A is *credulously* accepted iff $A \in \bigcup \mathcal{E}_i$, where $\mathcal{E}_i \in \texttt{EXT}(\mathcal{AF})$ with $1 \leq i \leq n$
- A is *rejected* iff $A \notin \bigcup \mathcal{E}_i$ where $\mathcal{E}_i \in \texttt{EXT}(\mathcal{AF})$ with $1 \leq i \leq n$.

Example 2.3. In the framework of Example 2.1 there are two preferred extensions:$\{C\},\{A,E\}$. Notice that both extensions have different arguments; hence, we can say that all the arguments are credulously accepted under the preferred semantics. Note also that the grounded extension is the empty set, it means that there is no sceptically accepted argument. We can observe that there are no stable extensions, it means that all the arguments are rejected under this semantics.

2.1.2 LOGIC-BASED ARGUMENTATION

An abstract AF has a set of atomic arguments (i.e., their internal structure is not considered) and a set of attacks whose nature is not explicitly defined. Nevertheless, we need to study arguments taking into account their internal structure and the nature of the attacks between them. Next, we will study arguments that are built from a knowledge base and are based on a given logic.

In argumentation literature, there are two major categories of instantiations of Dung's abstract argumentation. The first category includes approaches where arguments are built from a Tarskian logic (TARSKI, 1956), and the second one includes approaches that use rule-based systems.

Tarskian logics can be defined as a pair $\langle \mathcal{L}, \mathtt{CN} \rangle$ where $\mathcal{L}$ is a set of well-formed formulae and $\mathtt{CN}$ is a consequence operator, which is defined as a function: $\mathtt{CN} : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}$. Thus, given a set of formulae $\Phi \subseteq \mathcal{L}$, $\mathtt{CN}(\Phi)$ returns the set of formulas that are logical consequences of Φ. For this category, the consequence operator $\mathtt{CN}$ has to satisfy a set of axioms stated in (TARSKI, 1956). In the following definition and example, we will represent formulae by ϕ and atoms by a, b, and c.

Definition 2.6. (Axioms for Tarskian logics) A Tarskian logic is a pair $\langle \mathcal{L}, \mathtt{CN} \rangle$ where $\mathcal{L}$ is a set of formulae and $\mathtt{CN} : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}$ is a consequence operator, which verifies the following axioms:

$\Phi \subseteq \mathtt{CN}(\Phi)$	(Expansion)
$\mathtt{CN}(\mathtt{CN}(\Phi)) = \mathtt{CN}(\Phi)$	(Idempotence)
$\mathtt{CN}(\Phi) = \bigcup_{\Psi \subseteq_f \Phi} \mathtt{CN}(\Psi)$	(Finiteness)
$\mathtt{CN}(\{\phi\}) = \mathcal{L}$ for some $\phi \in \mathcal{L}$	(Absurdity)
$\mathtt{CN}(\emptyset) \neq \mathcal{L}$	(Coherence)

Notation: $\Psi \subseteq_f \Phi$ means that Ψ is a finite subset of Φ.

Example 2.4. Let $\langle \mathcal{L}, \mathtt{CN} \rangle$ be a propositional logic and $\Phi = \{a, a \rightarrow b, c, c \rightarrow \neg b\}$. Then, $\mathtt{CN}(\Phi) = \{a, a \wedge a, b, \neg b, \neg a \vee b, ...\}$.

Some works that use Tarskian logics are: (AMGOUD; BESNARD, 2010) and (AMGOUD; BESNARD, 2009) study AFs built under any Tarskian logic; in (BESNARD; HUNTER, 2001), (BESNARD; HUNTER, 2009), (GOROGIANNIS; HUNTER, 2011), the frameworks are based on classical logic; in (KAKAS et al., 2014), the framework is based on argumentation logic, which is an extension of classical propositional logic.

Next, we present a definition of an argument based on classical logic. This definition was extracted from (BESNARD; HUNTER, 2009). Symbol $\vdash$ denotes classical consequence relation, $\bot$ denotes contradiction, and $\Phi \vdash \bot$ denotes that Φ is contradictory or inconsistent.

Definition 2.7. (Argument based on classical logic) Let Φ be a set of formulae, an argument is a pair $\langle \Phi', \phi \rangle$ such that:

- $\Phi' \nvdash \bot$,
- $\Phi' \vdash \phi$,
- Φ' is a minimal[2] subset of Φ.

Example 2.5. (Extracted from (BESNARD; HUNTER, 2009)) Let $\Phi = \{a, a \rightarrow b, c \rightarrow \neg b, c, d, d \rightarrow b, \neg a, \neg c\}$. Some arguments that can be constructed from Φ are:

(1) $\langle \{a, a \rightarrow b\}, b \rangle$
(2) $\langle \{c, c \rightarrow \neg b\}, \neg b \rangle$
(3) $\langle \{d, d \rightarrow b\}, b \rangle$
(4) $\langle \{\neg a\}, \neg a \rangle$
(5) $\langle \{a \rightarrow b\}, \neg a \wedge b \rangle$
(6) $\langle \{\neg c\}, d \rightarrow \neg c \rangle$

Regarding rule-based systems, these distinguish between facts, strict rules, and defeasible rules. A strict rule encodes strict information that has no exception, whereas a defeasible rule expresses general information that may have exceptions. Their underlying logic language, denoted $\mathcal{L}_{lit}$, usually consists of a set of literals[3]. Thus, let $l, l_1, ..., l_n$ represent literals in $\mathcal{L}_{lit}$ and $r_1, r_2, ...$ name rules. Facts are elements of $\mathcal{L}_{lit}$, strict rules are of the form $r = l_1, ..., l_n \rightarrow l$, and defeasible rules are of the form $r = l_1, ..., l_n \Rightarrow l$. In what follows, $\texttt{HEAD}(r) = l$ denotes the head of a rule, and $\texttt{BODY}(r) = \{l_1, ..., l_n\}$ denotes the body of the rule. It is assumed that the body of every strict/defeasible rule is finite and not empty.

[2] Minimal means that there is no $\Phi'' \subset \Phi'$ such that $\Phi'' \vdash \phi$ (HUNTER, 2010).
[3] Literals are atoms or negation of atoms (the negation of an atom a is denoted $\neg a$).

Definition 2.8. (Theory) A theory is a triple $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$ where $\mathcal{F} \subseteq \mathcal{L}_{lit}$ is a set of facts, $\mathcal{S}$ is a set of strict rules, and $\mathcal{D}$ is a set of defeasible rules.

Now, let us show how new information is produced from a given theory. This happens when (strict and/or defeasible) rules are fired. In (AMGOUD; BESNARD, 2013), the authors define the notion of derivation schema which is used to generate arguments from a given theory and a set of properties that rule the consequence operator $\mathtt{CN}$.

Definition 2.9. (Derivation schema) Let $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$ be a theory and $l \in \mathcal{L}_{lit}$. A derivation schema for l from $\mathcal{T}$ is a finite sequence $T = \{(l_1, r_1), ..., (l_n, r_n)\}$ such that:

- $l_n = l$
- for $i = 1...n$,
  - $l_i \in \mathcal{F}$ and $r_i = \emptyset$, or
  - $r_i \in \mathcal{S} \cup \mathcal{D}$ and $\mathtt{HEAD}(r_i) = l_i$ and $\mathtt{BODY}(r_i) \subseteq \{l_1, ..., l_{i-1}\}$

$\mathtt{SEQ}(T) = \{l_1, ..., l_n\}$
$\mathtt{FACTS}(T) = \{l_i \mid i \in \{1, ..., n\}, r_i = \emptyset\}$
$\mathtt{STRICT}(T) = \{r_i \mid i \in \{1, ..., n\}, r_i \in \mathcal{S}\}$
$\mathtt{DEFE}(T) = \{r_i \mid i \in \{1, ..., n\}, r_i \in \mathcal{D}\}$
$\mathtt{CN}(\mathcal{T})$ denotes the set of all literals that have a derivation schema from $\mathcal{T}$, i.e., the consequences drawn from $\mathcal{T}$.

The set of properties that rule the operator $\mathtt{CN}$ are the following:

Definition 2.10. (Properties of $\mathtt{CN}$) Let $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$ be a theory.

- $\mathcal{F} \subseteq \mathtt{CN}(\mathcal{T}) \subseteq \mathcal{F} \cup \{\mathtt{HEAD}(r) | r \in \mathcal{S} \cup \mathcal{D}\} \subseteq \mathcal{L}_{lit}$
- If $\mathcal{T}$ is finite, then $\mathtt{CN}(\mathcal{T})$ is finite
- $\mathcal{F} = \emptyset$ iff $\mathtt{CN}(\mathcal{T}) = \emptyset$
- If T is a derivation schema from $\mathcal{T}$, $\mathtt{SEQ}(T) \subseteq \mathtt{CN}(\mathcal{T})$

The first property states that all the facts of a theory are part of the consequences drawn from the theory. It also states that the set of consequences drawn from the theory are part of the conclusions of the strict/defeasible rules along with the facts. The rule's conclusions that do not have a derivation schema are those whose literals of their bodies have not been derived from the theory. Finally, it states that all the consequences are literals of the language.

Example 2.6. Let $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$ be a theory such that $\mathcal{F} = \{penguin, bird\}$, $\mathcal{S} = \{r_1 : penguin \rightarrow \neg fly\}$, and $\mathcal{D} = \{r_2 : bird \Rightarrow fly\}$. From $\mathcal{T}$, we can obtain the following derivations:

- $T_1 = \{(penguin, \emptyset)\}$
- $T_2 = \{(bird, \emptyset)\}$
- $T_3 = \{(penguin, \emptyset), (\neg fly, r_1)\}$
- $T_4 = \{(bird, \emptyset), (fly, r_2)\}$

Thus, $\texttt{CN}(\mathcal{T}) = \{penguin, bird, \neg fly, fly\}$.

In rule-based systems, the notion of consistency is defined as follows.

Definition 2.11. (Consistency) A set $\mathcal{L}'_{lit} \subseteq \mathcal{L}_{lit}$ is consistent iff $\nexists l, l' \in \mathcal{L}'_{lit}$ such that $l = \neg l'$. It is inconsistent otherwise.

Some works that rely on rule-based systems are: (PRAKKEN; SARTOR, 1997) and ASPIC (AMGOUD et al., 2004), (MODGIL; PRAKKEN, 2014), its extended version ASPIC+ (PRAKKEN, 2010), and Delp (GARCIA; SIMARI, 2004).

In this thesis, we are interested in instantiating Dung's AF by rule-based systems. We start by defining the structure of the arguments. An argument is composed of two parts: a support and a conclusion. It is constructed from the set of facts, strict rules, and defeasible rules of a theory by using a consequence operator $\texttt{CN}$, which is ruled by the properties established in Definition 2.10. There is more than one way of defining rule-based arguments. The following definition is, to the best of our knowledge, the most recent one and is the first one that is based on derivation schemas. This definition was extracted from (AMGOUD; BESNARD, 2013).

Definition 2.12. (Argument) Let $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$ be a theory. An argument, constructed from $\mathcal{T}$, is a pair $A = \langle T, l \rangle$ such that:

- $l \in \mathcal{L}_{lit}$
- T is a derivation schema for l from $\mathcal{T}$
- $\texttt{SEQ}(T)$ is consistent
- $\nexists \mathcal{T}' \subset (\texttt{FACTS}(T), \texttt{STRICT}(T), \texttt{DEFE}(T))$ such that $l \in \texttt{CN}(\mathcal{T}')$. This means that T is minimal.

T is called the support of the argument and l its claim or conclusion. As for notation, $\texttt{CONC}(A) = l$ denotes the conclusion of the argument and $\texttt{SUPPORT}(A) = T$ denotes its support.

The consistency condition does not allow using inconsistent sets as supports, since an argument should be based on coherent premises. The second condition specifies that the claim

is deduced from the support, and the last condition guarantees that the support consists only of relevant information.

Example 2.7. (Cont. Example 2.6) The following arguments can be built from $\mathcal{T}$:

- $A_1 : \langle\{(bird, \emptyset)\}, bird\rangle$
- $A_2 : \langle\{(penguin, \emptyset)\}, penguin\rangle$
- $A_3 : \langle\{(bird, \emptyset), (fly, r_2)\}, fly\rangle$
- $A_4 : \langle\{(penguin, \emptyset), (\neg fly, r1)\}, \neg fly\rangle$

Next, we present the rule-based version of Example 2.5 in order to better differentiate the arguments that can be generated by using a Tarskian logic from the arguments that can be generated by using a rule-based system.

Example 2.8. Let $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D}\rangle$ where $\mathcal{F} = \{a, c, d, \neg a, \neg c\}$ and $\mathcal{S} = \{a \to b, c \to \neg b, d \to b\}$. Some arguments that can be constructed from $\mathcal{T}$ are:

(1) $\langle\{(a, \emptyset), (b, a \to b)\}, b\rangle$
(2) $\langle\{(c, \emptyset), (\neg b, c \to \neg b)\}, \neg b\rangle$
(3) $\langle\{(d, \emptyset), (b, d \to b)\}, b\rangle$
(4) $\langle\{(\neg a, \emptyset)\}, \neg a\rangle$

Notice that these arguments have their equivalent arguments based on classical logic. However, arguments (5) and (6) that were generated in Example 2.5 ($\langle\{a \to b\}, \neg a \wedge b\rangle$ and $\langle\{\neg c\}, d \to \neg c\rangle$) cannot be generated by using the rule-based approach.

Regarding the attacks between arguments, some rule-based approaches define only one kind of attack and others work with two or three kinds of attacks. The system *DeLP* uses only rebuttal as attack relation (GARCIA; SIMARI, 2004). *Rebuttal* occurs when the claims of two arguments are conflicting. ASPIC uses, in addition to rebuttal, undercut and undermine. *Undercut* attacks the application of defeasible rules; thus, it occurs when the conclusion of an argument contradicts an inference rule used to build the other argument. Before presenting undermine, it is important to mention that in ASPIC, there are two kinds of facts: axioms and ordinary premises. Axioms are certain facts that cannot be attacked and ordinary premises are uncertain facts and so they can be attacked. *Undermine* occurs when the claim of an argument is conflicting with the ordinary premise of another argument.

At this point, it is important to distinguish between an attack and a successful attack (or defeat). Since we work with strict and defeasible rules, this has impact on the arguments built on them. Thus, we have to consider that conclusions of defeasible rules are always defeasible

whereas conclusions of strict rules are always strict. This means that an argument whose claim is the conclusion of a strict rule cannot be defeated by an argument whose claim is the conclusion of a defeasible rule. However, an argument whose claim is defeasible can be defeated by an argument whose claim is strict. Finally, when two arguments with defeasible (or strict) claims attack each other, there is no successful attack.

Once the arguments and attacks in rule-based systems have been defined, the next step is to define the AF. Like in the abstract argumentation approach, it is composed of a set of arguments and the attack relation between them.

Definition 2.13. (Argumentation framework for rule-based systems) An AF defined over a theory $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$ is a pair $\mathcal{AF}_r = \langle \texttt{ARG}(\mathcal{T}), \texttt{att} \rangle$ where $\texttt{ARG}(\mathcal{T})$ is the set of all the arguments defined from $\mathcal{T}$ and $\texttt{att} \subseteq \texttt{ARG}(\mathcal{T}) \times \texttt{ARG}(\mathcal{T})$ is the attack relation.

As in abstract argumentation, the arguments of rule-based systems can be evaluated using the semantics of Dung or the labeling-based semantics. Thus, Definitions 2.2, 2.3, and 2.4 also hold for these arguments. Therefore, an AF for rule-based systems also returns a set of extensions under a given semantics where each argument has a status determined by Definition 2.5. Since each argument has a structure, then a set of conclusions or claims is also returned.

Definition 2.14. (Justified conclusions) Let $\mathcal{AF}_r = \langle \texttt{ARG}(\mathcal{T}), \texttt{att} \rangle$ be an AF for rule-based systems and $\{\mathcal{E}_1, ..., \mathcal{E}_n\}$ its extensions under a given semantics.

- $\texttt{CONCS}(\mathcal{E}_i) = \{\texttt{CONC}(A) | A \in \mathcal{E}_i\}$ (for all $1 \leq i \leq n$).
- $\texttt{Output} = \bigcap_{i=1...n} \texttt{CONCS}(\mathcal{E}_i)$.

$\texttt{CONCS}(\mathcal{E}_i)$ denotes the justified conclusions for a given extension $\mathcal{E}_i$ and $\texttt{Output}$ denotes the conclusions that are supported by at least one argument in each extension.

The obtained extensions sometimes lead to irregular results. In order to detect and avoid such behavior, a set of rationality postulates[4] have been proposed in (CAMINADA; AMGOUD, 2007) in order to judge the quality of rule-based systems. These postulates are: *direct consistency*, *indirect consistency*, and *closure*. Next, we explain each one of the postulates[5].

- **Direct consistency**: A rule-based system violates this postulate when the set of justified conclusions is logically inconsistent. This means that $\exists l, l' \in \texttt{OUTPUT}$ such that $l = \neg l'$.

[4] In (MAUTZ; SHARAF, 1961), a postulate is defined as follows: "*Postulates are generally defined as basic assumptions that cannot be verified. They serve as a basis for inference and a foundation for a theoretical structure that consists of propositions deduced from them*".

[5] For a more detailed description and discussion the reader may consult (CAMINADA; AMGOUD, 2007).

- **Closure:** Closure means that the output of a rule-based system should be closed under the set of strict rules. That is, if there is a strict rule $l' \rightarrow l$ and l' is a justified conclusion, then l should also be a justified conclusion.

- **Indirect consistency:** Indirect consistency means that the closure under the set of strict rules of the set of justified conclusions should be consistent. In some cases, a rule-based system does not violate the direct consistency; however, when the closure is applied to its output, the new resultant set is inconsistent. Suppose that $\texttt{OUTPUT} = \{l_1, l_2, l_3\}$ and there is a strict rule $l_2 \rightarrow \neg l_3$. Suppose also that the system is not closed under strict rules, then the closure of output is: $\mathcal{C}l_{\mathcal{S}}(\texttt{OUTPUT}) = \{l_1, l_2, l_3, \neg l_3\}$, which is clearly logically inconsistent.

Ideally, a rule-based system should behave according to these postulates; however, depending on the application and the semantics it may not occur. Some well-known rule-based systems violate these postulates. For example, the defeasible argumentation system for legal reasoning of (PRAKKEN; SARTOR, 1997), Delp (GARCIA; SIMARI, 2004), ASPIC (AMGOUD et al., 2004), ASPIC+ (PRAKKEN, 2010), and the rule-based system for practical reasoning of (AMGOUD et al., 2011) satisfy the postulates when it works with stable semantics; however, they do not satisfy all the postulates when it works with preferred semantics.

2.2 BELIEF-BASED GOAL PROCESSING MODEL

The BDI model, developed by Bratman (BRATMAN, 1987), is possibly the best-known model of practical reasoning agents. According to Bratman, the rational behavior of humans cannot be analyzed just in terms of beliefs and desires; the notion of intention is needed. Thus, an intention is considered more than a mere desire; it is something the agent is committed to. The process through which a desire becomes an intention is named intention formation or goal processing and has two stages in BDI models: (i) desires, which are potential influences of an action, and (ii) intentions, which are desires the agent is committed to and that are achieved through the execution of a certain plan.

An extended model for goal processing has been proposed in (CASTELFRANCHI; PAGLIERI, 2007). This is the Belief-based Goal Processing Model, which we call the BBGP model. The authors propose a four-stage goal processing model, where the stages are: (i) activation, (ii) evaluation, (iii) deliberation, and (iv) checking. They claim that this extended model may have relevant consequences for the analysis of what an intention is and may better explain how an intention becomes what it is. These characteristics are especially useful when

the agents need to explain and justify why a given desire became an intention and why another one did not. For example, recall the scenario of a natural disaster presented in Chapter 1, where a set of robot agents wander an area in search of people needing help. When a person is seriously injured he/she must be taken to the hospital, otherwise he/she must be sent to a shelter. After the rescue work, the robot agents can be asked for an explanation of why a wounded person was sent to the shelter instead of taking him/her to the hospital, or why the robot decided to take to the hospital a person *x* first, instead of taking another person *y*.

Unlike Bratman's theory, where desires and intentions are different mental states, Castelfranchi and Paglieri argue that intentions share many of the properties of the desires; hence, in their approach both desires and intentions are considered as goals at different stages of processing. Consequently, four different statuses for a goal are defined: (i) active (desire), (ii) pursuable, (iii) chosen and (iv) executive (intention).

Figure 4 shows a general schema of the goal processing stages and the status of a goal after passing each stage. In this formalization, a status before the active one is also considered, it is called sleeping status[6]. Additionally, we include the cancelled status. It means that a goal can be dropped under some circumstances, which are presented and discussed in Section 3.4.

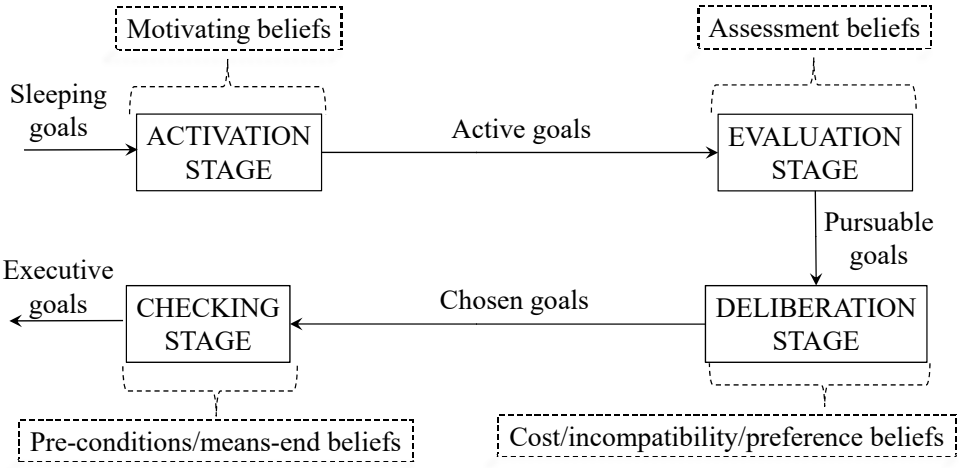


Figure 4: Schema of the goal processing stages and the status of a goal after passing each stage.

Next, we present an overview of the BBGP model of Castelfranchi and Paglieri.

1. **Activation stage:** This is the first stage of the model. Goals that pass this stage are called *active goals*, which is the same as desires. When we say that a goal passes the stage, we mean that there is a set of conditions that are fulfilled, which allows a goal to become active. These conditions are based on a set of beliefs called *motivating beliefs*.

[6]Sleeping is a status of a goal that is proposed in (CASTELFRANCHI, 2008) to refer to goals that have not been activated yet.

When a motivating belief holds, its supported goal becomes active. The authors classify motivating beliefs in: (i) triggering beliefs, which activate goals on the basis of a pre-established association, for example, if you are hungry, then you activate goal "*going to eat*"; and (ii) conditional beliefs, which are related to the conditional nature of the goals, for example, if it is Christmas, then you activate the conditional goal "*buying a Panettone*".

There is no additional information about the nature of both kinds of motivating beliefs and there are few examples. Thus, we model the triggering beliefs as strict rules whereas the conditional beliefs as defeasible rules.

2. **Evaluation stage:** In this stage, the conditions that may refrain the advance of a goal are evaluated. Goals that pass this stage are called *pursuable goals*. A goal only becomes pursuable when there is not a refrain condition. Such conditions are called assessment beliefs and are divided in three types:

   - *Self-realization beliefs*, which connect two kinds of goals. One of the goals does not need the direct intervention of the agent, this is called self-realized, and the other goal depends on the achievement of the self-realized goal. Thus, only the achievement of the self-realized goal will refrain the activation of the other goal. For example, the agent has the self-realized goal "*receiving his payment every month from his employer*". If he does not receive such payment then he activates the goal "*taking actions against his employer*". In other words, the fact that a self-realized goal is achieved refrains the activation of another goal.
   - *Satisfaction beliefs*, which are related to maintenance goals[7]. For example, the belief that an agent is married, prevents him of pursuing the goal "*marrying*".
   - *Impossibility beliefs*, which concern with situations that make impossible to pursue a goal. For example, the belief that the agent is now in Curitiba, prevents the goal "*attending a show in Lima in one hour*".

   In this thesis, we will work with impossibility beliefs because we do not plan to deal with maintenance or self-realized goals and because we want to focus on assessment beliefs that evaluate already active goals. We can notice that although the evaluation stage is the second one in the BBGP model, it can be overlapped with the activation stage and it may prevent the activation of goals.

[7] Maintenance goals define states that must remain true in time (HINDRIKS; RIEMSDIJK, 2007).

3. **Deliberation stage:** In this stage the goals are no longer analysed individually but together. The aim of this stage is to identify the possible incompatibilities or conflicts between goals and then choose those goals that are the most valuable for the agent. Goals that pass this stage are called *chosen goals*. This stage is based on the following beliefs:

    - *Cost beliefs*, which are concerned with the costs an agent may utilize as a consequence of achieving a certain goal. This belief is directly related to the use of internal resources. For example, cleaning a room measuring $100m^2$ can take one hour of the agent's time, but for doing this task he would also need a broom and water. In the former case, time is an internal resource of the agent and in the latter both the broom and water are resources of the environment. External resources are not evaluated in this stage because they are part of the environment.
    - *Incompatibility beliefs*, which are concerned with the different forms of incompatibility between goals that lead an agent to choose among them. These beliefs can be related to cost beliefs, in which case the incompatibility is due to resources. For example, if an agent needs one hour for cleaning a room and he has three goals: "*cleaning the living room*", "*cleaning the kitchen*", and "*cleaning the bathroom*", definitely there is time conflicts between these goals. Other type of conflict appears when two goals cannot be pursued concurrently, it is called terminal incompatibility. For example, cleaning the kitchen and at the same time fixing an antenna is not possible.
    - *Preference beliefs* which are applied to incompatible goals with the aim of establishing a precedence order that determines which goal will become chosen. There are two sub-classes of these beliefs: (i) value beliefs, which are related to the subjective value of a certain goal, given the current interests of the agent. For example, cleaning the living room is more important than cleaning the kitchen because there is a party tonight; and (ii) urgency beliefs, which are concerned with when (if ever) a given goal will expire, for example, recharging its battery is more urgent for a robot than cleaning a room with the remaining battery.

4. **Checking stage:** The aim of this stage is to evaluate whether the agent knows and is capable of performing the required actions to achieve a chosen goal; in other words, if the agent has a plan and he is capable of executing it. Goals that pass this stage are called *executive goals*. The following beliefs are evaluated in this stage:

    - *Precondition beliefs,* which can be divided in two sub-classes: (i) incompetence beliefs, which are concerned with both the basic know-how and competence, and

the sufficient skills and abilities needed to reach the goal, and (ii) lack of conditions beliefs, which are concerned with external conditions, opportunities, and resources.

- *Means-end beliefs,* When the agent is competent to achieve a certain chosen goal, he must evaluate whether it has the necessary instruments for executing a plan.

2.3 SUMMARY

In this chapter, we have briefly introduced the main concepts on which this thesis is based. We have begun by introducing basic abstract argumentation concepts. Then, we have moved to structured arguments based on logics. We have specially focused on rule-based systems because our proposal is an instantiation of these systems. We have also presented an overview of the Belief-based Goal Processing model (BBGP).

In the next chapter, we will present our proposal of formalization of an agent based on the BBGP model. We will also investigate if our proposal satisfies the rationality postulates described in this Chapter.

3 FORMALIZATION OF THE BELIEF-BASED GOAL PROCESSING MODEL

This chapter presents an argument-based formalization for the BBGP model proposed in (CASTELFRANCHI; PAGLIERI, 2007). This formalization focuses on the progress of the goals of an agent since they become active goals (desires) until they become executive goals (intentions), including the conditions under which they can be cancelled.

3.1 INTRODUCTION

One key problem in BDI architectures is that the relation and interplay between beliefs and goals is not clear (CASTELFRANCHI; PAGLIERI, 2007). The BBGP model clarifies these processing and structural relationships, making explicit the function of beliefs in the goal processing as diachronic and synchronic supports. Besides, the use of this model provides an agent with the ability to make justified and consistent decisions, and to choose actions in the same manner. Thus, the agent may choose which intention to commit to based on their beliefs, which act as reasons and are of different types in each processing stage.

Argumentation is employed to support the passage of the goals from their initial status until the last one. We use argumentation since it allows assessing the justifications that back up a conclusion; in this case, arguments provide justifications for or against the passage of a goal from one status to other. Considering that one of the aims of the BBGP model is to better explain how an intention (an executive goal) emerges, argumentation is an excellent approach for doing this. Besides, an argument can put together both the supporting beliefs and the supported goal in just one structure, facilitating future analysis.

Thus, the arguments act as filters between one stage and another and guide the passage of goals (diachronic support) and can be saved for future analysis (synchronic support). Four types of arguments are defined; each one is associated with each stage of the goal processing cycle. On the one hand, arguments in activation, deliberation and checking stage act as supports for a goal to pass to the next stage; thus, if there is at least one acceptable supporting argument for a given goal, it will become active, chosen or executive, respectively. On the other hand,

arguments in the evaluation stage act as attacks, preventing a goal from passing to the next stage. Thus, if there is at least one acceptable attacking argument for a given goal, it will not become pursuable.

The chapter is organized as follows. In Section 3.2, we present the knowledge representation used by the agents in this formalization and define the building blocks used to construct arguments. Section 3.3 is devoted to the definition of arguments, attacks and the argumentation process. Section 3.4 presents a framework containing particular information about the goal processing for each goal and a general framework with information about the entire process. In Section 3.5, we study and detail the properties we expect of BBGP-based agents. Section 3.6 illustrates the performance of this proposal in the robot agents scenario. In Section 3.7, we discuss a few related works. Finally, Section 3.8 presents the conclusion of this chapter.

3.2 KNOWLEDGE REPRESENTATION AND BUILDING BLOCKS

We start by presenting the underlying logical language that will be used. Let $\mathcal{L}_{lit}$ be a set of literals in first-order logical language which are used to represent the mental states of the agent. In $\mathcal{L}_{lit}$ we can distinguish a set of n-ary ($n \geq 1$) predicates $\mathcal{P}$, a set of constant symbols $\mathcal{C}$, and a set of variables $\mathcal{V}$. Literals are defined as positive or negative atoms where an atom is an n-ary predicate. We will represent ground atoms by lower case Roman letters ($a, b, c, d, ...$), non-ground formulae by Greek letters ($\varphi, \psi, ...$), and variables with Roman letters ($x, y, ...$). We use symbols $\wedge$,$\vee$, $\rightarrow$, and $\neg$ to denote the logical connectives conjunction, disjunction, implication, and negation, respectively.

From $\mathcal{L}_{lit}$, we can mainly distinguish the following finite sets:

- The set $\mathcal{F}$, which denotes the beliefs of the agent, is a subset of ground literals from the language $\mathcal{L}_{lit}$
- The set $\mathcal{G}$, which denotes the goals of the agent, is a subset of ground atoms from the language $\mathcal{L}_{lit}$.
- The set $\mathcal{RES}$, which denotes the resources of the agent, is a subset of ground atoms from the language $\mathcal{L}_{lit}$.

Sets $\mathcal{F}$, $\mathcal{G}$, and $\mathcal{RES}$ are pairwise disjoint.

Other important structures are the rules, which express the relation between the beliefs and the goals. The rules can be classified mainly in standard rules and non-standard rules. The

former are made up of beliefs in both their premises and their conclusions and the latter are made up of beliefs in their premises and goals or beliefs about goals in their conclusions. Both standard and non-standard rules can be strict (denoted by $\rightarrow$) or defeasible rules (denoted by $\Rightarrow$).

3.2.1 ACTIVATION AND EVALUATION STAGES

Recall that along the goals processing stages, a goal can be in one of the following states: active, pursuable, chosen or executive. In any of these states, a goal is represented by a ground atom. However, before a goal becomes active, it has the form of a non-ground atom; in this case, we call it a sleeping goal.

Definition 3.1. (Sleeping goal) A sleeping goal is a non-ground atom of the form $Goal\_Name(x_1,...,x_i...,x_n)$ such that $x_1,...,x_i...,x_n \in \mathcal{V}$.

Let $\mathcal{G}_s$ be the set of all sleeping goals.

When a goal is activated, it means that variables $x_1,...,x_i...,x_n$ have been unified with a given set of constants of $\mathcal{C}$. Therefore, from that moment on, a goal is represented by a ground atom and becomes part of the set $\mathcal{G}$ of the goals of the agent.

Next, we define standard rules and the rules involved in the activation and evaluation stages.

Definition 3.2. (Standard rules and rules for the activation an evaluation) Let α be a non-ground literal, φ_i a set of non-ground literals (for $1 \leq i \leq n$), and $\psi \in \mathcal{G}_s$ be a non-ground atom that represents a goal.

- A ***standard rule*** is an expression of the form $\bigwedge \varphi_i \rightarrow \alpha$ or $\bigwedge \varphi_i \Rightarrow \alpha$.
- An ***activation rule*** r_{ac} is an expression of the form $\bigwedge \varphi_i \rightarrow \psi$ or $\bigwedge \varphi_i \Rightarrow \psi$.
- An ***evaluation rule*** r_{ev} is an expression of the form $\bigwedge \varphi_i \rightarrow \neg\psi$ or $\bigwedge \varphi_i \Rightarrow \neg\psi$.

The premise of an activation and the evaluation rules is a finite set of non-ground literals whose variables have to be unified with constants in the beliefs of $\mathcal{F}$ in order to activate the sleeping goal ψ in the case of an activation rule and to refrain goal ψ of becoming pursuable in the case of evaluating rules. Let $\mathcal{R}_{st}$ be the set of all standard rules, $\mathcal{R}_{ac}$ be the set of all activation rules, and $\mathcal{R}_{ev}$ be the set of all evaluation rules.

It is important to emphasize that if there is an evaluation rule for a certain goal ψ, there must be an activation rule for it as well. This is because evaluation rules are part of the second

stage; hence, these rules can only refrain an already active goal. Therefore, it is necessary to have an activation rule that first activates it. However, the opposite is not strictly necessary as evaluation rules are not required for a goal to pass to the next stage. Indeed, the absence of evaluation rules allows a goal to always become pursuable.

3.2.2 DELIBERATION AND CHECKING STAGES

Standard, activation and evaluation rules are designed and entered by the programmer of the agent, and their content is dependent on the application domain. Otherwise, rules in deliberation and checking stages are pre-defined and no new rules of these types can be defined by the user. Before presenting the deliberation and the checking rules, let us define the beliefs that made up these kinds of rules.

Unlike the beliefs that support the first two stages, the beliefs for the deliberation and checking stages are beliefs that express something about the goals and are closely related to the agent's plans for achieving such goals. Thus, we denote with $\mathcal{PS}$ the set of plans of the agent where each plan p has the following form: $p = g : PC \leftarrow PB$ encoding a plan-body program PB for handling a goal g when the context condition PC is satisfied, where PC is a conjunction of beliefs and resources.

Both the deliberation and the checking stages are divided in two parts. In the case of the deliberation stage, the first part involves the evaluation of incompatibilities among pursuable goals and the second part is about determining the most valuable goals from the set of incompatible ones. Regarding the checking stage, the first part is about the agent's know-how, which is the same as saying whether or not the agent has at least a plan for achieving a goal, and the second one involves determining if the context of these plans is satisfied.

We believe that it is necessary to explain some details about these two stages. Although the checking stage occurs after the deliberation one, the plans associated with each goal are taken into account in both stages. In the deliberation stage, the plans are used in order to determine the emergence of incompatibilities and in the checking one, it should be verified if there is at least one plan for a goal to become executive. However, we can notice that the existence of plans for the goal has to be verified in the deliberation stage, since when there is no plan for a goal, it is not possible to determine if it has or not incompatibilities with other goals. Hence, the belief that expresses the existence of plans is generated in the deliberation stage but it is used, to generate the respective argument, in the checking stage. Note that when there is no plan for a goal, it will not pass the deliberation stage, i.e., it will not become chosen.

Recall that the deliberation stage concerns with the costs that an agent expects to sustain as a consequence of pursuing a certain goal. This is related to the necessary resources the agent needs to perform a plan that allows him to achieve such goal. Thus, the quantity of resources an agent has can be enough for achieving a goal; nevertheless, when two or more goals need the same resources, it is possible that some conflicts arise. When we say that two or more goals need the same resources we refer to the plans associated to such goals. Hence, we will compare the plans – of different goals – to determine if a conflict exists between them and then we will identify the conflict between goals based on their plans.

In order to deal with resource conflict we first define a semantic inference that works exclusively for reasoning about resources and their availability. Consider that the agent has a set of the available resources $\mathcal{RES}_{sum}$. Let Φ be a conjunction of ground atoms that represent resources that are associated to different plans. It is important to remark that the literals of Φ must represent the same type of resource. We say that $\Phi \in \mathcal{RES}_{sum}$, if $\mathcal{RES}_{sum}$ has the necessary amount of resources the ground atoms in Φ represent. In other words, if the sum of the required resources is less than the total amount of such resource in $\mathcal{RES}_{sum}$.

Definition 3.3. (Inference for resources) Let $\mathcal{RES}_{sum}$ be the set of available resources of the agent. $\mathcal{RES}_{sum}$ satisfies a nested formula Φ, denoted $\mathcal{RES}_{sum} \models_r \Phi$, recursively as follows:

- For elementary Φ, $\mathcal{RES}_{sum} \models_r \Phi$ iff $\Phi \in \mathcal{RES}$
- $\mathcal{RES}_{sum} \models_r \Phi \wedge \Psi$ iff $\mathcal{RES}_{sum} \models_r \Phi$ and $\mathcal{RES}_{sum} \models_r \Psi$
- $\mathcal{RES}_{sum} \models_r not\ \Phi$ iff $\mathcal{RES}_{sum} \not\models_r \Phi$

For instance, let $\Phi = \{battery(50) \wedge battery(80)\}$ and $\Phi' = \{oil(30) \wedge oil(40)\}$. Consider that $\mathcal{RES}_{sum} = \{(battery(100)), (oil(100))\}$. For resource battery we have that $\mathcal{RES}_{sum} \not\models_r \Phi$, which means that there is no enough battery. However, for oil we have that $\mathcal{RES}_{sum} \models_r \Phi'$ because the agent has enough oil for the two plans associated to each resource requirement.

The following notation will be used to differentiate the need of a given resource in different plans. We write res_p to represent the need of resource res in plan p, $res_{p'}$ to represent the need of resource res in a plan p', and so on.

Definition 3.4. (Resource conflict between plans) Let $p, p' \in \mathcal{PS}$ be two plans for achieving different goals, such that $p = g : PB \leftarrow PC, p' = g' : PB' \leftarrow PC'$. We say that there is a conflict between p and p' when:

- $res_p, res_{p'} \in \mathcal{RES}$ such that res_p is part of PC and $res_{p'}$ is part of PC',
- $\Phi = res_p \wedge res_{p'}$,
- $\mathcal{RES}_{sum} \models_r not\ \Phi$.

When there is a conflict between two plans of two different goals, it does not necessarily mean that both goals are conflicting, unless they are the unique plans for each goal. We say that two goals are conflicting when all their plans are conflicting; hence, there is no way of achieve them.

Definition 3.5. (Resource conflict between goals) Let $g', g'' \in \mathcal{G}$ be two goals, $\mathcal{PS}' \subseteq \mathcal{PS}$ the plans that handle goal g', and $\mathcal{PS}'' \subseteq \mathcal{PS}$ the plans that handle goal g'' such that $\mathcal{PS}' \cap \mathcal{PS}'' = \emptyset$. We say that g' and g'' are conflicting when $\forall p \in \mathcal{PS}'$ and $\forall p' \in \mathcal{PS}''$, it holds that $\mathcal{RES}_{sum} \models_r \ not\ res_p \wedge res_{p'}$.

Next definition presents the beliefs that are generated during deliberation and checking stages.

Definition 3.6. (Beliefs for deliberation and checking) Let g be the representation of a goal in $\mathcal{G}$. The first two beliefs are used in the deliberation stages and the last two in the checking stage.

- A ***non-incompatible belief*** is an expression of the form $\neg has\_incompatibility(g)$. It means that goal g has no conflicts with other pursuable goals.
- A ***value belief*** is an expression of the form $most\_valuable(g)$. It means that goal g is the most preferable goal of a set of incompatible ones.
- A ***competence belief*** is an expression of the form $has\_plans\_for(g)$. It means that there exist at least one plan for goal g.
- A ***condition belief*** is an expression of the form $satisfied\_context(g)$. It means that at least the context of one plan for achieving goal g is satisfied.

After a belief is generated, it has to be added to $\mathcal{F}$. We want to point out that the agent may also generate or perceive by communication the negation of these beliefs, which may serve for generating attacks to the arguments generated in these stages.

After having defined these beliefs, we can present the rules for deliberation and checking stages. We use strict rules to represent the rules of these two stages.

Definition 3.7. (Rules for deliberation and checking) Let g be the representation of a goal in $\mathcal{G}$. The first two rules are used in the deliberation stage and the last one in the checking stage.

- $r^1_{de} : \neg has\_incompatibility(g) \rightarrow chosen(g)$
- $r^2_{de} : most\_valuable(g) \rightarrow chosen(g)$
- $r_{ch} = has\_plans\_for(g) \wedge satisfied\_context(g) \rightarrow executive(g)$

Let $\mathcal{R}_{de}$ stand for the set of deliberation rules and $\mathcal{R}_{ch}$ stand for the set of checking rules.

3.2.3 AGENT ARCHITECTURE

We begin by defining the theory on which a BBGP-based agent is constructed. Let us recall that standard, activation, and evaluation rules may be either strict or defeasible rules whereas the deliberation and checking rules are always strict rules.

Definition 3.8. (Theory for a BBGP-based agent) A theory is a triple $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$ such that:

- $\mathcal{F}$ is the set of beliefs of the agent,
- $\mathcal{S} = \mathcal{R}'_{st} \cup \mathcal{R}'_{ac} \cup \mathcal{R}'_{ev} \cup \mathcal{R}_{de} \cup \mathcal{R}_{ch}$ is the set of strict rules, and
- $\mathcal{D} = \mathcal{R}''_{st} \cup \mathcal{R}''_{ac} \cup \mathcal{R}''_{ev}$ is the set of defeasible rules.

where $\mathcal{R}_{st} = \mathcal{R}'_{st} \cup \mathcal{R}''_{st}$, $\mathcal{R}_{ac} = \mathcal{R}'_{ac} \cup \mathcal{R}''_{ac}$, and $\mathcal{R}_{ev} = \mathcal{R}'_{ev} \cup \mathcal{R}''_{ev}$. It holds that $\mathcal{R}'_{st} \cap \mathcal{R}''_{st} = \emptyset$, $\mathcal{R}'_{ac} \cap \mathcal{R}''_{ac} = \emptyset$, and $\mathcal{R}'_{ev} \cap \mathcal{R}''_{ev} = \emptyset$.

We can now define the architecture of an intelligent agent based on the BBGP model.

Definition 3.9. (BBGP-based Agent) A BBGP-based agent is a tuple $\langle \mathcal{T}, \mathcal{RES}_{sum}, \mathcal{G}, \mathcal{PS} \rangle$ where:

- $\mathcal{T}$ is the theory of the agent,
- $\mathcal{RES}_{sum}$ is a resource summary, which contains the information about the available amount of every resource of the agent.
- $\mathcal{G} = \mathcal{G}_a \cup \mathcal{G}_p \cup \mathcal{G}_c \cup \mathcal{G}_e \cup \mathcal{G}_{canc}$ where $\mathcal{G}_a$ stands for the set of active goals, $\mathcal{G}_p$ stands for the set of pursuable goals, $\mathcal{G}_c$ stands for the set of chosen goals, $\mathcal{G}_e$ stands for the set of executive goals, and $\mathcal{G}_{canc}$ stands for the set of cancelled goals. Finally, the following condition must hold: $\mathcal{G}_x \cap \mathcal{G}_y = \emptyset$, for $x, y \in \{a, p, c, e, canc\}$ with $x \neq y$.
- $\mathcal{PS}$ is the set of plans, where each plan has the following form: $g : PC \leftarrow PB$ encoding a plan-body program PB for handling an goal g when the context condition PC is satisfied.

The agent is also equipped with a set of functions, which determine the generation of the beliefs for the deliberation and checking stages.

Definition 3.10. (Functions)

- EVAL_COMPET : $\mathcal{G}_p \rightarrow \mathcal{G}_p \times 2^{\mathcal{PS}}$. This function evaluates the competence for a goal, i.e. if there is a plan for such goal. It takes as input the set of pursuable goals and returns those ones that have at least one plan associated along with such plan(s). For all these goals, a competence belief has to be generated.
- EVAL_INCOMP : $\mathcal{G}'_p \rightarrow \mathcal{G}_{incomp}$ such that $\mathcal{G}_{incomp} = 2^{\mathcal{G}'_p} \times 2^{\mathcal{G}'_p} \times ... \times 2^{\mathcal{G}'_p}$. The incompabilities evaluation function takes as input the set pursuable goals and returns subsets of incompatible goals, taking into account the different resources. For instance, if EVAL_INCOMP returns $\{\{g, g'\}, \{g'', g'''\}\}$, it means that goals g and g' are incompatible due to a certain resource *res* and goals g'' and g''' are incompatible due to another resource *res*$'$. This function takes into account Definition 3.5 to determine the conflicting goals. A non-incompatibility belief has to be generated for each goal that does not belong to any subset returned by this function.
-PREF :$\mathcal{G} \rightarrow [0,1]$ is a function that returns the preference value of a given goal such that 0 stands for the minimum preference value and 1 for the maximum one. We use $\succ$ to denote the preference of a goal over another one.
- EVAL_VALUE : $\mathcal{G}_{incomp} \rightarrow 2^{\mathcal{G}''_p}$, it takes as input the set of subsets of incompatible goals and returns the most valuable ones of each subset, for which the preference value of each goal is used. In other words, function EVAL_VALUE calls function PREF in order to obtain the values of the incompatible goals. For all the goals returned by EVAL_VALUE, a value belief has to be created.
-EVAL_CONTEXT : $\mathcal{G}_c \times 2^{\mathcal{PS}} \rightarrow 2^{\mathcal{G}_c}$, it takes as input the result of EVAL_COMPET and returns the set of chosen goals that have the context of at least one of their associated plans satisfied. For each of these goals, a condition belief has to be generated. For the rest of the goals a negation of the condition belief has to be generated.

Once the main building blocks of the agent architecture have been defined, the argumentation process can be presented.

3.3 ARGUMENTATION PROCESS

This section is devoted to the argumentation process that is carried out in each stage in order to determine which goals pass to the next stage. This argumentation process can be decomposed into the following steps: (i) constructing arguments, (ii) determining conflicts among arguments, (iii) evaluating the acceptability of arguments, and (iv) defining the justified conclusions.

3.3.1 ARGUMENTS

There are mainly two categories of arguments. The first category justifies or attacks beliefs, while the other category justifies or attacks the passage of a goal from one stage to another.

The first category of arguments is already studied in argumentation literature, principally for handling inconsistency in knowledge bases. These arguments are called epistemic in (HARMAN, 2004). In this work, such arguments are built from the base $\mathcal{F}$ and the set of standard rules $\mathcal{R}_{st}$. In order to represent this constraint, let $\mathcal{T}' = \langle \mathcal{F}, \mathcal{R}'_{st}, \mathcal{R}''_{st} \rangle$ be the sub-theory used for constructing epistemic arguments. From now on, A, B, C and their primed versions will stand for arguments.

Definition 3.11. (Epistemic argument) Let $\mathcal{T}' = \langle \mathcal{F}, \mathcal{R}'_{st}, \mathcal{R}''_{st} \rangle$ be a sub-theory of $\mathcal{T}$. An epistemic argument constructed from $\mathcal{T}'$ is a pair $A = \langle T, \varphi \rangle$ such that:

(1) $\varphi \in \mathcal{L}_{lit}$
(2) T is a derivation schema for φ from $\mathcal{T}'$,
(3) $\mathtt{SEQ}(T)$ is consistent and T must be minimal.

$\mathtt{ARG}_{ep}$ denotes the set of all epistemic arguments that can be built from the sub-theory $\mathcal{T}'$. As for notation, $\mathtt{CONC}(A) = \varphi$ and $\mathtt{SUPPORT}(A) = T$ denote the conclusion and the support of the epistemic argument A, respectively.

The second category of arguments represents the reasons for a goal change its status. There is a set of arguments for each stage of the BBGP model. These arguments are built from the beliefs of the agent and the sets of rules of each stage. Like in the notation for rules, we use the subscript *ac* for denoting activation arguments, *ev* for denoting evaluation arguments, *de* for denoting deliberation arguments, and *ch* for denoting checking arguments. We also define a sub-theory $\mathcal{T}'' = \langle \mathcal{F}, \mathcal{S}', \mathcal{D}' \rangle$ where $\mathcal{S}' = \mathcal{S} \setminus \mathcal{R}'_{st}$ and $\mathcal{D}' = \mathcal{D} \setminus \mathcal{R}''_{st}$.

Definition 3.12. (Stage arguments) $\mathcal{T}'' = \langle \mathcal{F}, \mathcal{S}', \mathcal{D}' \rangle$ be a sub-theory of $\mathcal{T}$. Arguments built in each stage are represented by a tuple $A = \langle T, g \rangle$ such that:

(1) $g \in \mathcal{G}$
(2) For the evaluation and evaluation stages: T is a derivation schema for g from $\mathcal{T}''$
For the deliberation stage: T is a derivation schema for $\mathit{chosen}(g)$ from $\mathcal{T}''$
For the checking stage: T is a derivation schema for $\mathit{executive}(g)$ from $\mathcal{T}''$
(3) $\mathtt{SEQ}(T)$ is consistent T must be minimal.

$\texttt{ARG}_{ac}$, $\texttt{ARG}_{ev}$, $\texttt{ARG}_{de}$, and $\texttt{ARG}_{ch}$ denote the set of all activation, evaluation, deliberation, and checking arguments, respectively, which can be built from $\mathcal{T}''$.

Notice that each argument may have a set of sub-arguments.

Definition 3.13. (Sub-argument) An argument $\langle T', \varphi' \rangle$ is a sub-argument of $\langle T, \varphi \rangle$ iff $\texttt{FACTS}(T') \subseteq \texttt{FACTS}(T), \texttt{STRICT}(T') \subseteq \texttt{STRICT}(T)$, and $\texttt{DEFE}(T') \subseteq \texttt{DEFE}(T)$.

Let $\texttt{SUB}(A)$ denote the set of all sub-arguments of A.

3.3.2 ATTACKS AMONG ARGUMENTS

Arguments built from the the theory $\mathcal{T}$ constitute a cause for a goal changes its status. However, it is not a proof that the the goal should adopt another status. The reason is that an argument can be attacked by other arguments. Next, we will investigate the different kinds of conflicts among the arguments.

There are two kinds of attacks: (i) the attacks between epistemic arguments, and (ii) the mixed attacks, in which an epistemic argument attacks a stage argument.

An epistemic argument can be attacked by another epistemic argument for two main reasons: (i) they have contradictory conclusions (this is known as rebuttal), the conclusion of an epistemic argument contradicts an element of the support of another epistemic argument (this is know as undermining or less conservative undercutting in (BESNARD; HUNTER, 2009)). This attack is defined over $\texttt{ARG}_{ep}$ and is captured by the binary relation $\texttt{att}_{ep} \subseteq \texttt{ARG}_{ep} \times \texttt{ARG}_{ep}$. We denote with $(A,B) \in \texttt{att}_{ep}$ the attack relation between arguments A and B. This means that the epistemic argument A attacks the epistemic argument B.

Definition 3.14. (Rebuttal) An epistemic argument $\langle T', \varphi' \rangle$ is a rebuttal for an epistemic argument $\langle T, \varphi \rangle$ if $\varphi = \neg\varphi'$.

Definition 3.15. (Undercut)[1] An undercut for an epistemic argument $\langle T, \varphi \rangle$ is an epistemic argument $\langle T', \neg\varphi' \rangle$ where $\varphi' \in \texttt{FACTS}(T)$.

In the case of activation, evaluation, deliberation, or checking arguments, these can be attacked by an epistemic argument when the conclusion of the epistemic argument contradicts a belief of the support of the stage argument. This attack is defined over $\texttt{ARG}_{ep}$ and $\texttt{ARG}_{ac}$, $\texttt{ARG}_{ev}$, $\texttt{ARG}_{de}$, or $\texttt{ARG}_{ch}$; and is captured by the binary relation $\texttt{att}_{mx} \subseteq \texttt{ARG}_{ep} \times \texttt{ARG}_{x}$

[1] For the sake of simplicity, instead of calling this attack "less conservative undercut", we call it "undercut". We do not use the term "undermining" because we do not work with axioms and ordinary premises.

($x \in \{ac, ev, de, ch\}$). We denote with $(A,B) \in \mathtt{att}_{mx}$ the attack relation between arguments A and B. This means that the epistemic argument A attacks the stage argument B.

Definition 3.16. (Mixed undercut) An undercut for an activation (evaluation, deliberation or checking) argument $A = \langle T, g \rangle$ is an epistemic argument $\langle T', \neg\varphi \rangle$ where $\varphi \in \mathtt{FACTS}(T)$.

3.3.3 EVALUATING THE ACCEPTABILITY OF ARGUMENTS

This evaluation is important because it determines which goals pass from one stage to the next. First, we define an AF and then we show how the evaluation is done. It is generated a different AF for each stage of the BBGP model, which includes epistemic arguments and stage arguments. In order to differentiate the AFs, we use the same notation we have used for arguments.

Definition 3.17. (Argumentation framework) An argumentation framework $\mathcal{AF}_x$ is a pair $\mathcal{AF}_x = \langle \mathtt{ARG}, \mathtt{att} \rangle$ ($x \in \{ac, ev, de, ch\}$) such that:

- $\mathtt{ARG} = \mathtt{ARG}_x \cup \mathtt{ARG}'_{ep} \cup \mathtt{SUBARGS}$, where $\mathtt{ARG}_x$ is the set of arguments generated for the activation, evaluation, deliberation, or checking stage, $\mathtt{ARG}'_{ep} = \{A \mid A \in \mathtt{ARG}_{ep}$ and $(A,B) \in \mathtt{att}_{mx}$ or $(A,C) \in \mathtt{att}_{ep}\}$, where $B \in \mathtt{ARG}_x$ and $C \in \mathtt{ARG}'_{ep}$, and $\mathtt{SUBARGS} = \bigcup_{A \in \mathtt{ARG}_x, A \in \mathtt{ARG}'_{ep}} \mathtt{SUB}(A)$.
- $\mathtt{att} = \mathtt{att}_{ep} \cup \mathtt{att}_{mx}$, where $\mathtt{att}_{ep}$ is the set of attacks between epistemic arguments in $\mathtt{ARG}'_{ep}$ and $\mathtt{att}_{mx}$ is the set of mixed attacks of arguments in $\mathtt{ARG}'_{ep}$ to arguments in $\mathtt{ARG}_x$.

The next step is to evaluate the arguments that make part of the AF by taking into account the attacks among them. The aim is to obtain a subset of ARG without conflicting arguments. In order to obtain it, we use an *acceptability semantics* (DUNG, 1995). The idea of a semantics is that given an AF, it determines zero or more sets of acceptable arguments (i.e., non-conflicting arguments), which are also called *extensions*. In our case, a semantics will determine if a given goal will become (i) active, which happens when at least one activation argument, supporting it, belongs to an extension, (ii) pursuable, which happens when no evaluation argument, attacking the goal, belongs to an extension, (iii) chosen, which happens when at least one deliberation argument, supporting it, belongs to an extension, or (iv) executive, which happens when at least one checking argument, supporting it, belongs to an extension.

In order to reason with a semantics, one has to take either a credulous or a sceptical perspective. That is, an argument is accepted with respect to a semantics if the argument is part of at least one extension returned by the semantics (the credulous perspective) or if the argument

is part of all the extensions returned by the semantics (the sceptical perspective). This means that depending on the semantics that will be used, the reasoning of the agent can be considered either credulous or sceptical.

Next, the main definitions about the semantics introduced by Dung are recalled.

Definition 3.18. (Semantics) Let $\mathcal{AF}_x = \langle \texttt{ARG}, \texttt{att} \rangle$ be an AF (with $x \in \{ac, ev, de, ch\}$) and $\mathcal{E} \subseteq \texttt{ARG}$:

- $\mathcal{E}$ is ***conflict-free*** iff there is not $A, A' \in \mathcal{E}$ such that A attacks A'.
- $\mathcal{E}$ ***defends*** an argument A iff for each argument $A' \in \texttt{ARG}$, if A' attacks A, then there exist an argument $A'' \in \mathcal{E}$ such that A'' attacks A'.
- $\mathcal{E}$ is ***admissible*** iff it is conflict-free and defends all its elements.
- A conflict-free $\mathcal{E}$ is a ***complete extension*** iff we have $\mathcal{E} = \{A | \mathcal{E} \text{ defends } A\}$.
- $\mathcal{E}$ is a ***preferred extension*** iff it is a maximal (w.r.t the set inclusion) complete extension.
- $\mathcal{E}$ is a ***grounded extension*** iff is a minimal (w.r.t. set inclusion) complete extension.

Depending on the type of reasoning of the agent about the semantics, the status of acceptability of an argument varies.

Definition 3.19. (Argument status) Let $\mathcal{AF}_x = \langle \texttt{ARG}, \texttt{att} \rangle$ be an AF (with $x \in \{ac, ev, de, ch\}$), $\mathcal{E}_1, ..., \mathcal{E}_n$ its extensions under a given semantics, and $A \in \texttt{ARG}$ an argument.

- A is *sceptically accepted* iff $A \in \mathcal{E}_i$, $\forall \mathcal{E}_i$, with $i = 1, ..., n$.
- A is *credulously accepted* iff $\exists \mathcal{E}_i$, such that $A \in \mathcal{E}_i$.

In order to determine which semantics is the most adequate in the context of the problem we are tackling, let us present the cases illustrated in Figure 5, where arguments A, B, and C are epistemic arguments and argument D is a stage argument. We denote with $\mathcal{E}_p$ the set of preferred extensions and with $\mathcal{E}_g$ the set of grounded extensions.

- **Case (a):** This is a very basic case, where an epistemic argument attacks a stage argument. There is only one preferred extension, which is the same as the grounded extension: $\mathcal{E}_p = \mathcal{E}_g = \{A\}$. We can say that A is sceptically accepted with respect to both semantics.
- **Case (b):** We have two epistemic arguments that attack each other and one of them attacks a stage argument. There are two preferred extensions $\mathcal{E}_p = \{\{B\}, \{A, D\}\}$ and one grounded extension $\mathcal{E}_g = \{\}$. In both semantics there is no sceptically accepted argument; however, in preferred semantics all the arguments are credulously accepted.

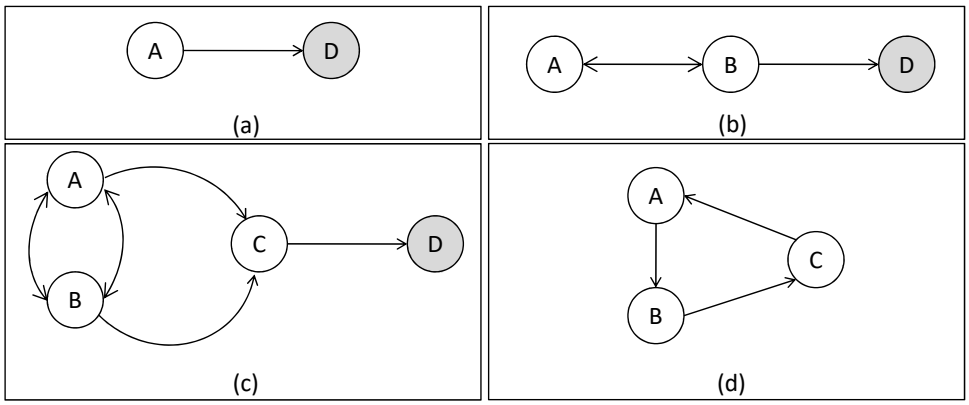


Figure 5: Four possible configurations for analyzing the preferred and grounded semantics. Arguments A, B, and C are epistemic arguments and argument D is a stage argument. We have used white coloured circles for epistemic arguments and grey coloured circles for the stage argument.

- **Case (c):** We have three epistemic arguments and some attacks between them; besides one epistemic argument attack one stage argument. There are two preferred extensions $\mathcal{E}_p = \{\{B,D\},\{A,D\}\}$ and one grounded extension $\mathcal{E}_g = \{\}$. In $\mathcal{E}_g$ there is no accepted argument; however, in $\mathcal{E}_p$ D is a sceptically accepted argument and arguments B and A are credulously accepted.

- **Case (d):** We only have epistemic arguments and some attacks between them. There is one preferred extension, which is the same as the grounded extension: $\mathcal{E}_p = \mathcal{E}_g = \{\}$. In this case, there is no accepted argument neither sceptically nor credulously.

Against this background, we state the following:

1. Although it is true that both semantics always return an extension, in some cases it is the empty extension. In the case of the grounded semantics, given its extremely sceptical nature, it is more notorious. In the context of the problem that is being tackled, an empty extension means that the status of no goal will change and there is no reasons that support it. We can also note that preferred semantics also return an empty extension in case (d), even when this semantics has a credulous nature. However, in this case there is no stage arguments involved, which means that the status of a goal would not change either. Therefore, it is not desirable to use a semantics that has an extremely sceptical nature, like the grounded one.

2. In cases (b) and (c) we can observe that there are two preferred extensions. The difference is that in case (b) all the arguments are credulously accepted and in case (c), arguments A and B are credulously accepted and argument D is sceptically accepted. In case (b),

one of the extensions has the stage argument and the other does not, and in case (c) both extensions have the stage argument but with different epistemic arguments. In both cases, one of the extensions has to be selected. Since the aim is the status of goals, it is clear that an extension that has as element a stage argument is preferred than an extension that only has epistemic arguments. Therefore, we think that preferred semantics have a more adequate behavior for dealing with the tackled problem. We also want to point out that extensions that include stage arguments are more preferred than extensions with only epistemic arguments. Thus, the status of an argument is not determinant for the selection of an extension. For example, consider that $\mathcal{E}_p = \{\{A,D\},\{A,C\}\}$, in this case argument A is sceptically accepted and the others are credulously accepted. If we prioritize the sceptical status, we could choose $\{A,C\}$ and no stage argument would be accepted.

3. Finally, we can note that in case (a) there is a sceptically accepted argument, which is an epistemic argument. In this case, there is no way for the stage argument to be accepted; however, we have an accepted argument that gives a reason for this situation and this is completely valid.

We have analysed the three possible scenarios that may occur when applying preferred and grounded semantics. In the first scenario (Case (a)), both semantics return the same non-empty extension, in the second scenario (cases (b) and (c)) the semantics return different extensions, and in the last scenario (Case (d)) both semantics return the empty extension. There is no other possible scenario. From this analysis, we can conclude that preferred semantics have a better behavior than grounded ones. This mainly occurs because even though a grounded extension always exist it can be the empty set. Preferred extension may also be empty; however, it does not have consequences on any stage argument.

Since it is likely that there are more than two preferred extensions, next we propose some criteria that can be taken into account in order to select a preferred extension.

- Given or more preferred extensions, those with stage arguments should be selected.
- When there are more than two preferred extensions with stage arguments, it should be selected the extension that allows the passage of the greater number of goals.
- When there are more than two preferred extensions with stage arguments, it should be selected the extension that allows to maximize the preference value of the goals.

Selecting an extension is important because the arguments that belong to this extension

determine the goals that pass (or not) to the next stage. We call the arguments that belong to the selected extension of the accepted arguments.

Definition 3.20. (Accepted argument) Let $\mathcal{AF}_x = \langle \texttt{ARG}, \texttt{att} \rangle$ be an AF for stage x (with $x \in \{ac, ev, de, ch\}$), $\mathcal{E}_p$ the set of extensions under the preferred semantics, and $\mathcal{E} \in \mathcal{E}_p$ the selected preferred extension. An argument $A \in \texttt{ARG}$ is accepted iff $A \in \mathcal{E}$.

The last step of an argumentation process consists in determining the set of justified conclusions. We call justified conclusion the conclusion that is supported by at least one argument of the selected extension. In this approach, this set may be formed by beliefs (claims of epistemic arguments) and goals (claims of stage arguments). The set of justified goals is especially important because these goals will pass from one stage to the next.

Definition 3.21. (Justified goals) Let $\mathcal{E}$ be a selected preferred extension. For the activation, deliberation, and checking stages, the passage of a goal g is justified iff $\exists\ A \in \mathcal{E}$ such that $\texttt{CONC}(A) = g$. For the evaluation stage, the passage of a goal g is justified iff $\nexists\ A \in \mathcal{E}$ such that $\texttt{CONC}(A) = g$.

3.4 GLOBAL STRUCTURES

So far we have defined the necessary beliefs, rules and arguments for the goal processing. In this section, we present the *goal life-cycle* that describes the states and transition relationships of goals. We also present a *memory record* that saves the cognitive path of each goal and an *internal control structure* in charge of supervising goal processing as a whole.

Figure 6 shows all the possible transitions of a goal from its sleeping status until it becomes executive, which happens when all the conditions are favorable for it, i.e. when there is at least one acceptable activation argument supporting it, there is no evaluation argument attacking it, there is a deliberation argument supporting it and a checking one also supporting it. Notice that in the life-cycle of goals the cancelled status is also considered. Thus, a goal is cancelled when:

- It is deactivated, which happens when the agent receives new information that lets him generate new epistemic arguments such that after the calculation of the preferred extensions from the $\mathcal{AF}_{ac}$, the activation argument(s) that supported it are no longer accepted.

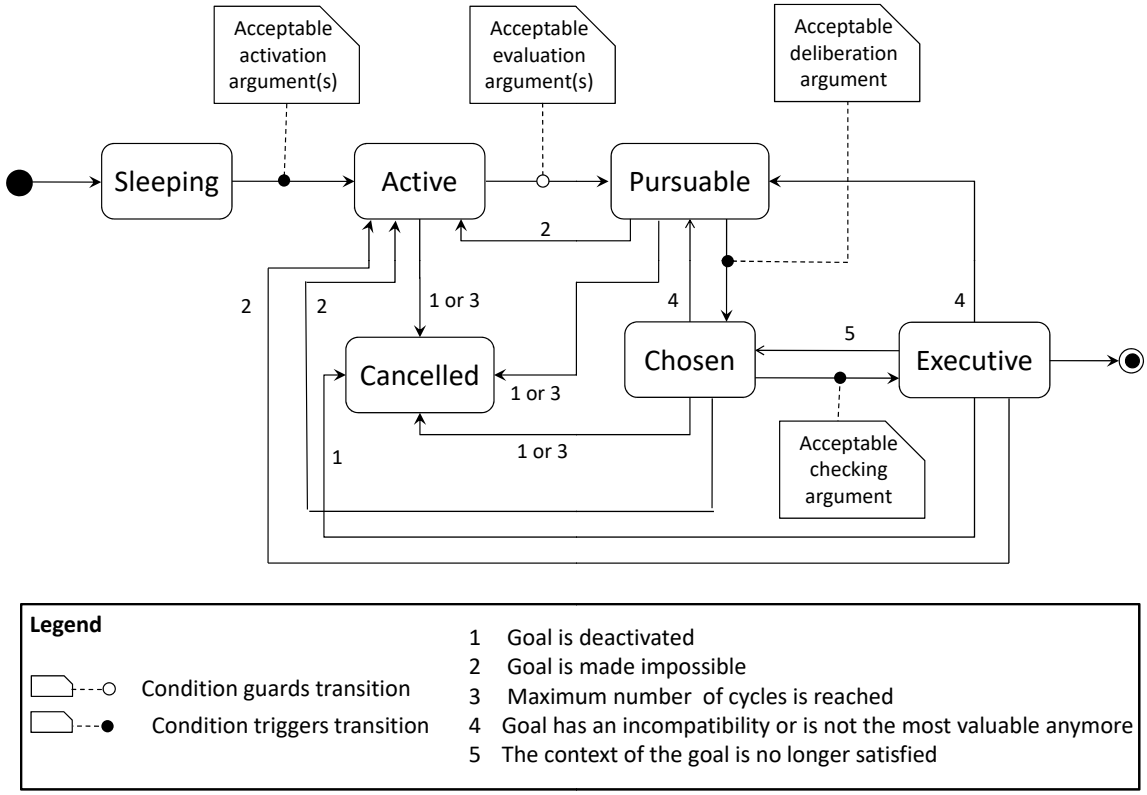


Figure 6: Life-cycle of goals.

- The maximum number of cycles for stay in a state is reached; it means that a goal cannot stay in a certain status during the whole life of the agent. This is even more important when the status of the goal is chosen since some resources could be reserved for it and, if it never becomes executive, such resources could have been used to reach other goals. When this happens, the agent generates a belief $max(cycles)$ and an epistemic argument based on it: $\langle\{(max(cycles),\emptyset)\},max(cycles)\rangle$.

Observe that a pursuable or a chosen goal can be moved back to the active status when the agent receives new information that leads to the recalculation of the preferred extensions in the evaluation stage and the selected preferred extension includes an evaluation argument that attacks such goal. We can also observe that an executive goal can go back to the chosen status, this happens when (i) the executive goal has not been executed yet and (ii) new knowledge leads to the construction of an argument or arguments that attack a checking argument and after the recalculation of the preferred extension these attacker arguments are accepted.

The agent needs to store information about the progress of his goals – considering the life-cycle of the goals – in order to have a record about all the changes in the statuses of his goals and the causes of these changes. We could say that each AF stores these causes in each

accepted argument. Therefore, we need to save both the status of the goal and the AF that supports this status. However, the AFs may be updated after each cycle, which may alter the selected preferred extensions. Thus, the saved AF is the one that supports such status. Since the saved AF is focused on a given goal g, it has to include the stage arguments whose claims are g, the epistemic arguments that attack and defend such stage arguments, and the sub-arguments of such stage and epistemic arguments.

Definition 3.22. (Sub-argumentation framework) Let $\mathcal{AF}_x = \langle \texttt{ARG}, \texttt{att} \rangle$ (with $x \in \{ac, ev, de, ch\}$) be the an AF, $g \in \mathcal{G}$ a goal, and $\texttt{ARG}' \subseteq \texttt{ARG}$ a set of arguments such that $\texttt{ARG}'$ includes all the arguments related to goal g. We denote by $\texttt{att} \otimes \texttt{ARG}'$ the subset of att that involves just the arguments in $\texttt{ARG}'$. An argumentation framework $\mathcal{AF}'_x = \langle \texttt{ARG}', \texttt{att}' \rangle$ is a *sub-AF* of $\mathcal{AF}_x$, denoted $\mathcal{AF}'_x \sqsubseteq \mathcal{AF}_x$, iff $\texttt{ARG}' \subseteq \texttt{ARG}$ and $\texttt{att}' = \texttt{att} \otimes \texttt{ARG}'$.

Next, we define the structure that stores the causes of the change of the status of a goal, which must be updated after a new transition occurs. We can see this structure as a table record, where each row saves the status of a goal along with the AF that supports such status.

Definition 3.23. (Goal memory record) Let $\mathcal{AF}_x = \langle \texttt{ARG}, \texttt{att} \rangle$ be an AF (with $x \in \{ac, ev, de, ch\}$), $\mathcal{AF}'_x \sqsubseteq \mathcal{AF}_x$ a sub-AF, and $g \in \mathcal{G}$ a goal. The memory record $\mathcal{MR}_g$ for goal g is a set of ordered pairs $(\texttt{STA}, \texttt{REASON})$ such that:

- $\texttt{STA} \in \{a, p, c, e, canc, \text{not } a, \text{not } p, \text{not } c, \text{not } e\}$ where $\{a, p, c, e, canc\}$ represent the status g attains due to the arguments in REASON whereas $\{\text{not } a, \text{not } p, \text{not } c, \text{not } e\}$ represent the status g cannot attain due to the arguments in REASON.
- $\texttt{REASON} = \mathcal{AF}'_x$ such that $\mathcal{AF}'_x$ is the sub-AF whose selected preferred extension supports the current status of g. When g becomes cancelled due to the maximum number of cycles is being reached, $\texttt{REASON} = \langle \{(max(cycles), \emptyset)\}, max(cycles) \rangle$.

Let $\mathcal{MR}^+$ be the set of all memory records and $\texttt{NUM\_REC} : \mathcal{MR}^+ \rightarrow \mathbb{N}$ a function that returns the number of records of a given $\mathcal{MR}$.

The goal memory record is a structure linked with each agent's goal. However, a structure that controls the dynamic behavior of the system during the life-cycle of all goals is also necessary.

Definition 3.24. (Goal processing controller) The goal processing controller is represented by a tuple $\mathcal{GC} = \langle \mathcal{G}, \mathcal{MR}^+, \eta, \mu, \mathcal{M}onitor \rangle$, such that:

- $\mathcal{G}$ is the set of goals,
- $\mathcal{MR}^+$ is the set of all $\mathcal{MR}$,
- η is the maximum number of cycles a goal can remain in a stage,
- $\mu : \mathcal{G} \rightarrow \mathcal{MR}^+$ maps a given goal to its individual memory register,
- $\mathcal{M}onitor$ is a module in charge of supervising the status of the goals and keeping the consistency in the goal processing system by executing the following tasks:

  1. Moving goals from $\mathcal{G}_a$, $\mathcal{G}_p$, $\mathcal{G}_c$ to $\mathcal{G}_{canc}$, after η is reached.
  2. Releasing resources when a chosen goal is moved from $\mathcal{G}_c$ to $\mathcal{G}_{canc}$, $\mathcal{G}_a$ or $\mathcal{G}_p$.
  3. Moving deactivated goals from $\mathcal{G}_a$, $\mathcal{G}_p$, $\mathcal{G}_c$ or $\mathcal{G}_e$ to $\mathcal{G}_{canc}$. A goal g is activated when there is one or more acceptable activation arguments supporting it. It is important to highlight that executive goals that are moved to $\mathcal{G}_{canc}$ are those that have not been executed yet. This condition is also valid for other tasks that involve executive goals.
  4. Moving goals from $\mathcal{G}_p$, $\mathcal{G}_c$ or $\mathcal{G}_e$ to $\mathcal{G}_a$.
  5. Moving goals from $\mathcal{G}_c$ or $\mathcal{G}_e$ to $\mathcal{G}_p$.
  6. Moving goals from $\mathcal{G}_e$ to $\mathcal{G}_c$.

This structure is unique for each agent and the parameter η must be defined by the programmer. Every task that $\mathcal{M}onitor$ performs is important for the consistent and correct functioning of the goal processing system.

3.5 PROPERTIES

In this section, we demonstrate how our approach satisfies the desirable properties specified in (CASTELFRANCHI; PAGLIERI, 2007), that is, diachrony and synchrony. We also make an analysis about the rationality postulates satisfied by this approach. These postulates, proposed in (CAMINADA; AMGOUD, 2007), are used to judge the quality of argumentation systems that are rule-based and are an instance of the abstract argumentation.

Notation: Hereafter we use subscripts to refer to the arguments that make part of the four AFs. We use the same subscripts employed for AFs. Thus, *ep, ac, ev, de,* and *ch*

denote epistemic, activation, evaluation, deliberation, and checking arguments, respectively. We employ the same notation to refer to the selected preferred extensions.

We first prove that the change of status of goals is always supported by arguments. In other words, we prove that the approach satisfies the property of diachrony. The diachronic behavior of the approach is summarized in Proposition 3.1. For a better comprehension, we explain the intended meaning of each item of such proposition:

- Both items (1.a) and (1.b) concern the active status of a goal. The former holds that a goal becomes active when there is at least one accepted activation argument and the latter is related to condition 2 of the life-cycle of goals; thus, it mandates that a goal returns to its active status when there is at least one accepted evaluation argument.

- Both items (2.a) and (2.b) concern the pursuable status of a goal. The former holds that a goal becomes pursuable when there is not an accepted evaluation argument that refrains its passage to the next stage. The latter is related to condition 4 of the life-cycle of goals; thus, it mandates that a goal returns to its pursuable status when there exist at least one accepted epistemic argument attacks an (previously accepted) deliberation argument. This means that such epistemic argument(s) is the support for a goal to become pursuable again.

- Both items (3.a) and (3.b) concern the chosen status of a goal. The former holds that a goal becomes chosen when there is an accepted deliberation argument. The latter is related to condition 5 of the life-cycle of goals; thus, it mandates that a goal returns to its chosen status when there is at least one accepted epistemic argument attacks an (previously accepted) checking argument. This means that such epistemic argument(s) gives a reason for a goal become chosen again.

- Item (4) concerns the executive status of a goal. It holds that a goal becomes executive when and there is an accepted checking argument. In this case, there is not a case (b) because this is the last status of a goal.

- Both items (5.a) and (5.b) concern the cancelled status of a goal. The former is related to condition 1 of the life-cycle of goals, i.e., a goal is deactivated due to an attack of an accepted epistemic argument. If there was more than an activation argument that triggered the activation of the goal, none of them must be part of the current selected preferred extension. The latter one is related to the condition 5, it means that the maximum number of cycles has been reached.

Proposition 3.1. (Diachrony) Let $g \in \mathcal{G}$ be a goal, $\mathcal{AF}_{ac}$, $\mathcal{AF}_{ev}$, $\mathcal{AF}_{de}$, and $\mathcal{AF}_{ch}$ the four argumentations frameworks involved in that goal's processing and $\mathcal{E}_{ac}$, $\mathcal{E}_{ev}$, $\mathcal{E}_{de}$, and $\mathcal{E}_{ch}$ the selected preferred extensions obtained from each of them, respectively.

(1.a) If $g \in \mathcal{G}_a$, then $\exists A_{ac} \in \mathcal{E}_{ac}$ such that $\texttt{CONC}(A_{ac}) = g$.

(1.b) If $g \in \mathcal{G}_a$, then $\exists A_{ev} \in \mathcal{E}_{ev}$ such that $\texttt{CONC}(A_{ev}) = g$.

(2.a) If $g \in \mathcal{G}_p$, then $\nexists A_{ev} \in \mathcal{E}_{ev}$ such that $\texttt{CONC}(A_{ev}) = g$.

(2.b) If $g \in \mathcal{G}_p$, then $\exists A_{ep} \in \mathcal{E}_{de}$ such that $(A_{ep}, A_{de}) \in \texttt{att}$, $\texttt{CONC}(A_{de}) = g$, and $\texttt{att}$ is the attack relation of $\mathcal{AF}_{de}$.

(3.a) If $g \in \mathcal{G}_c$, then $\exists A \in \mathcal{E}_{de}$ such that $\texttt{CONC}(A) = g$.

(3.b) If $g \in \mathcal{G}_c$, then $\exists A_{ep} \in \mathcal{E}_{ch}$ such that $(A_{ep}, A_{ch}) \in \texttt{att}$, $\texttt{CONC}(A_{ch}) = g$, and $\texttt{att}$ is the attack relation of $\mathcal{AF}_{ch}$.

(4) If $g \in \mathcal{G}_e$, then $\exists A_{ch} \in \mathcal{E}_{ch}$ such that $\texttt{CONC}(A_{ch}) = g$.

(5.a) If $g \in \mathcal{G}_{canc}$, then $\forall A_{ac} \exists A_{ep} \in \mathcal{E}_{ac}$ such that $(A_{ep}, A_{ac}) \in \texttt{att}$, $\texttt{CONC}(A_{ac}) = g$, and $\texttt{att}$ is the attack relation of $\mathcal{AF}_{ac}$.

(5.b) If $g \in \mathcal{G}_{canc}$, then $\exists A_{ep} \in \mathcal{E}_{ev}, \mathcal{E}_{de}$, or $\mathcal{E}_{ch}$ such that $A_{ep} = \langle \{(max(cycles), \emptyset)\}, max(cycles) \rangle$.

The proof of this proposition and the proofs of the next propositions and theorems are given in an appendix at the end of the document.

The second property of this approach is related to the synchrony, which means that the agents save a cognitive path for explaining how a given goal reach out to its current status. Notice that for satisfying diachrony there should exist an argument that justifies the change in the status of a goal. For satisfying the property synchrony, there should be a set of arguments, which represent the cognitive path, that allow the agent to explain in detail the whole process since the goal was in its sleeping status until its current status. This explanation is saved in each AF, where arguments represent reasons in favor and against the change of the status of a goal. We can distinguish two kinds of explanations:

1. **Partial explanation:** It is given by the selected preferred extension, which contains the stage argument (or arguments) that make a goal change its status and other arguments that support this arguments, i.e., that do not attack it. Recall that when a goal becomes

pursuable, the preferred extension contains only epistemic arguments. Thus, from a partial perspective the agent can explain why a given goal is in its current status giving only positive reasons.

2. **Complete explanation:** For giving a complete explanation about a given goal, the agent takes into account the arguments of the selected preferred extension and the other arguments of the AF related to the goal, i.e., the arguments whose claim is the goal and the epistemic arguments that attack them or defend them. Thus, he can give a more detailed explanation by including not only the positive reason but also those negative ones and how these were defeated.

Next proposition summarizes the aforementioned. We use the goal memory record in order to obtain the cognitive path that allows the goal to reach out to its current status. For the complete explanation, we consider the union of all of the sub-AFs that are saved in REASON and for the partial explanation, we take into account the selected preferred extension that can be obtained from each sub-AF.

Proposition 3.2. (Synchrony) Let $g \in \mathcal{G}$ be a given goal, $\mathcal{MR}_g$ the memory record of goal g, and $\mathcal{AF}_{ac}$, $\mathcal{AF}_{ev}$, $\mathcal{AF}_{de}$, and $\mathcal{AF}_{ch}$ the four argumentation frameworks involved in the goal processing. Consider that $x \in \{a, p, c, e, canc\}$ denotes the status of g.

For complete explanation: If $g \in \mathcal{G}_x$, then $\exists \mathcal{AF}_t$ such that $\mathcal{AF}_t = \bigcup_{i=1}^{i=\texttt{NUM\_REC}(\mathcal{MR}_g)} \texttt{REASON}_i$, where $\texttt{REASON}_i \sqsubseteq \mathcal{AF}_{ac}$, $\texttt{REASON}_i \sqsubseteq \mathcal{AF}_{ev}$, $\texttt{REASON}_i \sqsubseteq \mathcal{AF}_{de}$, $\texttt{REASON}_i \sqsubseteq \mathcal{AF}_{ch}$, or $\texttt{REASON}_i = \langle \{(max(cycles), \emptyset)\}, max(cycles) \rangle$.

For partial explanation: If $g \in \mathcal{G}_x$, then $\exists \texttt{ARG}'$ such that $\texttt{ARG}' = \bigcup_{i=1}^{i=\texttt{NUM\_REC}(\mathcal{MR}_g)} \mathcal{E}_i$, where $\mathcal{E}_i$ is the selected preferred extension obtained from $\texttt{REASON}_i$.

The next analysis is very important. This shows that the proposed argumentation system satisfies the "direct consistency" rationality postulate proposed in (CAMINADA; AMGOUD, 2007). In order to analyse our approach with respect to this rationality postulate, let us recall the definition of justified conclusions.

Definition 3.25. (Justified conclusions) Let $\mathcal{AF}_x = \langle \texttt{ARG}, \texttt{att} \rangle$ (for $x \in \{ac, ev, de, ch\}$) be an AF and $\{\mathcal{E}_1, ..., \mathcal{E}_n\}$ its extensions under the preferred semantics.

- $\texttt{CONCS}(\mathcal{E}_i) = \{\texttt{CONC}(A) | A \in \mathcal{E}_i\}$ (for all $1 \leq i \leq n$).
- $\texttt{Output} = \bigcap_{i=1...n} \texttt{CONCS}(\mathcal{E}_i)$.

$\texttt{CONCS}(\mathcal{E}_i)$ denotes the justified conclusions for a given extension $\mathcal{E}_i$ and Output denotes the conclusions that are supported by at least one argument in each extension.

An argumentation system satisfies direct consistency if its set of justified conclusions and the different sets of conclusions corresponding to each extension are consistent. This property is important in our approach because it guarantees that the beliefs that support the passage of the goals to other status are consistent and therefore, the agent pursues goals based on consistent supporting beliefs.

Proposition 3.3. (Direct consistency) Let $\mathcal{AF}_x = \langle \texttt{ARG}, \texttt{att} \rangle$ be an AF constructed from the theory $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$. Let $\mathcal{E}_1, ..., \mathcal{E}_n$ be the set of extensions under the preferred semantics. $\mathcal{AF}_x$ satisfies direct consistency iff:

(1) $\texttt{CONCS}(\mathcal{E}_i)$ is consistent for each $1 \leq i \leq n$.

(2) `Output` is consistent.

3.6 APPLICATION: RESCUE ROBOTS SCENARIO

In this section, we present the application of the proposed approach to the rescue robots scenario. Thus, we use the argumentation-based BBGP model to support the processing of some goals of a robot agent, which is performing rescue tasks.

Firstly, let us present the starting mental states of the robot agent, let us call him `BOB`:

$\mathcal{G}_s = \{\mathsf{g}^1_\mathsf{s}, \mathsf{g}^2_\mathsf{s}, \mathsf{g}^3_\mathsf{s}\}$
$\mathcal{G} = \{\}$, which means that $\mathcal{G}_a = \{\}$, $\mathcal{G}_p = \{\}$, $\mathcal{G}_c = \{\}$, $\mathcal{G}_e = \{\}$, and $\mathcal{G}_{canc} = \{\}$
$\mathcal{R}_{st} = \{\mathsf{r}^1_\mathsf{st}, \mathsf{r}^2_\mathsf{st}, \mathsf{r}^3_\mathsf{st}, \mathsf{r}^4_\mathsf{st}\}$, $\mathcal{R}_{ac} = \{\mathsf{r}^1_\mathsf{ac}, \mathsf{r}^2_\mathsf{ac}, \mathsf{r}^3_\mathsf{ac}\}$, $\mathcal{R}_{ev} = \{\mathsf{r}^1_\mathsf{ev}, \mathsf{r}^2_\mathsf{ev}\}$,
$\mathcal{R}_{de} = \{\mathsf{r}^1_\mathsf{de}, \mathsf{r}^2_\mathsf{de}\}$, and $\mathcal{R}_{ch} = \{\mathsf{r}_\mathsf{ch}\}$
$\mathcal{F} = \{\mathsf{b}_1, \mathsf{b}_2, \mathsf{b}_3, \mathsf{b}_4, \mathsf{b}_5, \neg\mathsf{b}_6, \mathsf{b}_8, \mathsf{b}_9, \mathsf{b}_{10}, \mathsf{b}_{11}, \mathsf{b}_{12}, \mathsf{b}_{13}\}$

The detail of each set is presented below:

Sleeping goals	
- $\mathsf{g}^1_\mathsf{s} = \mathsf{take\_hospital(x)}$	*//take a person x to the hospital*
- $\mathsf{g}^2_\mathsf{s} = \mathsf{go(x,y)}$	*//go to zone (x,y)*
- $\mathsf{g}^3_\mathsf{s} = \mathsf{send\_shelter(x)}$	*//send a person x to the shelter*
Standard Rules	
- $\mathsf{r}^1_\mathsf{st} = \mathsf{new\_supply(x)} \rightarrow \mathsf{available(x)}$	*//if there is a new supply x, then x is available*
- $\mathsf{r}^2_\mathsf{st} = \mathsf{has\_fract\_bone(x)} \Rightarrow \mathsf{injured\_severe(x)}$	*//if x has a fractured bone, then x is severely injured*
- $\mathsf{r}^3_\mathsf{st} = \mathsf{fract\_bone(x,arm)} \Rightarrow \neg\mathsf{injured\_severe(x)}$	*//if the fractured bone is in the arm, then x is not severely injured*
- $\mathsf{r}^4_\mathsf{st} = \mathsf{open\_fracture(x)} \rightarrow \mathsf{injured\_severe(x)}$	*//if x has an open fracture, then x is severely injured*

Activation rules	
- $r^1_{ac} = \text{injured\_severe}(x) \rightarrow \text{take\_hospital}(x)$	*//if person x is severely injured, then take x to the hospital*
- $r^2_{ac} = \neg\text{injured\_severe}(x) \rightarrow \text{send\_shelter}(x)$	*//if person x is not severely injured, then send x to the shelter*
- $r^3_{ac} = \text{asked\_for\_help}(x,y) \Rightarrow \text{go}(x,y)$	*//if* BOB *is asked for help in zone* (x,y)*, then go to that zone*
Evaluation rules	
- $r^1_{ev} = \text{greater}(\text{weight}(x),80) \rightarrow \neg\text{take\_hospital}(x)$	*//If person x weights more than 80 kilos, then it is not possible to take him/her to the hospital*
- $r^2_{ev} = \neg\text{available}(\text{bed},x) \rightarrow \neg\text{take\_hospital}(x)$	*//If there is no available bed for x, then it is not possible to take x to hospital*
Deliberation rules	
- $r^1_{de} : \neg\text{has\_incompatibility}(g) \rightarrow \text{chosen}(g)$	
- $r^2_{de} : \text{most\_valuable}(g) \rightarrow \text{chosen}(g)$	
Checking rule	
- $r_{ch} = \text{has\_plans\_for}(g) \wedge \text{satisfied\_context}(g) \rightarrow \text{executive}(g)$	
Beliefs	
- $b_1 = \text{be\_operative}(\text{me})^2$	*//*BOB *is operative*
- $b_2 = \text{has\_fract\_bone}(\text{man\_32})$	*//There is a 32-year-old man with a fractured bone*
- $b_3 = \text{fract\_bone}(\text{man\_32},\text{arm})$	*//The 32-year-old man has a fractured arm.*
- $b_4 = \text{asked\_for\_help}(2,6)$	*//There is an aid request in slot (2,6).*
- $b_5 = \text{open\_fracture}(\text{man\_32})$	*//The 32-year-old man has an open fracture.*
- $\neg b_6 = \neg\text{available}(\text{bed},\text{man\_32})$	*//There is no an available bed.*
- $b_8 = \text{new\_supply}(\text{bed})$	*//There is a new supply.*
- $b_9 = \text{weight}(\text{man\_32},70)$	*//man_32 weights 70 kg.*

Beliefs b_{10}, b_{11}, b_{12}, and b_{13} are generated by applying the functions related to deliberation an checking stages. Even though such functions are applied at the beginning of a reasoning cycle, we will show in detail both the functions and the beliefs in the subsections devoted to such stages.

Thus, we have that the theory of agent BOB is: $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$ where $\mathcal{F} = \{b_1, b_2, b_3, b_4, b_5, \neg b_6, b_8, b_9, b_{10}, b_{11}, b_{12}, b_{13}\}$, $\mathcal{S} = \{r^1_{st}, r^4_{st}, r^1_{ac}, r^2_{ac}, r^1_{ev}, r^2_{ev}, r^1_{de}, r^2_{de}, r_{ch}\}$, and $\mathcal{D} = \{r^2_{st}, r^3_{st}, r^3_{ac}\}$.

3.6.1 ACTIVATION STAGE

Based on the current mental state of agent BOB, seven epistemic arguments (denoted by subscript ep) and four activation arguments (denoted by subscript ac), that are related to the activation stage, can be generated. Table 1 shows these arguments, which are constructed from $\mathcal{T}$ by unifying constants of the beliefs with variables of standard and activation rules. Notice that the conclusion of epistemic arguments $\mathsf{A}^7_{\mathsf{ep}}$ and $\mathsf{A}^9_{\mathsf{ep}}$ is $\mathsf{severe\_injure(man\_32)}$, which we call b_7. Figure 7 shows the attacks that arise between arguments. Notice that there are rebuttals between epistemic arguments and undercuts to activation arguments.

Let us now define the AF for this stage: $\mathcal{AF}_{ac} = \langle\{\mathsf{A}^1_{\mathsf{ep}}, \mathsf{A}^2_{\mathsf{ep}}, \mathsf{A}^3_{\mathsf{ep}}, \mathsf{A}^4_{\mathsf{ep}}, \mathsf{A}^5_{\mathsf{ep}}, \mathsf{A}^7_{\mathsf{ep}}, \mathsf{A}^1_{\mathsf{ac}}, \mathsf{A}^2_{\mathsf{ac}}, \mathsf{A}^3_{\mathsf{ac}}, \mathsf{A}^4_{\mathsf{ac}}\}, \{(\mathsf{A}^7_{\mathsf{ep}}, \mathsf{A}^8_{\mathsf{ep}}), (\mathsf{A}^8_{\mathsf{ep}}, \mathsf{A}^7_{\mathsf{ep}}), (\mathsf{A}^8_{\mathsf{ep}}, \mathsf{A}^9_{\mathsf{ep}}), (\mathsf{A}^9_{\mathsf{ep}}, \mathsf{A}^8_{\mathsf{ep}}), (\mathsf{A}^7_{\mathsf{ep}}, \mathsf{A}^4_{\mathsf{ac}}), (\mathsf{A}^8_{\mathsf{ep}}, \mathsf{A}^3_{\mathsf{ac}}), (\mathsf{A}^8_{\mathsf{ep}}, \mathsf{A}^2_{\mathsf{ac}}), (\mathsf{A}^9_{\mathsf{ep}}, \mathsf{A}^4_{\mathsf{ac}})\}\rangle$. The next step is to evaluate the acceptability of the arguments. We first apply the preferred semantics to $\mathcal{AF}_{ac}$, the result is: $\mathcal{E}_p = \{\mathsf{A}^2_{\mathsf{ep}}, \mathsf{A}^3_{\mathsf{ep}}, \mathsf{A}^4_{\mathsf{ep}}, \mathsf{A}^5_{\mathsf{ep}}, \mathsf{A}^7_{\mathsf{ep}}, \mathsf{A}^9_{\mathsf{ep}}, \mathsf{A}^1_{\mathsf{ac}}, \mathsf{A}^2_{\mathsf{ac}}, \mathsf{A}^3_{\mathsf{ac}}\}$. Therefore, we have that the set of justified conclusions is: $\mathtt{CONCS}(\mathcal{E}_p) = \mathtt{Output} = \{\mathsf{b}_2, \mathsf{b}_3, \mathsf{b}_4, \mathsf{b}_5, \mathsf{b}_7, \mathsf{g}_1, \mathsf{g}_2\}$. Notice that the set of justified goals is $\{\mathsf{g}_1, \mathsf{g}_2\}$. This means that robot agent BOB activates goals $\mathsf{g}_1 = \mathsf{go}(2,6)$ and $\mathsf{g}_2 = \mathsf{take\_hospital(man\_32)}$ but he does not activate goal $\mathsf{g}_3 = \mathsf{send\_shelter(man\_32)}$. Figure 8 shows the graph of $\mathcal{AF}_{ac}$ and highlights the accepted arguments.

$\mathcal{C}$	$\mathcal{F}$	$\mathcal{V}$	$\mathcal{R}$	ARG
	b_2			$\mathsf{A}^2_{\mathsf{ep}} = \langle\{(\mathsf{b}_2, \emptyset)\}, \mathsf{b}_2\rangle$
	b_3			$\mathsf{A}^3_{\mathsf{ep}} = \langle\{(\mathsf{b}_3, \emptyset)\}, \mathsf{b}_3\rangle$
	b_4			$\mathsf{A}^4_{\mathsf{ep}} = \langle\{(\mathsf{b}_4, \emptyset)\}, \mathsf{b}_4\rangle$
	b_5			$\mathsf{A}^5_{\mathsf{ep}} = \langle\{(\mathsf{b}_5, \emptyset)\}, \mathsf{b}_5\rangle$
man_32	b_2	x	$\mathsf{r}^2_{\mathsf{st}}$	$\mathsf{A}^7_{\mathsf{ep}} = \langle\{(\mathsf{b}_2, \emptyset), (\mathsf{b}_7, \mathsf{r}^2_{\mathsf{st}})\}, \mathsf{b}_7\rangle$
man_32, arm	b_3	x, y	$\mathsf{r}^3_{\mathsf{st}}$	$\mathsf{A}^8_{\mathsf{ep}} = \langle\{(\mathsf{b}_3, \emptyset), (\neg\mathsf{b}_7, \mathsf{r}^3_{\mathsf{st}})\}, \neg\mathsf{b}_7\rangle$
man_32	b_5	x	$\mathsf{r}^4_{\mathsf{st}}$	$\mathsf{A}^9_{\mathsf{ep}} = \langle\{(\mathsf{b}_5, \emptyset), (\mathsf{b}_7, \mathsf{r}^4_{\mathsf{st}})\}, \mathsf{b}_7\rangle$
2,6	b_4	x, y	$\mathsf{r}^3_{\mathsf{ac}}$	$\mathsf{A}^1_{\mathsf{ac}} = \langle\{(\mathsf{b}_4, \emptyset), (\mathsf{g}_1, \mathsf{r}^3_{\mathsf{ac}})\}, \mathsf{g}_1\rangle$
man_32	b_2	x	$\mathsf{r}^2_{\mathsf{st}}, \mathsf{r}^1_{\mathsf{ac}}$	$\mathsf{A}^2_{\mathsf{ac}} = \langle\{(\mathsf{b}_2, \emptyset), (\mathsf{b}_7, \mathsf{r}^2_{\mathsf{st}}), (\mathsf{g}_2, \mathsf{r}^1_{\mathsf{ac}})\}, \mathsf{g}_2\rangle$
man_32	b_5	x	$\mathsf{r}^4_{\mathsf{st}}, \mathsf{r}^1_{\mathsf{ac}}$	$\mathsf{A}^3_{\mathsf{ac}} = \langle\{(\mathsf{b}_5, \emptyset), (\mathsf{b}_7, \mathsf{r}^4_{\mathsf{st}}), (\mathsf{g}_2, \mathsf{r}^1_{\mathsf{ac}})\}, \mathsf{g}_2\rangle$
man_32	b_3	x	$\mathsf{r}^3_{\mathsf{st}}, \mathsf{r}^2_{\mathsf{ac}}$	$\mathsf{A}^4_{\mathsf{ac}} = \langle\{(\mathsf{b}_3, \emptyset), (\neg\mathsf{b}_7, \mathsf{r}^3_{\mathsf{st}}), (\mathsf{g}_3, \mathsf{r}^2_{\mathsf{ac}})\}, \mathsf{g}_3\rangle$

Table 1: Arguments generated for the activation stage.

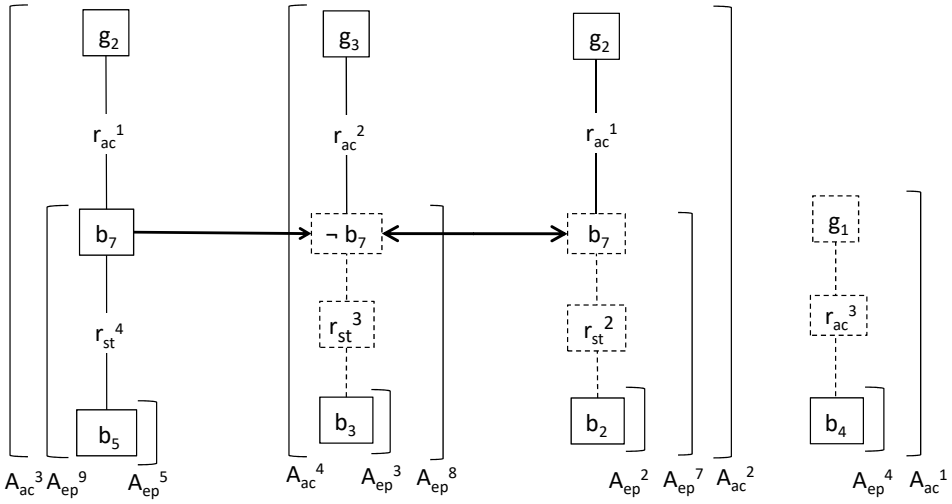


Figure 7: Attacks between arguments of the activation stage. Solid lines denote strict rules and dashed lines denote defeasible rules. Recall that the conclusion of a strict rule is also strict and the conclusion of a defeasible rule is also defeasible. Finally, recall that arguments with strict claims defeat arguments with defeasible claims.

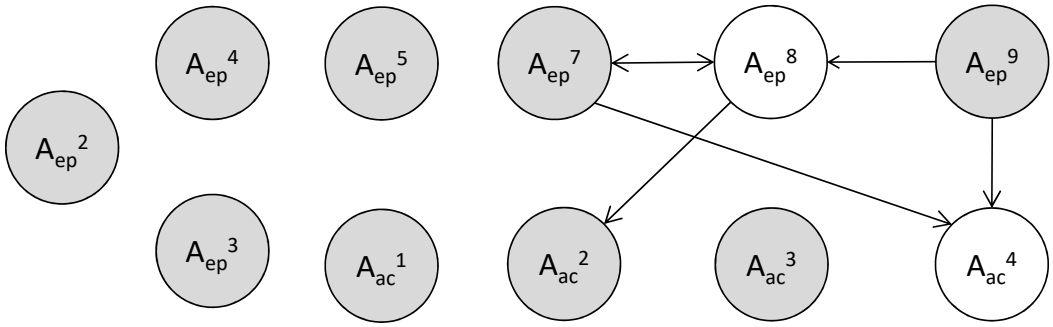


Figure 8: Graph of the argumentation framework $\mathcal{AF}_{ac}$. Nodes represent arguments and edges attacks between arguments. Grey coloured nodes represent the arguments that are part of the selected preferred extension.

3.6.2 EVALUATION STAGE

Let us start by presenting the updates made on the goals base of agent BOB.

$\mathcal{G}_a = \{g_1, g_2\}, \mathcal{G}_p = \{\}, \mathcal{G}_c = \{\}, \mathcal{G}_e = \{\}$, and $\mathcal{G}_{canc} = \{\}$

On the basis of the mental state of agent BOB three epistemic arguments and one evaluation argument can be constructed. Table 2 shows these arguments, which may be constructed by unifying constants of the beliefs with variables of standard and evaluation rules.

The AF for this stage is $\mathcal{AF}_{ev} = \langle\{A^6_{ep}, A^{10}_{ep}, A^{11}_{ep}, A^1_{ev}\}, \{(A^6_{ep}, A^{11}_{ep}), (A^{11}_{ep}, A^6_{ep}), (A^{11}_{ep}, A^1_{ev})\}\rangle$. We have two preferred extensions for $\mathcal{AF}_{ev}$: $\mathcal{E}_p = \{\{A^{10}_{ep}, A^{11}_{ep}\}, \{A^6_{ep}, A^{10}_{ep}, A^1_{ev}\}\}$. Since the second preferred extension refrains a goal of becoming pursuable, the agent chooses the first preferred extension. Figure 9(a) shows the attacks between the arguments of the $\mathcal{AF}_{ev}$

$\mathcal{C}$	$\mathcal{F}$	$\mathcal{V}$	$\mathcal{R}$	ARG
	$\neg b_6$			$A^6_{ep} = \langle\{(\neg b_6, \emptyset)\}, \neg b_6\rangle$
	b_8			$A^{10}_{ep} = \langle\{(b_8, \emptyset)\}, b_8\rangle$
bed	b_8	x	r^1_{st}	$A^{11}_{ep} = \langle\{(b_8, \emptyset), (b_6, r^1_{st})\}, b_6\rangle$
man_32	$\neg b_6$	x	r^2_{ev}	$A^1_{ev} = \langle\{(\neg b_6, \emptyset), (\neg g_2, r^2_{ev})\}, \neg g_2\rangle$

Table 2: Arguments generated for the evaluation stage.

and Figure 9(b) shows the graph of $\mathcal{AF}_{ev}$. Since there is no evaluation argument that belongs to the selected extension, we can say that the passage of both currently active goals is justified. Therefore, both $g_1 = go(2,6)$ and $g_2 = take\_hospital(man\_32)$ are now pursuable goals.

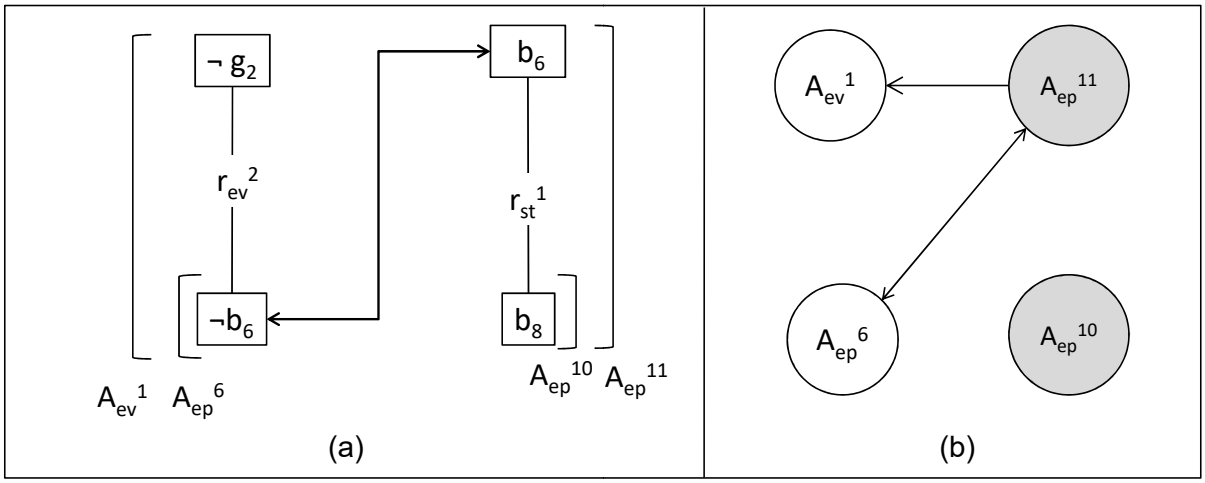


Figure 9: (a) Attack between arguments generated in the evaluation stage. (b) Graph of the argumentation framework $\mathcal{AF}_{ev}$. Nodes represent arguments and edges attacks between arguments. The grey coloured nodes represent the arguments that are part of the selected preferred extension.

3.6.3 DELIBERATION STAGE

Let us start by presenting the bases that are useful in this stage and the update of the goals base of agent BOB:

$\mathcal{G}_a = \{\}$, $\mathcal{G}_p = \{g_1, g_2\}$, $\mathcal{G}_c = \{\}$, $\mathcal{G}_e = \{\}$, and $\mathcal{G}_{canc} = \{\}$, recall that $g_2 \succ g_1$

$\mathcal{RES} = \{bat(20), bat(30), bat(40), bat(50), oil(20), oil(30)\}$

$\mathcal{RES}_{sum} = \{bat(80), oil(70)\}$

$\mathcal{PS} = \{p_1, p_2, p_3\}$ such that:

$p_1 = g_1 : \{bat(40) \wedge be\_operative(me)\} \leftarrow \{...\}$

$p_2 = g_2 : \{bat(50) \wedge be\_operative(me)\} \leftarrow \{...\}$

$p_3 = g_3 : \{bat(5) \wedge be\_operative(me)\} \leftarrow \{...\}$

Let us recall that deliberation stage has two parts, the first one concerns with the incompatibility evaluation between pursuable goals and the second one concerns with determining the most valuable goals. In each of these parts the agent generates supporting beliefs that are used to construct arguments. It is in this stage that the agent applies some functions that guide his decisions. Next, the set of steps the agent performs in order to generate the beliefs for constructing the arguments of this stage.

1. $\texttt{EVAL\_COMPET}(\mathsf{g}_1, \mathsf{g}_2) = \{(\mathsf{g}_1, \mathsf{p}_1), (\mathsf{g}_2, \mathsf{p}_2)\}$. This means that each pursuable goal has at least one plan that allows the agent to achieve it. Thus, two competence beliefs are generated for each goal: $\mathsf{b}_{10} = \mathsf{has\_plans\_for}(\mathsf{g}_2)$, $\mathsf{b}_{11} = \mathsf{has\_plans\_for}(\mathsf{g}_1)$.

2. $\texttt{EVAL\_INCOMP}(\mathsf{g}_1, \mathsf{g}_2) = \{\{\mathsf{g}_1, \mathsf{g}_2\}\}$. This occurs because $\Phi = \mathsf{bat}(40) \wedge \mathsf{bat}(50)$ and $\mathcal{R}_{sum} \models_r \mathit{not}\ \Phi$, that is, there is not enough resources for performing plans p_1 and p_2. Since each goal has only one plan associated, it means that both goals are incompatible. Therefore, no non-incompatibility belief is generated.

3. $\texttt{EVAL\_VALUE}(\{\mathsf{g}_1, \mathsf{g}_2\}) = \{\mathsf{g}_2\}$. This result is based on the preference relation between goals g_1 and g_2. Thus, a value belief for goal g_2 is generated: $\mathsf{b}_{12} = \mathsf{most\_valuable}(\mathsf{g}_2)$.

On the basis of the mental state of agent BOB one epistemic arguments and one deliberation argument are constructed. Table 3 shows these arguments. Notice that we treat the goal as a constant in order to keep the coherence with the logic language.

$\mathcal{C}$	$\mathcal{F}$	$\mathcal{V}$	$\mathcal{R}$	ARG
	b_{12}			$\mathsf{A}^{12}_{\mathsf{ep}} = \langle\{(\mathsf{b}_{12}, \emptyset)\}, \mathsf{b}_{12}\rangle$
$'\mathsf{take\_hospital(man\_32)}'$	b_{12}	x	$\mathsf{r}^2_{\mathsf{de}}$	$\mathsf{A}^1_{\mathsf{de}} = \langle\{(\mathsf{b}_{12}, \emptyset), (\mathsf{chosen}(\mathsf{g}_2), \mathsf{r}^2_{\mathsf{de}})\}, \mathsf{g}_2\rangle$

Table 3: Arguments generated for the deliberation stage

The AF for this stage is $\mathcal{AF}_{de} = \langle\{\mathsf{A}^{12}_{\mathsf{ep}}, \mathsf{A}^1_{\mathsf{de}}\}, \{\}\rangle$. In this case, there is no attacks between the arguments and we have only one preferred extension for $\mathcal{AF}_{de}$: $\mathcal{E}_p = \{\mathsf{A}^{12}_{\mathsf{ep}}, \mathsf{A}^1_{\mathsf{de}}\}$. We have that a deliberation argument belongs to the extension, so we can say that the passage of goal g_2 is justified. Therefore, $\mathsf{g}_2 = \mathsf{take\_hospital(man\_32)}$ is now a chosen goal.

3.6.4 CHECKING STAGE

Once again, let us start by presenting the bases of agent BOB are useful in this stage:

$\mathcal{G}_a = \{\}, \mathcal{G}_p = \{\mathsf{g}_1\}, \mathcal{G}_c = \{g_2\}, \mathcal{G}_e = \{\}$, and $\mathcal{G}_{canc} = \{\}$
$\mathcal{PS} = \{\mathsf{p}_1, \mathsf{p}_2, \mathsf{p}_3\}$ such that:
$\mathsf{p}_1 = \mathsf{g}_1 : \{\mathsf{bat}(40) \wedge \mathsf{be\_operative}(\mathsf{me})\} \leftarrow \{...\}$
$\mathsf{p}_2 = \mathsf{g}_2 : \{\mathsf{bat}(50) \wedge \mathsf{be\_operative}(\mathsf{me})\} \leftarrow \{...\}$
$\mathsf{p}_3 = \mathsf{g}_3 : \{\mathsf{bat}(5) \wedge \mathsf{be\_operative}(\mathsf{me})\} \leftarrow \{...\}$

Let us recall that the checking stage has two parts, the first one concerns with the existence of plans for achieving chosen goals and the second one is about determining if the contexts of these plans are satisfied. In each of these parts the agent generates supporting beliefs that are used to construct arguments. It is in this stage that the agent applies some functions that guide his decisions. Next, the set of steps the agent performs in order to generate the beliefs for constructing the arguments of this stage.

1. Let us recall that the existence of plan was already verified in the deliberation stage. Thus, two beliefs were generated and are already part of $\mathcal{F}$: $\mathsf{b}_{10} = \mathsf{has\_plans\_for}(\mathsf{g}_2)$, $\mathsf{b}_{11} = \mathsf{has\_plans\_for}(\mathsf{g}_1)$.

2. Since g_2 is the only one chosen goal, the evaluation of the context of a plan is done over p_2. Thus, $\texttt{EVAL\_CONTEXT}((\mathsf{g}_2, \{\mathsf{p}_2\})) = \{\mathsf{g}_2\}$, which means that the context of plan p_2 is satisfied. Hence, $\mathsf{b}_{13} = \mathsf{satisfied\_context}(\mathsf{g2})$ is generated.

Based on the mental state of agent BOB two epistemic arguments and one checking argument are constructed. Table 4 presents these arguments.

$\mathcal{C}$	$\mathcal{F}$	$\mathcal{V}$	$\mathcal{R}$	ARG
	b_{10}			$\mathsf{A}^{13}_{\mathsf{ep}} = \langle\{(\mathsf{b}_{10}, \emptyset)\}, \mathsf{b}_{10}\rangle$
	b_{13}			$\mathsf{A}^{11}_{\mathsf{ep}} = \langle\{(\mathsf{b}_{13}, \emptyset)\}, \mathsf{b}_{13}\rangle$
$'\mathsf{take\_hospital}(\mathsf{man\_32})'$	$\mathsf{b}_{10}, \mathsf{b}_{13}$	x	r_{ch}	$\mathsf{A}^{1}_{\mathsf{ch}} = \langle\{(\mathsf{b}_{10}, \emptyset), (\mathsf{b}_{13}, \emptyset), (\mathsf{executive}(\mathsf{g}_2), \mathsf{r}_{\mathsf{ch}})\}, \mathsf{g}_2\rangle$

Table 4: Arguments generated for the checking stage.

The AF for this stage is $\mathcal{AF}_{ch} = \langle\{\mathsf{A}^{13}_{\mathsf{ep}}, \mathsf{A}^{14}_{\mathsf{ep}}, \mathsf{A}^{1}_{\mathsf{ch}}\}, \{\}\rangle$. In this case, we have only one preferred extension for $\mathcal{AF}_{ch}$: $\mathcal{E}_p = \{\mathsf{A}^{13}_{\mathsf{ep}}, \mathsf{A}^{14}_{\mathsf{ep}}, \mathsf{A}^{1}_{\mathsf{ch}}\}$. We have that a checking argument belongs to the extension, so we can say that the passage of goal g_2 is justified. Therefore, $\mathsf{g}_2 = \mathsf{take\_hospital}(\mathsf{man\_32})$ is now an executive goal.

At last, we present the final configuration of $\mathcal{G}$: $\mathcal{G}_a = \{\}$, $\mathcal{G}_p = \{\mathsf{g}_1\}$,$\mathcal{G}_c = \{\}$, $\mathcal{G}_e = \{g_2\}$, and $\mathcal{G}_{canc} = \{\}$.

3.6.5 PARTIAL AND COMPLETE EXPLANATIONS

In the Introduction of this Chapter we have claimed that a BBGP-based agent is able to make justified decisions. That is why we also want to demonstrate the explaining power of our approach. First of all, let us show the individual memory records of goals g_1, g_2, and g_3. Table 5 shows the sub-AFs that allow goal g_1 to become pursuable, it also shows it cannot become chosen because there is no deliberation argument that supports its change of status.

STA	REASON
a	$\mathcal{AF}^{g_1}_{ac} = \langle\{\mathsf{A}^1_{\mathsf{ac}}, \mathsf{A}^4_{\mathsf{ac}}\}, \{\}\rangle$
p	$\mathcal{AF}^{g_1}_{ev} = \langle\{\}, \{\}\rangle$
not c	$\mathcal{AF}^{g_1}_{de} = \langle\{\}, \{\}\rangle$

Table 5: Individual memory record for goal $\mathsf{g}_1 = \mathsf{go}(2,6)$.

In Table 6, we have all the sub-AFs that allow goal g_2 to become executive. Notice that g_1 becomes pursuable because no evaluation argument refrains its passage to the deliberation stage; on the contrary, in sub-AF $\mathcal{AF}^{g_2}_{ev}$ there is an evaluation argument against g_2, which is attacked by an epistemic argument. Thus, g_2 becomes pursuable due to the defence of one of the epistemic arguments of the sub-AF.

STA	REASON
a	$\mathcal{AF}^{g_2}_{ac} = \langle\{\mathsf{A}^2_{\mathsf{ac}}, \mathsf{A}^3_{\mathsf{ac}}, \mathsf{A}^7_{\mathsf{ep}}, \mathsf{A}^2_{\mathsf{ep}}, \mathsf{A}^3_{\mathsf{ep}}, \mathsf{A}^9_{\mathsf{ep}}, \mathsf{A}^5_{\mathsf{ep}}, \mathsf{A}^8_{\mathsf{ep}}\}, \{(\mathsf{A}^8_{\mathsf{ep}}, \mathsf{A}^2_{\mathsf{ac}}), (\mathsf{A}^7_{\mathsf{ep}}, \mathsf{A}^8_{\mathsf{ep}}), (\mathsf{A}^8_{\mathsf{ep}}, \mathsf{A}^7_{\mathsf{ep}}), (\mathsf{A}^9_{\mathsf{ep}}, \mathsf{A}^8_{\mathsf{ep}})\}\rangle$
p	$\mathcal{AF}^{g_2}_{ev} = \langle\{\mathsf{A}^1_{\mathsf{ev}}, \mathsf{A}^{11}_{\mathsf{ep}}, \mathsf{A}^6_{\mathsf{ep}}\}, \{(\mathsf{A}^{11}_{\mathsf{ep}}, \mathsf{A}^1_{\mathsf{ev}}), (\mathsf{A}^{11}_{\mathsf{ep}}, \mathsf{A}^6_{\mathsf{ep}}), (\mathsf{A}^6_{\mathsf{ep}}, \mathsf{A}^{11}_{\mathsf{ep}})\}\rangle$
c	$\mathcal{AF}^{g_2}_{de} = \langle\{\mathsf{A}^1_{\mathsf{de}}, \mathsf{A}^{12}_{\mathsf{ep}}\}, \{\}\rangle$
e	$\mathcal{AF}^{g_2}_{ch} = \langle\{\mathsf{A}^1_{\mathsf{ch}}, \mathsf{A}^{13}_{\mathsf{ep}}, \mathsf{A}^{11}_{\mathsf{ep}}\}, \{\}\rangle$

Table 6: Individual memory record for goal $\mathsf{g}_2 = \mathsf{take\_hospital}(\mathsf{man\_32})$.

Finally, Table 7 shows that goal g_3 cannot become active. Even though there is an activation argument, it is attacked by two epistemic arguments. Thus, after applying the semantics the activation argument is not part of the preferred extension.

For the sake of simplicity, suppose that rescue robots can communicate with humans by means of natural language. Now, suppose that at the end of a rescue day, BOB is interrogated

STA	REASON
not a	$\mathcal{AF}_{ac}^{g_3} = \langle\{\mathsf{A}_{\mathsf{ac}}^4, \mathsf{A}_{\mathsf{ap}}^8, \mathsf{A}_{\mathsf{ap}}^2, \mathsf{A}_{\mathsf{ap}}^3, \mathsf{A}_{\mathsf{ap}}^5, \mathsf{A}_{\mathsf{ap}}^7, \mathsf{A}_{\mathsf{ap}}^9\}, \{(\mathsf{A}_{\mathsf{ap}}^7, \mathsf{A}_{\mathsf{ac}}^4), (\mathsf{A}_{\mathsf{ap}}^9, \mathsf{A}_{\mathsf{ac}}^4), (\mathsf{A}_{\mathsf{ap}}^8, \mathsf{A}_{\mathsf{ap}}^7), (\mathsf{A}_{\mathsf{ap}}^7, \mathsf{A}_{\mathsf{ap}}^8), (\mathsf{A}_{\mathsf{ap}}^9, \mathsf{A}_{\mathsf{ap}}^8)\}\rangle$

Table 7: Individual memory record for goal $\mathsf{g}_3 = \mathsf{send\_shelter}(\mathsf{man\_32})$.

with the following question: *Why have you taken to the hospital* man_32 *instead of sending him to the shelter?* BOB can give a partial explanation or a complete explanation. Next, we show both of them:

PARTIAL EXPLANATION

BOB only uses the arguments that are part of the selected preferred extension. Thus, he answers with: $\{\mathsf{A}_{\mathsf{ep}}^2, \mathsf{A}_{\mathsf{ep}}^3, \mathsf{A}_{\mathsf{ep}}^5, \mathsf{A}_{\mathsf{ep}}^7, \mathsf{A}_{\mathsf{ep}}^9, \mathsf{A}_{\mathsf{ac}}^2, \mathsf{A}_{\mathsf{ac}}^3\}$. In natural language he would give the following answer: man_32 *had a fractured bone (*$\mathsf{A}_{\mathsf{ep}}^2$*), the fractured bone was of his arm (*A_{ep}^3*), and it was an open fracture (*$\mathsf{A}_{\mathsf{ep}}^5$*); therefore, he was severely injured (*$\mathsf{A}_{\mathsf{ep}}^7, \mathsf{A}_{\mathsf{ep}}^9$*). Since he was severely injured I took him to the hospital (*$\mathsf{A}_{\mathsf{ac}}^2, \mathsf{A}_{\mathsf{ac}}^3$*).*

BOB can also use the preferred extension of $\mathcal{AF}_{ac}^{g_3}$. In this case, he gives the reasons for not sending man_32 to the shelter. Thus, this he answers with: $\{\mathsf{A}_{\mathsf{ep}}^2, \mathsf{A}_{\mathsf{ep}}^3, \mathsf{A}_{\mathsf{ep}}^5, \mathsf{A}_{\mathsf{ep}}^7, \mathsf{A}_{\mathsf{ep}}^9, \mathsf{A}_{\mathsf{ac}}^4\}$. In natural language he would give the following answer: man_32 *had a fractured bone (*$\mathsf{A}_{\mathsf{ep}}^2$*), the fractured bone was of his arm (*$\mathsf{A}_{\mathsf{ep}}^3$*), and it was an open fracture (*$\mathsf{A}_{\mathsf{ep}}^5$*). A fractured bone might be considered a severe injury (*$\mathsf{A}_{\mathsf{ep}}^7$*), but since it it was an open fracture it was indeed a severe injury (*$\mathsf{A}_{\mathsf{ep}}^9$*).*

There are two observations that can be done about this last answer. First, $\mathsf{A}_{\mathsf{ep}}^7$, unlike $\mathsf{A}_{\mathsf{ep}}^9$, is built on a defeasible rule; hence, it is a not a determinant reason to conclude that the injure is severe. Second, this explanation does not clarify completely the reasons for not sending the man to the shelter.

COMPLETE EXPLANATION

In this case, BOB uses the sub-AFs of his individual memory records. Thus, he uses $\mathcal{AF}_{ac}^{g_2} = \langle\{\mathsf{A}_{\mathsf{ep}}^2, \mathsf{A}_{\mathsf{ep}}^5, \mathsf{A}_{\mathsf{ep}}^7, \mathsf{A}_{\mathsf{ep}}^8, \mathsf{A}_{\mathsf{ep}}^9, \mathsf{A}_{\mathsf{ac}}^2, \mathsf{A}_{\mathsf{ac}}^3\}, \{(\mathsf{A}_{\mathsf{ep}}^8, \mathsf{A}_{\mathsf{ac}}^2), (\mathsf{A}_{\mathsf{ep}}^7, \mathsf{A}_{\mathsf{ep}}^8), (\mathsf{A}_{\mathsf{ep}}^8, \mathsf{A}_{\mathsf{ep}}^7), (\mathsf{A}_{\mathsf{ep}}^9, \mathsf{A}_{\mathsf{ep}}^8)\}\rangle$ to justify why he decided to take man_32 to the hospital. In natural language, this would be the answer: man_32 *had a fractured bone (*$\mathsf{A}_{\mathsf{ep}}^2$*), the fractured bone was of his arm (*$\mathsf{A}_{\mathsf{ep}}^3$*), and it was an open fracture (*$\mathsf{A}_{\mathsf{ep}}^5$*). Given that he had a fractured bone, he might be considered severe injured (*$\mathsf{A}_{\mathsf{ep}}^7$*); however, since such fracture was of his arm, it might not be considered a severe injure (*$\mathsf{A}_{\mathsf{ep}}^8$*). Finally, I noted that it was an open fracture, which determines – without exception – that it was a severe injury (*$\mathsf{A}_{\mathsf{ep}}^9$*). For these reasons I took him to the hospital (*$\mathsf{A}_{\mathsf{ac}}^2, \mathsf{A}_{\mathsf{ac}}^3$*).*

The answer above answers only half the question. Let's see now the complete reason for not sending him to the shelter, which is indeed a complement of the above answer. Thus, he uses the sub-AF related to goal g_3: $\mathcal{AF}_{ac}^{g_3} = \langle\{\mathsf{A}_{\mathsf{ap}}^2, \mathsf{A}_{\mathsf{ap}}^3, \mathsf{A}_{\mathsf{ap}}^5, \mathsf{A}_{\mathsf{ap}}^7, \mathsf{A}_{\mathsf{ap}}^8, \mathsf{A}_{\mathsf{ap}}^9, \mathsf{A}_{\mathsf{ac}}^4\}, \{(\mathsf{A}_{\mathsf{ap}}^7, \mathsf{A}_{\mathsf{ac}}^4), (\mathsf{A}_{\mathsf{ap}}^9, \mathsf{A}_{\mathsf{ac}}^4), (\mathsf{A}_{\mathsf{ap}}^8, \mathsf{A}_{\mathsf{ap}}^7), (\mathsf{A}_{\mathsf{ap}}^7, \mathsf{A}_{\mathsf{ap}}^8), (\mathsf{A}_{\mathsf{ap}}^9, \mathsf{A}_{\mathsf{ap}}^8)\}\rangle$. In natural language, this would be the answer: man_32 *had a fractured bone ($\mathsf{A}_{\mathsf{ep}}^2$) and the fractured bone was of his arm ($\mathsf{A}_{\mathsf{ep}}^3$). Given that he had a fractured bone, he might be considered severely injured ($\mathsf{A}_{\mathsf{ep}}^7$); however, since such fracture was of his arm, it might not be considered a severe injury ($\mathsf{A}_{\mathsf{ep}}^8$). Since the injury is not severe, the man might be sent to the shelter; however, I noted that it was an open fracture, which determines – without exception – that it was a severe injury ($\mathsf{A}_{\mathsf{ep}}^9$). This last reason refutes the action of sending him to the shelter.*

Note that the complete explanation is more accurate, especially when the agent has to clarify why he did not send the man to the shelter. Note also that both complete explanations are complementary. We can say that depending on the question, the agent can use part of the entire individual memory record. The agent may also use more than one individual memory record in order to give satisfactory answers.

3.7 RELATED WORK

Many works have highlighted the benefits of using argumentation in multiagent settings; however, there are still few works that use argumentation inside the agent architecture. In (KAKAS et al., 2004), argumentation is used to resolve conflicts that arise. The implementation of the argumentation process is based on Logic Programming without Negation as Failure (LPwNF). Kakas et al. (2008) present the logic foundations for an agency model called KGP (Knowledge, Goals, and Plan). In this model, both goal decision and cycle theories for internal control are done through argumentation. In (BERARIU, 2014), argumentation is used to maintain consistency of the belief base of BDI agents. Lastly, a fully integrated argumentation-based agent (called ABA) architecture with a highly modular structure is developed in (KAKAS et al., 2011).

Argumentation has also been used for generating goals and plans. Amgoud (2003) proposes an argumentation-based framework for dealing with conflicting desires. She uses the argumentation concepts proposed by Dung to determine the set of intentions of an agent from a set of contradictory desires. Amgoud e Kaci (2005) study the generation of bipolar goals in argumentation-based negotiation. The authors claim that goals have two different sources: (i) from beliefs and (ii) from other goals. They propose explanatory and instrumental arguments to justify the adoption of goals. These arguments have a similar function to the activation

arguments of this thesis. Finally, Rahwan e Amgoud (2006) use argumentation for generating both desires and plans. They propose three different AFs, one for arguing about beliefs, other for arguing about desires, and the third one for arguing about what plan has to be used to achieve a desire.

As shown, most of the related works use argumentation in some specific parts of the reasoning cycle of an agent. Only Kakas et al. (2011) aims to create a fully argumentation-based agent architecture. In this thesis, the proposal is to use argumentation for the goal processing (also called intention formation process). Amgoud (2003) partially shares our objective, starting from a set of conflicting desires; argumentation is used to resolve such conflicts and to decide which of them will become intentions. We start from a set of sleeping goals, which can be conflicting or not and use argumentation to filter the set of goals that will advance to the next stage. The main difference is the path a desire has to go over until it becomes an intention, which in our work is more fine-grained as it includes not only conflicts but impossibilities and checks whether there are conditions for a goal be achieved.

3.8 SUMMARY

In this chapter, we have presented an an argumentation-based formalization for the goal processing model proposed in (CASTELFRANCHI; PAGLIERI, 2007). First, we have defined some types of beliefs, rules, and some functions taking into account the features described in the abstract model. The beliefs and rules are the building blocks for constructing the arguments, which act as filters by specifying what goals can pass from one stage to the next or not, and at the same time as reasons that support or attack such transitions.

In the second part, we showed the argumentation process for each one of the stages. We defined two types of arguments and the attacks between these arguments. We made an analysis in order to determine which semantics better behaves in this context. We concluded that preferred semantics is the most adequate because it does not return empty extensions – in this context – and its credulous nature allows the agent to have more than one option to proceed, in this sense, we have proposed a set of possible criteria the agent can consider to make a decision.

In order to endow the agent with a managing structure in charge of keeping the consistency of the whole system of goal processing, we defined two global structures. One of these structures – the memory record – is important to fulfil the synchrony property. We also have made an analysis to prove that the system also fulfils the diachrony property and the direct con-

sistency property. Thus, we demonstrate that our argumentation-based proposal complies with the main properties stipulated in (CASTELFRANCHI; PAGLIERI, 2007) and with one of the rationality properties dictated in (CAMINADA; AMGOUD, 2007).

In the next Chapter, we will study how BBGP-based agents – involved in a persuasive negotiation dialogue – can calculate the strength of rhetorical arguments. Since the extended agent has to engage in a negotiation, we will extend the agent architecture by equipping the agent with structures that store knowledge about his possible opponents. From such knowledge, the negotiating agent will be able to generate and calculate the strength of threats, rewards, and appeals.

4 CALCULATION OF THE STRENGTH OF RHETORICAL ARGUMENTS

This chapter presents a model for the calculation of the strength of threats, rewards, and appeals, which are a kind of argument that is used in negotiation dialogues – more specificaly in persuasive negotation (RAMCHURN et al., 2003) – when a proponent agent tries to persuade his opponent to accept a proposal.

4.1 INTRODUCTION

Negotiation is a key form of interaction among agents (generally called proponent and opponent) that can be used for resolving conflicts and reaching agreements. Argumentation has been used in some works of negotiation because it allows an agent to exchange additional information, which can be used for justifying his proposals (SIERRA et al., 1998),(AMGOUD et al., 2000),(RAHWAN et al., 2003),(DIMOPOULOS; MORAITIS, 2011).

Arguments used in negotiation dialogues are generally explanatory and allow agents to argue about their beliefs or other mental attitudes during the negotiation process (RAHWAN et al., 2003). Nevertheless, there are other types of arguments that may act as persuasive elements, these ones are called rhetorical arguments[1], and are the following:

- *Threats*, which carry out sanctions when the opponent does not agree with the proposal sent by the proponent and can be regarded as somewhat risky not only for the opponent but also for the proponent, because if the proponent backs down from a threat, he may appear to be weak or vacilating (WALTON, 2005).
- *Rewards*, which are used when a proponent agent wants to entice an opponent to do a certain action by offering to do another action as a reward or by offering something that the opponent needs.

[1] When an agent uses rhetorical arguments to back their proposals, the negotiation is called persuasive negotiation (RAMCHURN et al., 2003).

- *Appeals*, which try to persuade the opponent by offering a reward; however, this recompense is not a consequence of an action of the proponent. Thus, this kind of argument is an advantageous option for both the proponent and the opponent. If the proponent does not have a recompense to offer, he can appeal to one goal of the opponent that does not need the proponent's intervention. Appeals can be seen as self-rewards (AMGOUD; PRADE, 2004).

Notice that threats have a negative nature whereas both rewards and appeals have a positive nature. Some authors claim that threats are the rhetorical arguments with most persuasive power (SYCARA, 1990) (KRAUS et al., 1998). According to them, threatening an important goal of an opponent is the most effective of the arguments. Nevertheless, according to (RAMCHURN et al., 2003), it also depends on other factors like the convenience of the proposal for the proponent. Therefore, depending on the context and the interests of the proponent, a reward or an appeal may be more effective than a threat. In this chapter, we study how to calculate the strength of threats, rewards, and appeals. This strength value can be used by the proponent agent as a comparative measure that allows him to decide which argument may be the most effective one.

The chapter is organized as follows. In Section 4.2, we extend the BBGP-based agent such that the new BBGP-agent has the capacity of generating rhetorical arguments and calculating their strength values. Section 4.3 is devoted to the logical definition of rhetorical arguments. Section 4.4 presents the strength calculation model. It includes the analysis of the criteria that will be considered and the steps of the model. In Section 4.5, we study and detail the properties that describe the behavior of the calculation model. Section 4.6 illustrates the performance of the model in the rescue robot agents scenario and the software agents scenario. In Section 4.7, we discuss a few related work. Finally, Section 4.8 summarizes this chapter.

4.2 KNOWLEDGE REPRESENTATION AND NEGOTIATING AGENTS

In order to represent the threats, rewards, and appeals, we use the same knowledge representation defined in Chapter 3. This means that $\mathcal{L}_{lit}$ is our underlying logical language and we use derivation schemas as consequence operator. Our negotiating agents – proponent and opponent – are BBGP-based agents, which are extended in order to support the construction of rhetorical arguments and calculation of the strength of such rhetorical arguments.

Definition 4.1. (Negotiating BBGP-based agent) A negotiating BBGP-based agent is a tuple $\langle \texttt{AGT}, \mathcal{O}pp, \mathcal{GO}, \mathcal{S}_{\mathcal{O}pp}, \mathcal{S}_{\mathcal{GO}}, \mathcal{A}, \mathcal{AO}, \texttt{REP} \rangle$ such that:

- $\texttt{AGT} = \langle \mathcal{T}, \mathcal{RES}_{sum}, \mathcal{G}, \mathcal{PS} \rangle$ includes the components of a BBGP-based agent.
- $\mathcal{O}pp$ is the opponents base, whose elements are constants of $\mathcal{C}$;
- $\mathcal{GO} = \mathcal{GO}_a \cup \mathcal{GO}_p \cup \mathcal{GO}_c \cup \mathcal{GO}_e \cup \mathcal{GO}_{canc}$ is the set of the opponent's goals. This base has the same structure as the base of the proponent's goals, i.e. $\mathcal{GO}_a$ is the set of the active opponent's goals, $\mathcal{GO}_p$ the set of the pursuable ones, $\mathcal{GO}_c$ the set of the chosen ones, $\mathcal{G}_e$ is the set of the executive ones, and $\mathcal{GO}_{canc}$ is the set of the cancelled ones. These sets are pairwise disjoint: $\mathcal{G}_x \cap \mathcal{G}_y = \emptyset$ for $x, y \in \{a, p, c, e, canc\}$ and $x \neq y$.
- $\mathcal{S}_{\mathcal{O}pp}$ is a set of tuples $(op, \texttt{THRES}, L_{\mathcal{GO}})$ such that $op \in \mathcal{O}pp$, $\texttt{THRES} \in [0,1]$ is the value of the threshold of the opponent[2], and $L_{\mathcal{GO}} = TH_{\mathcal{GO}} \cup RW_{\mathcal{GO}} \cup AP_{\mathcal{GO}}$ is the set of goals of opponent op such that these goals can be threatanable ($go \in TH_{\mathcal{GO}}$), threatanable ($go \in RW_{\mathcal{GO}}$), or appealable ($go \in AP_{\mathcal{GO}}$). It holds that $\forall go \in L_{\mathcal{GO}}, go \in \mathcal{GO}$, this means that if a goal is in the goals list of an opponent – $L_{\mathcal{GO}}$ – it is also in the opponent's goal set $\mathcal{GO}$. It also holds that $TH_{\mathcal{GO}}$, $RW_{\mathcal{GO}}$, and $AP_{\mathcal{GO}}$ are pairwise disjoint. Finally, let $\texttt{TH\_GO}(op) = TH_{\mathcal{GO}}$, $\texttt{RW\_GO}(op) = RW_{\mathcal{GO}}$, and $\texttt{AP\_GO}(op) = AP_{\mathcal{GO}}$ be three functions that return the sets of threatanable, threatanable, and appealable goals of op, respectively.
- $\mathcal{S}_{\mathcal{GO}}$ is a set of pairs $(go, \texttt{IMP})$ such that $go \in \mathcal{GO}$ and $\texttt{IMP} \in [0,1]$ represents the importance value of go.
- $\mathcal{A}$ is the base of the proponent's actions, whose elements are ground atoms.
- $\mathcal{AO}$ is the base of the opponent's actions, whose elements are ground atoms. The role of action in our calculation model will be further explained below.
- $\mathcal{A}_{val}$ is a set of pairs (ac, val) such that $ac \in \mathcal{A}$ or $ac \in \mathcal{AO}$ is an action and $val \in [0,1]$ is a real number that represents the value of action ac where zero means that ac is not valuable at all whereas one is the maximum value of an action. Let $\texttt{VALUE}(ac) = val$ be a function that returns the value of a given action ac.
- $\texttt{REP}$ is the reputation value of the proponent, which is visible for any other agent.

When a proponent agent employs a rhetorical argument to try to convince an opponent to do a certain action, it can be seen as a goal of him. For this reason, goals in $\mathcal{G}$ are divided in

[2]The threshold is a value used in the strength calculation model. This is better explained in Section 4.4.

(i) goals that the agent himself has to perform actions to achieve them, and (ii) goals that need the opponent involvement to be achieved, for example, the goal of agent TOM is that agent BOB helps him. For this goal to be achieved, it is necessary that BOB executes the required action. This type of goal is called *outsourced*.

Definition 4.2. (Outsourced goal) An outsourced goal g is an expression of the form $g(op, ac)$, such that, $op \in \mathcal{O}pp$ and ac represents an action that op has to perform. Let $\texttt{FIRST}(g) = op$ and $\texttt{SECOND}(g) = ac$ be the functions that return each component of the outsourced goal g, respectively.

Next, we present some assumptions on which we are based.

1. We assume that a negotiating agent has in advance the necessary information for generating rhetorical arguments and for calculating their strengths. This information is related to the opponent's goals, the status of these goals, the opponent's actions, and the values of these actions. In order to obtain such information, agent can gather information about his opponent(s). This approach is known as opponent modelling. By making use of opponent modelling, it is possible to represent necessary information about the opponent, which may be used during the negotiation encounter. Opponent modeling can be performed either online or offline, it depends on the availability of past data. Regarding offline models, these are created before the negotiation starts by using previously obtained data from earlier negotiations. Whereas online models are constructed from knowledge that is gather during a single negotiation encounter.

   In (BAARSLAG et al., 2016), the authors present a survey about some techniques of opponent modeling that are based on learning. Such techniques include Bayesian learning, non-linear regression, kernel density estimation, and artificial neural networks. Other works about opponent modelling with focus on argumentation are (HADJINIKOLIS et al., 2013)(RIENSTRA et al., 2013)(HADJINIKOLIS et al., 2015), (HUNTER, 2015).

2. We assume that there is no uncertainty about this information. This means that the information about the opponent is accurate and correct.

4.3 THREATS, REWARDS, AND APPEALS

In this section, we present the logical definitions of the rhetorical arguments that are being studied in this thesis.

4.3.1 THREATS

The use of threats is a well-known strategy in negotiation (SYCARA, 1990; SIERRA et al., 1998; RAMCHURN et al., 2003). According to (AMGOUD; PRADE, 2005), two forms of threats can be distinguished:

1. You should do ‘a’, otherwise I will do ‘b’ or
2. You should not do ‘a’, otherwise I will do ‘b’

The first case happens when the proponent needs the opponent to do ‘a’ and the opponent refuses. Then, the proponent threatens the opponent to do ‘b’ which will have bad consequences for the opponent, while, the second case happens when the opponent insists on performing certain action ‘a’. Then, the proponent threatens to do ‘b’, which is an action that goes against the interests of the opponent.

Example 4.1. Let us recall the Consumer Complaint Website scenario where agent `CONSUMER` wants agent `COMPANY` to refund the total price of a ticket. Agent `CONSUMER` constructs the following threats:

- th_1 : *You should refund the total price of the ticket, otherwise I will never buy a ticket in your company anymore.*
- th_2 : *You should refund the total price of the ticket, otherwise I will destroy your reputation in social networks.*
- th_3 : *You should refund the total price of the ticket, otherwise I will take legal actions against your company.*

In this example, the action ‘a’ required by the proponent `CONSUMER` is that agent `COMPANY` refunds the total price of a ticket and the action ‘b’, where one goal of `COMPANY` is threatened, is that `CONSUMER` will never buy a ticket again, or `CONSUMER` will destroy the reputation of `COMPANY`, or `CONSUMER` will take legal actions against `COMPANY`. The three threats correspond to the the first form of threat. An example of the second form of threat would be the following:

- th_4 : *You should not destroy my reputation, otherwise I will denounce you for defamation and I will ask for a payment of civil reparations amounting to $1000 in favor of me.*

This threat may be constructed by agent COMPANY to try to convince agent CONSUMER not insist on his complaint.

Based on these two forms of threats and the given examples, we can say that a threat is mainly made up of two goals:

- **An opponent's goal:** It is the goal of the opponent that is being threatened by the proponent. It is a goal that the opponent wants to achieve or maintain. For example, "*maintaining a good reputation*", "*gaining customers fidelity*", and "*avoiding legal problems*".
- **An outsourced goal of the proponent:** It is the goal of the proponent that needs the opponent involvement to be achieved. For example, "*getting that* COMPANY *refunds my ticket's money*".

Following, we present the formal definition of a threat. This is based on the definition given in (AMGOUD; PRADE, 2004), with some modifications that consider the mental states of the negotiating BBGP-based agent and the rule-based approach.

Definition 4.3. (Threat) Let $\mathcal{T}$ be the theory of a negotiating BBGP-based agent, $\mathcal{G}$ be his goals base, and $\mathcal{GO}$ be his opponent's goals base. A threat constructed from $\mathcal{T}$, $\mathcal{G}$ and $\mathcal{GO}$ is a triple $th = \langle T, g, go \rangle$, where:

- $go \in \mathcal{GO}$ and $go \in \texttt{TH\_GO}(\texttt{FIRST}(g))$,
- $g \in \mathcal{G}$,
- $T \cup \neg\texttt{SECOND}(g)$ is a derivation schema for $\neg go$ from $\mathcal{T}$,
- $\texttt{SEQ}(T)$ is consistent,
- T is minimal.

Let us call T the support of the threat, g its conclusion and go is the threatened goal.

According to this definition, a threat is constructed under the hypothesis that the proponent's goal will not be achieved, which has a negative effect not only on the proponent but also on the opponent. The negative effect on the proponent is obviously that he does not achieve a goal and the negative effect on the opponent is that he will not achieve one of his goals either. Thus, both agents need each other to achieve their goals.

Example 4.2. Let us formalize one of the threats of Example 4.1. Consider the following components of the mental state of agent CONSUMER:

$$\mathcal{T} = \{\mathcal{F}, \mathcal{S}, \mathcal{D}\} \text{ such that } \mathcal{S} = \{\neg \textit{ refund(money)} \rightarrow \neg \textit{buy\_again}(\textit{ticket}), \neg \textit{buy\_again}(\textit{ticket}) \rightarrow \neg \textit{gain}(\textit{costu\_fidel})\}$$

$\mathcal{O}pp = \{\texttt{COMPANY}\}$

$\mathcal{G} = \{g\}$ such that $g = get(\texttt{COMPANY}, \text{`}refund(money)\text{'})$ is an outsourced goal

$\mathcal{GO} = \{go_1\}$ such that $go_1 = gain(costu\_fidel)$

$\mathcal{S}_{\mathcal{O}pp} = (\texttt{COMPANY}, \texttt{THRES}, \{go_1\})$ such that $go \in TH_{\mathcal{GO}}$

The following threat can be generated:

$th_1 = \langle T_1, g, go_1 \rangle$ such that

$$T_1 \cup \neg\texttt{SECOND}(g) = \{(\neg\ refund(money), \emptyset),$$
$$(\neg buy\_again(ticket), \neg\ refund(money) \rightarrow \neg buy\_again(ticket)),$$
$$(\neg gain(costu\_fidel), \neg buy\_again(ticket) \rightarrow \neg gain(costu\_fidel))\}$$

4.3.2 REWARDS AND APPEALS

Rewards and appeals are also used during a negotiation dialogue as positive persuasive elements (SIERRA et al., 1997; SHI et al., 2006; FLOREA; KALISZ, 2007; RAMCHURN et al., 2007). Both rewards and appeals result in a clear benefit for the opponent agent.

We start by talking about rewards. Two forms of rewards can be distinguished (AMGOUD; PRADE, 2005):

1. If you do 'a', then I will do 'b'
2. If you do not do 'a', then I will do 'b'

The first case happens when the proponent needs the opponent to do 'a' and the opponent refuses. Then, the proponent offers the opponent to do 'b' which will have positive consequences for the opponent. While, the second case happens when the opponent insists on performing certain action 'a'. Then, the proponent offers to do 'b', which is an action that goes in favor of the interests of the opponent.

Example 4.3. Let us recall the Consumer Complaint Website scenario. However, now suppose that COMPANY is trying to offer something positive to CONSUMER:

- rw_1 : *If you agree with the 20% of refund, we will give you 10,000 miles.*
- rw_2 : *If you agree with the 20% of refund, we will sell you an executive ticket for the price of a economic one for any national destination.*

- rw_3 : *If you agree with the 20% of refund, we will give you our service of assistance for elderly for free for any destination.*

In this example, since CONSUMER sent a threat to COMPANY, he tries to negotiate a 20% of refund along with an additional reward.

Regarding the appeals, we can distinguish the two following forms:

1. If you do 'a', then you can get 'c'
2. If you do not do 'a', then you can get 'c'

The first case happens when the proponent needs the opponent to do 'a' and the opponent refuses. Then, the proponent alleges that by performing 'a' the opponent can be benefited with 'c'. Notice that in appeals, the action 'a' allows the opponent to get 'c', while in rewards the action 'b' allows the opponent to get such 'c'.

Example 4.4. Let us recall the Rescue Robots scenario where agent TOM is trying to convince agent BOB of helping him. Agent TOM constructs the following appeals:

- ap_1 : *If you help me, you can win utility points.*
- ap_2 : *If you help me, you can recharge your battery since the workshop is next to this zone.*
- ap_3 : *If you help me, you can fix your sensor since the workshop is next to this zone.*

We can construct rewards and appeals in the same way as we construct threats. This means that rewards and appeals are also based on an opponent's goal and on an outsourced goal of the proponent.

Below, we present the formal definition of rewards and appeals. This is also based on the definition given in (AMGOUD; PRADE, 2004), with the necessary modifications that consider the mental states of the negotiating BBGP-based agent.

Definition 4.4. (Reward/Appeal) Let $\mathcal{T}$ be the theory of a negotiating BBGP-based agent, $\mathcal{G}$ be his goals base, and $\mathcal{GO}$ be his opponent's goals base. A reward/appeal constructed from $\mathcal{T}$, $\mathcal{G}$, and $\mathcal{GO}$ is a triple $\langle T, g, go \rangle$, where:

- $go \in \mathcal{GO}$,
- For rewards: $go \in \texttt{RW\_GO}(\texttt{FIRST}(g))$ and for appeals: $go \in \texttt{AP\_GO}(\texttt{FIRST}(g))$,

- $g \in \mathcal{G}$,
- $T \cup \texttt{SECOND}(g)$ is a derivation schema for go from $\mathcal{T}$,
- $\texttt{SEQ}(T)$ is consistent,
- T is minimal.

Let us call T the support of the reward/appeal, g its conclusion and go is the rewardable/appealable goal. Furthermore, let RHETARG denote the set of threats, rewards, and appeals that an agent can construct from his theory $\mathcal{T}$.

Example 4.5. Let us formalize one of the rewards of Example 4.3. Consider the following components of the mental state of agent COMPANY:

$\mathcal{T} = \{\mathcal{F}, \mathcal{S}, \mathcal{D}\}$ such that $\mathcal{S} = \{accept(refund\_20) \rightarrow gain(miles)\}$

$\mathcal{O}pp = \{\texttt{CONSUMER}\}$

$\mathcal{G} = \{g_1\}$ such that $g_1 = get(\texttt{CONSUMER}, \text{`}accept(refund\_20)\text{'})$

$\mathcal{GO} = \{go_4\}$ such that $go_4 = gain(miles)$

$\mathcal{S}_{\mathcal{O}pp} = (\texttt{CONSUMER}, \{go_4\})$ such that $go_4 \in RW_{\mathcal{GO}}$

The following reward can be generated:

$rw_1 = \langle T_1, g_1, go_4 \rangle$ such that

$$T_1 \cup \texttt{SECOND}(g_1) = \{(accept(refund\_20), \emptyset), \\ (gain(miles), accept(refund\_20) \rightarrow gain(miles))\}$$

Example 4.6. Let us formalize one of the appeals of Example 4.4. Consider the following components of the mental state of agent TOM:

$\mathcal{T} = \{\mathcal{F}, \mathcal{S}, \mathcal{D}\}$ such that $\mathcal{S} = \{help\_with(debris) \rightarrow gain(util\_points)\}$

$\mathcal{O}pp = \{\texttt{BOB}\}$

$\mathcal{G} = \{g_3\}$ such that $g_3 = get(\texttt{BOB}, \text{`}help\_with(debris)\text{'})$

$\mathcal{GO} = \{go_8\}$ such that $go_8 = gain(util\_points)$

$\mathcal{S}_{\mathcal{O}pp} = (\texttt{BOB}, \{go\})$ such that $go \in AP_{\mathcal{GO}}$

The following appeal can be generated:

$ap_1 = \langle T_1, g_3, go_8 \rangle$ such that

$$T_1 \cup \texttt{SECOND}(g_3) = \{(help\_with(debris), \emptyset), \\ (gain(util\_points), help\_with(debris) \rightarrow gain(util\_points))\}$$

4.4 STRENGTH CALCULATION MODEL

In this section, we start by analysing the necessary criteria for evaluating the strength of threats, rewards, and appeals. Next we detail the steps the proponent agent follows in order to obtain the strength values of the arguments he generates.

4.4.1 PRE-CONDITIONS OF CREDIBILITY AND PREFERABILITY

According to (GUERINI; CASTELFRANCHI, 2006), a rhetorical argument has to meet some pre-conditions in order the proponent can reach a negotiation favorable to him. Consequently, the chosen rhetorical argument has to belong to the set of rhetorical arguments that meet such pre-conditions. These pre-conditions are related to the credibility of the proponent agent and to the preferability of the opponent's goal regarding the requested action.

4.4.1.1 CREDIBILITY

According to (GUERINI; CASTELFRANCHI, 2006; CASTELFRANCHI; GUERINI, 2007), there exists a goal cognitive structure when an proponent utters an influencing sentence to an opponent. Figure 10 shows this cognitive structure, which was extracted from (CASTELFRANCHI; GUERINI, 2007). When a proponent agent x utters a sentence for an opponent agent y about his intention of performing an action ax, his first goal (G1) is that agent y believes that agent x is going to benefit or damage him. His second goal (G2) is that agent y has the intention of performing (or not) an action ay. At last, his third goal (G3) is that agent y performs (or not) action ay. Thus, we can notice that the first goal of the proponent agent is related to his credibility. In other words, when a proponent agent wants to persuade an opponent agent, the opponent has to believe that he (the proponent) is credible.

In this work, in order to evaluate the credibility of the proponent, we take into account the following concepts:

1. **The proponent's reputation:** Reputation can be defined as a social notion associated with how trustworthy an individual is within a society. The estimate value of reputation is formed and updated over time with the help of different sources of information. Several computational models of reputation consider that reputation can be estimated based on two different sources: (i) the direct interactions and (ii) the information provided by other members of the society about experiences they had in the past (e.g., (YU; SINGH, 2000;

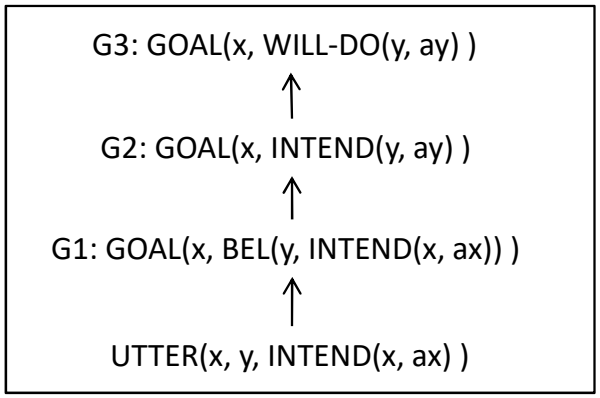


Figure 10: Goal cognitive structure of Castelfranchi and Guerini for persuasive speech acts. x denotes the proponent agent, y the opponent agent, ax the action of agent x, and ay the action of agent y.

SABATER; SIERRA, 2001; PINYOL; SABATER-MIR, 2013)). Another works about trust and reputation are (FALCONE; CASTELFRANCHI, 2001, 2004).

In this work, reputation can be seen as the "social" notion – within an agents society – about how trustworthy the proponent is with respect to fulfil his threats, rewards, and appeals. In other words, it is an evidence of the proponent's past behavior with respect to his opponents. We assume that this value is already estimated and it is not private information. Thus, reputation value of the proponent is known by any other agent. It means that when the proponent begins a negotiation with other agent (his opponent), this one is conscious of the reputation of the proponent. We also assume that the proponent has only one reputation value for the three kinds of rhetorical arguments.

The reputation value of a proponent agent P is represented by a real number: $\texttt{REP}(P) \in [0,1]$ where zero represents the minimum reputation value and one the maximum reputation value.

2. **The opponent's credibility threshold:** It is used to indicate the lowest value of the proponent's reputation so that the opponent considers a rhetorical argument credible. Thus, the credibility threshold of an opponent agent O is represented by a real number: $\texttt{THRES}(O) \in [0,1]$ where zero represents the minimum threshold value and one the maximum threshold value.

   A low threshold denotes a trusting (or easier to be persuaded) opponent and a high threshold denotes a mistrustful opponent, i.e., more difficult to be persuaded. We assume that the proponent knows the values of the thresholds of his possible opponents and stores these values.

The proponent evaluates his own credibility – in the eyes of his opponent – by comparing both values: the proponent's reputation and the opponent's threshold. When the reputation value is greater than or equal to the opponent's threshold, it means that the proponent believes that the opponent considers him (the proponent) credible; otherwise, the opponent believes that the proponent is not credible.

Definition 4.5. (Proponent's credibility) Let P be a proponent agent, $\texttt{REP}(P)$ his reputation, and $\texttt{THRES}(O)$ the threshold of his opponent O. P is credible if $\texttt{REP}(P) \geqslant \texttt{THRES}(O)$; otherwise, O does not believe that P is credible.

4.4.1.2 PREFERABILITY

The second pre-condition a proponent agent has to meet in order to attain a favourable negotiation is the *preferability* (GUERINI; CASTELFRANCHI, 2006). This pre-condition is based on the relation between the opponent's goal and the action he is required to perform. According to (GUERINI; CASTELFRANCHI, 2006), the opponent's goal must be more valuable for him (the opponent) than performing the required action.

We first present the criteria that will be evaluated in order to estimate how valuable a goal is for the opponent. Below, we analyze each criteria and indicate how the value of the opponent's goal will be estimated.

1. **Importance of the opponent's goal:** It is related to how meaningful the goal is for the opponent. The value of the importance of a given goal *go* is a real number represented by: $\texttt{IMP}(go) \in [0,1]$ where zero means that the goal is not important at all, and one is the maximum importance of the goal.

   The more important a goal is for the opponent, the more threatenable, rewardable, or appealable this goal is.

2. **Effectiveness of the opponent's goal:** It is related to the degree to which an opponent's goal is successful for persuasion and it is based on the status of the goal in the intention formation process. Let us recall that we are working with BBGP-based agents; therefore, the goals base of the opponent is divided in five sub-sets: active goals, pursuable goals, chosen goals, executive goals, and cancelled goals. A goal is close of being achieved when its status is chosen or executive and it is far of being achieved when its status is active or pursuable. Thus, depending on its status, a goal can be considered more or less threatenable, rewardable, or appealable. Let us analyse each case:

- **Threatenable goal:** Recall that threats have a negative nature. In terms of the status of a goal it means that a threat may make a goal go back to a previous status. In this work, we assume that every threatened goal will become cancelled. Therefore, a goal is considered more threatenable when its status is executive and less threatenable when its status is active. This is because an agent has more to lose when an executive goal is threaten than when an active goal is threaten. Regarding a cancelled goal, it is not threatenable at all.
- **Rewardable and appealable goal:** In this case, both rewards and appeals have a positive nature. In terms of the status of a goal it means that a reward/appeal may make a goal go forward to an advanced status. In this work, we assume that every rewarded/appealed goal will become executive. Therefore, a goal is considered more rewardable/appealable when its status is cancelled and less rewardable/appealable when its status is chosen. This is because an agent has more to win when a cancelled goal is rewarded/appealed than when a chosen goal is rewarded/appealed. Executive goals cannot be rewarded/appealed because the proponent has nothing to offer that makes them go forward. Therefore, executive goals are not rewardable/appealable at all.

The value of the effectiveness of a goal go depends on the argument that is built from it. We denote by $\texttt{arg}(go) \in \{th, rw, ap\}$ the type of argument that can be built where th means that the type of argument is a threat, rw means that the type of argument is a reward, and ap means that the type of argument is an appeal. The effectiveness of an opponent's goal go is represented by $\texttt{eff}(go) \in \{0, 0.25, 0.5, 0.75, 1\}$ such that zero means that go is not effective at all and one means that go is completely effective. The effectiveness of an opponent's goal is evaluated as follows:

$$\texttt{EFF}(go) = \begin{cases} 0 & \begin{array}{l}\text{if } \texttt{arg}(go) = th \text{ and } go \in \mathcal{GO}_{canc}, \text{ or} \\ \text{if } \texttt{arg}(go) = rw \text{ and } go \in \mathcal{GO}_{e}, \text{ or} \\ \text{if } \texttt{arg}(go) = ap \text{ and } go \in \mathcal{GO}_{e}\end{array} \\ \hline 0.25 & \begin{array}{l}\text{if } \texttt{arg}(go) = th \text{ and } go \in \mathcal{GO}_{a}, \text{ or} \\ \text{if } \texttt{arg}(go) = rw \text{ and } go \in \mathcal{GO}_{c}, \text{ or} \\ \text{if } \texttt{arg}(go) = ap \text{ and } go \in \mathcal{GO}_{c}\end{array} \\ \hline 0.5 & \begin{array}{l}\text{if } \texttt{arg}(go) = th \text{ and } go \in \mathcal{GO}_{p}, \text{ or} \\ \text{if } \texttt{arg}(go) = rw \text{ and } go \in \mathcal{GO}_{p}, \text{ or} \\ \text{if } \texttt{arg}(go) = ap \text{ and } go \in \mathcal{GO}_{p}\end{array} \end{cases}$$

$$\texttt{EFF}(go) = \begin{cases} & \text{if } \texttt{arg}(go) = th \text{ and } go \in \mathcal{GO}_c, \text{ or} \\ 0.75 & \text{if } \texttt{arg}(go) = rw \text{ and } go \in \mathcal{GO}_a, \text{ or} \\ & \text{if } \texttt{arg}(go) = ap \text{ and } go \in \mathcal{GO}_a \\ \hline & \text{if } \texttt{arg}(go) = th \text{ and } go \in \mathcal{GO}_e, \text{ or} \\ 1 & \text{if } \texttt{arg}(go) = rw \text{ and } go \in \mathcal{GO}_{canc}, \text{ or} \\ & \text{if } \texttt{arg}(go) = ap \text{ and } go \in \mathcal{GO}_{canc} \end{cases}$$

Based on the importance and the effectiveness of a opponent's goal it can be estimated how valuable this goal is. Thus, the worth of an opponent's goal is is represented by $\texttt{WORTH} : \mathcal{GO} \rightarrow [0,1]$ and it is estimated as follows.

Definition 4.6. (Worth of the opponent's goal) Let go be an opponent's goal, $\texttt{IMP}(go)$ its importance, and $\texttt{EFF}(go)$ its effectiveness. The equation for calculating the worth of go is:

$$\texttt{WORTH}(go) = \frac{\texttt{IMP}(go) + \texttt{EFF}(go)}{2} \tag{3}$$

We use the average value in order to obtain the final value of the worth a of an opponent's goal because we consider that both criteria are equally significant to make the calculation and they do not overlap each other, since each of them characterizes a different aspect of the goal. We also want to keep the values of the worth of the goal in the same interval than the values of the both criteria, namely importance and effectiveness.

So far, we have analysed the criteria to estimate how valuable an opponent's goal is. In order to evaluate the pre-condition preferability, the proponent should also know the value the opponent gives to the required action in order to compare both values. If the value of the opponent's goal is greater than the value of the required action then, the argument that uses that goal is considered preferable. Let us explain it with human examples. During an assault, the thief threatens the victim with the following sentence: "*If do not give me your bag, I hurt you*". In this situation, it is rational to think that the physical well-being is above all. Hence, the value of the goal of the victim (the opponent) is greater than the value of the required action. In another scenario, we have a boss that is trying to convince one of his employees to work on Saturdays with the following reward: "*If you work every Saturday, then I give you a chocolate*". In this situation, it is reasonable to believe that that the value of the opponent's goal is not grater than the value of the required action.

In the proposal of (GUERINI; CASTELFRANCHI, 2006), the authors claim that a threat (reward or appeal) can be considered **convincing** when it is credible and preferable. How-

ever, depending on the scenario, the proponent agent may or not have information about the real value of an action for his opponent. Thus, we can divide the scenarios in: (i) *fully informed scenarios*, in which the proponent knows both the value of the actions for his opponent and the value the opponent's goals; and (ii) *partially informed scenarios*, in which the proponent only know the value of the opponent's goals. Therefore, the preferability of a given goal can only be evaluated in fully informed scenarios. An example of this kind of scenario may be the rescue robot scenario. In partially informed scenarios the preferability cannot be evaluated; however, the proponent agent has the value of the opponent's goal. While it is true that the proponent will not know if an argument is convincing, he can base on the value of its strength to decide which argument to choose and send to his opponent.

For fully informed scenarios, the preferabililty of a given opponent's goal is determined in the following definition.

Definition 4.7. (Preferability of an opponent's goal) Let $go \in \mathcal{GO}$ be an opponent's goal and $ac \in \mathcal{AO}$ an opponent's action. Goal go is preferable if $\texttt{WORTH}(go) > \texttt{VALUE}(ac)$; otherwise, it is not preferable.

Notice that in the rescue robots scenario, the base of actions may be the same for all of them. Therefore, we may have that $\mathcal{A} = \mathcal{AO}$.

4.4.2 STEPS OF THE MODEL

In the previous sub-section we have studied the pre-conditions for a rhetorical argument be consider convincing. In other words, if a rhetorical argument meets the pre-conditions previously presented, then the proponent agent believes that he is able to convince his opponent to perform the requested action. Assuming that the agent has more than one convincing rhetorical argument, he still needs a way to compare such arguments. Therefore, a strength value for each argument is still necessary. Thus, in this sub-section we will study how these pre-conditions can be combined to obtain the strength of each argument. We propose a strength calculation model based on the previously studied pre-conditions and that can be applied when the proponent has only one possible opponent or when the proponent can choose an opponent to send an argument. The output of this proposal is a set of rhetorical arguments with their respective strength values.

The first step of the calculation model is related to the proponent's credibility. Let us recall that an opponent agent has a credibility threshold that indicates the lowest value of the reputation of the proponent agent so the arguments of him may be considered credible. Let us

suppose that a proponent agent P – with reputation value $\texttt{REP}(P) = 0.6$ – has two opponents O_1 and O_2 and let $\texttt{THRES}(O_1) = 0.5$, $\texttt{THRES}(O_2) = 0.8$ be the thresholds of agents O_1 and O_2, respectively. Since $\texttt{REP}(P) \geqslant \texttt{THRES}(O_1)$, the proponent P can continue in the process of evaluation of the strength of the arguments addressed to O_1, but it does not occur for agent O_2 because $\texttt{REP}(P) \ngeqslant \texttt{THRES}(O_2)$.

When the proponent is considered credible by his opponent(s), the next step of the model is related to the preferability notion. It is important to highlight that only when the proponent believes that he is credible he can continue to the next step. Thus, regarding preferability two possibilities can be distinguished:

1. *Fully informed scenarios*: We can differentiate two sets of arguments. One set includes the arguments that are constructed using a preferable opponent's goal and the other set includes the arguments that are constructed using non-preferable goals. All the arguments of the first set are considered convincing. This means that any of these arguments can make the opponent performs the required action. The strength value of these arguments is considered an *absolute value*. Thus, in this kind of scenarios the agent is sure that will convince his opponent if the first set has at least one argument; otherwise the agent is sure that will not convince his opponent.

2. *Partially informed scenarios*: In this kind of scenario, we only have one set of arguments whose strength values are considered *relative values* because the agent is not sure about the convincing power of his arguments.

We can notice that the preferability concept has a direct impact on the assurance of the agent that his arguments are convincing or not. However, the preferability itself is not a value that represents the strength of an argument. Since the preferability evaluation is based on the worth of the opponent's goal, we will use this value in order to rank the arguments addressed to a given opponent. Thus, we will call this value the *basic strength* of an argument.

Definition 4.8. (Basic strength) Let $A = \langle T, g, go \rangle$ be a rhetorical argument, the basic strength of A is obtained by applying the same formula used for calculating the worth of go.

$$\texttt{ST\_BASIC}(A) = \texttt{WORTH}(go) = \frac{\texttt{IMP}(go) + \texttt{EFF}(go)}{2} \tag{4}$$

A direct consequence of the above definition is that the value of the basic strength of a rhetorical argument is a real number between 0 and 1. Formally:

Property 4.1. Let $A = \langle T, g, go \rangle$ be a rhetorical argument. Since the value of the importance of $go \in [0,1]$ and the effectiveness value of go is also between 0 and 1, then $\texttt{ST\_BASIC}(A) \in [0,1]$, where 0 represents the minimum value and 1 represents the maximum value of the basic strength.

The basic strength is useful for ranking the arguments; however, for a more accurate value of the strength of the arguments, we can also take into account the credibility value. We will call this value the *combined strength* of an argument.

Before presenting the formula for calculating the combined strength of an argument, let us analyse de following situation. P is a proponent agent and O his opponent, let $\texttt{REP}(P) = 0.6$ be the reputation of agent P and $\texttt{THRES}_1(O) = 0.5$ and $\texttt{THRES}_2(O) = 0.2$ be two possible thresholds of O. We can notice that $\texttt{THRES}_1$ reflects a less credulous attitude than $\texttt{THRES}_2$; thus, although P is credible in both cases, the "accurate" value of P's credibility is different for each case since the difference between $\texttt{REP}(P)$ and $\texttt{THRES}_1$ is less than the difference between $\texttt{REP}(P)$ and $\texttt{THRES}_2$. Therefore, the credibility value of P should have an impact on the calculation of the strength of the arguments because the higher the difference between the threshold value and the reputation value is, the higher the credibility of the proponent is.

We use next Equation to calculate the **"accurate" value of the credibility** of P with respect to an opponent O, whose threshold is $\texttt{THRES}(O)$.

$$\texttt{ACCUR\_CRED}(P,O) = \texttt{REP}(P) - \texttt{THRES}(O) \tag{5}$$

This value is used to obtain the combined strength of the arguments. Thus, the combined strength of an argument depends on the basic strength of the argument and the "accurate" value of the proponent's credibility.

Definition 4.9. (Combined strength) Let $A = \langle T, g, go \rangle$ be a rhetorical argument and $O \in \mathcal{O}pp$ be an opponent whose threatened/rewarded/appealed goal is go. The combined strength of A is obtained by applying:

$$\texttt{ST\_COMB}(A) = \texttt{ST\_BASIC}(A) \times \texttt{ACCUR\_CRED}(P,O) \tag{6}$$

We can say that the value of the combined strength of an argument is a real number that is between zero and the product of the basic strength times the proponent reputation value. Thus, the combined strength has its maximum value when the opponent's threshold is zero and the basic strength of the argument is maximal. Formally:

Property 4.2. Let $A = \langle T, g, go \rangle$ be a rhetorical argument – whose basic strength is $\texttt{ST\_BASIC}(A)$ – and $\texttt{REP}(P)$ be the value of the proponent's reputation. The combined strength of A is a real number that is in the following interval $\texttt{ST\_COMB}(A) \in [0, \texttt{ST\_BASIC}(A) \times \texttt{REP}(P)]$.

Figure 11 depicts a work-flow of our approach for the strength evaluation. In summary, first of all, the proponent has to evaluate his credibility with respect to his opponent, if he is not credible enough he stops the process; otherwise, he continues. Depending on the type of scenario, the preferability is evaluated or not. Besides, the agent may choose to take into account the accuracy, in such case the combined strength is calculated; otherwise, he only calculates the basic strength.

4.5 ANALYSIS OF THE PROPOSAL

In this section, we analyse the behavior of the basic and the combined strength equations in all the possible scenarios that arise considering the different values of the importance and the effectiveness of the goal. Besides, we present a set of criteria for evaluating our proposal.

4.5.1 SCENARIOS FOR THE BASIC CALCULATION

Ten possible scenarios for the basic strength calculation are shown in Table 8.Scenario 1 illustrates the behavior of the equation of basic strength (Equation 5) with extreme opposite values, and the result is that both basic strengths have the same value. Hereafter, we assume that the values of the first column of each scenario are associated to argument A and the values of the second column are associated to argument B.

If we only consider the importance, the argument B would be stronger than the argument A; however, when the effectiveness of the opponent's goal is also taken into account the value of the strength is the same. The next proposition generalizes this situation:

Proposition 4.1. Let $A = \langle T, g_1, go_1 \rangle$ and $B = \langle T, g_2, go_2 \rangle$ be two rhetorical arguments, if $\texttt{IMP}(go_1) = \texttt{EFF}(go_2)$ and $\texttt{IMP}(go_2) = \texttt{EFF}(go_1)$, then $\texttt{ST\_BASIC}(A) = \texttt{ST\_BASIC}(B)$.

Scenario 2 shows that the values of the importance and the effectiveness of the opponent's goal of argument A are greater than the values of the importance and the effectiveness of the opponent's goal of argument B. The result is that the basic strength of A is greater than the basic strength of B. In the Scenario 3, the situation is the opposite, that is, the values for

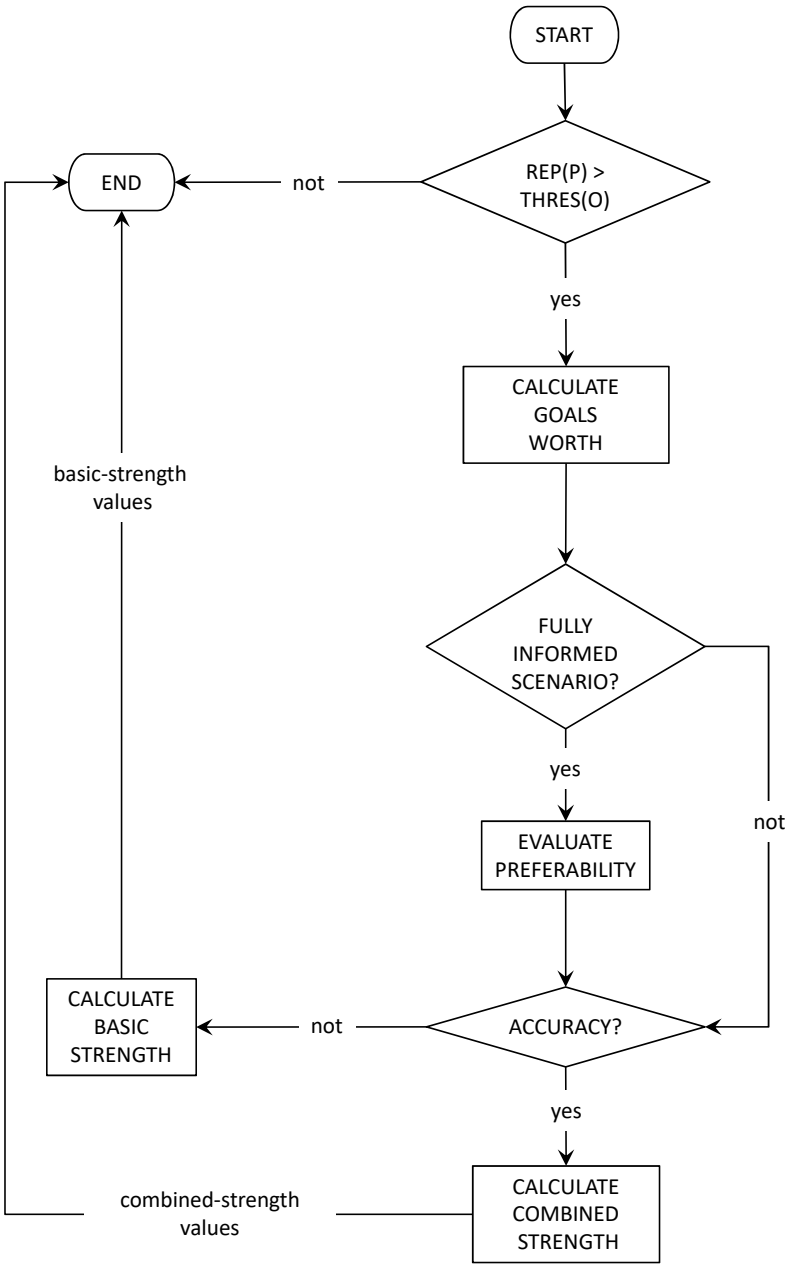


Figure 11: Work-flow of the proposed strength calculation model.

calculating the basic strength of A are less than the values for calculating the basic strength of B. The result is that the basic strength of A is less than the basic strength of B. The following proposition generalizes these situations:

Proposition 4.2. Let $A = \langle T, g_1, go_1 \rangle$ and $B = \langle T, g_2, go_2 \rangle$ be two rhetorical arguments. If $\texttt{IMP}(go_1) > \texttt{IMP}(go_2)$ and $\texttt{EFF}(go_1) > \texttt{EFF}(go_2)$, then $\texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B)$.

Scenarios 4, 5, 6, 7, and 8 show the basic strength value for each one of the different effectiveness values of an opponent's goal. Notice that, the more effective an opponent's goal is, the higher the values of both lower and upper bounds are. Therefore, when an opponent's goal

	Scenario 1		**Scenario 2**		**Scenario 3**	
	A	*B*	*A*	*B*	*A*	*B*
IMP(go)	0	1	0.8	0.4	0.4	0.45
EFF(go)	1	0	0.5	0.25	0.5	0.75
ST_BASIC	0.5	0.5	0.65	0.33	0.45	0.6

	Scenario 4	**Scenario 5**	**Scenario 6**	**Scenario 7**	**Scenario 8**	
	A	*B*	*A*	*B*	*A*	*B*
IMP(go)	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	
EFF(go)	0	0.25	0.5	0.75	1	
ST_BASIC	[0,0.5]	[0.125, 0.625]	[0.25,0.75]	[0.375, 0.875]	[0.5,1]	

	Scenario 9		**Scenario 10**			
	A	*B*	*C*	*D*		
IMP(go)	0.7	0.4	0.7	0.1		
EFF(go)	0.25	0.75	0.25	0.75		
ST_BASIC	0.48	0.58	0.48	0.43		

Table 8: Representative scenarios of the basic strength calculation.

is more effective, the strength of the argument made of it increases, even though the importance value is low. For instance, when the importance value is zero and the effectiveness is 0, the minimum strength value is zero (Scenario 4); however, when the effectiveness 0.75 the strength value is at least 0.375 (Scenario 7). The following proposition describes this behavior:

Proposition 4.3. Let $A = \langle T, g_1, go_1 \rangle$ and $B = \langle T, g_2, go_2 \rangle$ be two rhetorical arguments. Two cases can be distinguished:

1. If $\texttt{EFF}(go_1) > \texttt{EFF}(go_2)$ and $\texttt{IMP}(go_1) = \texttt{IMP}(go_2)$, then $\texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B)$
2. If $\texttt{EFF}(go_1) = \texttt{EFF}(go_2)$ and $\texttt{IMP}(go_1) > \texttt{IMP}(go_2)$, then $\texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B)$.

Now, let us analyse scenarios 9 and 10. The value of the effectiveness of the opponent's goal is the same in arguments *A* and *C* (it is 0.25) and it is the same in arguments *B* and *D* (it is 0.75). Otherwise, the value of the importance is the same in arguments *A* and *C* (it is 0.7) but it is different in threats *B* and *D* (it is 0.4 in *B* and 0.1 in *D*). Notice that in both scenarios, the value of the importance of the opponent's goal in arguments *A* and *C* is greater than the importance in arguments *B* and *D*, respectively, and the value of the effectiveness in arguments *A* and *C* is less than the value of the effectiveness in arguments *B* and *D*, respectively. However, even though there is a pattern between the values of importance and between the values of effectiveness, the value of the basic strength does not follow a pattern. Thus, in the Scenario 9, the basic strength of *A* is less than the basic strength of *B*, and in the Scenario 10, the basic strength of *C* is greater than the basic strength of *D*.

We can notice that this behavior depends on the value of the importance. Besides, we have noticed that it also depends on the difference between the effectiveness value of the opponent's goals. The following proposition better describes this behavior:

Proposition 4.4. Let $A = \langle T, g_1, go_1 \rangle$ and $B = \langle T, g_2, go_2 \rangle$ be two rhetorical arguments. Three cases can be distinguished considering the difference between the effectiveness values of the opponent's goal:

1. Let $\texttt{EFF}(go_1) = \texttt{EFF}(go_2) + 0.25$.

   If $\texttt{IMP}(go_1) \geq (\texttt{IMP}(go_2) - 0.25)$, then $\texttt{ST\_BASIC}(A) \geq \texttt{ST\_BASIC}(B)$.

   Otherwise, if $\texttt{IMP}(go_1) < (\texttt{IMP}(go_2) - 0.25)$, $\texttt{ST\_BASIC}(A) < \texttt{ST\_BASIC}(B)$.

2. Let $\texttt{EFF}(go_1) = \texttt{EFF}(go_2) + 0.5$.

   If $\texttt{IMP}(go_1) \geq (\texttt{IMP}(go_2) - 0.5)$, then $\texttt{ST\_BASIC}(A) \geq \texttt{ST\_BASIC}(B)$.

   Otherwise, if $\texttt{IMP}(go_1) < (\texttt{IMP}(go_2) - 0.5)$, $\texttt{ST\_BASIC}(A) < \texttt{ST\_BASIC}(B)$.

3. Let $\texttt{EFF}(go_1) = \texttt{EFF}(go_2) + 0.75$.

   If $\texttt{IMP}(go_1) \geq (\texttt{IMP}(go_2) - 0.75)$, then $\texttt{ST\_BASIC}(A) \geq \texttt{ST\_BASIC}(B)$.

   Otherwise, if $\texttt{IMP}(go_1) < (\texttt{IMP}(go_2) - 0.75)$, $\texttt{ST\_BASIC}(A) < \texttt{ST\_BASIC}(B)$.

4.5.2 SCENARIOS FOR THE COMBINED CALCULATION

For the combined calculation, the accurate credibility value is taken into account. In order to study the behavior of Equation 7, we analyze the resultant values when $\texttt{REP}(P) = 1$ and $\texttt{THRES}(O)$ varies from 0 to 1 (see Table 9).

Consider the following scenario, one of the thresholds is greater than the other one and the basic strength is the same:

- $\texttt{THRES}_1(O) = 0.7$ and $\texttt{ST\_BASIC}(A) = 0.4$, hence $\texttt{ST\_COMB}(A) = 0.12$;
- $\texttt{THRES}_2(O) = 0.4$ and $\texttt{ST\_BASIC}(B) = 0.4$, hence $\texttt{ST\_COMB}(B) = 0.24$;

Note that since $\texttt{THRES}_1(O) > \texttt{THRES}_2(O)$, the combined strength of B is greater than the combined strength of A. This happens because the higher the threshold is, the lower the value of the accurate credibility is. This illustrate that a less credulous opponent is also less influential. Formally:

		ST_BASIC									
THRES(P)	ACCUR_CRED	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
0	**1**	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	**0.9**	0.09	0.18	0.27	0.36	0.45	0.54	0.63	0.72	0.81	0.9
0.2	**0.8**	0.08	0.16	0.24	0.32	0.40	0.48	0.56	0.64	0.72	0.8
0.3	**0.7**	0.07	0.14	0.21	0.28	0.35	0.42	0.49	0.56	0.63	0.7
0.4	**0.6**	0.06	0.12	0.18	0.24	0.30	0.36	0.42	0.48	0.54	0.6
0.5	**0.5**	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.5
0.6	**0.4**	0.04	0.08	0.12	0.16	0.20	0.24	0.28	0.32	0.36	0.4
0.7	**0.3**	0.03	0.06	0.09	0.12	0.15	0.18	0.21	0.24	0.27	0.3
0.8	**0.2**	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.2
0.9	**0.1**	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
1	**0**	0	0	0	0	0	0	0	0	0	0

Table 9: Values for the combined strength when $\texttt{REP}(P) = 1$

Proposition 4.5. Let A and B be two rhetorical arguments, and $\texttt{THRES}_1(O)$ and $\texttt{THRES}_2(O)$ represent two thresholds, each one associated to arguments A and B, respectively. If $\texttt{ST\_BASIC}(A) = \texttt{ST\_BASIC}(B)$ and $\texttt{THRES}_1(O) > \texttt{THRES}_2(O)$, then $\texttt{ST\_COMB}(A) < \texttt{ST\_COMB}(B)$.

Table 10 shows three scenarios that illustrate what happens when both parameters are different.

	Scenario 1		**Scenario 2**		**Scenario 3**	
	A	B	A	B	A	B
ST_BASIC	0.7	0.5	0.7	0.5	1	0.2
THRES(O)	0.6	0.8	0.6	0.3	0.7	0.1
ST_COMB	0.28	0.1	0.28	0.35	0.3	0.18

Table 10: Scenarios where the threshold value and the basic strength are different for each threat. We use $\texttt{REP}(P) = 1$ for calculating the value of the combined strength.

In the scenario 1, the basic strength of argument A is greater than the basic strength of argument B, and the threshold value of argument A is less than the threshold value of argument B. The result is that the value of the combined strength of argument A is greater than the combined strength of argument B. We have observed that this behaviour is always the same, i.e. when the basic strength of a given threat is greater than the basic strength of another one, but its threshold is less than the other one, then it is the strongest one in terms of combined strength. The following proposition presents this situation:

Proposition 4.6. Let A and B be two rhetorical arguments, and $\texttt{THRES}_1(O)$ and $\texttt{THRES}_2(O)$ be two different thresholds, each one associated to arguments A and B, respectively. If $\texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B)$ and $\texttt{THRES}_1(O) < \texttt{THRES}_2(O)$, then $\texttt{ST\_COMB}(A) > \texttt{ST\_COMB}(B)$.

In the scenarios 2 and 3, both the basic strength and the threshold value of argument A are greater than the basic strength and the threshold value of argument B. However, note that in the Scenario 2 the combined strength of argument A is less than the combined strength of argument B and in the Scenario 3 is the opposite, that is, the combined strength of argument A is greater than the combined strength of argument B. The following proposition better describes this behavior:

Proposition 4.7. Let A and B be two rhetorical arguments, and $\texttt{THRES}_1(O)$ and $\texttt{THRES}_2(O)$ be two different thresholds, each one associated to arguments A and B, respectively. Given $\texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B)$ and $\texttt{THRES}_1(O) > \texttt{THRES}_2(O)$, one of the following situations occur:

1. If [3,4] $\frac{\texttt{ST\_BASIC}(A)}{\texttt{ACCUR\_CRED}_2} < \frac{\texttt{ST\_BASIC}(A)-\texttt{ST\_BASIC}(B)}{\texttt{ACCUR\_CRED}_2-\texttt{ACCUR\_CRED}_1}$, then $\texttt{ST\_COMB}(A) > \texttt{ST\_COMB}(B)$.
2. If $\frac{\texttt{ST\_BASIC}(A)}{\texttt{ACCUR\_CRED}_2} > \frac{\texttt{ST\_BASIC}(A)-\texttt{ST\_BASIC}(B)}{\texttt{ACCUR\_CRED}_2-\texttt{ACCUR\_CRED}_1}$, then $\texttt{ST\_COMB}(A) < \texttt{ST\_COMB}(B)$.
3. If $\frac{\texttt{ST\_BASIC}(A)}{\texttt{ACCUR\_CRED}_2} = \frac{\texttt{ST\_BASIC}(A)-\texttt{ST\_BASIC}(B)}{\texttt{ACCUR\_CRED}_2-\texttt{ACCUR\_CRED}_1}$, then $\texttt{ST\_COMB}(A) = \texttt{ST\_COMB}(B)$.

4.5.3 EVALUATION CRITERIA

In this subsection, we analyze the model of strength evaluation proposed in this work. We will base our analysis on the study done by Carl Hovland at Yale University, which is detailed in (DEMIRDÖĞEN, 2010). One of the main contributions of this study is a specification of a set of characteristics that have a positive influence on an individual during a persuasion dialogue. These characteristics are related to (i) the source of the persuasive communication, (ii) the aspects of the message, and (iii) the characteristics of the receiver and of the message context. It is important to highlight that these characteristics are closely related to those presented by Aristotle in Rhetoric, i.e. ethos, pathos and logos (BRAET, 1992; HIGGINS; WALKER, 2012).

[3] Recall that $\texttt{ACCUR\_CRED}(P,O) = \texttt{REP}(P) - \texttt{THRES}(O)$.

[4] It is clear that $\texttt{ACCUR\_CRED}_2 - \texttt{ACCUR\_CRED}_1$ is a positive number because it holds that $\texttt{THRES}_1(O) > \texttt{THRES}_2(O)$. Let us demonstrate it: $\texttt{ACCUR\_CRED\_PART}_2 > \texttt{ACCUR\_CRED}_1 = (1 - \texttt{THRES}_2(O)) > (1 - \texttt{THRES}_1(O)) = -\texttt{THRES}_2(O) > -\texttt{THRES}_1(O) = \texttt{THRES}_2(O) < \texttt{THRES}_1(O)$.

Thus, the evaluation criteria for the validation of our proposal are the following:

1. The source (i.e. the proponent agent) should have some persuasive characteristic as credibility, appeal, or similarity with the opponent.

2. The persuasive message (i.e. the rhetorical argument) should take into account the receiver (i.e. the opponent) from the emotional[5] and rational point of view and should be organized.

3. The context of the dialogue should be considered and the characteristics of the receiver should be taken into account in order to find out which receivers are more likely to be persuaded.

Below, we explain to what extent the proposed model accomplish these criteria:

1. In our proposal, the proponent's credibility is evaluated before the strength calculation is performed. We use an opponent's threshold that indicates if the proponent's credibility value is enough or not to continue with the calculation. Regarding the appeal, this characteristic is more related to the physical aspect of the proponent, and for this reason it is out of the scope of this research, since we work only with software agents. On the other hand, the similarity is related to the social environment[6], which was not addressed in this work, but we think it is interesting to include it in future research because it could indeed affect the strength of a rhetorical argument.

2. Regarding the content of rhetorical arguments, this was defined in a logical form and some conditions were established. Thus, on the one hand, the minimality condition prevents the emergence of repetitive content that could cause disorganization in the content. On the other hand, the consistency condition leads to a valid and correct logical inference. Besides, the content of a rhetorical argument takes into account the opponent from a rational point of view since it considers an opponent's goal, and the calculation of the strength is based on two main aspect of it, namely its importance and its effectiveness. The emotional aspects of the opponent have not been considered, but it could also be a good future direction.

[5] According to (MEYER, 2006), an emotional agent is an artificial system that is designed in such a manner that emotions play a role. Thus, the emotional state the agent may determine his actions or part of his actions.

[6] A social environment is a communication environment in which agents interact in a coordinated manner (ODELL et al., 2002).

3. In relation to the last criterion, the context is not taken into account; however, the receiver is considered, specifically, the importance of the goal of the receiver, the effectiveness of this goal, and the value of the required action. We can also use the combined strength to compare how persuadable an opponent is with respect to other one because the lower a threshold is, the more credulous an opponent is; and the higher a threshold is, the less credulous an opponent is.

4.6 APPLICATION

In this section, we present the application of the proposed model to both scenarios the rescue robots scenario and the software agents scenario. We will use the threats, rewards, and appeals that were introduced in Chapter 1. We first show the logical formalization and them we make the basic strength calculation of each of them.

4.6.1 SOFTWARE AGENTS SCENARIO

In this scenario, we work with threats and rewards. Recall that this scenario is an example of partially informed scenario; therefore, the value of the arguments strength is considered relative.

THREATS

Once again, let us recall the threats that agent CONSUMER can construct to try to convince agent COMPANY to refund the money of his ticket.

- th_1 : *You should refund the total price of the ticket, otherwise I will never buy a ticket in your company anymore.*
- th_2 : *You should refund the total price of the ticket, otherwise I will destroy your reputation in social networks.*
- th_3 : *You should refund the total price of the ticket, otherwise I will take legal actions against your company.*

Next, we present the mental state of agent CONSUMER and the logical formalization of each threat. Hereafter, we omit some elements from the mental state because they are not necessary.

$\texttt{CONSUMER} = \langle \mathcal{T}, \mathcal{G}, \mathcal{O}pp, \mathcal{GO}, \mathcal{S}_{\mathcal{O}pp}, \mathcal{S}_{\mathcal{GO}}, \mathcal{A}, \mathcal{AO}, \texttt{REP} \rangle$ where:

$\mathcal{T} = \{\mathcal{F}, \mathcal{S}, \mathcal{D}\}$ such that

$\mathcal{S} = \{\neg$ *refund(money)* $\rightarrow \neg buy\_again(ticket)$,

$\neg$ *refund(money)* $\rightarrow destroy(rep\_social\_net)$,

$\neg$ *refund(money)* $\rightarrow take(legal\_actions)$,

$\neg buy\_again(ticket) \rightarrow \neg gain(costu\_fidel)$,

$destroy(rep\_social\_net) \rightarrow \neg have(good\_rep)$,

$take(legal\_actions) \rightarrow \neg avoid(legal\_probs)\}$

$\mathcal{G} = \{g\}$ such that $g = get(\texttt{COMPANY}, \text{`}refund(money)\text{'})$

$\mathcal{O}pp = \{\texttt{COMPANY}\}$

$\mathcal{GO}_a = \{go_3\}, \mathcal{GO}_c = \{go_1\}, \mathcal{GO}_e = \{go_2\}$ such that $go_1 = gain(costu\_fidel)$, $go_2 = have(good\_rep)$, and $go_3 = avoid(legal\_probs)$

$\mathcal{S}_{\mathcal{O}pp} = \{(\texttt{COMPANY}, 0.75, \{go_1, go_2, go_3\})\}$ such that $\texttt{THRES}(\texttt{COMPANY}) = 0.75$, and $\{go_1, go_2, go_3\} \in TH_{\mathcal{GO}}$

$\mathcal{S}_{\mathcal{GO}} = \{(go_1, 0.85), (go_2, 0.85), (go_3, 0.7)\}$, and $\texttt{REP} = 0.8$.

From this mental state, the following threats can be generated:

$th_1 = \langle T_1, g, go_1 \rangle$ where:

$T_1 \cup \neg\texttt{SECOND}(g) = \{(\neg$ *refund(money)*$, \emptyset)$,

$(\neg buy\_again(ticket), \neg$ *refund(money)* $\rightarrow \neg buy\_again(ticket))$,

$(\neg gain(costu\_fidel), \neg buy\_again(ticket) \rightarrow \neg gain(costu\_fidel))\}$

$th_2 = \langle T_2, g, go_2 \rangle$ where:

$T_2 \cup \neg\texttt{SECOND}(g) = \{(\neg$ *refund(money)*$, \emptyset)$,

$(destroy(rep\_social\_net), \neg$ *refund(money)* $\rightarrow destroy(rep\_social\_net))$,

$(\neg have(good\_rep), destroy(rep\_social\_net) \rightarrow \neg have(good\_rep))\}$

$th_3 = \langle T_3, g, go_3 \rangle$ where:

$T_3 \cup \neg\texttt{SECOND}(g) = \{(\neg$ *refund(money)*$, \emptyset)$,

$(take(legal\_actions), \neg$ *refund(money)*$\rightarrow take(legal\_actions))$,

$(\neg avoid(legal\_probs), take(legal\_actions) \rightarrow \neg avoid(legal\_probs))\}$

According to the calculation model, firstly the credibility of CONSUMER has to be evaluated. Since $\texttt{REP}(\texttt{CONSUMER}) > \texttt{THRES}(\texttt{COMPANY})$ (that is, $0.8 > 0.75$), we can proceed to calculate the basic strength values of the threats generated by CONSUMER. Since there is only one possible opponent, then it is not necessary to make the calculation of the combined strength values. Table 11 shows the basic and combined values of the strength of the threats generated by agent CONSUMER, the combined strength is calculated considering that $\texttt{ACCUR\_CRED}(\texttt{CONSUMER}, \texttt{COMPANY}) = 0.8 - 0.75 = 0.05$.

GOAL	IMP(go)	STATUS	EFF(go)	ST_BASIC(th)	ST_COMB(th)	
go_1	0.85	chosen	0.75	0.8	0.04	th_1
go_2	0.85	executive	1	0.925	0.0463	th_2
go_3	0.7	active	0.25	0.475	0.0238	th_3

Table 11: Basic and combined strength values of the threats of agent CONSUMER in the software agents scenario.

Thus, we have that threat th_2 –whose opponent's goal is $have(good\_rep)$– is the strongest one and threat th_3 –whose opponent's goal is $avoid(legal\_probs)$– is the least strong threat.

REWARDS

Let us recall the rewards that agent COMPANY can construct to try to convince agent CONSUMER to accept only the 20% refund. If CONSUMER decides to send his strongest threat (i.e., threat th_2), COMPANY can construct a counter-threat to such threat. Thus, COMPANY would have both rewards and threats to support his position. In natural language, the rewards and threat that COMPANY can generate are:

- rw_1 : *If you agree with the 20% of refund, we will give you 10000 miles.*
- rw_2 : *If you agree with the 20% of refund, we will sell you an executive ticket for the price of a economic one for any national destination.*
- rw_3 : *If you agree with the 20% of refund, we will give you our service of assistance for elderly for free for any destination.*
- th_4 : *You should not destroy my reputation, otherwise I will denounce you for defamation and I will ask for a payment of civil reparations amounting to $1000 in favor of me.*

Next, we present the mental state of agent COMPANY and the logical formalization of the three rewards and the threat.

$\texttt{COMPANY} = \langle \mathcal{T}, \mathcal{G}, \mathcal{O}pp, \mathcal{GO}, \mathcal{S}_{\mathcal{O}pp}, \mathcal{S}_{\mathcal{GO}}, \mathcal{A}, \mathcal{AO}, \texttt{REP} \rangle$ where:

$\mathcal{T} = \{\mathcal{F}, \mathcal{S}, \mathcal{D}\}$ such that

$\mathcal{S} = \{accept(refund\_20) \rightarrow gain(miles),$

$accept(refund\_20) \rightarrow get\_discount(exec\_ticket),$

$accept(refund\_20) \rightarrow free(elderly\_assist),$

$\neg drop(destroy\_rep) \rightarrow make(denounce\_difam),$

$make(denounce\_difam) \rightarrow pay\_repar(1000)$,
$pay\_repar(1000) \rightarrow \neg avoid(extra\_budget)\}$

$\mathcal{G} = \{g_1, g_2\}$ such that $g_1 = get(\texttt{CONSUMER}, \text{`}accept(refund\_20)\text{'})$ and
$g_2 = get(\texttt{CONSUMER}, \text{`}drop(destroy\_rep)\text{'})$

$\mathcal{O}pp = \{\texttt{CONSUMER}\}$

$\mathcal{GO}_a = \{go_6\}, \mathcal{GO}_c = \{go_4, go_7\}, \mathcal{GO}_{canc} = \{go_5\}$ such that $go_4 = gain(miles)$, $go_5 = get\_discount(exec\_ticket)$, $go_6 = free(elderly\_assist)$, and $go_7 = avoid(extra\_budget)$

$\mathcal{S}_{\mathcal{O}pp} = \{(\texttt{CONSUMER}, 0.7, \{go_4, go_5, go_6, go_7\})\}$ such that $\texttt{THRES}(\texttt{CONSUMER}) = 0.7$, $\{go_4, go_5, go_6\} \in RW_{\mathcal{GO}}$, and $\{go_7\} \in TH_{\mathcal{GO}}$

$\mathcal{S}_{\mathcal{GO}} = \{(go_4, 0.8), (go_5, 0.7), (go_6, 0.4), (go_7, 0.9)\}$, and $\texttt{REP} = 0.9$

From this mental state, the following rewards and threat can be generated:

$rw_1 = \langle T_1, g_1, go_4 \rangle$ where:
$T_1 \cup \texttt{SECOND}(g_1) = \{(accept(refund\_20), \emptyset),$
$(gain(miles), accept(refund\_20) \rightarrow gain(miles))\}$

$rw_2 = \langle T_2, g_1, go_5 \rangle$ where:
$T_2 \cup \texttt{SECOND}(g_1) = \{(accept(refund\_20), \emptyset),$
$(get\_discount(exec\_ticket), accept(refund\_20) \rightarrow get\_discount(exec\_ticket))\}$

$rw_3 = \langle T_3, g_1, go_6 \rangle$ where:
$T_3 \cup \texttt{SECOND}(g_1) = \{(accept(refund\_20), \emptyset),$
$(free(elderly\_assist), accept(refund\_20) \rightarrow free(elderly\_assist))\}$

$th_4 = \langle T_4, g_2, go_7 \rangle$ where:
$T_4 \cup \neg\texttt{SECOND}(g_2) = \{(\neg drop(destroy\_rep), \emptyset),$
$(make(denounce\_difam), \neg drop(destroy\_rep) \rightarrow make(denounce\_difam))$
$(pay\_repar(1000), make(denounce\_difam) \rightarrow pay\_repar(1000))$
$(\neg avoid(extra\_budget), pay\_repar(1000) \rightarrow \neg avoid(extra\_budget))\}$

First of all, the credibility of COMPANY has to be evaluated. Since $\texttt{REP}(\texttt{COMPANY}) > \texttt{THRES}(\texttt{CONSUMER})$ (that is, $0.9 > 0.7$), we can proceed to calculate the basic strength values of the arguments generated by COMPANY. Like in previous case, since there is only one possible opponent, then it is not necessary to make the calculation of the combined strength values. Table 12 shows the basic and combined values of the strength of the rewards and the threat generated by agent COMPANY, the combined strength is calculated considering that $\texttt{ACCUR\_CRED}(\texttt{COMPANY}, \texttt{CONSUMER}) = 0.9 - 0.7 = 0.2$.

GOAL	IMP(go)	STATUS	EFF(go)	ST_BASIC(rw)	ST_COMB(rw)	
go_4	0.8	chosen	0.25	0.525	0.105	rw_1
go_5	0.7	cancelled	1	0.85	0.17	rw_2
go_6	0.4	active	0.75	0.575	0.115	rw_3
GOAL	IMP(go)	STATUS	EFF(go)	ST_BASIC(th)	ST_COMB(th)	
go_7	0.9	chosen	0.75	0.825	0.165	th_4

Table 12: Basic strength values of the rewards and the threat of agent COMPANY in the software agents scenario.

Thus, we have that reward rw_2 –whose opponent's goal is $get\_discount(exec\_ticket)$– is the strongest rhetorical argument and reward rw_1 –whose opponent's goal is $gain(miles)$– is the least strong argument. However, notice that the strength value of the unique threat is very close to the the strength value of reward rw_2. This means that depending on the strategy of the agent, he can choose to send a threat or a reward.

4.6.2 RESCUE ROBOTS SCENARIO

In this scenario, we work with appeals. Recall that this scenario is a sample of fully informed scenario; therefore, the concept of preferability is employed and the arguments strength is considered absolute.

APPEALS

In Chapter 1, we have introduced three appeals that agent TOM can generate in order to try to convince agent BOB to help him with a heavy debris. Let us recall these appeals:

- ap_1 : *If you help me, you can win utility points.*
- ap_2 : *If you help me, you can recharge your battery since the workshop is next to this zone.*
- ap_3 : *If you help me, you can fix your sensor since the workshop is next to this zone.*

Next, we present the mental state of agent TOM and the logical formalization of the three appeals.

$\texttt{TOM} = \langle \mathcal{T}, \mathcal{G}, \mathcal{O}pp, \mathcal{GO}, \mathcal{S}_{\mathcal{O}pp}, \mathcal{S}_{\mathcal{GO}}, \mathcal{A}, \mathcal{AO}, \texttt{REP} \rangle$ where:

$\mathcal{T} = \{\mathcal{F}, \mathcal{S}, \mathcal{D}\}$ such that

$\mathcal{S} = \{help\_with(debris) \rightarrow gain(util\_points),$

$help\_with(debris) \rightarrow go(workshop)$,
$go(workshop) \rightarrow recharge(battery)$,
$go(workshop) \rightarrow fix(sensor)$,
$\mathcal{G} = \{g_3\}$ such that $g_3 = get(\texttt{BOB}, \text{`}help\_with(debris)\text{'})$
$\mathcal{O}pp = \{\texttt{BOB}\}$
$\mathcal{GO}_a = \{go_9\}, \mathcal{GO}_p = \{go_8\}, \mathcal{GO}_c = \{go_{10}\}$ such that
$go_8 = gain(util\_points)$, $go_9 = recharge(battery)$, and $go_{10} = fix(sensor)$
$\mathcal{S}_{\mathcal{O}pp} = \{(\texttt{BOB}, 0.7, \{go_8, go_9, go_{10}\})\}$ such that
$\texttt{THRES}(\texttt{BOB}) = 0.7$, and $\{go_8, go_9, go_{10}\} \in AP_{\mathcal{GO}}$
$\mathcal{S}_{\mathcal{GO}} = \{(go_8, 0.7), (go_9, 0.9), (go_{10}, 0.75)\}$
$\mathcal{A} = \mathcal{AO} = \{help\_with(debris)\}$
$\mathcal{A}_{val} = \{(help\_with(debris), 0.55)\}$
$\texttt{REP} = 0.8$

From this mental state, the following appeals can be generated:

$ap_1 = \langle T_1, g_3, go_8 \rangle$ where:
$T_1 \cup \texttt{SECOND}(g_3) = \{(help\_with(debris), \emptyset),$
$(gain(util\_points), help\_with(debris) \rightarrow gain(util\_points))\}$

$ap_2 = \langle T_2, g_3, go_9 \rangle$ where:
$T_2 \cup \texttt{SECOND}(g_3) = \{(help\_with(debris), \emptyset),$
$(go(workshop), help\_with(debris) \rightarrow go(workshop))$
$(recharge(battery), go(workshop) \rightarrow recharge(battery))\}$

$ap_3 = \langle T_3, g_3, go_{10} \rangle$ where:
$T_3 \cup \texttt{SECOND}(g_3) = \{(help\_with(debris), \emptyset),$
$(go(workshop), help\_with(debris) \rightarrow go(workshop))$
$(fix(sensor), go(workshop) \rightarrow fix(sensor))\}$

Like in previous scenarios, the credibility of TOM has to be evaluated. We have that TOM is considered credible by BOB based on $\texttt{REP}(\texttt{TOM}) > \texttt{THRES}(\texttt{BOB})$ (that is, $0.8 > 0.7$). Thus, we can proceed to calculate the basic strength values of the appeals generated by TOM. Table 13 shows the basic and combined values of the strength of the appeals generated by agent TOM, the combined strength is calculated considering that $\texttt{ACCUR\_CRED}(\texttt{TOM}, \texttt{BOB}) = 0.8 - 0.7 = 0.1$.

Recall that $\texttt{ST\_BASIC}(ap) = \texttt{WORTH}(go)$ such that go is the opponent's goal that makes up the appeal ap. Thus, we can compare the values of the opponent's goals and the value of the required action. The result is the following:

GOAL	IMP(go)	STATUS	EFF(go)	ST_BASIC(ap)	ST_COMB(ap)	
go_8	0.7	pursuable	0.5	0.6	0.06	ap_1
go_9	0.9	active	0.75	0.825	0.083	ap_2
go_{10}	0.75	chosen	0.25	0.5	0.05	ap_3

Table 13: Basic strength values of the appeals of agent TOM in the rescue robots scenario.

$\texttt{WORTH}(go_8) > \texttt{VALUE}(help\_with(debris))$ (0.6>0.55)
$\texttt{WORTH}(go_9) > \texttt{VALUE}(help\_with(debris))$ (0.825>0.55)
$\texttt{WORTH}(go_{10}) \not> \texttt{VALUE}(help\_with(debris))$ (0.5<0.55)

Therefore, we have that appeals ap_1 and ap_2 are convincing ones whereas appeal ap_3 is not convincing. This means that the set of arguments that agent TOM can use during the negotiation dialogue has been reduced to two.

4.7 RELATED WORK

Kraus et al. (1998) present a set of axioms for rhetorical arguments generation. In these axioms, when the rule body is satisfied, a candidate rhetorical argument is generated. With respect to the strength of the arguments, the authors claim that a threat is the strongest rhetorical argument (compared to rewards and appeals); however, a calculation model is not defined.

Ramchurn et al. (2003) propose a model where the strength value of rhetorical arguments varies during the negotiation depending on the environmental conditions. For calculating the strength value of the argument, it is taken into account a set of world states an agent can be carried to by using a certain argument. The intensity of the strength values depends on the desirability of each of these states. For a fair calculation, an average over all possible states is used.

In (AMGOUD; PRADE, 2004), a formal definition of rhetorical arguments and a strength evaluation system are presented. For the evaluation of the strength of rhetorical argument, the certainty of the beliefs that are used for the generation of the argument and the importance of the opponents goal are considered. The same authors have other later articles about rhetorical arguments ((AMGOUD; PRADE, 2006)(AMGOUD; PRADE, 2005)). In these works, the calculation of the strength value is always done by taking into account the two criteria just mentioned. In our proposal, we made a profound analysis of the components of an argument and defined new criteria for calculating the strength values.

4.8 SUMMARY

In this chapter, we have presented a formalization of a negotiating agent ground on the BBGP-based agent. We have extended it so that the negotiating agent can generate threats, rewards, and appeals and is able to calculate the strength value of them. Since it is necessary to model the knowledge of the opponent (e.g., the opponent's goals and the importance and status of such goals), we assume that the agent knows in advance such information.

In the second part, we have presented the logic formalization of threats, rewards, and, appeals. Like in the previous chapter, we use derivation schemas as the consequence operator and we use it in the definition of the rhetorical arguments.

The third part of this chapter is the core of it. We start by studying the pre-conditions for an argument to be considered convincing. We base on the proposal of (GUERINI; CASTELFRANCHI, 2006), which states that the credibility of the proponent and the preferability of the opponent's goal over the value of the required actions are these two pre-conditions. We use the reputation of the proponent and the threshold of trust of the opponent to evaluate the credibility of the proponent and the opponent's goal importance and its status to evaluate the preferability. We do not directly use the status of an opponent's goal but we judge its effectiveness based on the type of rhetorical argument and the status itself. Based on the numerical values of these pre-conditions, we have proposed a model for evaluating and calculating the strength value of the rhetorical arguments. The model starts evaluating the credibility of the proponent agent. The proponent agent can proceed to the calculation of the rhetorical arguments only if he is considered credible by his opponent; otherwise, the process ends.

In the next chapter, we will formalize a persuasive negotiation dialogue. We will define the protocol that will rule the participation of the negotiating agents defined in that chapter. We will also propose a strategy that will guide the decisions of the agent regarding the argument that should be sent.

5 PERSUASIVE NEGOTIATION DIALOGUE: SIMULATIONS AND RESULTS

In this chapter, we present a negotiation model for persuasive negotiation dialogues and a set of experiments that aim to evaluate our proposal for calculating the strength of rhetorical arguments. Thus, we evaluate our proposal in terms of efficiency, more specifically in terms of negotiation cycles, number of arguments exchanged by the agents, and number of reached agreements.

5.1 INTRODUCTION

In (WALTON, 1992), the author defines a dialogue as follows:

> *"A dialogue is an exchange of speech acts between two speech partners in turn-taking sequence aimed at a collective goal. The dialogue is coherent to the extent that the individual speech acts fit together to contribute to this goal. As well, each participant has an individual goal in the dialogue, and both participants have an obligation in the dialogue, defined by the nature of their collective and individual goals."*

This definition begins by referring to speech acts, which are utterances that can be seen not just as information transmitters but as actions that may change the state of the world (AUSTIN; URMSON, 1962). In the case of dialogues between agents, the speech acts can also be seen as actions that may change the mind of the agents. In the same sentence, the author mentions that there is a turn-taking sequence. This sequence rules the participation of the agents during the dialogue. Finally, it is mentioned a collective goal. This means that the participant agents have a goal in common that governs the dialogue among them. According to the definition, an important characteristic of a dialogue is the coherence. This means that there exists a set of legal movements the agent can do during the dialogue, with movement we mean the next utterance the agent can emit. In dialogues between agents, such movements are ruled by a protocol. Besides the collective goal, the author also indicate that each participant agent has an individual goal, which impacts on the utterances he emits. The impact depends on the type

of dialogue he is engaged. At last, the author states that the agents have an obligation. These obligations are also dependant on the type of dialogue. For example, in a basic query-answer dialogue the expert agent has the obligation of responding the queries of other agents.

We can notice that some aspects of a dialogue and the agents depend on type of dialogue. A well known dialogue taxonomy is given in (WALTON; KRABBE, 1995). The authors divide dialogues into five different types (see Table 14, adapted from (WALTON; KRABBE, 1995)).

Type	**Initial situation**	**Participant's goal**	**Goal of the dialogue**
Persuasion	Conflict of opinions	Persuade the other(s)	Resolution of the conflict
Inquiry	General ignorance	Find and verify evidence	Prove (or disprove) hypothesis
Negotiation	Conflict of interests	Get the best bargain for himself	Making a deal
Information-seeking	Personal ignorance	Acquire or give information	Exchange information
Deliberation	Dilemma	Influencing the decision	Reach a decision

Table 14: Walton and Krabbe's classification of dialogues.

In this thesis, the participant agents (i.e., proponent and opponent) are engaged in persuasive negotiation dialogues. These are negotiation dialogues where the proposals of the participant agents are backed up by rhetorical arguments (RAMCHURN et al., 2003).

Let us recall the software agent scenario where agent `CONSUMER` acts in behalf of a passenger of an airline and agent `COMPANY` acts on behalf of the airline. The user of `CONSUMER` missed an international flight due to a schedule change and he wants the airline company reimburses him the total price of the ticket. In Chapter 1, the beginning of the conversation between the two agents was presented. Next, we recall it and add some new possible answers of the agents trying to defend their positions.

`CONSUMER`: Since I was not properly informed about the schedule change of my flight I ask for the reimbursement of the total cost of the ticket.

`COMPANY`: We are sorry, but we sent an e-mail about the schedule change to every passenger. According to our policies we only can refund the 20% of the total price of the ticket, without including taxes.

`CONSUMER`: You should refund the total price of the ticket, otherwise I will never buy a ticket in your company anymore.

COMPANY: A total refund is not possible. However, if you agree with the 20% of refund, we will give you free service of assistance for elderly for any destination.

CONSUMER: I am a lawyer; hence, I know my rights. A total refund is perfectly possible. Therefore, you should refund the total price of the ticket, otherwise I will take legal actions against your company.

Since COMPANY rejected the proposal of CONSUMER, both agents engage in a negotiation dialogue where each agent defends their interests by using threats, rewards, or appeals. On the one hand, the interest of CONSUMER is to get back the total value of his ticket, and on the other hand, the interest of COMPANY is to pay only a part of such value. CONSUMER begins the persuasive negotiation. He tries to convince COMPANY to accept his proposal by using a threat. In response, COMPANY defends his position and sends a reward to CONSUMER offering a free assistance service for an elder. CONSUMER defends again his position by using a threat. The dialogue can continue until both agents reach an agreement or until both agents have no useful arguments. With useful arguments, we mean arguments that are strong enough to convince the opponent.

In this chapter, we simulate this kind of dialogue. In these simulations, pairs of agents use the same mechanism for calculating the strength of their arguments. We consider two forms of strength calculation: (i) the strength calculation that is based only on the importance value of the opponent's goal and (ii) the strength calculation that is based on our proposal described in Chapter 4. We compare the results of both mechanisms in terms of negotiation cycles, number of arguments exchanged by the agents, and number of reached agreements.

This chapter is organized as follows. In next Section, we present a basic negotiation model for persuasive negotiation dialogues. Based on this model, in Section 5.3, we construct the dialogues for the scenarios of the rescue robots and the software agents. Section 5.4 is devoted the empirical evaluation of the strength calculation model proposed in Chapter 4 in accordance with the proposed negotiation model. Thus, we present three simulations and their respective results. Finally, Section 5.6 summarizes this chapter.

5.2 THE NEGOTIATION MODEL

In this section, we detail some important aspects about the negotiation model. Thus, we mainly focus on the notion of dialogue, on the negotiation protocol, and on modelling the response of the agents during a negotiation encounter.

1. **The participants:** There are only two participants by negotiation encounter. We call them P and O, where P represents the proponent agent and O the opponent agent.

2. **The communication language $\mathcal{L}_c$:** We use the well-known speech acts $\texttt{REQUEST}(x)$, $\texttt{REJECT}(x)$, and $\texttt{ACCEPT}(x)$. Recall that the requested action ac can be obtained from the outsourced goal of the proponent g: $ac = \texttt{SECOND}(g)$. Furthermore, we use specific speech acts for the rhetorical arguments as proposed in (AMGOUD; PRADE, 2006) and `WITHDRAW` to indicate that the agent withdraws from the negotiation. Thus, we have that $\mathcal{L}_c = \{\texttt{REQUEST}(\texttt{SECOND}(g)), \texttt{REJECT}(\texttt{SECOND}(g)), \texttt{ACCEPT}(\texttt{SECOND}(g)), \texttt{THREAT}(T,g,go), \texttt{REWARD}(T,g,go), \texttt{APPEAL}(T,g,go), \texttt{WITHDRAW}\}$.

3. **Move:** A move –made by any of the participant agents– is represented by a tuple $m = \langle id, ag, tg, sp \rangle$ where:

   - $id \in \mathbb{N}$ is the identifier of the move. For the sake of simplicity, we use m_{id} to refer to a given move;
   - $ag \in \{P, O\}$ is the agent that sends the utterance, i.e. the agent that carries out the move;
   - $tg \in \mathbb{N}$ is target of the move, i.e. a previous move to which this is directed;
   - $sp \in \mathcal{L}_c$ is the speech act performed in the move.

4. **Dialogue:** A dialogue D between two agents P and O is a finite sequence $D = \langle m_1, ..., m_n \rangle$, ruled by a negotiation protocol.

5. **Negotiation protocol:** We use an elementary protocol, where the two agents O and P make moves alternately. When the proposal made by the proponent is rejected then the agents exchange rhetorical arguments. The kind of the rhetorical argument is not taken into account. This means that the agents may use any type of argument to defend their interests.

   Since both agents may generate rhetorical arguments, it means that the rhetorical arguments of the proponent aim to defend the proponent's position and the rhetorical arguments of the opponent aim to defend the opponent's position. Therefore, it means that the "opponent's goal" used to construct the arguments depends on who is the opponent. It is clear that that the "opponent" of P is O, but it has to be clear that the "opponent" of O is P. For this reason, we use the following notation to differentiate the rhetorical arguments of each participant: g_i denotes the goal of agent i (for $i \in \{O, P\}$) and go_j denotes the opponent's goal of agent j for ($j \in \{O, P\}$).

Now, let us present the rules that govern the moves of the agents during the dialogue:

- $m_1 = \langle 1, P, 0, \texttt{REQUEST}(\texttt{SECOND}(g_P))\rangle$. It means that the first move is sent by the proponent agent and has to be a request.
- $m_2 = \langle 2, O, 1, \texttt{REJECT}(\texttt{SECOND}(g_P))\rangle$ or $m_2 = \langle 2, O, 1, \texttt{ACCEPT}(\texttt{SECOND}(g_P))\rangle$. It means that the second move is sent by the opponent agent and it is a negative or a positive answer, respectively.
- $\forall k > 1$, it holds that $tg(m_k) = k-1$. It means that the target of a move is always the previous move.
- $\forall k > 1$, it holds that if $ag(k-1) = j$, then $ag(k) = i$ (for $i \neq j$). It means that agents carry out moves alternately.
- If $m_2 = \langle 2, O, 1, \texttt{REJECT}(\texttt{SECOND}(g_P))\rangle$, then
  - $\forall (k > 2 \text{ and } k < n)$, it holds that $m_k = \langle k, i, k-1, RA\rangle$, where $RA \in \{\texttt{THREAT}(T, g_i, go_j), \texttt{REWARD}(T, g_i, go_j), \texttt{APPEAL}(T, g_i, go_j)\}$ (for $i \neq j$). It means that after a negative answer, the agents exchange only rhetorical arguments.

Besides the rules related to legal moves, it is important to define some rules about the end of the dialogue.

- If $m_k = \langle k, j, k-1, \texttt{ACCEPT}(\texttt{SECOND}(g_i))\rangle$, then dialogue D ends in an agreement and $n = k$.
- If $m_k = \langle k, j, k-1, \texttt{WITHDRAW}\rangle$, then dialogue D does not end in an agreement and $n = k$.

6. **Agent's response:** The following function rules the behavior of an agent regarding the proposal he receives from the other participant agent. Let $B \in \texttt{RHETARG}_j$ be the argument sent by agent j in move k and $A \in \texttt{RHETARG}_i$ be the argument that will be sent by agent i according the conservative strategy. The answer of agent i in move $k+1$ regarding the move k of agent j is formulated as follows:

$$\texttt{ANSWER}_i^{m_{k+1}}(m_k) = \begin{cases} \textit{send a rhetorical argument} & \textit{When } sp(m_k) \textit{ denotes an argument } B: \\ & \textit{if } \exists A \in \texttt{RHETARG}_i, \textit{ s.t.} \\ & \quad (1)\ \texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B), \textit{ or} \\ & \quad (2)\ \texttt{ST\_BASIC}(A) = \texttt{ST\_BASIC}(B) \textit{ and} \\ & \qquad A \textit{ is not the strongest argument} \\ \hline\hline \end{cases}$$

$$\texttt{ANSWER}_i^{m_{k+1}}(m_k) = \begin{cases} \textit{accept proposal} & \textit{When } sp(m_k) \textit{ denotes an argument } B: \\ & \textit{if } \forall A \in \texttt{RHETARG}_i,\ s.t. \\ & \texttt{ST\_BASIC}(A) < \texttt{ST\_BASIC}(B) \\ --------- & ---------- \\ \textit{withdraw} & (1)\ \textit{When } sp(m_k) \textit{ denotes an argument } B: \\ & \quad \textit{if } \exists A \in \texttt{RHETARG}_i,\ s.t. \\ & \quad \texttt{ST\_BASIC}(A) = \texttt{ST\_BASIC}(B) \textit{ and} \\ & \quad A \textit{ is the strongest argument} \\ & (2)\ \textit{When } sp(m_k) = \texttt{WITHDRAW}(B): \\ & \quad \textit{if } \nexists A \in \texttt{RHETARG}_i,\ s.t. \\ & \quad \texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B) \end{cases}$$

According to this function, there are three possible answers, which depend on the characteristics of the move an agent receives. Thus, the agent can receive either a move whose speech act (*sp*) is a rhetorical argument or a move whose speech act is a $\texttt{WITHDRAW}(B)$. When an agent receives this kind of speech act, it means that his opponent only has an argument (or arguments) that is as strong as the argument he sent in his previous participation and such argument is the strongest argument of the opponent. Let us exemplify this situation, suppose that in move m_5 agent P sent an argument (say A) whose strength value is 0.75. Agent O has an argument (say B) with the same strength value and such argument is his strongest argument; hence, in move m_6, agent O sends a withdraw indicating the argument that supports that decision. Thus, movement m_7 will depend on the arguments of P. Next, we describe the possible answers of the agents:

- The agent *sends a rhetorical argument* when he has an argument that is stronger than or as strong as the rhetorical argument sent by his opponent in the previous move. When the argument is as strong as the argument sent by his opponent, the agent sends it when it is not the strongest one of his set of arguments.
- The agent *accepts* the proposal when he does not have any rhetorical argument that is equal to or stronger than the rhetorical argument sent by his opponent. In this case, it also means that both agents reach an agreement.
- The agent *withdraws from the negotiation* when his strongest argument(s) is as strong as the rhetorical argument sent by his opponent in the previous move. Another situation occurs when the agent receives a withdraw from his opponent and he has not a stronger argument to be sent. In this case, i.e., both agents withdraw from

the negotiation, it means that the agents do not reach an agreement and the dialogue ends. If the agent has a stronger argument, then the dialogue continues.

7. **Strategy:** Regarding the strategy of the agent about which rhetorical argument he will send in his next move, we consider the following basic strategy:

   - **Conservative strategy:** In this strategy, the agents begin the negotiation with the weakest arguments and as negotiation advances they use stronger arguments. Recall that the use of the strongest threat is considered by some authors as the most persuasive argument; however, it also has disadvantages. For instance, if the proponent backs down from a threat, he may appear to be vacilating and this impacts on his reputation for future negotiations. In the same way, the use of the strongest rewards or appeals may also be disadvantageous for the proponent. In the case of rewards, a strong reward may represent a high investment by the proponent agent because he has to make an opponent's goal to become executive. In the case of the appeals, the fail of an appeal may also impact on the reputation of the agent.

5.3 APPLICATION OF THE NEGOTIATION MODEL

In this section, we present the application of the negotiation model to both scenarios the rescue robots scenario and the software agents scenario. We will use the threats, rewards, and appeals that were logically defined in the previous Chapter along with their respective strength values. Besides, we assume that all the participant agents are considered credible.

SOFTWARE AGENTS SCENARIO

First of all, let us recall the arguments generated by agents `CONSUMER` and `COMPANY` and their respective strength values, which are ordered from lowest to greatest one:

$th_3 = \langle T_3, g, go_3 \rangle$, 0.475 $\quad$ $rw_1 = \langle T_1, g_1, go_4 \rangle$, 0.525
$th_1 = \langle T_1, g, go_1 \rangle$, 0.8 $\quad$ $rw_3 = \langle T_3, g_1, go_6 \rangle$, 0.575
$th_2 = \langle T_2, g, go_2 \rangle$, 0.925 $\quad$ $rw_2 = \langle T_2, g_1, go_5 \rangle$, 0.85

Let us also recall the meaning of the goals:

$g = get(\texttt{COMPANY}, \text{`}refund(money)\text{'})$
$go_1 = gain(costu\_fidel)$
$go_2 = have(good\_rep)$
$go_3 = avoid(legal\_probs)$
$g_1 = get(\texttt{CONSUMER}, \text{`}accept(refund\_20)\text{'})$

$go_4 = gain(miles)$
$go_5 = get\_discount(exec\_ticket)$
$go_6 = free(elderly\_assist)$

Table 15 shows the dialogue between agents CONSUMER and COMPANY. In order to defend their positions and try to convince each other they use threats and rewards. At the end of the dialogue, agent COMPANY accepts refunding the total price of the ticket. Note that agent CONSUMER employs all of his arguments while agent COMPANY only employs two of his rewards. We can distinguish three phases during the dialogue: the starting phase, in which the proponent makes the request and receives a rejection; this leads to the second phase, i.e., the persuasive negotiation. Finally, the third part is related to the end of the dialogue. In this example, the dialogue ends favorable to CONSUMER since COMPANY accepts to refund the total price of the ticket.

	CONSUMER	COMPANY
start	$\langle 1, \texttt{CONSUMER}, 0, \texttt{REQUEST}(refund(money))\rangle$	$\langle 2, \texttt{COMPANY}, 1, \texttt{REJECT}(refund(money))\rangle$
persuasive negotation	$\langle 3, \texttt{CONSUMER}, 2, \texttt{THREAT}(T_3, g, go_3)\rangle$ $\langle 5, \texttt{CONSUMER}, 4, \texttt{THREAT}(T_1, g, go_1)\rangle$ $\langle 7, \texttt{CONSUMER}, 6, \texttt{THREAT}(T_2, g, go_2)\rangle$	$\langle 4, \texttt{COMPANY}, 3, \texttt{REWARD}(T_1, g_1, go_4)\rangle$ $\langle 6, \texttt{COMPANY}, 5, \texttt{REWARD}(T_2, g_1, go_5)\rangle$
end		$\langle 8, \texttt{COMPANY}, 7, \texttt{ACCEPT}(refund(money))\rangle$

Table 15: Dialogue between agents CONSUMER and COMPANY. We can distinguish three phases: the start of the dialogue, the persuasive negotiation part, and the end of the dialogue, which is favorable to CONSUMER.

RESCUE ROBOTS SCENARIO

For this scenario, we have constructed three appeals for agent TOM. In order to be able to generate a dialogue, we will consider that agent BOB also can use three appeals defend his position. Let us also recall that this scenario is defined as fully informed; hence, both agents attribute the same value for the required action. Recall that this action is $(help\_with(debris)$; in this application, we will assume that $\texttt{VALUE}(help\_with(debris)) = 0.55$.

Now, let us recall the arguments generated by agent TOM. We also present the arguments that agent BOB can use during negotiation and their respective strength values, which are ordered from lowest to greatest one:

$ap_3 = \langle T_3, g_3, go_{10} \rangle$, 0.5 $\quad$ $ap_4 = \langle T_4, g_4, go_{11} \rangle$, 0.825

$ap_1 = \langle T_1, g_3, go_8 \rangle$, 0.6 $\quad$ $ap_6 = \langle T_6, g_4, go_{12} \rangle$, 0.75

$ap_2 = \langle T_2, g_3, go_9 \rangle$, 0.825 $\quad$ $ap_5 = \langle T_5, g_4, go_{13} \rangle$, 0.2

Let us also recall the meaning of the goals used by agent TOM and present the meaning of the goals used by agent BOB:

$g = get(\texttt{BOB}, `help\_with(debris)')$

$go_8 = gain(util\_points)$

$go_9 = recharge(battery)$

$go_{10} = fix(sensor)$

$g_4 = get(\texttt{TOM}, `understand(cant\_help)')$

$go_{11} = get(group\_goal)$

$go_{12} = get(improve\_featu)$

$go_{13} = have(break)$

The appeals constructed by BOB are based on three goals of TOM that have not been described previously. Thus, go_{11} means that BOB is focused on a group goal that is convenient for all the participant robots, including TOM; go_{12} means that TOM can get his features improved so he will not need help anymore; and go_{13} means that TOM can take a break.

Table 16 show the dialogue between agents TOM and BOB. Notice that due the value of the required action, agent TOM starts the persuasive negotiation with appeal ap_1 instead of appeal ap_3. Unlike the dialogue in the software agents scenario, in this scenario the agents do not reach an agreement. Recall that this happens when their strongest arguments have the same value, so none of the agents have a persuasive enough argument. Like in previous scenario, we can distinguish three phases during the dialogue; however, the conditions that determine the end of this dialogue are different. In this case, the agents do not reach an agreement because their strongest arguments have the same value.

	TOM	BOB
start	$\langle 1, \texttt{TOM}, 0, \texttt{REQUEST}(help\_with(debris)) \rangle$	$\langle 2, \texttt{BOB}, 1, \texttt{REJECT}(help\_with(debris)) \rangle$
persua. neg.	$\langle 3, \texttt{TOM}, 2, \texttt{APPEAL}(T_1, g_3, go_8) \rangle$	$\langle 4, \texttt{BOB}, 3, \texttt{APPEAL}(T_5, g_4, go_{13}) \rangle$
	$\langle 5, \texttt{TOM}, 4, \texttt{APPEAL}(T_2, g_3, go_9) \rangle$	
end	$\langle 7, \texttt{TOM}, 6, \texttt{WITHDRAW} \rangle$	$\langle 6, \texttt{BOB}, 5, \texttt{WITHDRAW}(T_4, g_4, go_{11}) \rangle$

Table 16: Dialogue between agents TOM and BOB.

5.4 EXPERIMENTS

In this section, we present a set of experiments that aim to evaluate our proposal. For this evaluation, we compare our proposal with its closest alternative approach (i.e., (AMGOUD, 2003)(AMGOUD; PRADE, 2004)), which is based on the importance of the opponent's goal to determine the strength of a rhetorical argument. The environment is an abstract one involving just two agents. Under the premise that the opponent agent rejects the proposal sent by the proponent, the experiments start from the third move of the dialogue. The input for each experiment is a set of rhetorical arguments. The kind of the rhetorical arguments is not relevant for the experiments because these are focused on comparing the strength measurement model. Regarding the technical details, the experiments were implemented in C++ and the values of the importance and the effectiveness were generated randomly in the interval [0,1] and the set $\{0, 0.25, 0.5, 0.75, 1\}$, respectively. Thus, for each individual negotiation encounter, these values were always different. The values of the basic and the combined strength were calculated from these values. The answers given by the agents were ruled by the strategy defined in previous section. Finally, the output of the experiments is mainly the number of negotiation cycles, the number of exchanged arguments, and the number of reached agreements.

In our experiments, a single simulation run involves 1000 separate negotiation encounters between two agents. For all the negotiations, the agents were paired against agents that use the same mechanism of strength calculation. We call "BBGP-based agents" the agents that use the strength evaluation model proposed in Chapter 4 and "IMP-based agents" the agents that use the strength evaluation model based on the importance of the opponent's goal. We performed negotiations where agents generate 10, 25, 50, 100, 250, 500, 750, and 1000 rhetorical arguments. This means that an agent has at most 10, 25, 50, 100, 250, 500, 750, or 1000 arguments to defend his position. We make the experiments with different amounts of arguments in order to analyse the bias of the efficiency of our proposal. For each setting of number of arguments, the simulation was repeated 10 times. This makes a total of 10000 encounters for each setting. Finally, the experimental variables that were measured are: (i) the number of cycles taken to reach agreements, (ii) the number of agreements made, and (iii) the number of arguments (threats, rewards, appeals) used.

Next, we describe each one of the experiments that will be presented in the following subsections:

1. In experiment number one, we assume that both agents are credible in all the negotiation encounters. This means that all the negotiation encounters are carried out. We evaluate the efficiency of our proposal considering the basic equation for the calculation of the strength of a rhetorical argument. Thus, we compare how efficient a negotiation between BBGP-based agents is regarding a negotiation between IMP-based agents.

2. In experiment number two, we assume that the credibility of the agents varies in each negotiation encounter. Thus, in some encounters, both agents are credible, in others only one agent is credible, and in others none of the agents is credible. Like in Experiment 1, we will compare the efficiency of BBGP-based agents with the efficiency of IMP-based agents under these new conditions.

3. In experiment number three, we focus on the performance of BBGP-based agents. We compare the efficiency of BBGP-based agents that negotiate in fully informed scenarios with the efficiency of BBGP-based agents that negotiate in partially informed scenarios.

4. In experiment number four, we again focus on the performance of BBGP-based agents. Thus, we compare the behavior of the two ways of measuring the strength, namely the basic and the combined strength.

Procedure 1 presents the general steps that were considered in our experiments. Notice that the values of importance and effectiveness are different for each negotiation encounter. The value of the reputation of the agents was fixed in 0.8 because when they were generated randomly, the number of negotiation encounters was always very low. We noticed that the number of negotiation encounters was proportional to the reputation value so we decided to establish a reputation value that allowed us to better compare both approaches. The lines of the algorithm that represent the dialogue are run twice in the experiments because two dialogues are carried out. In Experiment 1 and Experiment 2, the dialogues are: (i) between the BBGP-based agents and (ii) between the IMP-based agents (in this dialogue, line 11 is never taken into account). For the Experiment 3, the lines that represent the dialogue are also run twice: (i) in one dialogue, it was considered the value of the required action –which was also generated randomly– so only the arguments whose basic strength was greater than the value of the required action were taken into account for the dialogue, and (ii) in the other dialogue, all the arguments were considered. Finally, for Experiment 4, there are also two dialogues: (i) in one dialogue, the agents use the basic strength for comparing their arguments, and (ii) in the other dialogue, the agents calculate and use the combined strength for comparing their arguments.

Procedure 1

Input: A set of rhetorical arguments

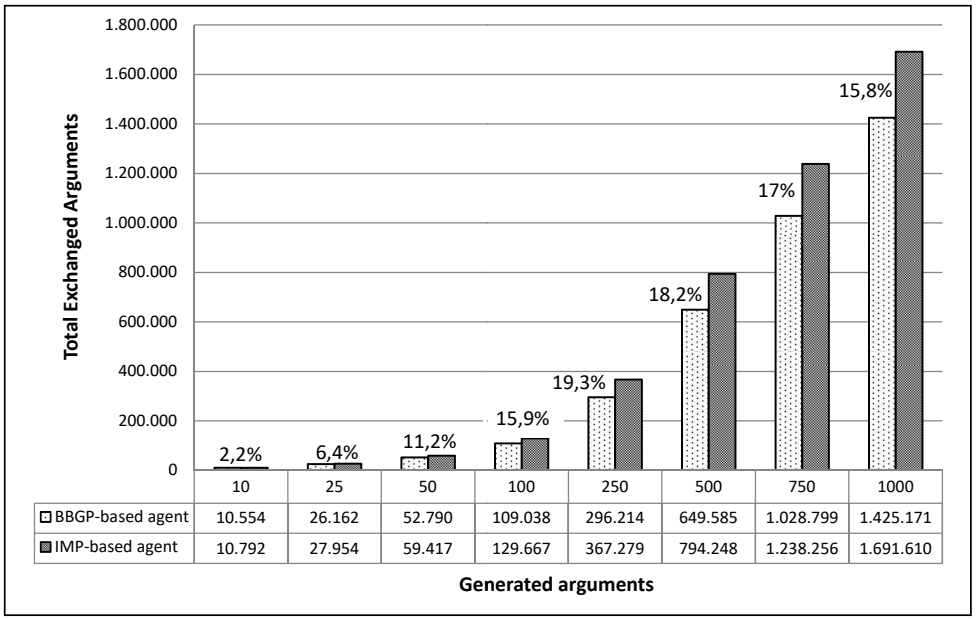
Output: Number of arguments sent by the proponent, number of exchanged arguments, and number of reached agreements

```
1:  for num_encounters = 1 to 1000 do
2:    /*——-Values generated for all the experiments——-*/
3:    Generate randomly the importance and effectiveness values
4:    Calculate the basic strength
5:    Sort the importance values from lowest to highest
6:    Sort the basic strength values from lowest to highest
7:    /*——-Values for the second experiment——-*/
8:    Set proponent/opponent reputation equal to 0.8
9:    Generate randomly the thresholds of proponent/opponent
10:   /*——-Dialogue between IMP-based agents begins——-*/
11:   Proponent sends an argument
12:   while exist available arguments do
13:     if Proponent/Opponent has a stronger argument than the received then
14:       Proponent/Opponent sends the argument
15:     else if Proponent/Opponent has an argument as strong as the received then
16:       if The candidate argument is the strongest argument then
17:         Proponent/Opponent sends WITHDRAW
18:       else
19:         Proponent/Opponent sends the candidate argument
20:       end if
21:     else
22:       Proponent/Opponent sends ACCEPT
23:     end if
24:   end while
25:   /*——-Dialogue between IMP-based agents ends——-*/
26:   if Proponent is credible then
27:     /*——-Dialogue between BBGP-based agents begins——-*/
28:     Proponent sends an argument
29:     while exist available arguments do
30:       if Proponent/Opponent has a stronger argument than the received then
31:         Proponent/Opponent sends the argument
32:       else if Proponent/Opponent has an argument as strong as the received then
33:         if The candidate argument is the strongest argument then
34:           Proponent/Opponent sends WITHDRAW
35:         else
36:           Proponent/Opponent sends the candidate argument
37:         end if
38:       else
39:         Proponent/Opponent sends ACCEPT
40:       end if
41:     end while
42:     /*——-Dialogue between BBGP-based agents ends——-*/
43:   end if
44: end for
```

5.4.1 EXPERIMENT 1

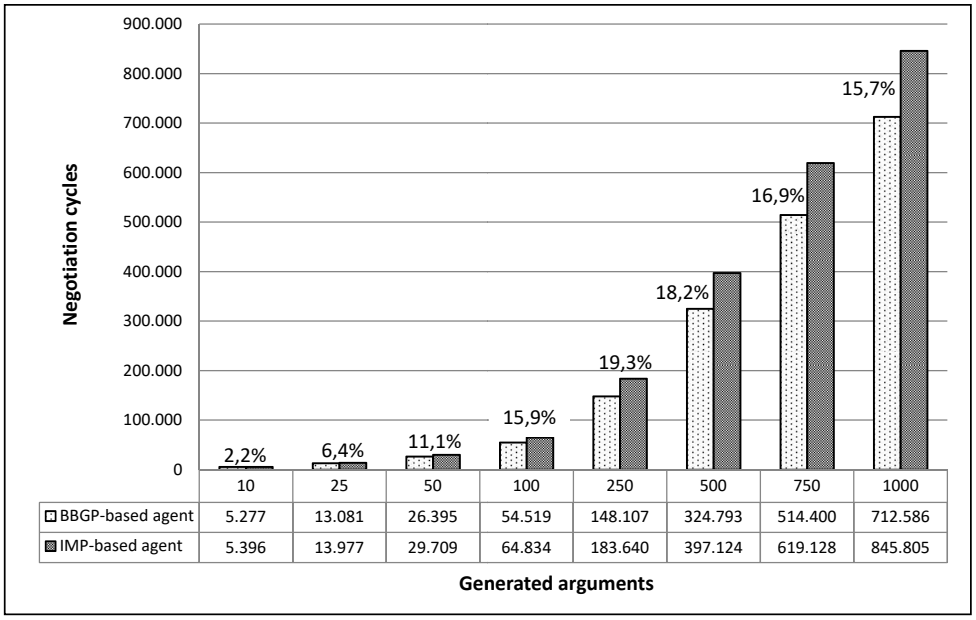
In this experiment, we compare the efficiency of our mechanism for calculating the strength of rhetorical arguments with the mechanism proposed in (AMGOUD, 2003), which is based on the importance value of the opponent's goal.

Figures 12, 13, and 14 show the behavior of the variables for negotiations between two BBGP-based agents and two IMP-based agents, which generate 10, 25, 50, 100, 250, 500, 750, and 1000 arguments. From these graphics, we can derive evidences that that our mechanism fares better than the other mechanism. This can be deduced from the fact that the evaluated variables always have a better behavior in negotiations between the BBGP-based agents than in negotiations between the IMP-based agents. Figure 12 shows the behavior of variable number of arguments exchanged during a negotiation. We can notice that BBGP-based agents always need fewer amounts of exchanged arguments to reach an agreement comparing with IMP-based agents. We can also notice that the greater the number of generated arguments the more significant the difference between the total number of exchanged arguments. The same happens with the variable cycles to reach agreements. In Figure 13, we can observe that BBGP-based agents reach an agreement in fewer cycles than IMP-based agents. This means that BBGP-based agents are more persuasive than IMP-agents since they need fewer cycles and less arguments to convince their opponents. Finally, Figure 14 shows that BBGP-based agent reach more agreements than IMP-based agents.



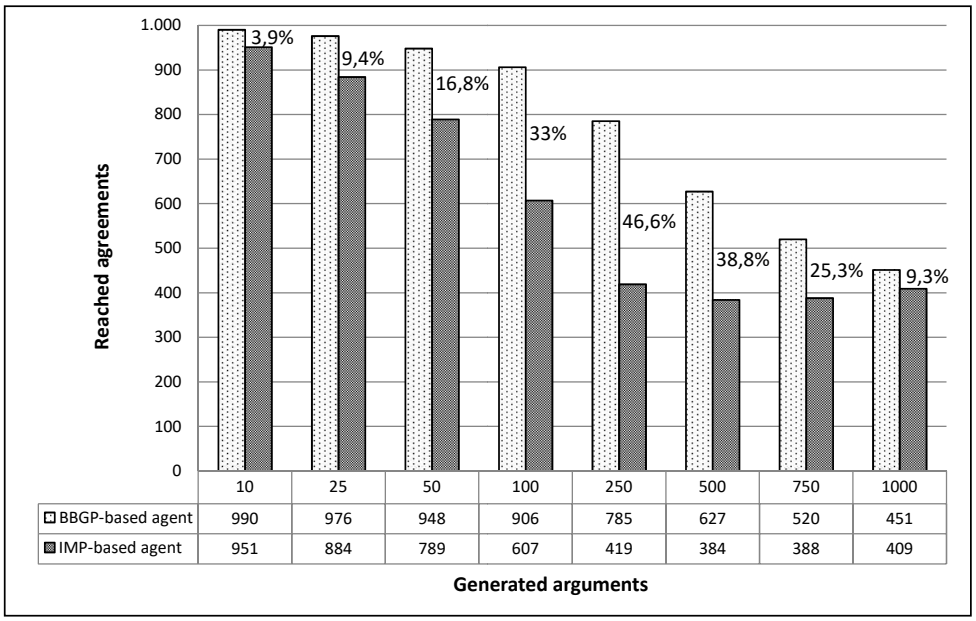
	10	25	50	100	250	500	750	1000
BBGP-based agent	10.554	26.162	52.790	109.038	296.214	649.585	1.028.799	1.425.171
IMP-based agent	10.792	27.954	59.417	129.667	367.279	794.248	1.238.256	1.691.610

Figure 12: Experiment 1: comparison of the variable *number of arguments exchanged*. Hereafter, the labels of each group denote the percentage difference between both values represented by the bars.



	10	25	50	100	250	500	750	1000
BBGP-based agent	5.277	13.081	26.395	54.519	148.107	324.793	514.400	712.586
IMP-based agent	5.396	13.977	29.709	64.834	183.640	397.124	619.128	845.805

Figure 13: Experiment 1: comparison of the variable *number of negotiation cycles*.



	10	25	50	100	250	500	750	1000
BBGP-based agent	990	976	948	906	785	627	520	451
IMP-based agent	951	884	789	607	419	384	388	409

Figure 14: Experiment 1: comparison of the variable *number of reached agreements*.

Regarding the variable reached agreements, we can note that BBGP-based agents achieve more numbers of agreements than IMP-based agents, which is also characteristic of the efficiency of our mechanism. Figure 15 shows percentages of negotiations that end with an agreement in both approaches. In negotiations between BBGP-based agents 78% of the negotiations end with an agreement whereas in negotiations between IMP-based agents only 60% end with an agreement.

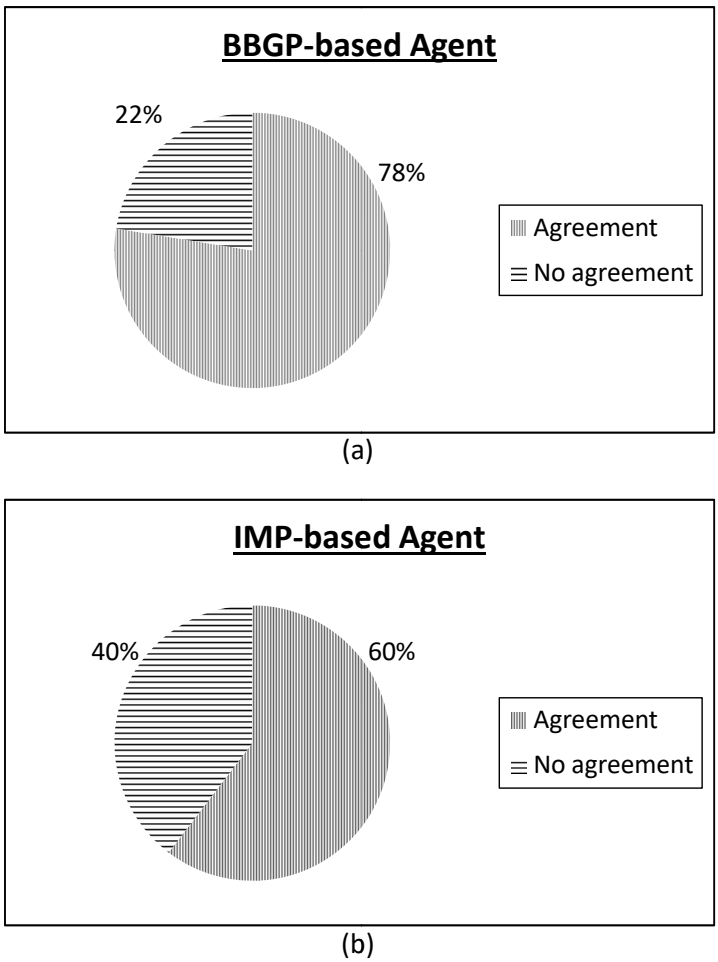


Figure 15: Percentage of negotiations that end in an agreement vs. percentage of negotiations that do not end in an agreement. (a) Comparison of these percentages for negotiations between BBGP-based agents. (b) Comparison of these percentages for negotiations between IMP-based agents.

5.4.2 EXPERIMENT 2

Unlike Experiment 1, in this experiment, we consider that there are some BBGP-based agents that are credible and others that are not. This leads to three possible situations:

1. Both BBGP-based agents are credible. In this case, both engage in a negotiation dialogue.
2. The proponent BBGP-based agent is credible, whereas the opponent BBGP-based agent is not credible. In this case, the proponent wins the negotiation because the opponent does not even calculate the strength of their arguments and has to accept to do the required action.
3. The proponent BBGP-based agent is not credible, whereas the opponent BBGP-based agent is credible. In this case, the the negotiation does not even begin.

Figures 16, 17, and 18 show the behavior of the variables for the negotiation enco-nunters. In the three graphics, we consider that both agents can generate 10, 25, 50, 100, 250,

500, 750, and 1000 arguments. Based on the results, we can state that –under this experiment configuration– our mechanism fares better than the other mechanism. Comparing with the results of Experiment 1, this is even more notorious since the BBGP-based agents do always not engage in a negotiation. Recall that for each experiment, we run 1000 negotiation encounters; however, BBGP-based agents only engage in a negotiation when either both are credible or the proponent is credible. We have run experiments taking into account different reputation values for the agents and we have noticed that the less the reputation value is the less the number of negotiation encounters is. This is quite rational because low reputation values mean that it is more difficult that agents engage in a negotiation. For the results presented in this experiment, we used a reputation value of 0.8 for both agents and the thresholds are generated randomly in the interval [0, 1] before each negotiation encounter.

The fact that BBGP-based agents do not engage in all the negotiations impacts on the experimental variables. Thus, in both Figure 16 and Figure 17, we can notice that the difference of the variables exchanged arguments and number of cycles is greater than in Experiment 1. However, we could believe that it may impact negatively on the variable number of reached agreements. Figure 18 shows the behavior of this variable, this figure also shows the number of negotiations the agents engage in. Thus, IMP-base agents participate in all the possible negotiations, i.e., 1000 negotiation encounters, while BBGP-based agents only participate in some negotiation encounters. We can observe that the behavior of the variable reached agreements changes as the number of generated arguments grows. Thus, when the agents generate 10 and 25 arguments, the IMP-based agents reach more agreements than BBGP-based agents; however, from the 50 generated arguments this behavior changes. From this point, BBGP-based agents reach more agreements than IMP-based agents.

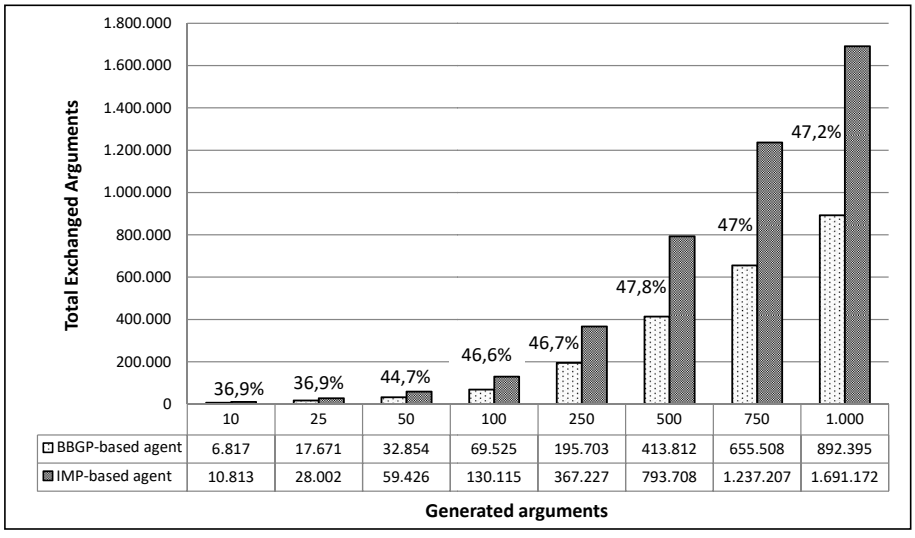


Figure 16: Experiment 2: comparison of the variable *number of arguments exchanged*.

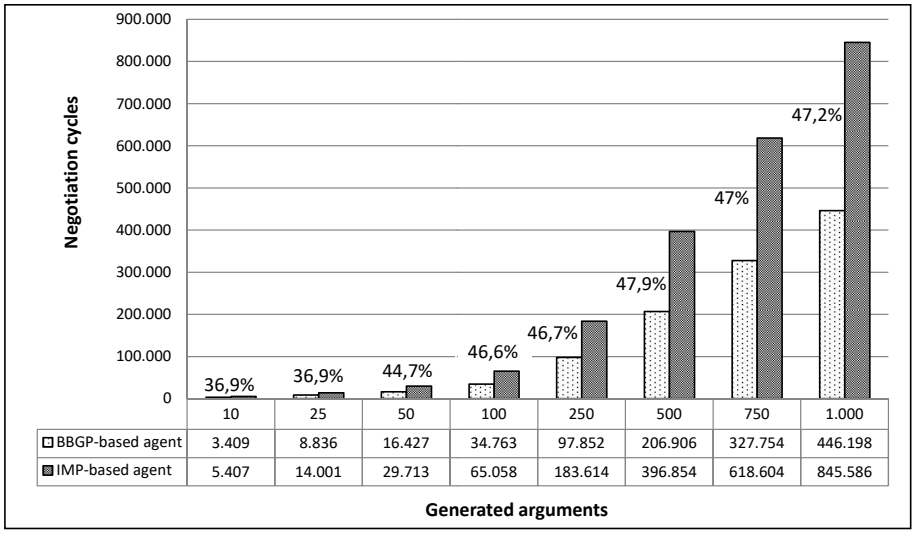


Figure 17: Experiment 2: comparison of the variable *number of negotiation cycles*.

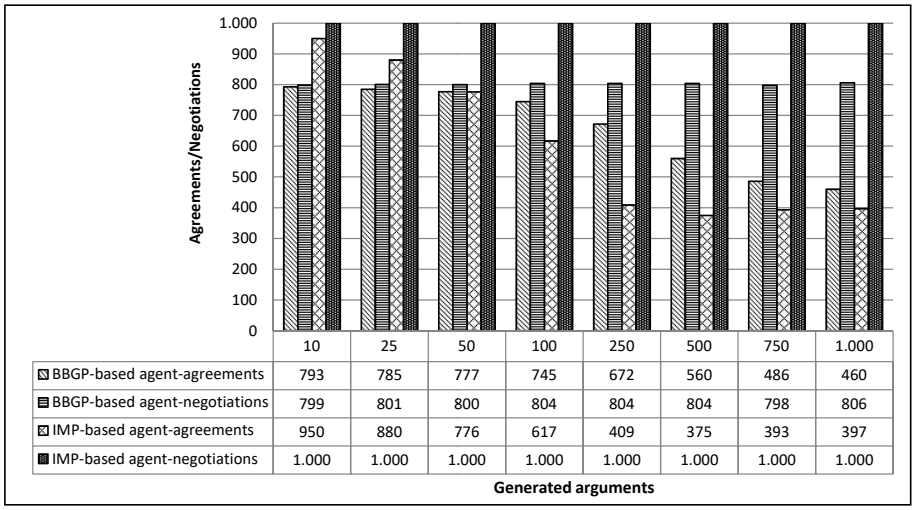


Figure 18: Experiment 2: comparison of the variable *number of reached agreements*.

In order to better observe the variable reached agreements, Figure 19 shows the percentages of reached agreements, non-reached agreements, and the percentage of the negotiations the BBGP-agents do not engage in. These values reflect the average behavior of the agents. Thus, for IMP-based agents the percentages of reached agreements and non-reached agreements are the same as in Experiment 1 whereas for BBGP-based agents these percentages change. Thus, the percentage of reached agreements is 66%, this means that this percentage drops 12 points down comparing with Experiment 1 and the percentage of non-reached agreements is 14%, this percentage also drops 8 points down comparing with Experiment 1. Figure 19(a) also shows the percentage of negotiation the BBGP-based agent do not engage in, which is 20%. Notice that, although the percentage of reached agreement dropped down, the BBGP-based agents reach more agreements than IMP-based agents.

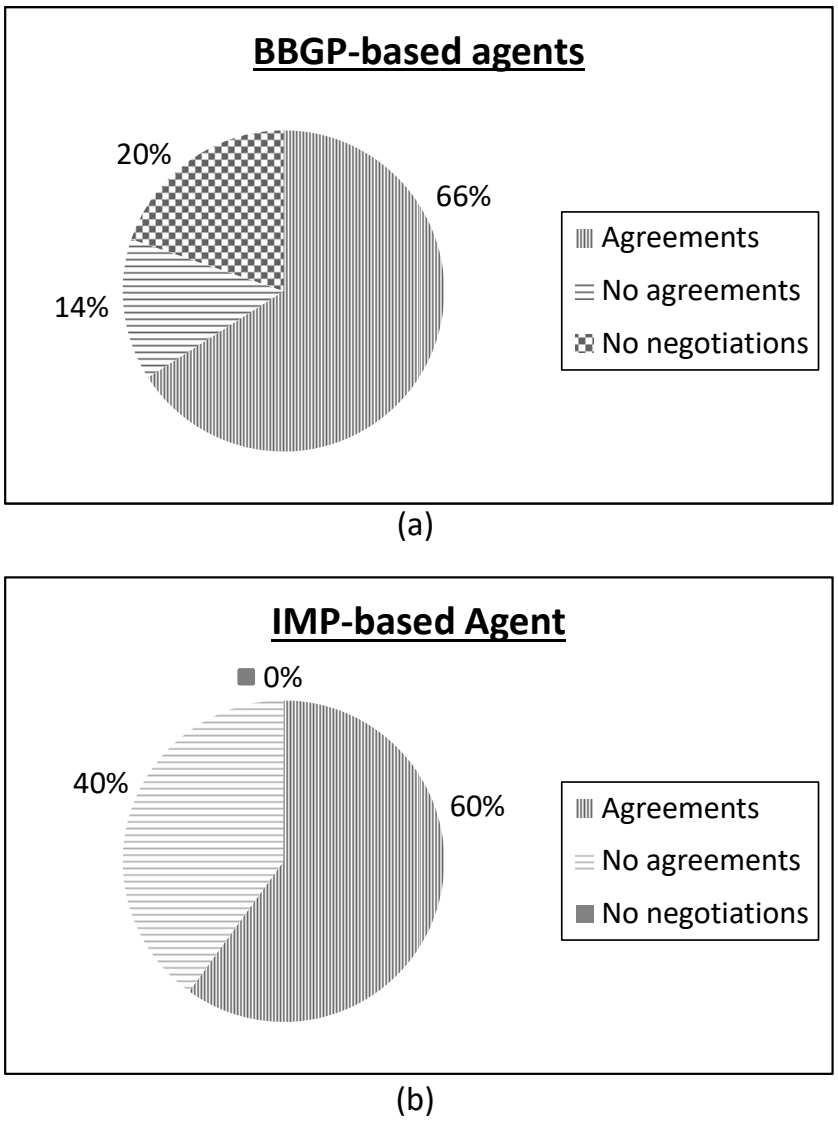


Figure 19: Percentage of negotiations that end in an agreement vs. percentage of negotiations that do not end in an agreement. (a) For BBGP-based agents. (b) For IMP-based agents.

5.4.3 EXPERIMENT 3

In this experiment, we focus on comparing the performance of the BBGP-based agents considering that they negotiate either in fully informed scenarios or in partially informed scenarios. Let us recall that in fully informed scenarios the proponent agent employs convincing arguments to try to convince his opponent whereas in partially informed scenarios the agent does not know which arguments are convincing and which are not.

The experimental variables that are taken into account in this experiment are: (i) number of arguments used by the proponent, (ii) total number of arguments exchanged during the negotiation, and (iii) number of negotiation cycles. In this experiment, the first variable is es-

pecially important given that in fully informed scenarios the proponent knows the value of the required action while in partially informed scenarios the proponent does not know this information. This fact impacts directly on the number of arguments used by the proponent because when he knows this value his persuasive strategy only includes those rhetorical arguments that fulfil the preferability condition, i.e., the value of these rhetorical arguments is greater than the value of the action. Thus, agents in fully informed scenarios employ a modified conservative strategy, in which their first rhetorical argument to be sent is the least valued preferable argument.

Figures 20, 21, and 22 shows the results of this experiment. On Figure 20, we can observe the behavior of the variable number of proponent arguments. The difference of amount of arguments used by the proponent is very notorious. In average, BBGP-agents in fully informed scenarios use 70,282 arguments whereas in partially informed scenarios, they use 205,179 arguments during the negotiation encounters. This means that in partially informed scenarios the amount of used arguments is almost three more times than the amount of arguments used in fully informed scenarios. The number of arguments used by the proponent has also an impact on the total number of exchanged arguments (see Figure 21). For this variable, on average, we have that the number of exchanged arguments in fully informed scenarios is 140,030 whereas in partially informed scenarios it is 274,926, which is almost double of arguments. Since the variable negotiation cycles depends on the total number of exchanged arguments, this variable has an equal behavior to the observed in that variable. Figure 22 shows the behavior of this variable.

We have defined fully informed scenarios under the premise that the value of the actions is the same for all the participant agents. We are conscientious that in a scenario of robot agents, this premise may be more easily satisfied than in scenarios involving humans. However, with this experiment we wanted to show that the preferability condition has a big impact on the number of arguments used during the negotiation encounters. We have considered that in partially informed scenarios, the proponent agents do not know the value of the actions for their opponents; nevertheless, we could consider that a proponent agent may employ the value he gives to his actions as a reference point to select their arguments. All this aiming at distinguishing convincing arguments and non-convincing arguments.

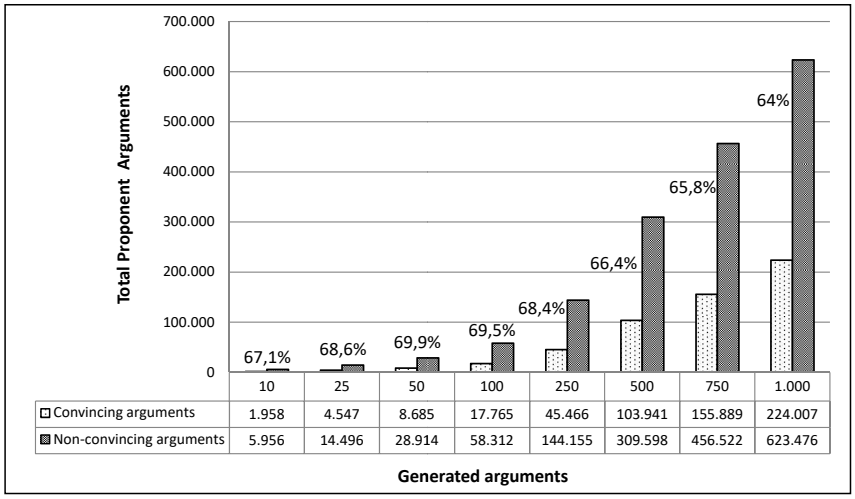


Figure 20: Experiment 3: comparison of the variable *number of arguments sent by the proponent*.

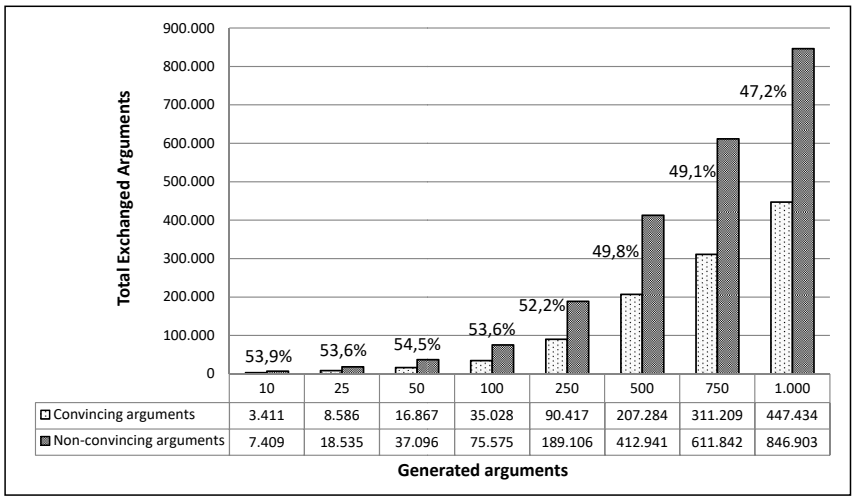


Figure 21: Experiment 3: comparison of the variable *number of exchanged arguments*.

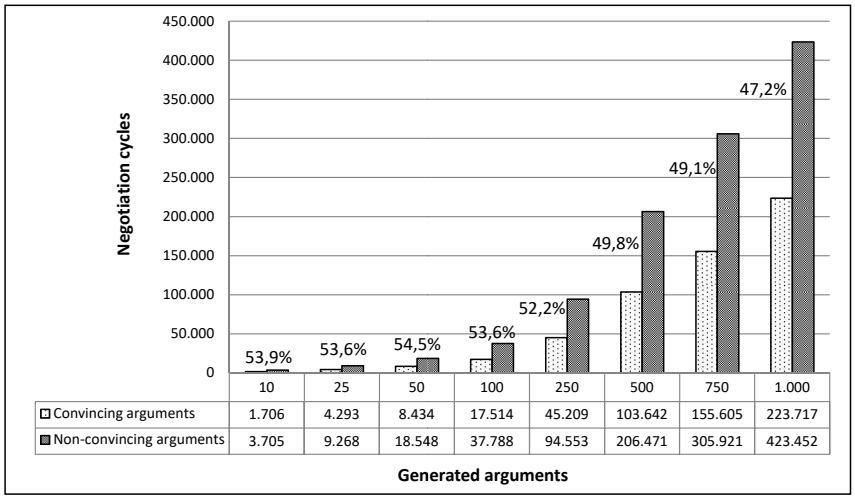


Figure 22: Experiment 3: comparison of the variable *number of negotiation cycles*.

5.4.4 EXPERIMENT 4

In this experiment, we again evaluate the performance of BBGP-based agents. However, in this experiment, we compare the performance of the basic strength to the performance of the combined strength. Let us recall that for calculating the basic strength, we only take into account the opponent's goal, whereas for calculating the combined strength, besides the opponents goal, we take into account the accurate credibility of the agent[1]. Thus, we want to know how the value of the accurate credibility impacts on the studied variables, i.e., the number of arguments sent by the proponent and the number of exchanged arguments during the dialogue.

The settings in this experiment are a little different from the settings in previous experiments. This is because we want to focus on the impact of the accurate credibility. Thus, the settings of this experiment are: each agent generates 10 arguments, the reputation of each agents is 1, the threshold of the proponent agent is always 0.8, the threshold of the opponent goes down from 0.8 to 0.55. Decreasing the oponent's threshold makes him more credulous and persuadable. We have run 6 scenarios:

- In the first scenario, both agents have the same reputation and threshold values, therefore, the value of the accurate credibility is the same for both agents (i.e., 1-0.8=0.2).
- In the second scenario, the threshold of the proponent agent is 0.8 and the threshold of the opponent is 0.75; hence, the value of the accurate credibility is different. Thus, the value of the accurate credibility that the proponent uses for calculating the combined strength of the arguments is 0.25, whereas the value of the accurate credibility that the opponent uses for calculating the combined strength of the arguments is 0.2. This means that there is a difference of 0.05 between both values of the accurate credibility and this also means that in the second scenario the opponent is more persuadable than in the first scenario.
- In the remaining scenarios, the threshold of the proponent is 0.8 and the threshold of the proponent goes down in 0.05 in each scenario. Thus, in the third scenario, the threshold of the opponent is 0.7, in the fourth scenario it is 0.65, in the fifth scenario it it 0.6, and in the sixth scenario it is 0.55. This means that the difference between the value of the accurate credibility increases. Thus, in the third scenario, the difference is 0.1, in the fourth scenario it is 0.15, in the fifth scenario it is 0.2, and in the last scenario it is 0.25.

[1]Recall that the reputation is an evidence of the proponent's past behavior of an agent with respect to his opponents. We assume that this value is already estimated and it is not private information; thus, the reputation value of an agent is visible for any other agent. On the other hand, the "accurate" value of the credibility of an agent P with respect to an opponent O –whose threshold is $\texttt{THRES}(O)$– is given by $\texttt{ACCUR\_CRED}(P,O) = \texttt{REP}(P) - \texttt{THRES}(O)$.

Table 17 shows a resume of these six scenarios, which includes the values of the reputation, threshold, and accurate credibility of agents proponent and opponent.

	PROPONENT			OPPONENT		
SCENARIO	REP	THRES	ACCUR_CRED	REP	THRES	ACCUR_CRED
#1	1	0.8	0.2	1	0.8	0.2
#2	1	0.8	0.2	1	0.75	0.25
#3	1	0.8	0.2	1	0.7	0.3
#4	1	0.8	0.2	1	0.65	0.35
#5	1	0.8	0.2	1	0.6	0.4
#6	1	0.8	0.2	1	0.55	0.45

Table 17: Experiment 4: Values of the reputation, threshold, and accurate credibility of agents proponent and opponent.

We have run 1000 negotiation encounters for each scenario and besides evaluating the variables of efficiency; we also compare the number of times that the proponent succeeds in persuading the opponent.

Figures 23 and 24 illustrate the results of this experiment. Figure 23 shows the behavior of the variable number of arguments sent by the proponent. When the calculation is done by applying the combined strength equation, we can notice that the greater the difference between the values of the accurate credibility, the fewer the number of arguments the proponent sends. On the other hand, when the calculation is done by applying the basic strength equation, the number of arguments is very similar. Figure 24 shows the behavior of the variable number of exchanged arguments. This result reaffirms that the performance of the combined strength calculation improves as the difference of the values of the accurate credibility increases.

In Figure 25, we compare the number of times that the proponent succeeds in persuading the opponent. Figure 25(a) shows the results for the first scenario, we can notice that the number of times the proponent succeeds is the same in both ways of calculating the strength. Figure 25(b) shows the results for the second scenario, where the difference between the values of the accurate credibility is 0.05. We can notice that when the basic strength calculation is applied, the percentages of succeed of proponent and opponent are balanced, whereas when the combined strength calculation is applied, the percentage of succeed of the proponent is high, we can even notice that it increased 40% with respect to the percentage of the first scenario.

In Figures 25(c) and 25(d), the percentage of succeed of the proponent is even more notorious. These figures correspond to the fourth and sixth scenarios, that is, when the difference between the values of the accurate credibility is 0.15 and 0.25, respectively. Indeed, in the fourth scenario the percentage of succeed of the proponent is almost 100% and in the last

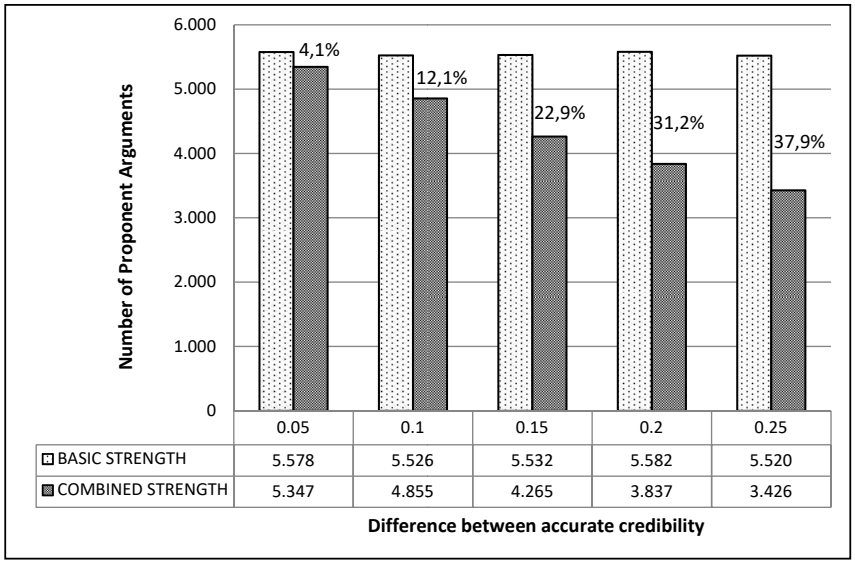


Figure 23: Experiment 4: comparison of the variable *number of arguments sent by the proponent*.

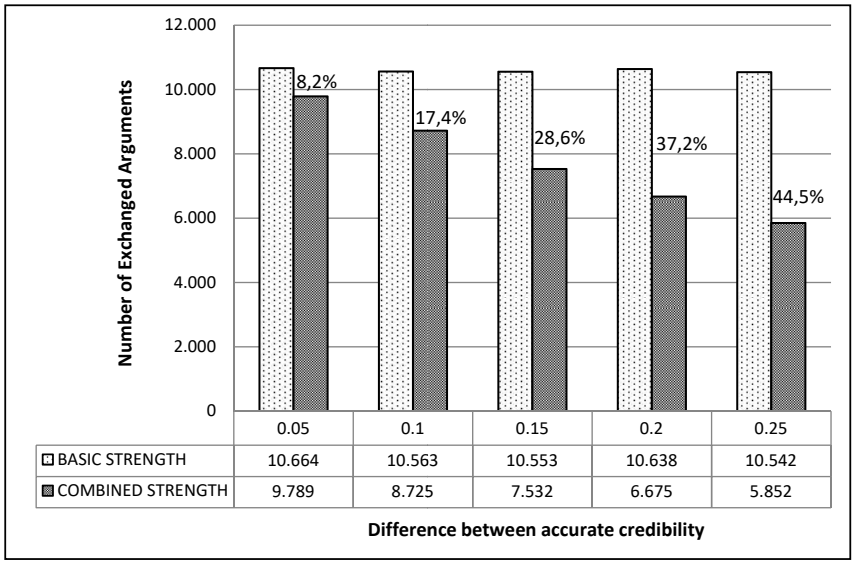
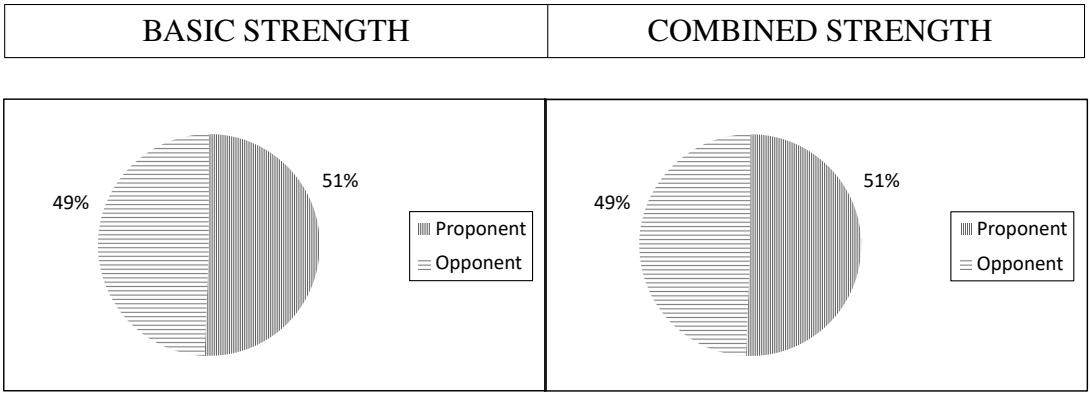
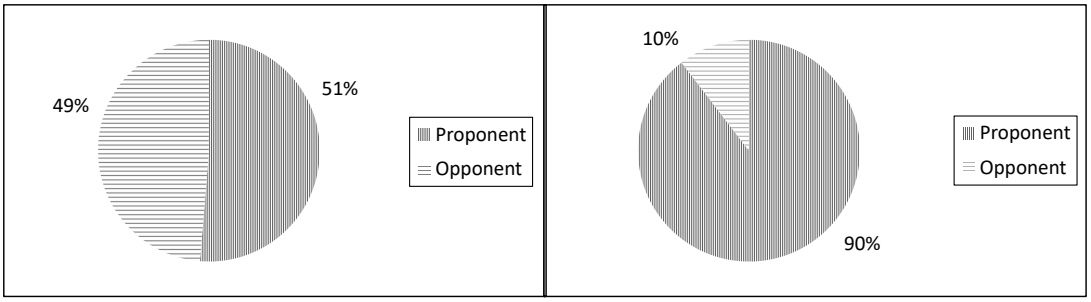


Figure 24: Experiment 4: comparison of the variable *number of exchanged arguments*.

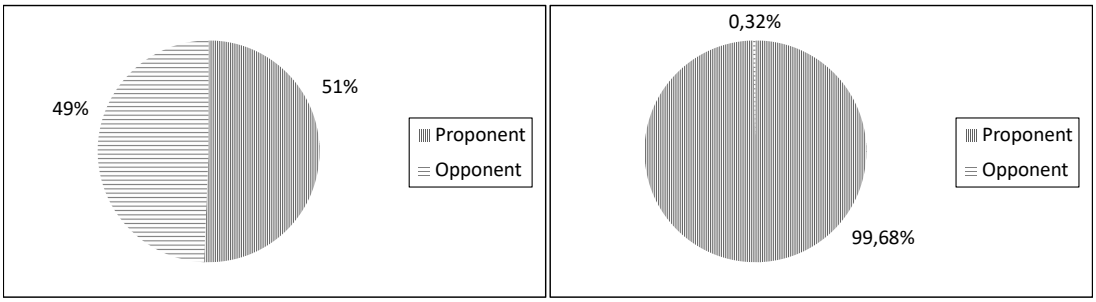
scenario it is 100%. These results show that if the difference between the values of the accurate credibility is equal to or greater than 0.15, the proponent always succeeds.



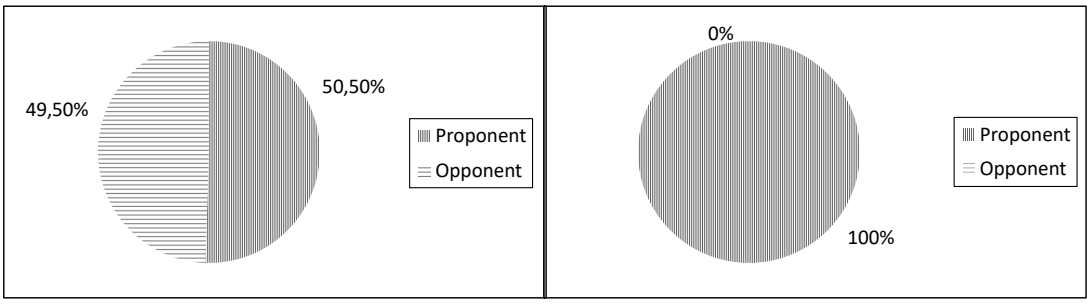
(a) The difference between the values of the accurate credibility is equal to 0.



(b) The difference between the values of the accurate credibility is equal to 0.05.



(c) The difference between the values of the accurate credibility is equal to 0.15.



(d) The difference between the values of the accurate credibility is equal to 0.25.

Figure 25: Experiment 4: Percentage of negotiations that end favourably for the proponent vs. percentage of negotiations that end favourably for the opponent.

5.5 DISCUSSION

In the previous section, the results of the experiments have shown that the model proposed in this thesis has a better performance compared to the approach that bases the calculation of the rhetorical argument on the importance of the opponent's goal. However, it is important to discuss some aspects of our approach.

While it is true that our approach has a better performance, it is also true that it is necessary to model further knowledge about the opponent. This need of further modelling may be seen as a weakness of the model; however, we can notice that in all the evaluated variables, the model is always more efficient and effective than the other approach. Indeed, when more criteria are employed both the efficiency and the effectiveness increase. Let us recall Experiment 2, IMP-based agents engage in all negotiation encounters (i.e., 1000 encounters) whereas BBGP-based agents only engage in a negotiation encounter when the proponent agent is credible enough (i.e., around 800 encounters). This would mean that IMP-based agents may always achieve more agreements than BBGP-based agents; nevertheless, the results show that the higher the number of generated arguments is, the more agreements the BBGP-agents achieve. The efficiency of the model is even more notorious when the scenario is a fully informed scenario. In this type of scenario, the number of negotiation cycles and the number of exchanged arguments is less than in partial informed scenarios. Thus, when we take into account the criterion credibility and the type of scenario, the efficiency of the model increases even more. Let us consider that the number of generated arguments is equal to 1000, the behavior of the variable *number of negotiation cycles* is the following:

- Experiment 1 (neither the credibility nor the type of scenario is taken into account): the number of negotiation cycles between BBGP-based agents is 712,586 whereas the number of negotiation cycles between IMP-based agents is 845,805.
- Experiment 2 (only the credibility is taken into account): the number of negotiation cycles between BBGP-based agents is 446,158 whereas the number of negotiation cycles between IMP-based agents is 845,586.
- Experiment 3 (both the credibility and the type of scenario are taken into account): the number of negotiation cycles between BBGP-based agents in a fully informed scenario is 223,717 whereas the number of negotiation cycles between BBGP-based agents in a partially informed scenario is 423,452.

Therefore, the proposed model needs further knowledge for the measurement of the strength; however, the results show that when the agents model this additional knowledge, it impacts positively on the evaluated variables.

Note that when the credibility is taken into account the number of negotiation cycles is reduced to a third part. This happens because only some negotiations are carried out, which impacts on the number of exchanged arguments and on the number of negotiation cycles. This situation also happens when the scenario is a fully informed scenario. In this case, it occurs because some arguments are not taken into account, such arguments are the non-convincing ones. However, when neither the credibility nor the convincing arguments is considered (Experiment 1), the proposed model is more efficient and effective anyway.

In the Experiment 1, we have observed that the number of negotiation cycles in the dialogues between BBGP-based agents is not always less than the number of negotiation cycles in the dialogues between IMP-based agents. Thus, we have observed three possible situations: (i) there were some dialogues between IMP-based agents that were more efficient than the dialogues between BBGP-based agents, (ii) some other dialogues –between IMP-based agents and between BBGP-based agents– had the same number of negotiation cycles, and (iii) in other dialogues, the proposed model was more efficient. Nevertheless, after running the 1000 negotiation encounters, the proposed model was more efficient in most cases. Since this behavior is not due to the use of credibility or/and the use of convincing arguments, we next present an example of a dialogue where the proposed model is more efficient. From this example, we will try to analyze and explain the efficient behavior of the basic strength calculation.

Table 18 presents the values of the importance, effectiveness, and basic strength that were modeled for agents proponent and opponent. Table 19, shows the values of the importance and the basic strength ordered from lowest to highest. Notice that these values are independent of each other, this means that the basic strength value of a given line l is not always obtained based on the importance value of the same line, i.e., l. These values are ordered from lowest to highest due to the strategy described in the negotiation model. According to that strategy, an agent starts the dialogue with his less strong rhetorical argument. We can notice that both IMP-based agents have a rhetorical argument whose basic strength is zero; in this case, the IMP-based agents start the dialogue with the next rhetorical argument whose strength value is different from zero. Finally, Table 20 shows both dialogues, the dialogue between IMP-based agents and the dialogue between BBGP-based agents. Notice that in both cases the agents reach an agreement; however, in the dialogue between IMP-based agents it is favorable for the opponent whereas in the dialogue between BBGP-based agents it is favorable for the proponent

agent. Also notice that in the dialogue between IMP-based agents the number of negotiation cycles is 5 whereas in the dialogue between BBGP-based agents it is 4. We have noticed that the variety of the effectiveness values has impact on it. Thus, if all the effectiveness values where the same, the number of negotiation cycles would be the same in both dialogues, while with different effectiveness values the number of negotiation cycles is also different.

PROPONENT'S MIND				OPPONENT'S MIND		
IMP	EFF	BASIC_STRENGTH		IMP	EFF	BASIC_STRENGTH
0.5842	1.00	0.7921		0.1485	0.00	0.0743
0.3366	1.00	0.6683		0.0792	1.00	0.5396
0.0792	0.50	0.2896		0.0000	1.00	0.5000
0.2970	0.25	0.2735		0.6733	0.75	0.7116
0.0000	0.50	0.2500		0.6832	0.00	0.3416
0.5644	0.25	0.4072		0.3960	0.25	0.3230
0.0693	0.75	0.4097		0.1485	1.00	0.5743
0.6535	1.00	0.8267		0.0891	0.00	0.0446
0.7525	0.75	0.7512		0.9703	0.50	0.7351
0.5149	0.00	0.2574		0.6733	0.00	0.3366

Table 18: Values of the importance, effectiveness, and basic strength that were modelled by agents proponent and opponent.

PROPONENT'S MIND			OPPONENT'S MIND	
IMP	BASIC_STRENGTH		IMP	BASIC_STRENGTH
0.0000	0.2500		0.0000	0.0446
0.0693	0.2574		0.0792	0.0743
0.0792	0.2735		0.0891	0.3230
0.2970	0.2896		0.1485	0.3366
0.3366	0.4072		0.1485	0.3416
0.5149	0.4097		0.3960	0.5000
0.5644	0.6683		0.6733	0.5396
0.5842	0.7512		0.6733	0.5743
0.6535	0.7921		0.6832	0.7116
0.7525	0.8267		0.9703	0.7351

Table 19: Values of the importance and the basic strength ordered from lowest to highest.

	IMP-BASED AGENTS			BBGP-BASED AGENTS	
	PROPONENT	OPPONENT		PROPONENT	OPPONENT
1	0.0693 –»			0.2500 –»	
2		«– 0.0792			«– 0.3230
3	0.0792 –»			0.4072 –»	
4		«– 0.0891			«– 0.500
5	0.2970 –»			0.6683 –»	
6		«– 0.3960			«– 0.7116
7	0.5149 –»			0.7512 –»	
8		«– 0.6733			«–ACCEPT
9	0.7525 –»				
10		«– 0.9703			
11	ACCEPT–»				

Table 20: Dialogue between IMP-based agents and dialogue between BBGP-based agent. The numbers represent the basic strength value of the exchanged rhetorical arguments.

5.6 SUMMARY

In this chapter, we have focused on the persuasive negotiation dialogue that is carried out between two negotiating agents. We have presented a negotiation model that includes a communication language, a negotiation protocol, a function that rules the agents response, and a strategy the agents follow during the dialogue to choose the next argument to be sent. In order to show the practical application of the negotiation model and to know what happens with the agents that are part of the scenarios that have been employed throughout this work, we have used this negotiation model to construct the dialogue between agents `CONSUMER` and `COMPANY` and the dialogue between agents `TOM` and `BOB`.

The protocol, the response function, and the strategy were specially useful for the empirical evaluation of the strength calculation model proposed in the previous chapter. These elements were necessary and complemented that proposal. Recall that the calculation model returns a set of arguments along with their respective strength values. It was necessary to define the negotiation model so the agents can engage in a negotiation and in this way, the strength values can be used to define the end of the negotiations.

Thus, the second part of this chapter was devoted to three experiments, which aim to evaluate four variables: (i) the number of exchanged arguments, (ii) the number of negotiation cycles, (iii) the number of reached agreements, and (iv) the number of arguments sent by the proponent. In all cases, we demonstrated that our proposed model fares batter than the calculation model that only takes into account the importance of the opponent's goal. Let us recall

that in our proposed model we consider the importance of the opponent's goal and also the effectiveness of it. Furthermore, we evaluate the credibility of the agents before the calculation task begins. Thus, we have noticed that the new criteria included in the calculation model have made our proposal more efficient than the model based only on one criteria.

6 FINAL REMARKS

This chapter concludes the thesis and presents several possible directions for future work.

6.1 CONCLUSION

We would like to begin by recalling the research question of this thesis, which will be answered in the following paragraphs.

What criteria should an intelligent agent take into account in order to measure the strength of a rhetorical argument and how should this measurement be done?

In this thesis, we have mainly studied how to measure the strength of the rhetorical arguments, i.e., threats, rewards, and appeals. With this aim, in Chapter 3, we have formalized an intelligent agent that is based on the Belief-based Goal Processing model proposed by Castelfranchi e Paglieri (2007). This model and its formalization are important because they allowed us to analyze the agent's goals from another perspective (the status of a goal), which was taken into account for the calculation of the strength of the rhetorical arguments.

In Chapter 4, we have extended the BBGP-based agent by endowing him with negotiating characteristics, more specifically, a BBGP-based negotiating agent is able to generate rhetorical arguments and calculate the strength values of that rhetorical arguments. Thus, we defined two ways for calculating the strength value: (i) the basic strength, which is calculated based on the importance and the effectiveness of the opponent's goal and (ii) the combined strength, which is calculated based on the basic strength and on the accurate credibility value of the proponent agent. Therefore, the criteria that are taken into account in order to measure the strength of rhetorical arguments are: (i) the importance of the opponent's goals, (ii) the effectiveness of the opponent's goal, and (iii) the accurate credibility of the proponent agent.

These two ways of measuring the strength of a rhetorical argument are part of a strength calculation model, which guides the measurement process based on the values of the criteria

and on the preferences of the agent. Thus, the process for strength calculation only starts if the proponent believes that he is credible enough (from the point of view of his opponent). The model also takes into account the type of scenario (i.e., fully informed scenario or partially informed scenario) and the preference of the agent about the accuracy of the calculation (i.e., the agent uses or not the accurate credibility value). We can say that this model answers the second part of the research question since it determines how the strength measurement should be done. Thus, the proposal of this thesis goes beyond of the two equations for calculating the basic and the combined strength values.

Regarding the stated aim of this research "*to propose a model for the calculation of the strength of rhetorical arguments, which, besides contemplating the importance of the opponent's goal, takes into account the status of this goal in the opponent's mind and the credibility of the proponent. The use of these criteria impacts on the results returned by the model. Thus, the results can be more efficient and effective than the results of the approach that is based only on the importance of the opponent's goal. The efficiency is measured in terms of two variables: (i) number of exchanged arguments and (ii) number of negotiation cycles, and the efficacy is measured in terms of the variable number of reached agreements*", we can say that this objective was attained. The model including new criteria has been proposed and its efficiency and efficacy have been measured and compared. The empirical results have shown that the proposed model fares better than the approach based only on the importance of the opponent's goal. Specifically, BBGP-based agents achieved more agreements in fewer numbers of cycles and with fewer exchanged arguments than the IMP-based agents.

6.2 FUTURE WORK

In Chapter 3, we have formalized an agent architecture based on the BBGP model. We have worked with certain beliefs and rules. One future work will consider that the knowledge base of the BBGP-based agents is pervaded with uncertainty. This in turn can be used to measure the strength of the arguments, which may lead to different results in the sets of goals of the agent. Another extension of the work is related to the kinds of incompatibilities that may exist between goals because they are determinant on the goal selection. We have included the resource incompatibility; however, the terminal incompatibility and the superfluity were not considered. Indeed, we are currently working on formalizing these kinds of incompatibilities and on studying how to integrate them in a unique framework. It is also interesting to further study the resource incompatibility due to the different types of resources that may be necessary.

Chapter 4 is devoted to the study of the strength calculation model. We have worked under the premise that the proponent agent knows in advance the information about his opponent. An interesting future work is to complement this model with the study of an adequate opponent modelling approach. We can also consider that the information of the opponent is uncertain, which may impact on directly the strength calculation. Besides, we have supposed that the agent has in advance the rules that allow him to construct the rhetorical arguments. If the agent is able to model the opponent information during the negotiation, then he should be able to generate new rules that allow him to generate new rhetorical arguments. In the proposed approach, there is no model of the environment or the context where the negotiation occurs, especially in terms of organizational structure. We believe that this information can influence the strength of the arguments and therefore on the persuasion power of the agents.

In Chapter 5, we have proposed a negotiation model and performed a set of simulations that aimed to test the performance of our approach. We have considered that both agents only exchange rhetorical arguments. Future work will endow the agents with the ability of generating attacks, such that, during the persuasive negotiation the agents may exchange both arguments and counter-arguments. It would also be interesting that the agents can ask for explanations and use explanatory arguments during the dialogue. Another work would be to study and propose a strategy that allows the agents to choose the “best” arguments to be sent.

BIBLIOGRAPHY

AMGOUD, L. A formal framework for handling conflicting desires. In: **Symbolic and Quantitative Approaches to Reasoning with Uncertainty**. [S.l.]: Springer, 2003. p. 552–563.

AMGOUD, L.; BESNARD, P. Bridging the gap between abstract argumentation systems and logic. In: SPRINGER. **International Conference on Scalable Uncertainty Management**. [S.l.], 2009. p. 12–27.

AMGOUD, L.; BESNARD, P. A formal analysis of logic-based argumentation systems. In: SPRINGER. **International Conference on Scalable Uncertainty Management**. [S.l.], 2010. p. 42–55.

AMGOUD, L.; BESNARD, P. A formal characterization of the outcomes of rule-based argumentation systems. In: SPRINGER. **International Conference on Scalable Uncertainty Management**. [S.l.], 2013. p. 78–91.

AMGOUD, L. et al. Towards a consensual formal model: inference part. In: **Technical report, ASPIC project, Deliverable D2.2: Draft formal semantics for inference and decision-making**. [S.l.: s.n.], 2004.

AMGOUD, L.; DEVRED, C.; LAGASQUIE-SCHIEX, M.-C. Generating possible intentions with constrained argumentation systems. **International Journal of Approximate Reasoning**, Elsevier, v. 52, n. 9, p. 1363–1391, 2011.

AMGOUD, L.; HAMEURLAIN, N. An argumentation-based framework for designing dialogue strategies. **Frontiers in Artificial Intelligence and Applications**, IOS Press, v. 141, p. 713–714, 2006.

AMGOUD, L.; KACI, S. On the generation of bipolar goals in argumentation-based negotiation. In: **Argumentation in Multi-Agent Systems**. [S.l.]: Springer, 2005. p. 192–207.

AMGOUD, L.; PARSONS, S.; MAUDET, N. Arguments, dialogue, and negotiation. In: **Proceedings of the 14th European Conference on Artificial Intelligence**. [S.l.]: IOS Press, 2000. p. 338–342.

AMGOUD, L.; PRADE, H. Threat, reward and explanatory arguments: generation and evaluation. In: **Proceedings of the 4th Workshop on Computational Models of Natural Argument**. [S.l.: s.n.], 2004. p. 73–76.

AMGOUD, L.; PRADE, H. Handling threats, rewards, and explanatory arguments in a unified setting. **International Journal of Intelligent Systems**, Wiley Online Library, v. 20, n. 12, p. 1195–1218, 2005.

AMGOUD, L.; PRADE, H. Formal handling of threats and rewards in a negotiation dialogue. In: **Argumentation in Multi-Agent Systems**. [S.l.]: Springer, 2006. p. 88–103.

AMGOUD, L.; PRADE, H. Using arguments for making and explaining decisions. **Artificial Intelligence**, Elsevier, v. 173, n. 3, p. 413–436, 2009.

AUSTIN, J. L.; URMSON, J. **How to Do Things with Words. The William James Lectures Delivered at Harvard University in 1955.[Edited by James O. Urmson.].** [S.l.]: Clarendon Press, 1962.

BAARSLAG, T. et al. Learning about the opponent in automated bilateral negotiation: a comprehensive survey of opponent modeling techniques. **Autonomous Agents and Multi-Agent Systems**, Springer, v. 30, n. 5, p. 849–898, 2016.

BENCH-CAPON, T. J.; DUNNE, P. E. Argumentation in artificial intelligence. **Artificial Intelligence**, Elsevier Science Publishers Ltd., v. 171, n. 10-15, p. 619–641, 2007.

BERARIU, T. An argumentation framework for BDI agents. In: **Intelligent Distributed Computing VII**. [S.l.]: Springer, 2014. p. 343–354.

BESNARD, P. et al. Introduction to structured argumentation. **Argument & Computation**, IOS Press, v. 5, n. 1, p. 1–4, 2014.

BESNARD, P.; HUNTER, A. A logic-based theory of deductive arguments. **Artificial Intelligence**, Elsevier, v. 128, n. 1, p. 203–235, 2001.

BESNARD, P.; HUNTER, A. **Elements of argumentation**. [S.l.]: MIT press Cambridge, 2008.

BESNARD, P.; HUNTER, A. Argumentation based on classical logic. **Argumentation in Artificial Intelligence**, Springer, n. PART 2, p. 133–152, 2009.

BIKAKIS, A.; ANTONIOU, G. Defeasible contextual reasoning with arguments in ambient intelligence. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 22, n. 11, p. 1492–1506, 2010.

BRAET, A. C. Ethos, pathos and logos in aristotle's rhetoric: A re-examination. **Argumentation**, Springer, v. 6, n. 3, p. 307–320, 1992.

BRATMAN, M. **Intention, plans, and practical reasoning**. [S.l.]: Harvard University Press, 1987.

CAMINADA, M. Semi-stable semantics. In: **Proceedings of the 1st Conference on Computational Models of Argument. COMMA 06**. [S.l.: s.n.], 2006. p. 121–130.

CAMINADA, M. A gentle introduction to argumentation semantics. **Lecture material, Summer**, 2008.

CAMINADA, M.; AMGOUD, L. On the evaluation of argumentation formalisms. **Artificial Intelligence**, Elsevier, v. 171, n. 5, p. 286–310, 2007.

CASTELFRANCHI, C. Reasons: Belief support and goal dynamics. **Mathware & Soft Computing**, v. 3, n. 1-2, p. 233–247, 2008.

CASTELFRANCHI, C.; GUERINI, M. Is it a promise or a threat? **Pragmatics & Cognition**, John Benjamins Publishing Company, v. 15, n. 2, p. 277–311, 2007.

CASTELFRANCHI, C.; PAGLIERI, F. The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions. **Synthese**, Springer, v. 155, n. 2, p. 237–263, 2007.

COSTE-MARQUIS, S.; DEVRED, C.; MARQUIS, P. Prudent semantics for argumentation frameworks. In: IEEE. **17th IEEE International Conference on Tools with Artificial Intelligence, 2005. ICTAI 05.** [S.l.], 2005. p. 5–pp.

ČYRAS, K.; TONI, F. Non-monotonic inference properties for assumption-based argumentation. In: SPRINGER. **International Workshop on Theorie and Applications of Formal Argumentation**. [S.l.], 2015. p. 92–111.

DEMIRDÖĞEN, Ü. D. The roots of research in (political) persuasion: Ethos, pathos, logos and the Yale studies of persuasive communications. **International Journal of Social Inquiry**, v. 3, n. 1, p. 189–201, 2010.

DIMOPOULOS, Y.; MORAITIS, P. Advances in argumentation based negotiation. **Negotiation and Argumentation in Multi-agent Systems: Fundamentals, Theories, Systems and Applications**, p. 82–125, 2011.

DUNG, P. M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. **Artificial intelligence**, Elsevier, v. 77, n. 2, p. 321–357, 1995.

DUNG, P. M.; MANCARELLA, P.; TONI, F. Computing ideal sceptical argumentation. **Artificial Intelligence**, Elsevier, v. 171, n. 10, p. 642–674, 2007.

EEMEREN, F. H. V. et al. Argumentation and artificial intelligence. In: **Handbook of Argumentation Theory**. [S.l.]: Springer, 2014. p. 615–675.

EEMEREN, F. H. V. et al. Fundamentals of argumentation theory. **LEA**, 1996.

FALCONE, R.; CASTELFRANCHI, C. Social trust: A cognitive approach. In: **Trust and deception in virtual societies**. [S.l.]: Springer, 2001. p. 55–90.

FALCONE, R.; CASTELFRANCHI, C. Trust dynamics: How trust is influenced by direct experiences and by trust itself. In: IEEE. **Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems.** [S.l.], 2004. p. 740–747.

FAN, X.; TONI, F. Assumption-based argumentation dialogues. In: AAAI PRESS. **Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume One**. [S.l.], 2011. p. 198–203.

FERRETTI, E. et al. A possibilistic defeasible logic programming approach to argumentation-based decision-making. **Journal of Experimental & Theoretical Artificial Intelligence**, Taylor & Francis, v. 26, n. 4, p. 519–550, 2014.

FLOREA, A. M.; KALISZ, E. Adaptive negotiation based on rewards and regret in a multi-agent environment. In: IEEE. **Symbolic and Numeric Algorithms for Scientific Computing, 2007. SYNASC. International Symposium on**. [S.l.], 2007. p. 254–259.

GARCIA, A. J.; SIMARI, G. R. Defeasible logic programming: An argumentative approach. **Theory and Practice of Logic Programming.**, v. 4, n. 1, p. 95–138, 2004.

GOROGIANNIS, N.; HUNTER, A. Instantiating abstract argumentation with classical logic arguments: Postulates and properties. **Artificial Intelligence**, Elsevier, v. 175, n. 9, p. 1479–1497, 2011.

GUERINI, M.; CASTELFRANCHI, C. Promises and threats in persuasion. **6th Workshop on Computational Models of Natural Argument**, p. 14–21, 2006.

HADJINIKOLIS, C.; MODGIL, S.; BLACK, E. Building support-based opponent models in persuasion dialogues. In: SPRINGER. **International Workshop on Theorie and Applications of Formal Argumentation**. [S.l.], 2015. p. 128–145.

HADJINIKOLIS, C. et al. Opponent modelling in persuasion dialogues. In: **Proceedings of the 23th International Joint Conference on Artificial Intelligence**. [S.l.: s.n.], 2013. p. 164–170.

HARMAN, G. **Practical aspects of theoretical reasoning**. [S.l.]: The Oxford Handbook of Rationality, 2004.

HIGGINS, C.; WALKER, R. Ethos, logos, pathos: Strategies of persuasion in social/environmental reports. In: ELSEVIER. **Accounting Forum**. [S.l.], 2012. v. 36, n. 3, p. 194–208.

HINDRIKS, K. V.; RIEMSDIJK, M. B. V. Satisfying maintenance goals. In: SPRINGER. **International Workshop on Declarative Agent Languages and Technologies**. [S.l.], 2007. p. 86–103.

HUNTER, A. Base logics in argumentation. In: **Proceedings of the 3rd Conference on Computational Models of Argument. COMMA 10**. [S.l.: s.n.], 2010. p. 275–286.

HUNTER, A. Modelling the persuadee in asymmetric argumentation dialogues for persuasion. In: **Proceedings of the 24th International Joint Conference on Artificial Intelligence**. [S.l.: s.n.], 2015. p. 3055–3061.

JAKOBOVITS, H.; VERMEIR, D. Robust semantics for argumentation frameworks. **Journal of Logic and Computation**, Oxford University Press, v. 9, n. 2, p. 215–261, 1999.

KAKAS, A. et al. Aba: Argumentation based agents. In: SPRINGER. **International Workshop on Argumentation in Multi-Agent Systems**. [S.l.], 2011. p. 9–27.

KAKAS, A. C. et al. Computational logic foundations of KGP agents. **Journal of Artificial Intelligence Research**, v. 33, p. 285–348, 2008.

KAKAS, A. C.; TONI, F.; MANCARELLA, P. Argumentation for propositional logic and nonmonotonic reasoning. In: **29 esimo Convegno Italiano di Logica Computazionale.** [S.l.: s.n.], 2014. p. 272–286.

KAKAS, A. C.; TORRONI, P.; DEMETRIOU, N. Agent planning, negotiation and control of operation. In: **Proceedings of the 16th European Conference on Artificial Intelligence. ECAI 04**. [S.l.: s.n.], 2004. p. 28–32.

KOK, E. M. et al. A formal argumentation framework for deliberation dialogues. In: SPRINGER. **International Workshop on Argumentation in Multi-Agent Systems**. [S.l.], 2010. p. 31–48.

KRAUS, S.; SYCARA, K.; EVENCHIK, A. Reaching agreements through argumentation: a logical model and implementation. **Artificial Intelligence**, Elsevier, v. 104, n. 1, p. 1–69, 1998.

MARTINEZ, M. V.; GARCÍA, A. J.; SIMARI, G. R. On the use of presumptions in structured defeasible reasoning. In: **Proceedings of the 4th Conference on Computational Models of Argument. COMMA 12**. [S.l.]: IOS Press, 2012. v. 245.

MAUTZ, R.; SHARAF, H. A. The philosophy of auditing. american accounting association. **Monograph No. 6. Sarasota, FL: American Accounting Association**, 1961.

MEYER, J.-J. C. Reasoning about emotional agents. **International Journal of Intelligent Systems**, Wiley Online Library, v. 21, n. 6, p. 601–619, 2006.

MODGIL, S.; PRAKKEN, H. The ASPIC+ framework for structured argumentation: a tutorial. **Argument & Computation**, Taylor & Francis, v. 5, n. 1, p. 31–62, 2014.

ODELL, J. J. et al. Modeling agents and their environment. In: SPRINGER. **International Workshop on Agent-Oriented Software Engineering**. [S.l.], 2002. p. 16–31.

PINYOL, I.; SABATER-MIR, J. Computational trust and reputation models for open multi-agent systems: a review. **Artificial Intelligence Review**, Springer, v. 40, n. 1, p. 1–25, 2013.

PRAKKEN, H. An abstract framework for argumentation with structured arguments. **Argument and Computation**, Taylor & Francis, v. 1, n. 2, p. 93–124, 2010.

PRAKKEN, H.; SARTOR, G. Argument-based extended logic programming with defeasible priorities. **Journal of applied non-classical logics**, Taylor & Francis, v. 7, n. 1-2, p. 25–75, 1997.

RAHWAN, I.; AMGOUD, L. An argumentation based approach for practical reasoning. In: **Proceedings of the 5th International Joint Conference on Autonomous Agents and Multi-agent Systems**. [S.l.: s.n.], 2006. p. 347–354.

RAHWAN, I. et al. Argumentation–based negotiation. **The Knowledge Engineering Review**, Cambridge Univ Press, v. 18, n. 04, p. 343–375, 2003.

RAHWAN, I.; SIMARI, G. R. **Argumentation in artificial intelligence**. [S.l.]: Springer, 2009.

RAMCHURN, S. D.; JENNINGS, N. R.; SIERRA, C. Persuasive negotiation for autonomous agents: A rhetorical approach. 2003.

RAMCHURN, S. D. et al. Negotiating using rewards. **Artificial Intelligence**, Elsevier, v. 171, n. 10-15, p. 805–837, 2007.

REED, C.; NORMAN, T. J. A roadmap of research in argument and computation. In: **Argumentation Machines**. [S.l.]: Springer, 2003. p. 1–13.

RIENSTRA, T.; THIMM, M.; OREN, N. Opponent models with uncertainty for strategic argumentation. In: **Proceedings of the 23th International Joint Conference on Artificial Intelligence**. [S.l.: s.n.], 2013. p. 332–338.

SABATER, J.; SIERRA, C. Regret: A reputation model for gregarious societies. In: **Proceedings of the 4th Workshop on Deception Fraud and Trust in Agent Societies**. [S.l.: s.n.], 2001. v. 70, p. 61–69.

SHI, B.; TAO, X.; LU, J. Rewards-based negotiation for providing context information. In: ACM. **Proceedings of the 4th International Workshop on Middleware for Pervasive and Ad-Hoc Computing.** [S.l.], 2006. p. 8.

SIERRA, C. et al. A framework for argumentation-based negotiation. In: SPRINGER. **International Workshop on Agent Theories, Architectures, and Languages**. [S.l.], 1997. p. 177–192.

SIERRA, C. et al. A framework for argumentation-based negotiation. In: **Intelligent Agents IV Agent Theories, Architectures, and Languages**. [S.l.: s.n.], 1998. p. 177–192.

SYCARA, K. P. Persuasive argumentation in negotiation. **Theory and decision**, Springer, v. 28, n. 3, p. 203–242, 1990.

TARSKI, A. The concept of truth in formalized languages. **Logic, semantics, metamathematics**, Oxford, v. 2, p. 152–278, 1956.

TONI, F. From logic programming and non-monotonic reasoning to computational argumentation and beyond. In: SPRINGER. **International Conference on Logic Programming and Nonmonotonic Reasoning**. [S.l.], 2017. p. 36–39.

VERHEIJ, B. Two approaches to dialectical argumentation: admissible sets and argumentation stages. In: **Proceedings of the Eight Dutch Conference on Artificial Intelligenc.** [S.l.: s.n.], 1996. v. 96, p. 357–368.

WALTON, D. **Fundamentals of critical argumentation**. [S.l.]: Cambridge University Press, 2005.

WALTON, D.; KRABBE, E. C. **Commitment in dialogue: Basic concepts of interpersonal reasoning**. [S.l.]: SUNY press, 1995.

WALTON, D. N. Types of dialogue, dialectical shifts and fallacies. **Argumentation illuminated**, SICSAT Amsterdam, p. 133–147, 1992.

WOOLDRIDGE, M.; JENNINGS, N. R. Intelligent agents: Theory and practice. **The knowledge engineering review**, Cambridge University Press, v. 10, n. 2, p. 115–152, 1995.

YU, B.; SINGH, M. P. A social mechanism of reputation management in electronic communities. In: SPRINGER. **International Workshop on Cooperative Information Agents**. [S.l.], 2000. p. 154–165.

APPENDIX A – PROOFS

A.1 PROOFS FOR CHAPTER 3 RESULTS

Proposition 3.3 (Direct consistency) Let $\mathcal{AF}_x = \langle \texttt{ARG}, \texttt{att} \rangle$ be an AF constructed from the theory $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$[1]. Let $\mathcal{E}_1, ..., \mathcal{E}_n$ be the set of extensions under the preferred semantics[2]. $\mathcal{AF}_x$ satisfies direct consistency iff:

(1) $\texttt{CONCS}(\mathcal{E}_i)$ is consistent for each $1 \leq i \leq n$.

(2) $\texttt{Output}$ is consistent.

Proof. By reduction *ab absurbo*. Assume that $\texttt{CONCS}(\mathcal{E}_i)$ is inconsistent. This means that $\exists b, \neg b \in \texttt{CONCS}(\mathcal{E}_i)$ where b and $\neg b$ are beliefs. This, in turns, means that $\exists A, B \in \mathcal{E}_i$ such that $(A,B), (B,A) \in \texttt{att}_{ep}$, in other words there is a rebuttal between A and B. The fact that there is a conflict between two arguments that belong to the same extension contradicts the premise of the proposition, which states that all $\mathcal{E}_i$ are preferred extensions and therefore, they are conflict-free extensions[3]. Since none of the extensions is inconsistent, then the intersection of them is not inconsistent either. □

A.2 PROOFS FOR CHAPTER 4 RESULTS

Proposition 4.1 Let $A = \langle T, g_1, go_1 \rangle$ and $B = \langle T, g_2, go_2 \rangle$ be two rhetorical arguments, if $\texttt{IMP}(go_1) = \texttt{EFF}(go_2)$ and $\texttt{IMP}(go_2) = \texttt{EFF}(go_1)$, then $\texttt{ST\_BASIC}(A) = \texttt{ST\_BASIC}(B)$.

Proof. This follows directly from the fact that the value of the basic strength of arguments (Equation 5) is the result of the arithmetic mean of two parameters: the importance and the effectiveness of the opponent's goal. Therefore, $\frac{\texttt{IMP}(go_1)+\texttt{EFF}(go_1)}{2} = \frac{\texttt{EFF}(go_2)+\texttt{IMP}(go_2)}{2}$ when $\texttt{IMP}(go_1) = \texttt{EFF}(go_2)$ and $\texttt{EFF}(go_1) = \texttt{IMP}(go_2)$. □

[1] The definition of theory is given in Definition 2.8

[2] The definition of preferred semantics is given in Definition 2.4

[3] The definition of conflict-freeness is given in Definition 2.2.

Proposition 4.2 Let $A = \langle T, g_1, go_1 \rangle$ and $B = \langle T, g_2, go_2 \rangle$ be two rhetorical arguments. If $\texttt{IMP}(go_1) > \texttt{IMP}(go_2)$ and $\texttt{EFF}(go_1) > \texttt{EFF}(go_2)$, then $\texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B)$.

Proof. This also follows from the fact that the value of the basic strength of an argument (Equation 5) is the result of the arithmetic mean of the importance and the effectiveness of the opponent's goal. Thus, the greater the values of the parameters are, the greater the arithmetic mean is. □

Proposition 4.3 Let $A = \langle T, g_1, go_1 \rangle$ and $B = \langle T, g_2, go_2 \rangle$ be two rhetorical arguments. Two cases can be distinguished:

1.If $\texttt{EFF}(go_1) > \texttt{EFF}(go_2)$ and $\texttt{IMP}(go_1) = \texttt{IMP}(go_2)$, then $\texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B)$

2.If $\texttt{EFF}(go_1) = \texttt{EFF}(go_2)$ and $\texttt{IMP}(go_1) > \texttt{IMP}(go_2)$, then $\texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B)$.

Proof. By reduction *ab absurbo*. Assume that $\texttt{ST\_BASIC}(A) \not> \texttt{ST\_BASIC}(B)$. Then $\texttt{ST\_BASIC}(A) \leq \texttt{ST\_BASIC}(B)$. By applying Equation 5, we can say that $\frac{\texttt{IMP}(go_1)+\texttt{EFF}(go_1)}{2} \leq \frac{\texttt{IMP}(go_2)+\texttt{EFF}(go_2)}{2}$. Since the denominators are the same, we can simplify the equations, then $\texttt{IMP}(go_1) + \texttt{EFF}(go_1) \leq \texttt{EFF}(go_2) + \texttt{IMP}(go_2)$.

For the first case, let us consider that $\texttt{IMP}(go_1) = \texttt{IMP}(go_2)$, which leads to $\texttt{EFF}(go_1) \leq \texttt{EFF}(go_2)$. This clearly contradicts the hypothesis that $\texttt{EFF}(go_1) > \texttt{EFF}(go_2)$.

For the second case, let us consider that $\texttt{EFF}(go_1) = \texttt{EFF}(go_2)$, which leads to $\texttt{IMP}(go_1) \leq \texttt{IMP}(go_2)$. This contradicts the hypothesis that $\texttt{IMP}(go_1) > \texttt{IMP}(go_2)$. □

Proposition 4.4 Let $A = \langle T, g_1, go_1 \rangle$ and $B = \langle T, g_2, go_2 \rangle$ be two rhetorical arguments. Three cases can be distinguished considering the difference between the effectiveness values of the opponent's goal:

1.Let $\texttt{EFF}(go_1) = \texttt{EFF}(go_2) + 0.25$.

If $\texttt{IMP}(go_1) \geq (\texttt{IMP}(go_2) - 0.25)$, then $\texttt{ST\_BASIC}(A) \geq \texttt{ST\_BASIC}(B)$.

Otherwise, if $\texttt{IMP}(go_1) < (\texttt{IMP}(go_2) - 0.25)$, $\texttt{ST\_BASIC}(A) < \texttt{ST\_BASIC}(B)$.

2.Let $\texttt{EFF}(go_1) = \texttt{EFF}(go_2) + 0.5$.

If $\texttt{IMP}(go_1) \geq (\texttt{IMP}(go_2) - 0.5)$, then $\texttt{ST\_BASIC}(A) \geq \texttt{ST\_BASIC}(B)$.

Otherwise, if $\texttt{IMP}(go_1) < (\texttt{IMP}(go_2) - 0.5)$, $\texttt{ST\_BASIC}(A) < \texttt{ST\_BASIC}(B)$.

3.Let $\texttt{EFF}(go_1) = \texttt{EFF}(go_2) + 0.75$.

If $\texttt{IMP}(go_1) \geq (\texttt{IMP}(go_2) - 0.75)$, then $\texttt{ST\_BASIC}(A) \geq \texttt{ST\_BASIC}(B)$.

Otherwise, if $\texttt{IMP}(go_1) < (\texttt{IMP}(go_2) - 0.75)$, $\texttt{ST\_BASIC}(A) < \texttt{ST\_BASIC}(B)$.

Proof. By reduction *ab absurbo*. Let us employ the following variable to refer to the difference of effectiveness for each case: $x \in \{0.25, 0.5, 0.75\}$. We will make the proof for the main conditional of the three cases. Assume that $\texttt{ST\_BASIC}(A) \ngeq \texttt{ST\_BASIC}(B)$. Then $\texttt{ST\_BASIC}(A) < \texttt{ST\_BASIC}(B)$. By applying Equation 5, we can say that: $\frac{\texttt{IMP}(go_1)+\texttt{EFF}(go_1)}{2} < \frac{\texttt{IMP}(go_2)+\texttt{EFF}(go_2)}{2}$.

Since the denominators are the same, we can simplify the equations, then: $\texttt{IMP}(go_1) + \texttt{EFF}(go_1) < \texttt{EFF}(go_2) + \texttt{IMP}(go_2)$.

Considering that $\texttt{EFF}(go_1) = \texttt{EFF}(go_2) + x$, we can replace $\texttt{EFF}(go_1)$:

$\texttt{IMP}(go_1) + \texttt{EFF}(go_2) + x < \texttt{EFF}(go_2) + \texttt{IMP}(go_2)$

$= \texttt{IMP}(go_1) < \texttt{EFF}(go_2) + \texttt{IMP}(go_2) - \texttt{EFF}(go_2) - x$

$= \texttt{IMP}(go_1) < \texttt{IMP}(go_2) - x$

This clearly contradicts the hypothesis that $\texttt{IMP}(go_1) \geq (\texttt{IMP}(go_2) - x)$. □

Proposition 4.5 Let A and B be two rhetorical arguments, and $\texttt{THRES}_1(O)$ and $\texttt{THRES}_2(O)$ be two different thresholds, each one associated to arguments A and B, respectively. If $\texttt{ST\_BASIC}(A) = \texttt{ST\_BASIC}(B)$ and $\texttt{THRES}_1(O) > \texttt{THRES}_2(O)$, then $\texttt{ST\_COMB}(A) < \texttt{ST\_COMB}(B)$.

Proof. By reduction *ab absurbo*. Assume that $\texttt{ST\_COMB}(A) \not< \texttt{ST\_COMB}(B)$. Then $\texttt{ST\_COMB}(A) \geqslant \texttt{ST\_COMB}(B)$. By applying Equation 6 and Equation 7, we can say that:

$$\texttt{ST\_BASIC}(A) \times (\texttt{REP}(P) - \texttt{THRES}_1(O)) \geqslant \texttt{ST\_BASIC}(B) \times (\texttt{REP}(P) - \texttt{THRES}_2(O))$$

Let us consider that $\texttt{ST\_BASIC}(A) = \texttt{ST\_BASIC}(B)$, then we can write:

$$\texttt{ST\_BASIC}(A) \times (\texttt{REP}(p) - \texttt{THRES}_1(O)) \leq \texttt{ST\_BASIC}(A) \times (\texttt{REP}(P) - \texttt{THRES}_1(O))$$

Let us assume that the value of $\texttt{ST\_BASIC}(A)$ is 1, then the result is:

$(\texttt{REP}(P) - \texttt{THRES}_1(O)) \geqslant (\texttt{REP}(p) - \texttt{THRES}_2(O))$

$= -\texttt{THRES}_1(O) \geqslant \texttt{REP}(P) - \texttt{THRES}_2(O) - \texttt{REP}(P)$

$= -\texttt{THRES}_1(O) \geqslant -\texttt{THRES}_2(O)$

$= \texttt{THRES}_1(O) \leqslant \texttt{THRES}_2(O)$

This contradicts the hypothesis that $\texttt{THRES}_1(O) > \texttt{THRES}_2(O)$. □

Proposition 4.6 Let A and B be two rhetorical arguments, and $\texttt{THRES}_1(O)$ and $\texttt{THRES}_2(O)$ be two different thresholds, each one associated to arguments A and B, respectively. If $\texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B)$ and $\texttt{THRES}_1(O) < \texttt{THRES}_2(O)$, then $\texttt{ST\_COMB}(A) > \texttt{ST\_COMB}(B)$.

Proof. By applying Equation 7 to calculate the combined strength of arguments A and B we obtain that $\texttt{ST\_BASIC}(A) \times (\texttt{REP}(P) - \texttt{THRES}_1(O))$ and $\texttt{ST\_BASIC}(B) \times (\texttt{REP}(P) - \texttt{THRES}_2(O))$, respectively. Recall that $\texttt{REP}(P) \in [0,1]$ and it is the same value for both arguments.

Given that $\texttt{THRES}_1(O) < \texttt{THRES}_2(O)$, we can say that $(\texttt{REP}(P) - \texttt{THRES}_1(O)) > (\texttt{REP}(P) - \texttt{THRES}_2(O))$.

Considering that $\texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B)$, we have that both multiplicands $\texttt{ST\_BASIC}(A)$ and $(\texttt{REP}(P) - \texttt{THRES}_1(O))$ are greater than $\texttt{ST\_BASIC}(\texttt{THRES}_2(O))$ and $(\texttt{REP}(P) - \texttt{THRES}_2(O))$, respectively. Since the product of two numbers is always greater than the product of other two numbers when the first ones are greater than the last ones, we can say that $\texttt{ST\_COMB}(A) > \texttt{ST\_COMB}(B)$ always hold. □

Proposition 4.7 Let A and B be two rhetorical arguments, and $\texttt{THRES}_1(O)$ and $\texttt{THRES}_2(O)$ be two different thresholds, each one associated to arguments A and B, respectively. Given $\texttt{ST\_BASIC}(A) > \texttt{ST\_BASIC}(B)$ and $\texttt{THRES}_1(O) > \texttt{THRES}_2(O)$, one of the following situations occur:

1. If $\frac{\texttt{ST\_BASIC}(A)}{\texttt{ACCUR\_CRED}_2} < \frac{\texttt{ST\_BASIC}(A) - \texttt{ST\_BASIC}(B)}{\texttt{ACCUR\_CRED}_2 - \texttt{ACCUR\_CRED}_1}$, then $\texttt{ST\_COMB}(A) > \texttt{ST\_COMB}(B)$.
2. If $\frac{\texttt{ST\_BASIC}(A)}{\texttt{ACCUR\_CRED}_2} > \frac{\texttt{ST\_BASIC}(A) - \texttt{ST\_BASIC}(B)}{\texttt{ACCUR\_CRED}_2 - \texttt{ACCUR\_CRED}_1}$, then $\texttt{ST\_COMB}(A) < \texttt{ST\_COMB}(B)$.
3. If $\frac{\texttt{ST\_BASIC}(A)}{\texttt{ACCUR\_CRED}_2} = \frac{\texttt{ST\_BASIC}(A) - \texttt{ST\_BASIC}(B)}{\texttt{ACCUR\_CRED}_2 - \texttt{ACCUR\_CRED}_1}$, then $\texttt{ST\_COMB}(A) = \texttt{ST\_COMB}(B)$.

Proof. Given $\frac{\texttt{ST\_BASIC}(A)}{\texttt{ACCUR\_CRED}_{O_2}} < \frac{\texttt{ST\_BASIC}(A) - \texttt{ST\_BASIC}(B)}{\texttt{ACCUR\_CRED}_2 - \texttt{ACCUR\_CRED}_1}$, we can say that it is equal to:

$\texttt{ST\_BASIC}(A) \times (\texttt{ACCUR\_CRED}_2 - \texttt{ACCUR\_CRED}_1) < \texttt{ACCUR\_CRED}_2 \times (\texttt{ST\_BASIC}(A) - \texttt{ST\_BASIC}(B)) =$

$(\texttt{ST\_BASIC}(A) \times \texttt{ACCUR\_CRED}_2 - \texttt{ST\_BASIC}(A) \times \texttt{ACCUR\_CRED}_1) < (\texttt{ACCUR\_CRED}_2 \times \texttt{ST\_BASIC}(A) - \texttt{ACCUR\_CRED}_2 \times \texttt{ST\_BASIC}(B)) =$

$-\texttt{ST\_BASIC}(A) \times \texttt{ACCUR\_CRED}_1 < -\texttt{ACCUR\_CRED}_2 \times \texttt{ST\_BASIC}(B) =$

$\texttt{ST\_BASIC}(A) \times \texttt{ACCUR\_CRED}_{O_1} > \texttt{ACCUR\_CRED}_2 \times \texttt{ST\_BASIC}(B) = \texttt{ST\_COMB}(A) > \texttt{ST\_COMB}(B)$

Due to the proofs for the other cases can be done in the same way and with same idea, we omit them. □