Personality in Argumentative Agents

Faculty of Science
Utrecht University
The Netherlands

Thesis for the Degree of Master of Science
Artificial Intelligence
45 ECTS

Author:
Marlon Edward Etheredge

Supervisor:
Prof. dr. mr. Henry Prakken

Second reader:
dr. Chris Janssen

16 October 2016

Abstract

With an increase in the amount of applications of argumentative agents comes an increasing demand for the adjustability of these agents according to the context of their applications. Previously, the behavior of argumentative agents was primarily based on game-theoretic approaches. In this thesis, the introduction of personality in these agents is investigated by a formalization of a model of personality, the implementation of such an agent and a method of modeling personalities of opponents.

Acknowledgements

I first would like to thank Prof. dr. mr. Henry Prakken for his supervision throughout this research. His guidance and meticulousness have proven to be invaluable and incredibly instructive to me throughout this research, both in my writing and the direction of this research.

I would also like to acknowledge dr. Chris Janssen as the second reader of this thesis.

I would like to show my appreciation for the support of my friends, colleagues and family, especially Jerrol, for his remarks and listening to my ideas. Lastly, I would like to thank Lindsey for her love, support and patience.

Contents

1 Introduction

Settling differences by discussion, although typically associated with human-beings, is an interesting task for software agents. These software agents, or simply agents, are autonomous software components that can perform tasks independently of a user of the software. In a multi-agent system, more than one of these agents perform tasks. An application of such a multi-agent system is an argumentation dialogue, where argumentative agents settle their differences. In this, agents would use argumentation techniques including, but not limited to, persuasion, deliberation and negotiation. Agents could, for instance, disagree on whether science endangers humanity or any other subject and convince their opponent of their stance. Not surprisingly, multi-agent systems of this form are a popular topic of research with a wide set of applications.

In the field of multi-agent systems dialogue systems are studied in terms of different types of dialogues, rules for participation and contribution in the dialogue and the role of arguments in the dialogue. Walton and Krabbe give a description of the different types of argumentation dialogues [WK95], among which *persuasion* is the dialogue type that has gained the most attention in the research of formal dialogue systems. In *persuasion* dialogues, the parties need to settle on a conflicting point of view, in *deliberation* dialogues, the parties have to decide upon some course of action. In his 2006 article, Prakken reviews formal systems that regulate these persuasion dialogues [Pra06]. Prakken's dialogue framework [Pra05] is a state-of-the art framework for argumentation dialogues and is amongst the last class of these frameworks. Such systems have found their way into many fields, including computer science and artificial intelligence. Argumentation has been used in decision making by Zhong et al. [Zho+14] where argumentation is used to explain best decisions and Van der Weide [Wei11], who uses argumentation in a system that aids fire commanders in decision making. In addition, argumentation has been used to resolve conflicts in firewalls [App+12]. Regarding argumentative agents, Parsons et al. investigate different attitudes agents can have [Par+03]. Opponent modelling is studied in this context to make a model of the opponent's possible knowledge [Rie+13; Had+12].

The process of the agent that makes the agent formulate arguments has attracted less interest from the scientific community. Moreover, investigating the introduction of personality in the argumentative agent is a new research topic in this field. Pre-

viously, the reasoning of argumentative agents was primarily optimized by means of game-theoretic approaches. While such agents are able to determine an optimal move to play in an argumentation dialogue given certain knowledge, this approach offers little contribution to adjustability. By modelling the personality of its opponent, an agent is able to optimize its strategy according to the opponent. Since an introduction of personality in argumentative agents is new, opponent modelling in the context of modelling the opponent's personality has not been studied. Investigating personalities in argumentative agents is, however, of interest, since it is expected that the behavior of the resulting agent is expected to be adjustable according to the application of the agent. Additionally, introducing personality in software agents allows for the creation of software agents that are more human-like than agents that follow a more systematic reasoning process. This in turn improves the compatibility of human tasks and tasks of artificial intelligence applications. Taking the research of Van der Weide as an example, introducing personality to the decision support system agent would allow for control over the behavior of the agent as well as optimization of the communication of the agent with the fire commander.

In this thesis a model for personalities of argumentative agents in persuasion dialogues will be introduced. This thesis focuses on persuasion dialogues, where the goal of each participant is to persuade their opponent into accepting a status, truth or untruth, of the topic of a dialogue. According to the configuration of the agent's personality the agent's behavior can be adjusted to suit a specific application and context. Since many different applications exist, the reasoning process of the agent should be adjustable according to the context of the agent's application. As an example, the personality of the agent could be configured to easily accept arguments by its opponent in scenarios that demand rapid decision making. In contrast, the agent could be configured to only accept arguments when the opponent uses irrefutable argumentation in scenarios where the correctness of the outcome of a dialogue is critical. In addition to the introduction of a model for personality in argumentative agents, this thesis will investigate the agent's ability to model its opponent. This opponent model can be used for the agent to optimize its strategy in an argumentation dialogue. The actual optimization of strategies by means of machine learning or mathematical optimization methods according to the opponent model is beyond the scope of this research and will be left as a future research topic.

The following dialogue is an example of an argumentation dialogue between two agents, Paul and Otto and will serve as a running example throughout this thesis:

Peter: Science endangers humanity. (Making a claim)

Otto: Why do you think that science endangers humanity? (Asking for support for the claim)

Peter: Since science brings about many new technologies that could potentially harm human-beings. (Providing support for the claim)

Otto: I agree with you that science brings about new technologies. (Conceding the provided support for the claim) But I disagree that this poses a threat to humanity, since science primarily introduces new technologies that improve the lives of human-beings. (Providing a counter argument)

Peter: Why do you think that these technologies improve the lives of human-beings? (Asking for support for the counter argument)

Otto: Since these technologies provide for a method of helping humans in situations where they would have been helpless otherwise. In addition, improving the lives of human-beings does not endanger humanity. (Providing support for the counter argument)

Peter: OK, I agree that science introduces new technologies that improve the lives of human-beings. Moreover, I agree that improving the lives of human-beings does not endanger humanity. (Conceding a claim)

By the introduction of personality, the agent can behave differently according to its personality. Based on the configuration of its personality, the agent could, throughout a dialogue, disprefer or prefer using certain utterances or certain arguments at particular times. In addition, the agent could, by analyzing the behavior of its opponent, adjust its strategy in an argumentation dialogue.

Introducing personality to argumentative agents and modelling the personality of the opponent gives rise to the following four research questions:

- How can personality be introduced to argumentative agents for persuasion dialogues?
- How can a model for personality in argumentative agents be devised that allows argumentative agents to reason according to a personality configuration?
- How can an argumentative agent featuring personality be implemented?
- How can an argumentative agent featuring personality model the personality of its opponent?

First, a theoretical model of personality in argumentative agents will be investigated. In addition to a theoretical model of personality, this research will give a description of how an agent featuring personality could be implemented. Kok introduces a testbed for deliberative argumentation dialogues that allows for experimenting with argumentative agent implementation [Kok13]. Kok, however, focuses his research on deliberation dialogues, which requires that Kok's software is adjusted such that the software can be used for persuasion dialogs. As opposed to a typical implementation of an argumentative agent, where the reasoning process of the agent would be implemented according to a fixed reasoning scheme, this thesis focuses on adding personality to such an argumentative agent. A *personality model* is used to direct an agent's reasoning, both by determining the preferred replies to moves given a particular stage of the argumentation dialogue and by providing a basis for the regulation of controlling the behavior of an argumentative agent according to the behavior of its opponent. By introducing personalities in argumentative agents, the behavior of the agent is expected to be less deterministic. Moreover, the agent is expected to be easily adjustable to reason according to a preferred style of argumentation.

This thesis contributes to the field of agent technology, multi-agent systems and computational argumentation. In addition this thesis implicitly addresses the "strong" versus "weak" artificial intelligence problem. Primarily, this research focuses on improving the reasoning of argumentative agents and the optimization of agent strategies. Agent technology focuses on software agents that learn and automate procedures and processes by introducing a degree of autonomy. Reasoning is such a mechanism that allows for software agents to automate procedures and processes. This research contributes to the research of agent reasoning in an argumentation context. Strong artificial intelligence refers to the introduction of concepts in artificial intelligence that we typically associate with human-beings. Personality can be seen as such a human concept. Therefore, this research implicitly addresses the "strong" versus "weak" artificial intelligence problem.

Chapter 2 introduces the background of this thesis, setting out the dialogue framework that is used as a basis for this research. For the introduction of a definition of personality in argumentative agents, this research will investigate the field of personality theory in psychology to make the personality of the agent resemble the personality of humans as closely as possible; this will be treated in chapter 3. The resulting description of personality forms the basis for a personality model that will be introduced in an argumentative agent and is focused towards the utterances that can be made in Prakken's persuasion dialogue, this personality model is introduced in chapter 4. For a description of an implementation of the argumentative agent, Kok's testbed is required to be adjusted such that the testbed can be used for persuasion dialogues, a description of the adjustments that need to be made is

given in chapter 5. The inner working of the argumentative agent is investigated in chapter 6. The source code of the software implementation is included in an online repository which can found at: `https://bitbucket.org/metheredge/baipd`. A method for modelling the opponent, based on the personality model that results from this research is discussed in chapter 7. Finally, the conclusion will explain the answers to the research questions that are put forward in this introduction in chapter 8.

2 Background

This chapter will introduce the foundation for argumentation of this thesis. A distinction is made between argumentation-based inference and argumentation-based dialogue. The first, treated in section 2.1, deals with arguments and relations between them. The latter, treated in section 2.2, describes how these arguments can be used in an argumentation dialogue and describes how participants in these dialogues are able to use these arguments in dialogues.

2.1 Argumentation Framework

Prakken introduces a formal framework for argumentation with structured arguments [Pra10]. The framework imposes a structure on arguments, but its abstractness allows for freedom of choice in the underlying logical language with a contrariness relation defined over it.

The framework, named $ASPIC^+$ after its progenitor $ASPIC$, defines arguments as inference trees formed by applying two kinds of inference rules; strict and defeasible rules. Naturally it follows that arguments can be attacked in three ways; attacking a premise, attacking a conclusion and attacking an inference. To resolve these attacks, preferences may be set, leading to three corresponding types of defeat: undermining, rebutting and undercutting. The framework is abstract by being applicable to any set of inference rules divided in strict and defeasible ones, it does, however, require use of a logical language that defines a contrary or contradictory relation over well-formed formulas in the language.

$ASPIC^+$ builds on previous work resulting from the $ASPIC$ project [Amg+06]. The resulting framework is a characterization of a set of tree-structured arguments ordered with a binary defeat relation, allowing for the instantiation of Dung's abstract formalism [Dun95] and making use of Dung semantics for the determination of acceptability status of the structured arguments.

2.1.1 Dung's Abstract Framework

Let us first define the concepts of Dung's abstract argumentation framework.

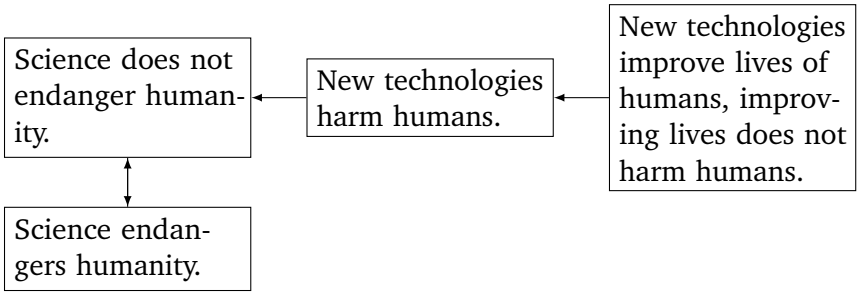


Fig. 2.1: A sample textual argumentation graph.

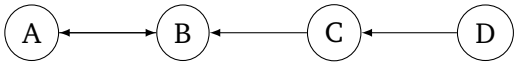


Fig. 2.2: A sample abstract argumentation graph.

Definition 2.1.1. An *abstract argumentation framework* (AF) is a pair $\langle \mathcal{A}, Def \rangle$. $\mathcal{A}$ is a set of arguments and $Def \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation of defeat. An argument A *defeats* an argument B iff $(A, B) \in Def$.

Definition 2.1.2. Let $\mathcal{B} \subseteq \mathcal{A}$.

- A set $\mathcal{B}$ is said to be *conflict-free* iff there exists no A_i, A_j in $\mathcal{B}$ such that A_i defeats A_j.
- A set $\mathcal{B}$ is said to *defend* an argument A_i iff for each argument $A_j \in \mathcal{A}$ it holds that if A_j defeats A_i, there exists A_k in $\mathcal{B}$ such that A_k defeats A_j.

Example 2.1.1. Let us consider a part of the dialogue between Otto and Peter presented in the introduction of this thesis as presented as a graph where the arrows indicates an argument attacking another argument.

The graph in figure 2.1 can be converted in a more abstract graph, where the arcs resemble attacks between arguments as seen in figure 2.2.

Definition 2.1.3. Let $\mathcal{B}$ be a conflict-free set of arguments, and let $\mathcal{F} : \wp(\mathcal{A}) \mapsto \wp(\mathcal{A})$ be a function such that $\mathcal{F}(\mathcal{B}) = \{A | \mathcal{B} \text{ defends } A\}$.

- $\mathcal{B}$ is *admissible* iff $\mathcal{B} \subseteq \mathcal{F}(\mathcal{B})$.
- $\mathcal{B}$ is a *complete extension* iff $\mathcal{B} = \mathcal{F}(\mathcal{B})$.
- $\mathcal{B}$ is a *grounded extension* iff it is the smallest (w.r.t. set-inclusion) complete extension.
- $\mathcal{B}$ is a *preferred extension* iff it is a maximal (w.r.t. set-inclusion) complete extension.

- $\mathcal{B}$ is a *stable extension* iff it is a preferred extension that defeats all arguments in $\mathcal{A} \setminus \mathcal{B}$.

Example 2.1.2. For the graph depicted in figure 2.2, the admissible sets are

- $\emptyset$,
- $\{A\}$,
- $\{D\}$,
- $\{A, D\}$ and
- $\{B, D\}$

, since

- $\mathcal{F}(\{A\}) = \{A, D\}$,
- $\mathcal{F}(\{D\}) = \{D\}$,
- $\mathcal{F}(\{B, D\}) = \{B, D\}$,
- $\mathcal{F}(\{A, D\}) = \{A, D\}$ and
- $\mathcal{F}(\emptyset) = \emptyset$.

It follows that the complete extensions are

- $\{D\}$,
- $\{A, D\}$ and
- $\{B, D\}$.

The grounded extension is $\{D\}$, while both the preferred and stable extensions are

- $\{A, D\}$ and
- $\{B, D\}$.

2.1.2 ASPIC$^+$

By combining the three possible ways of attack, Prakken is able to regard the $ASPIC^+$ framework as an instantiation of Dung's abstract framework. The framework still allows for different logical languages, different sets of inference rules for building arguments and different preference orderings and can therefore be regarded as an abstract framework. The framework allows for two types of reasoning, sound deductive reasoning and unsound defeasible but still rational reasoning.

The basis of the framework is that of an argumentation system, extending the notion of a proof system with a distinction between strict and defeasible inference rules. A preference ordering on the defeasible rules is added. The definitions in this section are taken from Prakken's paper [Pra10].

Definition 2.1.4. An *argumentation system* is a tuple $AS = (\mathcal{L}, ^-, \mathcal{R}, \leq)$ with

- $\mathcal{L}$ is a logical language,
- $^-$ is contrariness function from $\mathcal{L}$ to $\wp(\mathcal{L})$,
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of respectively strict and defeasible inference rules such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$,
- $\leq$ is a partial preorder on $\mathcal{R}_d$.

This contrariness function allows for the specification of non-symmetric conflict relations between formulas and contradictory formulas like 'harms humans' and 'saves lives', without the need for the axiom $\neg$(harmsHumans $\wedge$ savesLives).

Definition 2.1.5. Let $\mathcal{L}$, a set, be a logical language and $^-$ a contrariness function from L to $\wp(\mathcal{L})$. If $\varphi \in \bar{\psi}$ then if $\psi \notin \bar{\varphi}$ then φ is called a *contrary* of ψ, otherwise φ and ψ are called *contradictory*. *Contradictory* formulas a denoted by $\varphi = -\psi$.

Definition 2.1.6. Let $\mathcal{P} \subseteq \mathcal{L}$. $\mathcal{P}$ is *consistent* iff $\nexists \psi, \varphi \in \mathcal{P}$ such that $\psi \in \bar{\varphi}$, otherwise it is *inconsistent*.

Definition 2.1.7. Let $\varphi_1, \ldots, \varphi_n, \varphi$ be elements of $\mathcal{L}$.

- A *strict rule* is of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$, informally meaning that if $\varphi_1, \ldots, \varphi_n$ holds, then *without exception* it holds that φ.
- A *defeasible rule* is of the form $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$, informally meaning that if $\varphi_1, \ldots, \varphi_n$ holds, then it *presumably* holds that φ.

$\varphi, \ldots, \varphi_n$ are called *antecedents* of the rule and φ its *consequent*.

Arguments are constructed from a knowledge base and contain two kinds of formulas.

Definition 2.1.8. A *knowledge base* in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, \leq)$ is a pair $(\mathcal{K}, \leq')$ where $\mathcal{K} \subseteq \mathcal{L}$ and $\leq'$ is a partial preorder on $\mathcal{K} \setminus \mathcal{K}_n$. Here $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p$ where these subsets of $\mathcal{K}$ are disjoint and

- $\mathcal{K}_n$ is a set of (necessary) *axioms*. Intuitively, arguments cannot be attacked on their axiom premises.
- $\mathcal{K}_p$ is a set of *ordinary premises*. Intuitively, arguments can be attacked on their ordinary premises, and whether this results in defeat must be determined by comparing the attacked and the attacked premise.

Next, Prakken defines what is an *argument*. He makes use of the functions Prem that returns all the formulas of $\mathcal{K}$ (called *premises*) used to build the argument, Conc that returns its conclusion, Sub that returns all its sub-arguments, $\mathrm{DefRules}$ that returns all the defeasible inference rules of the argument and $\mathrm{TopRule}$ that returns the last inference rule used in the argument.

Definition 2.1.9. An *argument* A on the basis of a knowledge base $(\mathcal{K}, \leq')$ in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, \leq)$ is:

1. φ if $\varphi \in \mathcal{L}$ with:

   $\mathrm{Prem}(A) = \{\varphi\}$,

   $\mathrm{Conc}(A) = \varphi$,

   $\mathrm{Sub}(A) = \varphi$,

   $\mathrm{DefRules}(A) = \emptyset$,

   $\mathrm{TopRule}(A) =$ undefined.

2. $A_1 \ldots, A_n \to \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict rule

   $\mathrm{Conc}(A_1), \ldots, \mathrm{Conc}(A_n) \to \psi$ in $\mathcal{R}_s$,

   $\mathrm{Prem}(A) = \mathrm{Prem}(A_1) \cup \ldots \cup \mathrm{Prem}(A_n)$,

   $\mathrm{Conc}(A) = \psi$,

   $\mathrm{Sub}(A) = \mathrm{Sub}(A_1) \cup \ldots \cup \mathrm{Sub}(A_n) \cup \{A\}$,

$$\text{DefRules}(A) = \text{DefRules}(A_1) \cup \ldots \cup \text{DefRules}(A_n),$$

$$\text{TopRule}(A) = \text{Conc}(A_1), \ldots, \text{Conc}(A_n) \rightarrow \psi.$$

3. $A_1, \ldots, A_n \Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a defeasible rule $\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \Rightarrow \psi$ in $\mathcal{R}_d$.

$$\text{Prem}(A) = \text{Prem}(A_1) \cup \ldots \cup \text{Prem}(A_n),$$

$$\text{Conc}(A) = \psi,$$

$$\text{Sub}(A) = \text{Sub}(A_1) \cup \ldots \cup \text{Sub}(A_n) \cup \{A\},$$

$$\text{DefRules}(A) = \text{DefRules}(A_1) \cup \ldots \cup \text{DefRules}(A_n) \cup \{\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \Rightarrow \psi\},$$

$$\text{TopRule}(A) = \text{Conc}(A_1), \ldots, \text{Conc}(A_n) \Rightarrow \psi.$$

Example 2.1.3. $A = \varphi$, A is an argument using no rules of inference. For $r_s = \varphi \rightarrow \psi$, $B = \varphi, \varphi \supset \psi \rightarrow \psi$, B is an argument using a strict inference rule r_s from which it *without exception* follows that ψ. For $r_d = \varphi \Rightarrow \gamma$, $C = \varphi, \varphi \supset \gamma \Rightarrow \gamma$, C is an argument using a defeasible inference rule r_d from which it *presumably* follows that γ.

Four properties of arguments are defined, two that indicate the strictness of the argument and two that indicate the firmness of the argument.

Definition 2.1.10. An argument A is

- *strict* if $\text{DefRules}(A) = \emptyset$,
- *defeasible* if $\text{DefRules}(A) \neq \emptyset$,
- *firm* if $\text{Prem}(A) \subseteq \mathcal{K}_n$,
- *plausible* if $\text{Prem}(A) \not\subseteq \mathcal{K}_n$.

Prakken writes $S \vdash \varphi$ if there exists a strict argument for φ with all premises taken from S, and $S \mathrel{|\!\sim} \varphi$ if there exists a defeasible argument for φ with all premises taken from S.

Definition 2.1.11. Let $\mathcal{A}$ be a set of arguments. Then a partial preorder $\preceq$ on A is an *argument ordering* iff

1. if A is firm and strict and B is a defeasible or plausible, then $B \prec A$,

2. if $A = A_1, \ldots, A_n \rightarrow \psi$ then for all $1 \leq i \leq n$, $A \preceq A_i$ and for some $1 \leq i \leq n$, $A_i \preceq A$.

The first condition states that strict-and-firm arguments are stronger than all other arguments. The second condition states that a strict inference cannot make an argument weaker or stronger.

Definition 2.1.12. An *argumentation theory* is a triple $AT = (AS, KB, \preceq)$ where AS is an argumentation system, KB is a knowledge base in AS and $\preceq$ is an argument ordering on the set of all arguments that can be constructed from KB in AS.

Recall Dung's AF as defined in definition 2.1.1. By constructing arguments based on the knowledge base $\mathcal{K}$ and inference rules $\mathcal{R}$, we can now construct a corresponding AF.

Prakken defines three types of ways in which arguments can be attacked.

Definition 2.1.13. Three types of attack on arguments.

1. Argument A *undercuts* argument B (on B') iff $\mathrm{Conc}(A) \in \bar{B'}$ for some $B' \in \mathrm{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \psi$.

2. Argument A *rebuts* argument B on (B') iff $\mathrm{Conc}(A) \in \bar{\varphi}$ for some $B' \in \mathrm{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$. In such a case A *contrary-rebuts* B iff $\mathrm{Conc}(A)$ is a contrary of φ.

3. Argument A *undermines* B (on φ) iff $\mathrm{Conc}(A) \in \bar{\varphi}$ for some $\varphi \in \mathrm{Prem}(B) \setminus \mathcal{K}_n$. In such a case argument A *contrary-undermines* B iff $\mathrm{Conc}(A)$ is a contrary of φ or if $\varphi \in \mathcal{K}_a$.

Next, a notion of *successful attack* or *defeat* is defined.

Definition 2.1.14. Two types of successful attack.

1. Argument A *successfully rebuts* argument B if A rebuts B on B' and either A contrary-rebuts B' or $A \not\prec B'$.

2. Argument A *successfully undermines* B if A undermines B on φ and either A contrary-undermines B or $A \not\prec \varphi$.

The definition of successful attack leads to the definition of success.

Definition 2.1.15. An argument A *defeats* argument B iff no premise of A is an issue and A undercuts or successfully rebuts or successfully undermines B. Argument A *strictly defeats* argument B if A defeats B and B does not defeat A.

2.2 Prakken's Abstract Framework For Persuasion Dialogues

Prakken defines an abstract argumentation framework for argumentation dialogues with arbitrary argumentation-based logics with grounded semantics [Pra05]. The framework allows for the specification of different types of locutions, turn-taking rules, protocol and other structural decisions. The framework is flexible, while still providing for a basic structure an instantiation of the framework must adhere to.

Recall the dialogue between Peter and Otto from the introduction of this thesis. The goal of Prakken's abstract argumentation framework is providing participants in the dialogue to contribute utterances to the dialogue that further the goal of the dialogue. The following features of the nature of these dialogues can be observed:

- Different *locutions* can be used by participants, in this sample dialogue for persuasion; *claim*, *challenge*, *concede* and *retract*,
- arguments can be attacked by directly stating the opposite, while attacks are also possible on premises or used inference rules of the argument,
- participants can return to earlier choices and move alternative replies,
- participants may postpone replies.

First, the argumentation framework is defined as in [Pra05], to introduce the framework specialized for liberal dialogue systems in a later section.

2.2.1 Argumentation Framework

Let us define Prakken's argumentation framework formally. The following definitions were taken from Prakken's paper.

Dialogue systems have a *dialogue goal* and at least two *participants*, who can have various *roles*. Dialogue systems have two languages, the *communication language* specifying the locutions available to the participants, wrapped around the *topic language*. Sometimes a dialogue takes place in a *context* containing fixed and indisputable knowledge. The allowed moves at each point in a dialogue are governed by a *protocol*, the effects of utterances on the participants' commitments are specified as the *effect rules*. The outcome of a dialogue is defined by the *outcome rules* of the dialogue. Prakken typically makes a distinction between *turntaking*- and *termination rules* as part of the protocol.

Definition 2.2.1. A *dialogue system for argumentation* (*dialogue system* in sort) is a pair $\mathcal{L}, \mathcal{D}$, where $\mathcal{L}$ is a logic for defeasible argumentation and $\mathcal{D}$ is a dialogue system proper.

Originally, Prakken defined a logic for defeasible argumentation in his 2005 paper which evolved into $ASPIC^+$. Since $ASPIC^+$ is state-of-the-art and compatible with the argumentation framework, we take $ASPIC^+$ as a logic for defeasible argumentation.

Definition 2.2.2. A *dialogue system proper* is a triple $\mathcal{D} = \langle L_c, P, C \rangle$ with L_c (named the communication language) as a set of *locutions*, Pr as a *protocol* for L_c and C is a set of effect rules of locutions in L_c.

The communication language defines the available speech acts a participant can make use of when generating moves, while the protocol governs the allowed speech acts depending on a given dialogue. The effect rules determine for the speech acts available in the communication language what the effect is on the *commitments* of the participants. The communication language is defined as a set of locutions and two relations of attacking and surrendering replies defined on this set.

Definition 2.2.3. A *communication language* is a tuple $L_c = \langle S, R_a, R_s \rangle$. Here, S is a set of locutions and R_a and R_s respectively denote two binary relations of *attacking* and *surrendering* replies on S. Each $s \in S$ is of the form $p(c)$ where p is an element of a given set P of performatives and c is either a member of subset of L_t, or is a member of $Args$. Both R_a and R_s are considered being irreflexive and must adhere to the following conditions:

1. $R_a \cap R_s = \emptyset$
2. $\forall a, b, c : (a, b) \in R_a \Rightarrow (a, c) \notin R_s$
3. $\forall a, b, c : (a, b) \in R_s \Rightarrow (c, a) \notin R_a$

The function $att : R_s \longrightarrow \wp(R_a)$ assigns to each pair $(a, b) \in R_s$ one or more *attacking counterparts* $(c, b) \in R_a$.

Condition (1) states that a locution cannot be an attack and surrender at the same time, condition (2) states that a locutions cannot be an attack on one locution and a surrender to another locution, lastly (3) states that surrenders cannot be attacked.

Next, the definition of a move is given:

Definition 2.2.4. The set M of *moves* is defined as $\mathbb{N} \times \{P, O\} \times L_c^p \times \mathbb{N}$, where the four elements of a move m are denoted by:

- $id(m)$, indicating the *identifier* of m,
- $pl(m)$, indicating the *player* of m,
- $s(m)$, indicating the *speech act* of m,
- $t(m)$, indicating the *target* of m.

The set of *dialogues*, denoted by $M^{\leq\infty}$, is the set of all sequences $m_1, \ldots, m_i, \ldots, m_n$ from M such that:

- $id(m_i) = i$,
- $t(m_1) = 0$,
- for all $i > 1$ it holds that $t(m_i) = j$ for some m_j preceding m_i in the sequence.

The set of *finite dialogues*, denoted by $M^{<\infty}$, is the set of all finite sequences that satisfy these conditions. For any dialogue $d = m_1, \ldots, m_n$, the sequence $m_1, \ldots, m_i$ is denoted by d_i and d_0 denotes the empty dialogue.

The definition of dialogues enforces that only one speaker can speak at once. For move m and m' having $t(m) = id(m')$ we say that m *replies to* m', where m' is the *target* of m. While formally $t(m)$ denotes the identifier of a move, Prakken sometimes abuses notation to let $t(m)$ denote a move instead of its identifier. For $s(m)$ as an attacking reply to $s(m')$, it also holds that m is an attacking reply to m'.

A protocol assumes the existence of a turn-taking rule.

Definition 2.2.5. A *turn-taking function* T is a function $T : M^{<\infty} \longrightarrow \wp(\{P, O\})$ such that $T(\emptyset) = \{P\}$. A *turn* of a dialogue is a maximal sequence of stages in the dialogue where the same player moves.

When $T(d)$ is a singleton, Prakken omits the brackets. The definition allows the assignment of multiple participants to move next.

Next, the protocol should be defined.

Definition 2.2.6. A *protocol* on M is a function Pr with domain a non-empty subset of $M^{<\infty}$ taking subsets of M as values. The elements of $dom(Pr)$, the domain of Pr, are called the *legal finite dialogues*. The elements of $Pr(d)$ are called the moves allowed after d. A dialogue is said to be *terminated* if d is a legal dialogue and $Pr(d) = \emptyset$. A protocol Pr must satisfy the following condition: for all finite dialogues

d and moves m it holds that $d \in dom(Pr)$ and $m \in Pr(d)$ iff $d, m \in dom(Pr)$. All protocols must adhere to the following conditions for all moves m_i and all legal finite dialogues d:

If $m \in Pr(d)$ then:

- R_1: $pl(m) \in T(d)$.
- R_2: If $d \neq d_0$ and $m \neq m_1$, then $s(m)$ is a reply to $s(t(m))$ according to L_c.
- R_3: If m replies to m', then $pl(m) \neq pl(m')$.
- R_4: If there is a m' in d such that $t(m) = t(m')$ then $s(m) \neq s(m')$.
- R_5: For every $m' \in d$ that surrenders to $t(m)$, m is not an attacking counterpart of m'.

Rule R_1 states that a move is only valid if the move is played by the player who is allowed to move by the turn-taking function. Rule R_2 says that apart from the first move, a move should reply to an earlier move. Rule R_3 states that a player cannot reply to own moves. Rule R_4 states that, in case the player *backtracks*, meaning that the player moves an alternative reply to an earlier move, the content of the move should not be the same as the earlier move. Rules R_5 states that surrenders may not be revoked.

Lastly, a *commitment function* is introduced that assigns propositions as commitments to each player at different stages in the dialogue.

Definition 2.2.7. A *commitment function* is a function $C : M^{\leq\infty} \times \{P, O\} \longrightarrow \wp(L_t)$ such that $C_\emptyset(p) = \emptyset$. $C_d(p)$ denotes player p's commitments in dialogue d.

It is important to make clear the difference between *commitments* and *beliefs*. Where *commitments* are the propositions the player is expected to defend publicly, the player's *beliefs* are private and are expected to be not known by other players. Therefore, players' commitments can be used among others for determining when a dialogue ends, for example, the dialogue ends when a player gets committed to the initial claim of its opponent, or determining validity of a player's moves, for example, a player is not allowed to play a move that is contradicting its commitments.

2.2.2 Liberal Dialogue Systems

The abstract framework of Prakken can be used for any dialogue system that fits the framework. For persuasion, Prakken defines a framework specialized to what

he calls 'liberal dialogue systems'. This framework allows participants to move the following locutions:

- *Claim*, for making a claim,
- *challenge*, for challenging an opponent's claim,
- *argue*, for providing support for a claim,
- *concede*, for accepting an opponent's claim and
- *retract*, for retracting a claim.

Commitment rules are defined for this framework. However, these do not advocate the legal status of moves. In addition, turns by participants are not constrained in length, meaning that the participants may move as long as the move is a valid reply to some earlier move of the opponent.

2.2.3 Communication Language

The communication language will be defined such that the language allows for making claims, questioning propositions, conceding a claim, retracting a claim, providing support for an argument and attack by means of challenging premises. The resulting locutions each have appropriate commitment rules that define the effect on the commitments of players.

The table below contains the available types of speech acts in a liberal dialogue and presents the attacking and surrendering relations between the various types of speech acts.

Using these locutions, the running example dialogue can be converted into a dialogue as it would appear in a liberal dialogue system as following:

- P_1: *claim* endangersHumanity
- O_2: *why* endangersHumanity
- P_3: endangersHumanity *since* harmHumans
- O_4: *why* harmHumans
- P_5: harmHumans *since* newTechnologies

Acts	Attacks	Surrenders
claim φ	why φ	concede φ
why φ	argue A	retract φ
argue A	why φ($\varphi \in$ prem(A) argue B(B defeats A)	concede φ ($\varphi \in$ prem(A) or $\varphi =$ conc(A))
concede φ		
retract φ		

Tab. 2.1: Available speech act types

- O_6: *concede* newTechnologies
- O_7: *claim* ¬endangersHumanity
- P_8: *why* ¬endangersHumanity
- O_9: ¬endangersHumanity *since* helpingHumans
- P_{10}: *why* helpingHumans
- O_{11}: helpingHumans *since* newTechnologies
- P_{12}: *concede* helpingHumans
- O_{13}: ¬harmHumans *since* helpingHumans
- P_{14}: *retract* endangersHumanity

2.2.4 Commitment Rules

The following commitment rules are defined for the speech act types in table 2.1.

- If $s(m) =$ *claim*(φ) then $C_{pl}(d, m) = C_{pl}(d) \cup \{\varphi\}$
- If $s(m) =$ *why*(φ) then $C_{pl}(d, m) = C_{pl}(d)$
- If $s(m) =$ *concede*(φ) then $C_{pl}(d, m) = C_{pl}(d) \cup \{\varphi\}$
- If $s(m) =$ *retract*(φ) then $C_{pl}(d, m) = C_{pl}(d) - \{\varphi\}$
- If $s(m) =$ *argue* A then $C_{pl}(d, m) = C_{pl}(d) \cup prem(A) \cup \{conc(A)\}$

Where pl denotes the player of the move.

2.2.5 Turn-taking Function

The turn-taking function for liberal dialogue systems is such that a proponent starts the dialogue with a unique move, this move introduces the topic of the dialogue. After this first move, the opponent is assumed to reply, after this move the next speaker is always the speaker following the current speaker. This is captured in the turn-taking function

$$T_L : T(d_0) = P, T(d_1) = O, \text{ else } T(d) = \{P, O\}$$

This function ensures satisfaction of rule R_1.

2.2.6 Protocol

Two additional protocol rules are added to the protocol rules that are present in the abstract framework:

If $m \in P(d)$, then:

- R_6: If $d = \emptyset$, then $s(m)$ is of the form *claim*(φ) or *argue* A.
- R_7: If m concedes the conclusion of an argument moved in m', then m' does not reply to a *why* move.

Rule R_6 states that the first move should either present a claim or argument. The claimed proposition, or conclusion of the argument, is treated as the dialogue topic. Rule R_7 restricts concessions of an argument's conclusion to conclusions of counterarguments. This ensures that propositions are conceded at the place in which they were introduced.

Definition 2.2.8. A *dialogue system for liberal dialogues* is now defined as any dialogue system with L_c as specified in table 2.1, with turn-taking rule T_L and such that a move is legal iff it satisfies protocol rules R_1-R_7.

A dialogue is said to be terminated when no legal continuation is possible. Since the knowledge bases of players in the dialogue continually increase in terms of knowledge, it is unlikely that any player will run out of possible moves. For this reason it is necessary to determine additional conditions by which the dialogue is said to be terminated.

Definition 2.2.9. All attacking moves in a finite dialogue d are either *in* or *out* in d. Such a move m is *in* iff

1. m is surrendered in d; or else

2. all attacking replies to m are out,

otherwise m is *out*.

Definition 2.2.10. The status of the initial move m_1 of a dialogue d is *in favor of* P(O) and *against* O(P) iff m_1 is *in (out)* in d. We also say that m_1 favors, or is against p. Player p *currently wins* dialogue d if m_1 of d favors p.

Definition 2.2.11. A move m in a dialogue d is *surrendered* in d iff

- it is an *argue* A move and it has a reply in d that concedes A's conclusion; or else

- m has a surrendering reply in d.

Prakken has shown that for each finite dialogue d there is a unique dialogical status assignment. The 'current' winner of a dialogue can now be defined as follows:

Definition 2.2.12. For any dialogue d the proponent wins d if m_1 is *in*, otherwise the opponent wins d.

3 Personality Theory

> *If a thing exists, it exists in some amount; and if it exists in some amount, it can be measured.*
>
> **— E.L. Thorndike, 1914**

The first step in the introduction of personalities in argumentative agents is defining what *personality* is. This definition is attributed to the field of personality psychology. Winter and Barenbaum [WB99] give a comprehensive description of the field of personality psychology and its research. Personality psychologists take different stances in the way that personality should be investigated, including but not limited to biological, evolutionary and behaviorist perspectives. One such perspective is concerned with the study of so-called *personality traits*, which subdivide the concept of *personality* into measurable patterns of humans' behavior. Personality traits are considered to be relatively stable over age and situational changes [McC+02]. In general, personality is considered to be a fixed aspect of an individual. Different taxonomies exist, varying in the number of personality traits that make up the taxonomy and differing in the descriptions of personality traits. One popular personality theory includes five personality traits [JS99; MC08; Wig96], often referred to by the names five-factor model (FFM), "Big Five" or the acronym OCEAN and comprised of the personality traits:

- Openness to experience
- Conscientiousness
- Extraversion
- Agreeableness
- Neuroticism

Personality traits can be considered as discrete, indicating that a certain personality trait either exists, or does not exist in an individual. More common practice is to regard personality traits as being on a continuous scale, where for example extraversion–introversion indicates the personality trait as a scale.

The Eysenck Personality Inventory (EPI) [EE65] is a personality theory that describes three personality traits. The Myers-Briggs type indicator, often referred to by the acronym MBTI, is a popular personality test especially within companies. This test is based on Jung's taxonomy [Jun71]. The MBTI, however, lacks support of the majority of the scientific community, due to criticism regarding its reliability and validity [Pit93; Zem92; Boy95]. The HEXACO personality theory [Ash+04] partially overlaps with the FFM in the description of personality traits, but points to the existence of a sixth personality trait honesty–humiliation.

The current research will be based on the FFM personality theory. This theory remains to date the most popular and widely accepted theory and thus is used in a large body of research. The description of personality in our argumentative agent will, however, allow for adopting a different underlying personality theory.

3.1 The OCEAN model

FFM, also referred to by the acronym OCEAN, is a model describing personality in terms of as the name suggests, five personality traits. These personality traits are *openness to experience*, *conscientiousness*, *extraversion*, *agreeableness* and *neuroticism*. These five traits all characterize an individual by means of six particular facets. The following descriptions are taken from the NEO PI-R [CM92] personality inventory, keying the personality traits of the FFM.

3.1.1 Openness to Experience

The personality trait *openness to experience* describes an individual's active seeking and appreciation of experiences for their own sake.

This personality traits includes the facets *fantasy*, describing the receptivity to the inner world of imagination; *aesthetics*, describing a tendency to appreciate art and beauty; *feelings*, the openness to inner feelings and emotions; *actions*, which describes the openness to new experiences on a practical level; *ideas*, describing intellectual curiosity; *values*, which describes the readiness to re-examine own values and those of authority figures.

3.1.2 Conscientiousness

This trait describes the degree of organization, persistence, control and motivation in goal directed behaviour of an individual.

The corresponding facets are *competence*, which describes belief in own self-efficacy; *order*, describing personal organization; *dutifulness*, indicating the emphasis placed on importance of fulfilling moral obligations; *achievement striving*, which indicates the need for personal achievement and sense of direction; *self-discipline*, one's capacity to begin tasks and follow through to completion despite boredom or distractions; *deliberation*, the tendency to think things through before acting or speaking.

3.1.3 Extraversion

The personality trait *extraversion* describes an individual's quantity and intensity of energy directed outwards into the social world.

Facets that belong to this personality trait are *warmth*, one's interest in and friendliness towards others; *gregariousness*, describing the preference for the company of others; *assertiveness*, indicating social ascendancy and forcefulness of expression; *activity*, describing the pace of living; *excitement seeking*, the need for environmental stimulation; *positive emotions*, one's tendency to experience positive emotions.

3.1.4 Agreeableness

This personality trait relates to the kinds of interactions an individual prefers from compassion to tough mindedness.

The facets that agreeableness embodies are *trust*, describing a person's preference in believing others; *straightforwardness*, which describes the tendency of a person to be direct and frank in communication with others; *altruism*, which depicts a person's preference to be selfless; *compliance*, which characterizes a person's typical response to conflict in sense of cooperation versus competing; *modesty*, describing a person's tendency to be humble and other-focused in contrast with being arrogant and self-aggrandizing; *tender-mindedness*, which describes the attitude of sympathy for others.

3.1.5 Neuroticism

The personality trait *neuroticism* identifies individuals who are prone to psychological distress.

This personality trait describes the facets *anxiety*, the level of free-floating anxiety; *angry hostility*, the tendency to experience anger and related states such as frustration and bitterness; *depression*, the tendency to experience feelings of guilt, sadness, despondency and loneliness; *self-consciousness*, indicates the tendency to be shy

or anxious; *impulsiveness*, the tendency to act on craving and urges rather than reining them in and delaying gratification; *vulnerability*, one's general susceptibility to stress.

4 Personality Model

The last chapter laid the foundations of the personality of the argumentative agent. This chapter will introduce what will be called the *personality model* of the argumentative agent, which will describe the personality and mechanisms for reasoning based on the personality of the argumentative agent. Recall that the underlying personality theory used for this personality model is the FFM as introduced in the previous chapter. However, the personality model may be defined along other personality theories. The personality model consists of a number of components.

Personality vector The *personality vector* will associate with every personality facet included in the personality theory a strength that denotes the strength of that individual personality facet in the agent's personality.

Attitude The *attitude* of an agent acts as a condition that must be met before the agent is allowed to act in an argumentative dialogue.

Reasoning system The *reasoning system* of the agent allows the agent to act by making moves in an argumentative dialogue. The reasoning system will, based on the personality of the agent, define a preference ordering over possible actions an agent can perform in the dialogue.

The FFM personality theory describes a total of 30 personality facets associated with five personality traits. A number of these personality facets has no use in the description of personality of our argumentative agent; the personality theory as used in the personality model is thus reduced to only those personality facets that are useful for the context of our argumentative agent. The following sections will first introduce the interpretation of the FFM personality theory in the description of personality of the argumentative agent. Subsequently, the personality facets that were from the underlying personality theory are discussed.

4.1 Action Selection vs. Action Revision

The personality model makes a distinction between *action selection* and *action revision* personality facets and the personality vectors as the device for configuration of an agent's personality. Action selection makes for, as the name suggests, selection of

different actions throughout the dialogue. This means that the action selection process of the agent's reasoning system determines what speech act types are preferred at a given point in an argumentation dialogue based on the configuration of the individual personality facet strengths in the action selection personality vector of the agent. This means that the action selection facets directly influence the agent's preference for a particular speech act type like *claim* or *accept*. The agent's action revision process in its reasoning system does not directly influence the preference of speech act types, but rather indicates what *attitude* for a given speech act type is preferred. *Attitudes* will be introduced later in this chapter, but the intuitive definition of an agent *attitude* states that the *attitude* of an agent specifies under what conditions the agent is allowed to contribute a move to the dialogue with a particular speech act type. Attitudes will, for example, indicate that an agent is only allowed to move a *claim* move if the agent is allowed to generate an argument for the proposition. Thus, action selection makes the agent prefer speech act types, while action revision makes the agent prefer particular attitudes for those speech act types. Combined, both make sure that the agent not only prefers certain speech act types, but in addition is able to specify under what circumstances the agent agrees to contribute moves with a speech act type to the dialogue.

4.2 Agent Personality

The FFM describes personality in terms of traits and associated facets. IPIP [@Gol99] contains a comprehensive description of these traits and personality facets in the NEO PI-R personality inventory [CM92] which are typically used in personality tests. The NEO PI-R is a personality inventory that is intended to measure the personality traits of the FFM. For the current research, the IPIP is used to give a descriptive interpretation of the FFM personality facets in the agent that will later be used in the agent's reasoning system to allow for the agent to reason according to its personality.

The following sections will specify the interpretation of FFM personality facets in the argumentative agent. These facets are subdivided into two distinct sets, for *action selection facets* and for *action revision facets*. Remember that each facet is associated with one personality trait and that each personality trait has six associated facets. Since the facet are the fine-grained aspects of personality, facets form the basis of the personality model.

4.3 Action Selection Facets

Six action selection facets are distinguished: Self-consciousness, Assertiveness, Actions, Ideas, Values and Competence.

4.3.1 Self-consciousness

Self-consciousness "indicates the tendency to be shy or anxious". Based on this description the facet is determined to indicate an agent's tendency to not prefer to make claims and argue, and, therefore, not prefer the *claim* and *argue* speech acts. Additionally, the agent prefers to accept claims by opponents since the agent is less likely to stand up for itself. This interpretation is supported by the description in the IPIP description by a positively keyed description "Stumble over my words" and negatively keyed "Am able to stand up for myself".

4.3.2 Assertiveness

Assertiveness indicates an agent's "social ascendancy and forcefulness of expression". Based on this description, the assertive agent is expected to prefer making claims and, therefore, prefer the *claim* speech act. This interpretation is supported by description in the IPIP for positive keys "Take control of things", "Take charge", "Seek to influence others" and negatively keyed "Wait for others to lead the way", "Keep in the background", "Have little to say", "Don't like to draw attention to myself" and "Hold back my opinions".

4.3.3 Actions

Actions is described as an agent's "openness to new experiences on a practical level". This description is interpreted as the agent being open to information received from his opponent since the agent is open to new information. This indicates a preference for the *concede* speech act. Note that this does not mean that the agent is willing to accept every proposition an opponent claims, since the agent's stance towards acceptance is governed by its attitude. This interpretation is supported by the description in the IPIP, positively keyed "Interested in many things" and negatively keyed "Prefer to stick with things that I know", "Dislike changes", "Don't like the idea of change" and "Am attached to conventional ways".

4.3.4 Ideas

Ideas determines an agent's "intellectual curiosity". This is interpreted as an agent being interested in the support for a claim by its opponent, which indicates a preference for the *challenge* speech act since the agent essentially asks the opponent to provide additional support. The IPIP description of this facet supports this interpretation, positively keyed, "Can handle a lot of information", "Enjoy thinking about things" and negatively "Am not interested in abstract ideas", "Avoid philosophical discussions", "Have difficulty understanding abstract ideas", "Am not interested in theoretical discussions" and "Avoid difficult reading material".

4.3.5 Values

Values is described as an agent's "readiness to re-examine own values and those of authority figures". The description of this facet is interpreted as the agent being open to the possibility of another truth. This indicates a preference for the *retract* speech act. Again, note that this does not indicate that the agent is willing to retract all of its claims since the agent is only retracting once this is allowed by the selected attitude. This interpretation is supported by the description as present in the IPIP description, positively keyed "Believe that there is no absolute right or wrong" and negatively keyed "Believe in one true religion".

4.3.6 Competence

Competence indicates "belief in own self-efficacy". The description of this facet is interpreted as the agent being less open to accepting a truth that is not its personal goal. Meaning that the agent will strive to achieve its own goal at the expense of not accepting arguments by its opponent. This indicates the agent disfavoring the speech acts *retract* and *accept*. Although this might seem harsh, this does not mean that the agent is not willing to accept an argument by its opponent or retract its own claim. The agent disfavors these speech acts, however, if the agent is forced (since there is no other option available) to accept a claim or retract its own claim, the agent will still accept or retract as the last choice if this is allowed by the selected attitude. In addition, a competent agent prefers to support its propositions by moving *argue* moves. This interpretation is supported by the IPIP description, positively keyed, "Am sure of my ground".

4.4 Action Revision Facets

Ten action revision facets are distinguished: Achievement striving, Self-discipline, Deliberation, Activity, Trust, Straightforwardness, Modesty, Anxiety, Angry hostility and Depression.

4.4.1 Achievement Striving

Achievement striving is described as "the need for personal achievement and sense of direction". This description is interpreted as the agent preferring an attitude towards types of speech acts that either makes the agent support or claim some proposition that is in line with the agent's personal goal. Additionally, the agent prefers to select an attitude towards a type of speech act that defends the agent's personal goal. Lastly, the agent disfavors accepting and retracting propositions that are not in line with the agent's personal goal. This interpretation is supported by the IPIP description for a positive key "Go straight for the goal", "Turn plans into actions" and "Plunge into tasks with all my heart". The negative keys that support this interpretation are "Am not highly motivated to succeed" and "Do just enough work to get by".

4.4.2 Self-discipline

Self-discipline indicates "one's capacity to begin tasks and follow through to completion despite boredom or distractions". This description is interpreted as the agent preferring attitudes towards types of speech acts that require the agent to add to lines of dispute if the agent is able to. Meaning that the agent will continue to extend a line of dispute if the agent can do so. In other words, the agent prefers attitudes towards speech acts that disallow the agent to abandon certain lines of dispute. The interpretation is supported by the IPIP described, positively keyed, "Carry out my plans" and negatively keyed "Postpone decisions".

4.4.3 Deliberation

Deliberation is described as the agent's "tendency to think things through before acting or speaking". The description is interpreted for our purpose as the agent preferring attitudes towards types of speech acts that require the agent to make *well-motivated* moves. A move is considered to be *well-motivated* if the agent can provide a supporting argument for a proposition. This interpretation is supported by the IPIP descriptions, positively keyed, "Avoid mistakes", "Choose my words with care" and

negatively keyed "Jump into things without thinking", "Make rash decisions", "Like to act on a whim", "Rush into things", "Do crazy things" and "Act without thinking".

4.4.4 Activity

Activity is described as "the pace of living". The introduction of attitudes in argumentative agents may allow agents to not move at all since agents could select attitudes based on their personality that disallow the agent to move a certain type of speech act given a certain context, for every type of speech act. Therefore, the activity facet is interpreted as the preference of an agent to select attitudes towards types of speech acts that require the agent to make a move. The interpretation is supported by the IPIP description, positively keyed, "Am always busy" and negatively "Like to take it easy".

4.4.5 Trust

Trust is described as an agent's "preference in believing others". This facet is interpreted as the agent preferring attitudes towards types of speech acts that allow the agent to accept propositions by its opponent. This interpretation is supported by the description in the IPIP glossary, positively keyed, "Trust others", "Believe that others have good intentions", "Trust what people say", "Believe that people are basically moral" and "Believe in human goodness". Negatives keys that support this interpretation are "Distrust people", "Suspect hidden motives in others", "Am wary of others" and "Believe that people are essentially evil".

4.4.6 Straightforwardness

Straightforwardness is described as "the tendency of a person to be direct and frank in communication with others". This description is interpreted as the agent to prefer the selection of attitudes towards types of speech acts that disallow the agent to be incoherent, irrelevant or verbose. This interpretation is supported by the positive key of the IPIP description "Stick to the rules" and negative keys "Know how to get around the rules", "Cheat to get ahead", "Take advantage of others" and "Obstruct others' plans".

4.4.7 Modesty

Modesty is described as "a person's tendency to be humble and other-focused in contrast with being arrogant and self-aggrandizing". This description is interpreted as an agent's preference to select attitudes towards types of speech acts that dis-

allow the agent to make claims. In addition, the description is interpreted as an agent's preference to retract its own claims, accept claims by other players and the agent's disfavor to challenge other players' claims. The interpretation of this facet is supported by the positively keyed IPIP descriptions "Dislike being the center of attention", "Consider myself an average person" and "Seldom, toot my own horn". Negative keys that support this interpretation are "Believe that I am better than others", "Think highly of myself", "Have a high opinion of myself", "Know the answers to many questions" and "Make myself the center of attention".

4.4.8 Anxiety

Anxiety is described as "the level of free-floating anxiety". This facet is interpreted as the anxious agent preferring attitudes towards types of speech acts that allow the agent to accept and retract by non-well-motivated moves while disallowing the agent to claim or challenge. This interpretation is supported by the positive keys of the facet in the IPIP description "Worry about things", "Fear for the worst", "Am afraid of many things" and "Get stressed out easily". The negative keys that support this interpretation are "Am not easily bothered by things", "Am relaxed most of the times", "Am not easily disturbed by events" and "Don't worry about things that have already happened".

4.4.9 Angry Hostility

Angry hostility is described as "the tendency to experience anger and related states such as frustration and bitterness". This description is interpreted as the agent to prefer attitudes towards types of speech acts that allow the agent to obstruct his opponent. Here, obstruction can be seen as the agent claiming the contrary of some claimed proposition by its opponent, refusing to accept or retract, although the agent is morally required to accept some proposition or unnecessarily challenging propositions. This interpretation is supported by the following respectively positive and negative keys from the IPIP description "Get angry easily", "Get irritated easily", "Get upset easily", "Am often in a bad mood", "Lose my temper" and "Rarely get irritated", "Seldom get mad", "Am not easily annoyed", "Keep my cool", "Rarely complain".

4.4.10 Depression

Depression is described as "the tendency to experience feelings of guilt, sadness, despondency and loneliness". This description in argumentative agents is interpreted as the agent preferring to select attitudes towards types of speech acts that allow the agent to retract the agent's own claims. This description is supported by the IPIP

description, positively keyed, "Have a low opinion of myself" and "Have frequent mood swings". Negatively keyed, the descriptions "Feel comfortable with myself" and "Am very pleased with myself" support the interpretation.

4.5 Personality Facets Revisited

As can be observed from the descriptions of the personality facets of the last section, the personality facets that are included in the personality description of the argumentative agent are 16 personality facets. 14 personality facets were removed, since these personality facets either have overlapping descriptions, or their descriptions do not have a useful interpretation in the agent's personality.

The personality trait *openness to experience* includes the personality facet *fantasy*, which is described as "describing the receptivity to the inner world of imagination". Although this facet is beneficial in a generic theory of human personality, including this facet in a description of personality in argumentative agents does not seem beneficial. Agents that argue about certain topics are constrained by the fact that the purpose of these agents is argumentation and are therefore not expected to have a sense of imagination. The facet *aesthetics* seems non-beneficial; this facet is described as "describing a tendency to appreciate art and beauty". The context of the agent does not include the concepts "art" nor "beauty", therefore a description of this personality facet in our model seems superfluous. Another facet of this trait that is not considered is *feelings*, described as "the openness to inner feelings and emotions". Emotions are not considered part of this research.

The personality trait *conscientiousness* describes the facet *order*, described as "describing personal organization". As order is implicit in agents, it seems non-beneficial to include this facet. The same personality trait includes the personality facet *dutifulness*; this facet would be interpreted as an agent's preference to not cheat and stick to the rules, this interpretation overlaps with the interpretation of *straightforwardness* and is therefore removed from the model.

Warmth is a facet of the personality trait *extraversion* and indicates "one's interest in and friendliness towards others". This facet could be beneficial when introducing coalitions of argumentative agents. This is, however, beyond the scope of this research. The same holds for the facet *gregariousness*, describing "the preference for the company of others". The facet *excitement seeking* associated with the same personality trait and described by "the need for environmental stimulation" seems unuseful when respecting the context of our agent. *Positive emotions* is another facet of the personality trait *extraversion*, described as "one's tendency to experience positive emotions". Due to the constraints of our application, this facet is of no

use. Differentiating between positive or negative emotions has no meaning in this context.

The personality trait *agreeableness*, among others, describes the facets *altruism*, *compliance* and *tender-mindedness*. The first determines the "preference to be selfless". Argumentative agents are expected to be goal-oriented, which removes the use of this facet. The second facet describes "a person's typical response to conflict in sense of cooperation versus competing"; this is a facet that is useful when considering coalitions of argumentative agents. The last of these three facets *tender-mindedness* indicates "the attitude of sympathy for others". Much like *altruism*, this facet seems of no use.

Neuroticism includes the personality facet *impulsiveness* which in this context would be described as an agent's preference for inconsiderate argumentation. The *impulsiveness* personality facet is not considered in the personality model, since this facet overlaps with the description of *deliberation*.

The personality facet *vulnerability* as part of the personality trait *neuroticism* is described as "one's general susceptibility to stress". This facet is not considered beneficial for the description of personality of our agent.

Omitting the mentioned personality facets from the personality theory yields a simplified personality theory that is tailored towards a personality description of argumentative agents.

Openness	**Conscientiousness**	**Extraversion**	**Agreeableness**	**Neuroticism**
Actions	Competence	Assertiveness	Trust	Anxiety
Ideas	Achievement striving	Activity	Straightforwardness	Angry hostility
Values	Self-discipline		Modesty	Depression
	Deliberation			Self-consciousness

This revised personality theory will be used as the personality theory for the personality model. Any further references in this thesis to the personality theory will refer to this revised personality theory.

4.6 Personality Vector

To specify the strengths of the particular personality facets in the personality of the agent, we need a structure that associates a strength value for each of the personality facets in the personality theory. To this end, the *personality vector* is introduced. The personality vector will serve as the input to the agent's reasoning system and can be

regarded as the description of the agent's personality. Since the personality facets were subdivided into two sets, *action selection* facets and *action revision* facets, two distinct personality vectors will also co-exist.

Let $\mathcal{AS}$ denote the set of action selection personality facets and let $\mathcal{AR}$ denote the set of action revision personality facets. Moreover, let $\text{Strength}(f) \mapsto \mathbb{R}$ denote the strength of facet f in the personality of the agent where $f \in \mathcal{AS}$ or $f \in \mathcal{AR}$. The two distinct personality vectors can be defined as follows.

Definition 4.6.1. An *action selection personality vector* $\mathcal{PV}_{\mathcal{AS}}$ is a vector

$$[\text{Strength}(f_1), \text{Strength}(f_2), \ldots, \text{Strength}(f_n)]$$

Such that $n = |\mathcal{AS}|$ and $f_1, f_2, \ldots f_n \in \mathcal{AS}$.

Definition 4.6.2. An *action selection personality vector* $\mathcal{PV}_{\mathcal{AR}}$ is a vector

$$[\text{Strength}(f_1), \text{Strength}(f_2), \ldots, \text{Strength}(f_n)]$$

Such that $n = |\mathcal{AR}|$ and $f_1, f_2, \ldots f_n \in \mathcal{AR}$.

Based on the strengths of personality facets in the agent's personality, the reasoning system can determine the most preferred move given the context of an argumentative dialogue. The configuration of these personality vectors indicates the personality of the agent in combination with the descriptions of the personality facets in the model. This information is used to drive the agent's reasoning, based on the agent's specific configuration of its personality.

4.7 Agent Attitudes

Agent attitudes allow for the agent to specify under what conditions the agent is allowed to contribute a move to the dialogue that has a particular speech type as its content. Since attitudes specify these conditions for the different speech act types, attitudes are defined for each of the speech act types present in the argumentation framework. A simple example of an agent attitude could be an attitude that specifies that the agent is only allowed to make a *claim* for some proposition if the agent can support the proposition by an argument. Suppose that *claim* is a preferred speech act type for the agent, but the agent is not able to provide an argument for some proposition; then the agent would not be able to make the claim. The selection of attitudes is governed by the agent's action revision personality vector and thus allows for the selection of attitudes independent of the selection of preferred speech act types by means of the action selection personality vector. This division makes the reasoning process of the agent more rich in that, apart from a preference in

particular speech act types, the agent can be configured to have a certain stance towards contributing such a move to the dialogue.

The use of agent attitudes in the current research builds on previous work by Parsons et al. [Par+03], which describes three attitudes for *acceptance* and *assertion*.

Definition 4.7.1. Three *assertion attitudes* are defined:

- A *confident* agent can assert any proposition for which he can construct an argument,
- a *careful* agent can do so only if he can construct such an argument and cannot construct a stronger argument for the opposite conclusion and
- a *thoughtful* agent can do so only if he can construct a justified argument for the proposition (in terms of the inference relation of the underlying argumentation system).

A thoughtful agent thus only puts forward propositions that are, according to its knowledge, justified. A careful agent moves propositions that are not directly rebutted and a confident agent moves propositions as long as the agent can construct an argument. Parsons et al. show that if a player is thoughtful or careful, the assertions it can make are a subset of those that it could make were it confident. Moreover, the assertions a thoughtful player can make overlap with those it could make were it careful.

In addition to the assertion attitudes as defined above, three acceptance attitudes based on the definition by the authors are described:

Definition 4.7.2. Three *acceptance attitudes* are defined:

- A *credulous* agent accepts a proposition if he can construct an argument for it,
- a *cautious* agent does so only if in addition he cannot construct a stronger argument for the opposite conclusion and
- a *skeptical* agent does so only if he can construct a justified argument for the proposition.

Like for the three assertion attitudes, Parsons et al. show that if a player is skeptical or cautious, the assertions it can accept are a subset of those it could accept were it credulous. Moreover, if a player is skeptical, the assertions it can accept overlap with the set of assertions it could accept were it cautious.

Parsons et al. describe a protocol for persuasion that makes use of these attitudes, here player P is trying to persuade player O of a proposition p:

1. P *asserts* p.

2. O *accepts* p if its acceptance attitude allows, if not O *asserts* $\neg p$ if it is allows by its assertion attitude, or otherwise the player *challenges* p.

3. If O *asserts* $\neg p$, then goto 2 with the roles of the agents reversed and $\neg p$ in place of p.

4. If O has *challenged* then:

   a) P asserts S, the support of p;

   b) Goto 2 for each $s \in S$ in turn.

5. O accepts p if its acceptance attitude allows, or the dialogue terminates.

This protocol specifies the behavior of an agent as a static algorithm where the order of the speech act types is defined as a fixed sequence; *assertion*, *acceptance*, *challenge* before *argue*.

The personality of the agent and its reasoning system allow to have a variable order of speech act types. After all, the personality of the agent specifies the preference over speech act types and the attitude of the agent specifies whether the agent will play a move containing that speech act type or not. Moreover, the attitude selection of the agent is not fixed, but depending on the configuration of the agent's personality vector. This makes the behavior of the agent highly adjustable to the context, since the personality of the agent can be configured to let the agent behave differently. In addition, more information can be included into the reasoning system of the agent. One example of additional information that can be included is knowledge of the personality of the opponent, which will allow the agent to select different attitudes or prefer different speech act types based on the knowledge of its opponent's behavior. This allows the agent to change its strategy based on the personality of its opponent compared to the personality of itself.

The three attitudes that are defined by Parsons et al. for *assertion* and *acceptance* should be extended to allow for better usability in the personality model. In particular, the attitudes are extended such that agents are allowed to bullshit and lie, as defined by Hommes [Hom15]. Accordingly, for a more complete definition of the *acceptance-* and *assertion attitudes*, we extend definitions 4.7.1 and 4.7.2:

Definition 4.7.3. In addition to the *assertion attitudes* of definition 4.7.1, additional assertion attitudes are:

- A *spurious* agent can assert any proposition, regardless if the agent can construct an argument for the claim,
- a *deceptive* agent can assert any proposition for which the agent can construct an argument, in addition, the agent can assert any proposition for which the agent can construct an argument for the contrary and
- a *hesitant* agent cannot assert any proposition.

Definition 4.7.4. In addition to the *acceptance attitudes* of definition 4.7.2, additional acceptance attitudes are:

- A *faithful* agent accepts a proposition regardless if he can construct an argument for it and
- a *rigid* agent never accepts a proposition regardless if he can construct an argument for it.

In addition to *accept* and *claim*, the *why* and *retract* locutions available in Prakken's framework (see table 2.1) are also directed by an agent's personality. More specifically, certain personality traits and more precisely their facets, make an agent prefer to move certain locutions over other locutions. Let us define more types of attitudes, being *challenge attitudes* and *retraction attitudes*.

Definition 4.7.5. Five *challenge attitudes* are defined:

- A *judicial* agent can challenge any proposition for which he cannot construct an argument,
- a *suspicious* agent can challenge any proposition for which he cannot construct a stronger argument for the proposition than for the contrary,
- a *persistent* agent can challenge any proposition for which he cannot construct a justified argument,
- a *tentative* agent can challenge any proposition regardless if he can construct an argument for it and
- an *indifferent* agent cannot challenge any proposition.

Definition 4.7.6. Five *retraction attitudes* are defined:

- A *regretful* agent can retract any own proposition for which he can construct an argument for the contrary,
- a *sensible* agent can retract any own proposition for which he can construct a stronger argument for the contrary,
- a *retentive* agent can retract any own proposition for which he can construct a justified argument for the contrary,
- a *incongruous* agent can retract any own proposition regardless if he can construct an argument for the contrary and
- a *determined* agent can never retract any own proposition.

Let us finally define the *argue attitudes* associated with the *argue* locution.

Definition 4.7.7. Five *argue attitudes* are defined:

- A *hopeful* agent can provide support for any own proposition for which he can construct an argument,
- a *dubious* agent can provide support for any own proposition for which he can construct an argument and no stronger argument for the contrary,
- a *thorough* agent can provide support for any own proposition for which he can construct a justified argument,
- a *misleading* agent can provide support for any own proposition, regardless if he can construct an argument,
- a *fallacious* agent can provide support for any own proposition for which he can construct an argument, in addition, the agent can provide support for any own proposition for which the agent can construct an argument for the contrary and
- a *devious* agent cannot provide support for any own proposition.

4.8 Conclusion

This chapter introduced the fundamentals of the agent's personality by introduction of the *personality vector* and *attitudes*. The personality model of the agent is based on the FFM personality theory, allowing for the specification of the agent's personality in

terms of applying strength values to the particular facets of the FFM that are included in our personality model. For the personality model, two subsets of facets are defined, being the *action selection* personality facets and *action revision* personality facets. Both subsets have their own distinct personality vector that describes the agent's personality configuration in terms of these personality facets. To allow for the agent to reason according to its personality configuration, the agent needs an intermediate concept called an *attitude* that describes under what condition an agent is allowed to move a particular locution. The preference over attitudes is determined according to the agent's *action revision* personality facets, while the preference over the different available locutions is governed by the *action selection* facets. A large body of attitudes was introduced in this chapter. These attitudes, however, allow for modification and the addition or removal of attitudes according to the desired behavior of the agent. In addition, the selection of personality facets that are included in the personality model can be adjusted to either allow for more, or fewer personality facets according to the context of the agent.

5 Adjusting Kok's testbed for persuasion

This chapter will discuss the adjustments that are made to Erik Kok's testbed for argumentative agents. Kok's testbed is tailored towards argumentative agents participating in deliberation dialogues. Since the focus of the current research is persuasion dialogues, some structural changes are required to be made to the testbed to allow for persuasion dialogues.

Kok [Kok13] defines a framework and testbed named *BAIDD* for experimentation with argumentative agents in deliberation dialogues. The framework is based on Prakken's framework for persuasion dialogues but adjusted for deliberation dialogues. A framework named $ASPIC^-$ is used to express and evaluate arguments in the framework of Kok. $ASPIC^-$ is a slightly modified version of $ASPIC^+$ with some omissions for Kok's particular purpose.

Although the personality model is not limited to a particular dialogue system, the current research focuses on persuasion as a dialogue type. Because of this choice of dialogue type, the testbed needs to be modified to support Prakken's liberal dialogue system as was defined in chapter 2.

5.1 $ASPIC^+$ software implementation

Kok uses the $ASPIC^+$-compatible software implementation of South and Vreeswijk [@SV09]. South and Vreeswijk determine facts versus ordinary premises as well as strict and defeasible rules according to an assigned *degree of belief*. This *degree of belief* (DOB) is a real number in the range $(0, 1]$ where a higher degree of belief indicates a stronger inference (and a stronger argument). Kok defines the DOB to be either 0.5 (defeasible) or 1.0 (strict).

The implementation of South and Vreeswijk does not feature all types of attack (undercutting, rebuttal, undermining), but rather implements undermining as a special type of rebuttal. Moreover, it implements undercutting as rebuttal of atomic arguments, which requires every premise that is used in an argument to be actualized in an atomic argument before it can be used in an argument.

Acts	Attacks	Surrenders
propose o	why-propose o reject o	
why-propose o	argue A where $o \in prem(A)$	
reject o		
prefer o, o'		
prefer-equal o, o'		
skip		
inform φ		
argue A	argue B where B defeats A why ψ where $\psi \in prem(A)$	concede φ where $\varphi = conc(A)$ concede ψ where $\psi \in prem(A)$
why φ	argue A where $\varphi = conc(A)$	retract φ
concede φ		
retract φ		

Tab. 5.1: Available speech act types in Kok's deliberation framework

Kok mentions an issue of the South and Vreeswijk implementation of undercutting. This issue arises because every rule in a knowledge base automatically has its rule name added as a fact. This causes an undercutter of a rule to be automatically rebutted by the existence of the rule name as a fact.

5.2 Adjusting the Testbed for Persuasion

Deliberation and persuasion are two different types of dialogue types. The purpose of a dialogue of the first type is to decide on some course of action by providing arguments for a personal goal. The purpose of a dialogue of the second type is to resolve conflicting points of view by providing arguments for persuading the other party. In his framework, Kok supports dialogues of more than two agents, which is common in deliberation dialogues. This research however focuses on dialogues between two agents. Another significant difference in the types of dialogues resides in the structure of the dialogue tree. As arguments are added in a persuasion dialogue, these are each added as a response to an argument that is already present in the dialogue tree (except for the first argument, or root of the tree). For deliberation, multiple such trees exist, since multiple proposals exist for players with different personal goals. Depending on the status of the arguments that are *in* or *out* in the dialogue tree a dialogue outcome is determined by selecting a proposal that is in. Kok, in his deliberation framework implements this structure. For the purpose of this research only one dialogue tree should exist.

Another difference between Kok's framework and the framework as was introduced in chapter 2 is the addition of types of speech acts as can be seen in table 5.1. For

the purpose of this research the speech act types *propose*, *why-propose*, *reject*, *prefer*, *prefer-equal*, *skip* and *inform* remain unused and need to be removed.

Kok defines a *deliberation dialogue context* as.

Definition 5.2.1. A deliberation dialogue context $\mathcal{DK} = \langle AS, L_t, L_c, \mathcal{P}, \mathcal{A}, g_d \rangle$ consists of:

- An $ASPIC^-$ *argumentation system* $AS = (\mathcal{L}, ^-, \mathcal{R}, \leq)$, with $\mathcal{L}$ as the topic language.
- A communication language L_c.
- A protocol $\mathcal{P}$.
- A sequence of agents $\mathcal{A}\langle a_1, \ldots, a_i, \ldots a_n \rangle$.
- A mutual goal $g_d \in L_g$.

For our purpose, this dialogue context needs to be adjusted such that the argumentation system is an $ASPIC^+$ argumentation system. Additionally, the communication language needs to be the language as is described by the speech act types of table 2.1.

In essence, the framework of Kok that is implemented in BAIDD needs to be replaced by the liberal dialogue system of chapter 2. This entails the following adjustments.

- The turn-taking function defined by Kok needs to be replaced with T_L,
- the dialogical status of a move needs to be determined according to definition 2.2.9,
- the protocol needs to be replaced by the protocol for liberal dialogue systems,
- the dialogue outcome needs to be determined according to definition 2.2.12.

5.3 Introducing BAIPD

The required adjustments that were mentioned in the previous section have resulted in a modified version of Kok's BAIDD (BDI Agents Interacting in Deliberation Dialogues) platform tailored to persuasion dialogues, named BAIPD (BDI Agents Interacting in Persuasion Dialogues)[1]. The platform allows for experimenting with

[1]Available from: `https://bitbucket.org/metheredge/baipd`

implementations of software agents for persuasion dialogues. The platform will be used for the implementation of software agents for persuasion dialogues featuring personality by using the personality model that was introduced earlier in this thesis.

The following sections will describe BAIPD and the implementation details of the platform.

5.3.1 Structural Changes

The previous section mentioned the intrinsic differences in the structure of deliberation dialogues and persuasion dialogues. While in deliberation dialogues the participants move proposals that form new proposal trees in the dialogue, a persuasion dialogue only contains one such tree. For persuasion dialogues, the dialogue itself has a topic and the dialogical status is determined according to a claim move that is the first unique move played by the proponent of the dialogue topic. BAIPD removes the proposal structure of the deliberation dialogues and treats the dialogue itself as a tree structure, like proposals in BAIDD are treated as tree structures.

Additionally, persuasion lacks the need for goals and options. In deliberation dialogues, the participants have goals, either private or public, and options, alongside their beliefs. The participants deliberate to select a proposal that suits the participants' goals. This incentive for participants in persuasion dialogues is much simpler, where the participants move to persuade the other, and make the topic of the dialogue either *in* or *out*. For this reason, the goals, both private and public, and options where removed from BAIDD in forming the BAIPD platform.

BAIPD removes the majority of locutions that are present in the BAIDD platform, since most of the moves in the platform (as seen from table 5.1) are tailored to deliberation dialogues. However, the *claim* speech act is missing from BAIDD and has been introduced to the BAIPD platform. The *claim* locution is implemented as a locution that can be played by a participant, and has a proposition as its content.

Where deliberation dialogues make use of proposals moved by the participants, persuasion dialogues start by a proponent moving a move containing the *claim* speech act with the dialogue topic as its content. Moreover, when a dialogue starts in the BAIDD platform, agents are asked to join the dialogue, BAIPD omits this step.

5.3.2 Turn-taking

BAIDD features a turn-taking function that selects the next player from a the list of participants for the deliberation dialogue. In contrast, persuasion introduces

opponents and proponents of a dialogue topic, which results in the turn-taking function not simply returning the next player to move, but alternating between the proponents and opponents of the dialogue topic. The resulting turn-taking function as implemented in the BAIPD platform is sketched by the following algorithm:

Algorithm 1 Turn-taking function

```
1: if currentToMove is not defined then
2:     currentToMove ← nextProponent()
3: else
4:     if currentToMove is proponent then
5:         currentToMove ← nextOpponent()
6:     else
7:         currentToMove ← nextProponent()
8:     end if
9: end if
```

This function allows for the selection of the next player from either the set of opponents or proponents. Internally, the functions returning the next opponents or proponents keep track of the next player, respectively opponent or proponent, which has its turn. Although this research focuses on two player persuasion dialogues, the function allows for dialogues having more than two players.

The turn-taking function assigns a proponent as the first player to move. A test whether a participant is a proponent or opponent is included in the platform by testing whether the agent can generate an argument for the dialogue topic. The first player to move is expected to initially move a unique move containing the *claim* speech act with the dialogue topic as its content.

5.3.3 Outcome Selection

The outcome of the dialogue for a persuasion dialogue is specified as the *claim* move being either *in* or *out*, as was defined in definition 2.2.12. BAIDD features two termination rules, the first terminating the dialogue when no participants are present in the dialogue, while the second terminates the dialogue when every participant has skipped its turn. BAIPD re-uses both termination rules.

For determining the outcome of the dialogue, BAIPD implements the outcome rule for *conflict* resolution after Prakken (2006).

In BAIPD, w_t is defined as the distinct set of players of the set of *active attackers* of the the initial move of the dialogue. Where the list of *active attackers* is defined by a recursive function implemented by Kok in the BAIDD platform:

Algorithm 2 Active attackers function

```
1:  function FILLACTIVEATTACKERS(move, parentIsActiveAttacker)
2:      if move is attacking locution and move is in then
3:          addActiveAttacker(move)
4:          for child ∈ move.children do
5:              fillActiveAttackers(child, true)
6:          end for
7:      else if parentIsActiveAttacker then
8:          for child ∈ move.children do
9:              fillActiveAttackers(child, false)
10:         end for
11:     end if
12: end function
```

This function will return all moves in the dialogue that, in case the input move is *in*, make the move *in*, including the input move itself. In case the input move is *out*, the function will return the set of moves that make the input move *out*. The distinct set of players that played these moves for the initial move will contain every player that made the initial move either *in* or *out*. In case the initial move is *out*, the winners are thus specified as the opponents, while for the initial move being *in*, the winners are specified as the proponents.

Next, the outcome rules can be specified as follows:

Algorithm 3 Conflict resolution outcome rule

```
1: function CONFLICTRESOLUTION(dialogue)
2:     W_t ← d(fillActiveAttackers(dialogue.initialMove), true)
3:     winners ← W_t
4:     topicIsIn ← dialogue.isIn
5: end function
```

The set of winners is defined as W_t. Here, d is a function that returns the distinct set of players for a set of moves.

6 Reasoning

Up to now the personality model has been introduced in terms of the personality theory, interpretation of its personality facets for the argumentative agent and the available agent attitudes. The next step in our introduction of personality into the argumentative agent is an introduction of what will be called the *reasoning system* of the agent. We have seen that an agent's configuration of its personality consists of the specification of strengths for different personality facets that are part of the personality model. Two of these personality vectors exist, one for *action selection* that is used to compute a preference ordering over speech act types and one for *action revision* used to compute a preference ordering over the attitudes that belong to a speech act type. The agent's reasoning system takes care of determining the preference orderings based on the personality of the agent. For determining these preference orderings, the reasoning system takes the personality vectors of the agent as input, while the output of the reasoning system is a set of moves the agent will contribute to the dialogue. The reasoning system is composed of the following components.

Reasoning rules The reasoning system includes a set of *reasoning rules* that determine according to the strengths of personality facets in the corresponding personality vector an output value. This output value specifies for *action selection* a preference value for a speech act type. For *action revision* the output value specifies a preference value for an attitude.

Reasoning algorithm The reasoning algorithm utilizes the preference values resulting from the reasoning rules to generate moves that are played by the agent in a dialogue.

This chapter will describe the reasoning system and its components. Furthermore, a large section will be dedicated to the introduction of the reasoning rules as part of the reasoning system.

6.1 Mamdani Fuzzy Inference System

The reasoning rules that are part of the reasoning system are specified as fuzzy logic rules. A fuzzy inference system is used as the basis of the agent's reasoning

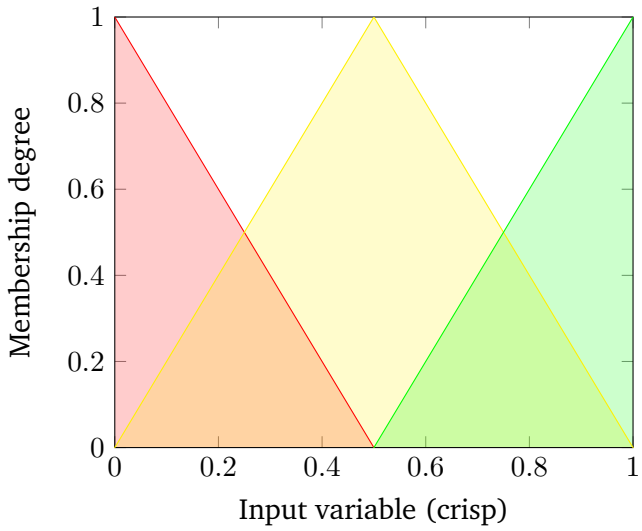


Fig. 6.1: Membership distribution of the three classes belonging to input variables in the reasoning system.

system, since the fuzzy rules that are included in the inference system allow for easy specification of the agent's reasoning process while offering control over the fuzziness of the application of these rules. This in turn allows for easy adjustment and modification of the agent's reasoning process and extension of the agent's personality facets and introduction of attitudes. Moreover, fuzzy inference allows for control over the agent's reasoning process since it allows for tweaking fuzzy membership classes, application of fuzzy logic operators and other aspects like aggregation. Mamdani and Assilian [MA75] describes a fuzzy inference system, which makes use of Zadeh's fuzzy calculus [Zad73]. In contrast with rules in boolean logic, where variables are denoted by the truth values true and false, variables in fuzzy logic denote a membership degree in a *fuzzy class*. These membership degrees are real numbers with values between 0, indicating a negative key, and 1, indicating a positive key. The input to fuzzy logic rules are named *crisp* values. In the reasoning system, three fuzzy classes are used; *low*, *med* and *high* for input variables as specified in the personality vectors. These three fuzzy classes are captured by a membership distribution function that indicates for a crisp input value. The degree of membership in each of the three classes as illustrated in figure 6.1. This process is called *fuzzification*. In this plot, the red shaded region resembles the *low* distribution, the yellow shaded region resembles the *med* distribution and the green shaded region resembles the *high* distribution. These fuzzy classes are chosen to resemble the key of the personality facets of the agent. Three classes are chosen to be able to match the agent's personality facets in three levels. The specification, however, allows for introducing more or even differently distributed fuzzy classes to allow for more fine-grained definition of reasoning rules.

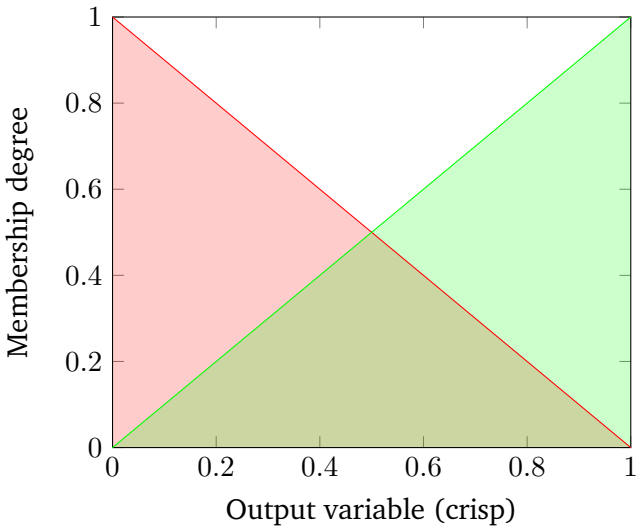


Fig. 6.2: Membership distribution of the two classes belonging to output variables in the reasoning system.

Example 6.1.1. Suppose $f_v \in \mathcal{AS}$ denotes the action selection personality facet *values* with a strength of $\text{Strength}(f_v) = 0.7$. Fuzzifying this crisp value yields the membership degrees

$$[\mu_{\text{low}}(\text{Strength}(f_v)) = 0, \mu_{\text{med}}(\text{Strength}(f_v)) = 0.6, \mu_{\text{high}}(\text{Strength}(f_v)) = 0.4]$$

Where μ denotes the membership function that resembles the membership distribution of figure 6.1.

The *fuzzification* of an input variable such as the strength of the facet *values* in the example above is simply a matter of computing the output value for the membership function provided the input variable. Likewise, two fuzzy classes for output variables are defined, being the *disfavored* and *favored* classes as illustrated in figure 6.2. Here, the red shaded region represents the *disfavored* class, while the green shaded region resembles the *favored* class.

A Mamdani fuzzy inference system consists of the following components:

Fuzzy class A fuzzy class, such as *low*, *med* and *high* in our case subdivides the real input into a fuzzy valuation.

Membership distribution Specifies for a crisp value a fuzzy membership degree in a class. For instance, in the example above *low* is a fuzzy class in the membership distribution over input variables. Membership distributions exist for input and output variables.

Fuzzy rule Specifies a conditional statement of the form

```
if x is a then y is b
```

where the variable x is a linguistic variable with its value coming from the membership distribution function containing a. The variable x is part of the input and makes up the *antecedent* of the rule. Likewise, y is a linguistic variable with its value coming from the membership distribution function containing b. The variable y is part of the output and makes up the *consequent* of the rule. In the reasoning system, fuzzy rules such as the following are included.

```
if ideas is high then challenge is favored
```

Fuzzy rules can specify various operators, including but not limited to; *and*, *or*, *not*, respectively conjunction, disjunction and negation.

Typically, the three operators *and*, *or* and *not* are implemented as $\min(x, y)$, $\max(x, y)$ and $1 - x$, where x and y are fuzzy membership degrees.

The fuzzy inference system uses the set of fuzzy rules using the fuzzy classes to compute an output variable for a given set of inputs. This output variable is computed as a membership degree in a fuzzy class, but can be converted to a crisp value by *defuzzification* of the output variable. Computing the output variable based on a set of input variables and a fuzzy rule is conducted by the following process:

Fuzzification Computing the membership degree in a fuzzy class for each of the input variables.

Rule evaluation Combining multiple input variables to determine the rule strength and computation of the fuzzy membership of the output variable according to the output membership distribution function and the rule strength.

Aggregation Combining multiple output variables to determine the output distribution.

Defuzzification Conversion of the membership degree in a fuzzy class of the output variable to a crisp output.

Figure 6.3 illustrates this process. The figure contains two rules

```
if x is med and y is high then z is disfavored
```

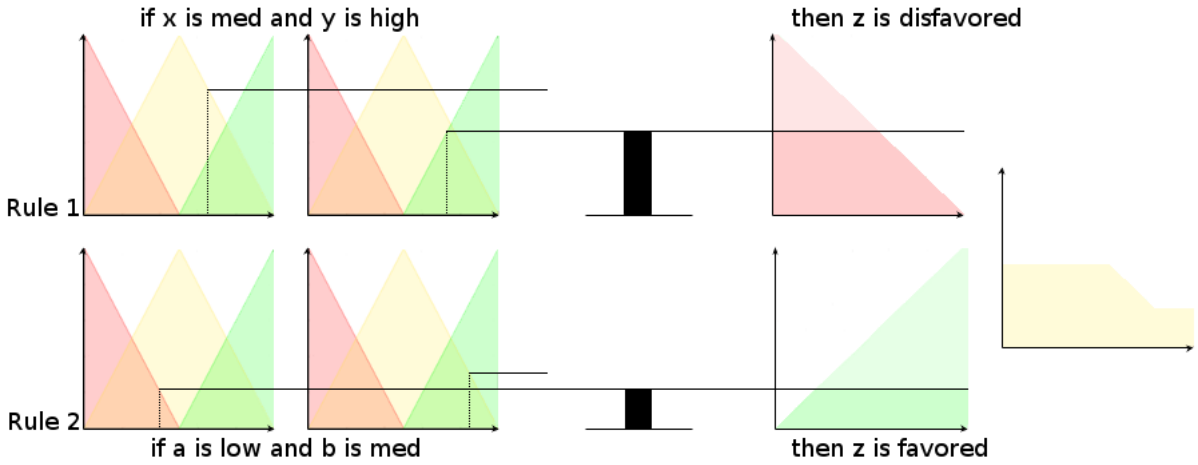


Fig. 6.3: Mamdani FIS for two rules with AND-operators.

```
if a is low and b is med then z is favored
```

that use the operator *and*. First, $x = 0.62$ provides for a fuzzy membership degree of 0.76 in the *med* class, $y = 0.67$ provides for a fuzzy membership degree of 0.34 in the *high* class. Next, the antecedent is used to compute the rule strength by application of the *and* operator, which yields $\min(x, y) = 0.34$. This process is repeated for the second rule. The output of both rules is combined into an output distribution that is defuzzified to compute a crisp output value. Various defuzzification methods exist, among which the center of maximum, the mean of maximum and the center of mass methods are the most popular.

6.2 Reasoning Rules

Fuzzy rules in the reasoning system are used to construct *reasoning rules*. These reasoning rules determine based on a number of input variables, which can be either strengths of *action selection* facets or strengths of *action revision* facets as stored in the personality vectors of the agent, the preference of respectively types of speech acts or attitudes associated with the speech act types. These reasoning rules make use of the fuzzy classes that were introduced last section, where *low*, *med* and *high* denote the fuzzy classes for the input variables and *disfavored* and *favored* denote the fuzzy classes for output variables. These fuzzy classes and variables are used to construct expressions as

```
if ideas is high then challenge is favored
```

which is an *action selection* reasoning rule stating that if the *ideas* facet has a high strength, the speech act type *why* should be favored. Likewise, reasoning rules for *action revision* can be constructed

```
if deliberation is not high then thoughtful is disfavored
```

indicating that in case the *deliberation* facet does not have a high strength, the attitude *thoughtful* should be disfavored. The syntax of reasoning rules can be written in BNF notation as:

```
<reasoning_rule> ::= if <proposition> then <proposition>
<proposition> ::= <disjunction> {and <disjunction>}
<disjunction> ::= <variable> {or <variable>}
<variable> ::= <attribute> is {not} <value>
```

Definition 6.2.1. Reasoning rules have the following properties:

For *action selection* reasoning rules:

- Has strengths of facets contained in the $\mathcal{PV}_{\mathcal{AS}}$ vector as its input variables.
- Has preference values for speech act types as its output variables.

For *action revision* reasoning rules:

- Has strengths of facets contained in the $\mathcal{PV}_{\mathcal{AR}}$ vector as its input variables.
- Has preference values for attitudes as its output variables.
- Each reasoning rule is associated with a speech act type.

Based on the reasoning rules that are included in the reasoning system of the argumentative agent, the agent is allowed to compute preference orderings over available speech act types and their associated attitudes. The agent bases its reasoning on the configuration of its personality, as these properties are mapped to input variables in the fuzzy inference system. Next, the fuzzy inference system is used to compute the fuzzy membership degree in the input classes and subsequent defuzzification of the computed output value as explained in the previous section.

One problem arises when including conjunction and disjunction operators. Recall that these operators are implemented as, respectively $\min(x, y)$ and $\max(x, y)$, where x and y denote membership degrees in fuzzy classes. Due to the nature of these

functions, the conjunction and disjunction operators will strictly require every input variable to match the fuzzy class as written in the reasoning rule.

Example 6.2.1. Suppose the following reasoning rule is included in the agent's reasoning system:

```
if a is high
and b is high
and c is low
and d is med
and e is high
then f is favored
```

Here $a, \ldots, e$ indicates input variables and f indicates an output variable to either *action selection* or *action revision* reasoning rules.

Now suppose the membership degrees of the variables $a, \ldots, e$ are such that:

	low	med	high
a	0.0	0.0	1.0
b	0.0	0.0	1.0
c	1.0	0.0	0.0
d	0.0	1.0	0.0
e	0.1	0.8	0.1

The value for f will be computed as $f = \min(1.0, \min(1.0, \min(1.0, \min(1.0, 0.1))))$ which results in a value of $f = 0.1$. This indicates a membership degree of 0.1 in the *favored* class.

Even though four out of five facets match perfectly with the condition in the reasoning rule, the preference for f is still highly disfavored. The same occurs with disjunction although in an opposite case.

Example 6.2.2. Suppose the following reasoning rule is included in the agent's reasoning system:

```
if a is high
or b is high
or c is low
```

```
or d is med
or e is high
then f is favored
```

Again, $a, \ldots, e$ would indicate input variables and f would indicate an output variable to either *action selection* or *action revision* reasoning rules.

Now suppose the membership degrees of the variables $a, \ldots, e$ are such that:

	low	med	high
a	0.0	0.0	1.0
b	1.0	0.0	0.0
c	0.0	0.0	1.0
d	1.0	0.0	0.0
e	1.0	0.0	0.0

The value for f will be computed as $f = \max(1.0, \max(0.0, \max(0.0, \max(0.0, 0.0))))$ which results in a value of $f = 1.0$. This would indicate a membership degree of 1.0 in the *favored* class.

In this case, although four out of five facets do not match the condition in the reasoning rule at all, the outcome for f is still highly favored. These are fairly extreme cases which could partially be solved by writing different reasoning rules. However, the notion of "conjunctionness" or "disjunctionness" for the conjunction and disjunction operators need to be definable. In other words, the aggregation process that evaluates the antecedent of the reasoning rules requires an aggregation operator that can be modelled to lie in between conjunction and disjunction, where a pure conjunction operator would require all criteria to be satisfied, while a pure disjunction operator would require at least one criteria to be satisfied.

This is where we introduce the ordered weighted aggregation (OWA) operator [Yag88] into the aggregation process. This operator uses a set of weights and an ordered set of input variables to determine the aggregated output. Depending on the weights in the set, the outcome of the operator can shift towards either conjunction or disjunction.

Definition 6.2.2. An ordered weighted aggregation operator of length n is a function $F : R_n \mapsto R$ that has a set of weights $W = \{w_1, \ldots, w_i, \ldots, w_n\}$ with $0 \leq w_i \leq 1$ and $w_1 + \ldots + w_n = 1$. This operator is used to aggregate a set of input variables $\{a_1, \ldots, a_n\}$ as

$$F(a_1, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j$$

where $\{a_1, \ldots, a_n\}$ is an ordered set of input variables.

By manipulation of the set of weights W, the operator can be adjusted. For instance, the conjunction and disjunction operators can be implemented.

Definition 6.2.3. The OWA operator $F(a_1, \ldots, a_n)_\vee$ is the operator such that $w_1 = 1$ and $w_i = 0$ if $i \neq 1$. The OWA operator $F(a_1, \ldots, a_n)_\wedge$ is the operator such that $w_1 = 0$ and $w_i = 1$ if $i \neq n$.

Yager defines the *orness* of the operator as the degree in which the OWA operator resembles a disjunction operator. To determine the weighting vector of an OWA operator based only on the orness ρ of the OWA operator, O'Hagan introduces a maximum entropy ordered weighting aggregation operator (MEOWA) [O'H87]. The resulting operator defines the weighting vector according to the solution of a constraint optimization problem.

The OWA operator is used in the reasoning system to control the strictness of the conjunction and disjunction operators in the reasoning rules. Where conjunction has an orness value of $1 - \rho$.

6.3 Introduction to Reasoning Rules

Now that the basics of fuzzy rules has been discussed, the reasoning rules that, together with the reasoning algorithm, make up the agent's reasoning system can be treated. These reasoning rules are implemented along the interpretation of the personality facets in the personality theory. By making use of the membership classes for input and output as we saw in the previous section, the personality configuration of the agent can be matched according to the strength of individual personality facets in the agent's personality configuration. In addition, the reasoning rules specify under what conditions the selection of attitudes or actions is either favored or disfavored. The reasoning rules that are introduced in this thesis are based on the current interpretation. These rules are, however, flexible and easy to modify and are expected to be adjusted according to what is expected of the agent in terms of its behavior and its context.

As we have seen earlier in this thesis, a distinction is made between *action selection* and *action revision*. The same also applies to reasoning rules. For *action selection*,

reasoning rules are typically more simple than the reasoning rules for *action revision* as the descriptions of the associated personality facets are easier and apply to favoring particular locutions in the argumentation framework. For *action revision*, reasoning rules are typically complex and rely on combinations of personality facets that need to have low, medium or high strength in the personality configuration of the agent. As explained earlier, the reasoning rules in this thesis are contestable depending on the particular agent and its context. The reasoning system that is introduced here should therefore be seen as a system that allows for the specification and implementation of an agent's reasoning process according to its personality and not as an irrefutable and exhaustive description of what the personality of the agent consists of. Reasoning rules having their origin in fuzzy logic supports this, as both the reasoning rules, membership classes and modus operandi of the inference system allow for modification in the agent's reasoning system.

Let us take the personality facet *achievement striving* as an example of how such a personality facet would be used in a reasoning rule. An intuitive description of the interpretation of this facet in the personality model would state that if the agent is achievement striving to some degree, the agent will prefer attitudes that make the agent win an argument. Conversely, if the agent is not achievement striving, the agent will prefer attitudes that make it easier for the opponent to win the dialogue.

Let us consider the effects of *achievement striving* on the attitude preference for *acceptance* as follows.

Acceptance	
High	Skeptical, Rigid
Med	Cautious
Low	Credulous, Faithful

This table should be read as "if achievement striving is high then skeptical is favored". When combining this with other personality facets, for instance *straightforwardness* as follows,

Acceptance	
High	Credulous, Cautious, Skeptical
Med	Faithful
Low	Rigid

we can start to see the personality facets that need to be combined to make up the reasoning rules for the various locution that are available in the argumentation framework.

6.4 Reasoning Rules for Action Selection

The first set of reasoning rules included in the reasoning system is the set of reasoning rules for *action selection*. As we have seen before, these rules make use of the strength values in $\mathcal{PV}_{\mathcal{AS}}$ and are used to determine a preference ordering over locutions available in the argumentation framework.

R_1 `if actions is high`
`or selfconsciousness is high`
`then acceptance is favored`

R_2 `if ideas is high`
`then challenge is favored`

R_3 `if values is high`
`then retraction is favored`

R_4 `if competence is high`
`then retraction is disfavored`

R_5 `if competence is high`
`then acceptance is disfavored`

R_6 `if selfconsciousness is high`
`then assertion is disfavored`

R_7 `if assertiveness is high`
`then assertion is favored`

These rules closely resemble the interpretation of the personality facets in chapter 4.

6.5 Reasoning Rules for Action Revision

The following sections will introduce the reasoning rules features in the implementation of the agent's reasoning system. Like the reasoning rules for *action selection*, these reasoning rules are based on the personality model as introduced in chapter 4.

6.5.1 Reasoning Rules for Assertion

A set of twelve reasoning rules is included in the reasoning process related to the *assert* locution. Recall that the output variables of these reasoning rules are associated with a preference value for the assertion attitudes *thoughtful*, *careful*, *confident*, *spurious*, *deceptive* and *hesitant* as introduced in chapter 4. The first three attitudes respectively require the agent to be able to construct an argument, construct an argument and not a stronger argument for the contrary and construct a justified argument before being allowed to make a claim. The *spurious* attitude does not require the agent to be able to construct an argument before making a claim. The *deceptive* attitude allows the agent to make a claim for a proposition that the agent knows is untrue. The *hesitant* attitude does not allow the agent to make a claim at all. The following twelve reasoning rules will describe the effect of the personality configuration on the agent's preference ordering.

R_1

```
if achievementstriving is high
  and selfdiscipline is high
  and straightforwardness is high
  and modesty is low
  and anxiety is low
  and activity is high
  and deliberation is high
  then thoughtful is favored
```

R_2

```
if achievementstriving is high
  and selfdiscipline is high
  and straightforwardness is high
  and modesty is low
  and anxiety is low
  and activity is high
  and deliberation is med
  then careful is favored
```

R_3

```
if achievementstriving is high
  and selfdiscipline is high
  and straightforwardness is high
  and modesty is low
  and anxiety is low
  and activity is high
  and deliberation is low
  then confident is favored
```

The personality facet *achievement striving* relates to the agent's preference for an attitude towards types of speech acts that either make the agent support or claim some proposition that is in line with the agent's personal goal. The *thoughtful*, *careful* and *confident* attitudes will help the agent to achieve its personal goal by making claims. Therefore a high value for this personality facet indicates a preference for these attitudes. *Self-discipline* makes agents prefer attitudes towards types of speech acts that require the agent to add to lines of dispute if the agent is able to. The interpretation in the personality model of this personality facet indicates that a high value for this personality facet makes the agent prefer the attitudes *thoughtful*, *careful* and *confident*. By allowing to make claims, although under certain conditions, these attitudes allow the agent to add to lines of dispute. The personality facet *straightforwardness* deals with the agent's preference to stick to the rules, displaying proper behavior and being direct in its communication. Since the attitudes *thoughtful*, *careful* and *confident* require the agent to be able to construct an argument before being allowed to make a claim for a certain proposition, the agent is required to be straightforward. *Modesty* in the personality model is interpreted as the agent disfavoring making claims. For the preference of the three attitudes *thoughtful*, *careful* and *confident*, it is required that this personality facet is low. *Anxiety* in the personality model disallows the agent to make claims, requiring this facet to be low for the attitudes *thoughtful*, *careful* and *confident*. Since the agent is active by making claims, the agent having high activity indicates a preference for the attitudes *thoughtful*, *careful* and *confident*. The facet *deliberation* distinguishes the three attitudes *thoughtful*, *careful* and *confident*. This facet is interpreted as preferring well-motivated moves in the personality model. Since the three mentioned attitudes respectively require well- to non-motivated moves, the facet should be high for preferring the *thoughtful* attitude, medium for preferring the *careful* attitude and low for the *confident* attitude. The definition of a move being *well-motivated* was given in chapter 4.

R_4

```
if deliberation is not high
```

```
  then thoughtful is disfavored
```

In case deliberation is not high, the *thoughtful* attitude is disfavored and the other attitudes *careful* and *confident* are favored more under the conditions of rules R_1, R_2 and R_3.

R_5
```
if deliberation is not med
  then careful is disfavored
```

R_6
```
if deliberation is not low
  then confident is disfavored
```

The same holds for deliberation being anything other than medium for the *careful* attitude and deliberation being not low for the *confident* attitude. Following the same reasoning of the description of rules R_1, R_2 and R_3.

R_7
```
if deliberation is low
  and straightforwardness is low
  and selfdiscipline is low
  and achievementstriving is low
  and activity is low
  and modesty is high
  and anxiety is high
  then hesitant is favored
```

We have seen the interpretation of *deliberation*. Since the *hesitant* attitude disallows the agent to make claims, low deliberation makes the agent prefer a *hesitant* attitude. Low *straightforwardness* is a requirement for a preference for the *hesitant* attitude, since this attitude indicates the agent preferring being indirect in its communication. By selecting a *hesitant* attitude, the agent is not allowed to make claims, resulting in not being able to add to lines of dispute. Therefore, a low value for *self-discipline* indicates a preference for the *hesitant* attitude. In addition, by not making claims, the agent shows low *activity* and the agent shows low *achievement striving*. A high value for *modesty* indicates a preference to not make claims in the interpretation of the facets as part of the personality model. Therefore, high *modesty* indicates a preference for the *hesitant* attitude. Lastly, *anxiety* contributes to the preference for the *hesitant* attitude, since the agent is not allowed to make claims when it is anxious.

R_8

```
if activity is not low
  and selfdiscipline is not low
  and achievementstriving is not low
  then hesitant is disfavored
```

Contrariwise, a strength that is not low for *straightforwardness* in the personality configuration of the agent indicates a preference for another attitude than *hesitant*. This inverse relation also holds for *self-discipline* and *activity*.

R_9

```
if straightforwardness is low
  and deliberation is low
  and selfdiscipline is not low
  and achievementstriving is high
  then spurious is favored
```

The *spurious* attitude allows the agent to make a claim when the agent has no supporting argument. This allows the agent to bullshit by making claims while the agent has no support for some proposition. If the agent is *achievement striving*, the agent is more likely to prefer this attitude, since this can be a strategy for the agent to achieve its personal goal. Selecting this attitude allows the agent, however, to make non-motivated moves, indicating that a low value for *deliberation* in the agent's personality configuration should prefer this attitude. The agent is active by making claims. Since the attitude allows the agent to display improper behavior by bullshitting, low *straightforwardness* is necessary for preferring this attitude. Since the agent makes claims, the agent should not be immodest or anxious for preferring this attitude. Although the agent's contribution to a line of dispute is improper, it does add to a line of dispute, resulting in a value that is not low for *self-discipline* preferring the *spurious* attitude.

R_{10}

```
if deliberation is not low
 then spurious is disfavored
```

Deliberation differentiates between the *spurious* and *deceptive* attitudes. While the first attitude allows the agent to not be deliberate at all to make a claim, the second attitude allows the agent to lie about some claim. Since the *deceptive* attitude requires to make a deliberate lie by being able to provide an argument for the contrary, the facet *deliberation* for this attitude cannot be low.

R_{11}

```
if straightforwardness is low
```

```
and deliberation is not low
and selfdiscipline is not low
and achievementstriving is high
and angryhostility is high
then deceptive is favored
```

In addition to the differentiating facet *deliberation*, *angry hostility* can be a factor in preferring the *deceptive* attitude. An angry agent can explain obstruction of the his opponent by claiming the opposite of some proposition by its opponent.

R_{12}
```
if deliberation is low
then deceptive is disfavored
```

6.5.2 Reasoning Rules for Acceptance

In addition to the reasoning rule for assertion as defined last section, ten reasoning rules for acceptance are to be defined. These reasoning rules determine the agent's preference ordering over the acceptance attitudes *credulous*, *cautious*, *skeptical*, *faithful* and *rigid*. These attitudes respectively allow the agent to accept a claim by its opponent if the agent can construct an argument for the claimed proposition, cannot construct a stronger counterargument for the proposition and can construct a justified argument for the claimed proposition. The *faithful* acceptance attitude allows the agent to accept a claim regardless if the agent can construct an argument for the proposition. Lastly, the *rigid* attitude disallows the agent to accept a claim by its opponent.

R_1
```
if achievementstriving is low
  and deliberation is low
  and trust is high
  and modesty is not low
  and activity is high
  then credulous is favored
```

R_2
```
if achievementstriving is med
  and deliberation is med
  and trust is med
  and modesty is not low
  and activity is high
```

```
  then cautious is favored
```

R_3

```
if achievementstriving is high
  and deliberation is high
  and trust is low
  and modesty is not low
  and activity is high
  then skeptical is favored
```

Achievement striving plays a role in determining the agent's preference for the *credulous*, *cautious* and *skeptical* attitudes. In case the value for this facet is low in the personality configuration of the agent, the agent will accept a claim by its opponent more easily, requiring that the agent is less achievement striving. The *cautious* attitude requires the agent to be more critical when accepting a claim by its opponent. For this reason, the agent will be required to be more achievement striving. Lastly, for the agent to prefer the *skeptical* attitude, the agent will be required to be able to construct a strong argument for the proposition as claimed by its opponent. Recall that the *achievement striving* facet is interpreted in the personality model as having the agent prefer to strive to achieve its personal goal. By easily accepting a claim by its opponent, the agent does not strive to achieve its personal goal, but rather contributes to its opponent's goal. For the same reason the facets *deliberation* and *trust* are expected to have a value in a range from low to high respectively for the *credulous*, *cautious* and *skeptical* attitudes. For the agent to prefer each of these three attitudes, the agent is expected to be modest to some extent, since the interpretation of the facet *modesty* in the personality model describes that modesty increases the agent's preference towards acceptation. By moving a *concede* move, the agent is active. Trust differentiates between the three attitudes, as a *credulous* attitude will require the agent to be more trusting, since the agent is more easily accepting a proposition. While for a *skeptical* attitude, the agent is expected to be less trusting since the agent is less easily accepting a proposition.

R_4

```
if achievementstriving is not low
  and deliberation is not low
  and trust is not high
  then credulous is disfavored
```

In case the value for the facet *achievement striving* is anything other than low, the agent should rather prefer a difference attitude than the *credulous* attitude following the reasoning of rules R_1, R_2 and R_3. The same, albeit with other values, holds for the facets *deliberation* and *trust*.

R_5
```
if achievementstriving is not med
  and deliberation is not med
  and trust is not med
  then cautious is disfavored
```

R_6
```
if achievementstriving is not high
  and deliberation is not high
  and trust is not low
  then skeptical is disfavored
```

Similar to rule R_4, the values for *achievement striving*, *deliberation* and *trust* should be suited for the attitudes *cautious* and *skeptical*. In case of a mismatch, the agent should prefer a different attitude.

R_7
```
if achievementstriving is low
  and activity is high
  and trust is high
  and modesty is high
  and anxiety is high
  then faithful is favored
```

Recall that the *faithful* attitude allows the agent to accept a claim by its opponent, regardless of its knowledge about the associated proposition. For the agent to prefer the attitude, the agent should trust its opponent, as indicated by a high value for the facet *trust*. An additional reason for the agent to prefer this attitude is being anxious, as indicated by a high value for the facet *anxiety*. *Modesty* determines if the agent accepts a claim by its opponent. Since this attitude allows the agent to accept a claim without a supporting argument, the agent effortlessly accepts an opponent's claim. Therefore, a high value for the facet *modesty* indicates a preference for the *faithful* attitude. As we have seen before, the agent is active by moving an accept move. Since the agent effortlessly accepts the claim by its opponent, the agent again does not contribute towards achieving its personal goal.

R_8
```
if achievementstriving is not low
```

```
and trust is not high
and modesty is not high
then faithful is disfavored
```

In case the personality configuration does not meet the above conditions, the agent should rather prefer a different attitude.

R_9

```
if achievementstriving is high
 and activity is low
 and trust is low
 and straightforwardness is low
 and modesty is low
 and angryhostility is high
 then rigid is favored
```

By selecting the *rigid* attitude, the agent is not allowed to accept the claim of its opponent at all. By not accepting at all, the agent ultimately obstructs its opponent. It is, however, required for the agent to be non-active. Another reason for the agent to not accept claims by its opponent is when the value for *trust* in the personality configuration of the agent is low. In this case the agent refuses to accept, since it is not trusting. By not accepting, the agent is non-straightforward, a low value for *straightforwardness* thus indicates a preference for the *rigid* attitude. Similar to the reasoning rules defined earlier, *modesty* in this reasoning rule should have a low value to prefer the *rigid* attitude since not accepting is preferred for modest agents according to the interpretation in the personality model. Since the agent does not move, the *active* facet is expected to have a low value. Finally, having a high value for the *angry hostility* facet can be a reason for the agent to select the *rigid* attitude, since this can be a reason for the agent to obstruct its opponent.

R_{10}

```
if achievementstriving is not high
 and trust is not low
 and straightforwardness is not low
 and modesty is not low
 then rigid is disfavored
```

Similar to reasoning rule R_9, in case the personality configuration does not match with the above defined reasoning rule, let the agent prefer a different attitude.

6.5.3 Reasoning Rules for Challenge

In addition to the reasoning rules we have seen in the last few sections, reasoning rules for challenge need to be defined. These reasoning rules help determine the agent's preference over the attitudes *judicial*, *suspicious*, *persistent*, *tentative* and *indifferent*, associated with the *why* speech act type. The first three attitudes respectively allow the agent to challenge an opponent's claim under the condition that the agent cannot construct an argument, cannot construct a stronger argument for the proposition than for the contrary and cannot construct a justified argument for the proposition. The *tentative* attitude allows the agent to challenge some claim, regardless if the agent can construct an argument for the proposition. Lastly, the *indifferent* attitude never allows the agent to challenge a claim by its opponent. Ten reasoning rules for challenge based on the personality model will now be introduced.

R_1

```
if achievementstriving is high
  and deliberation is low
  and activity is high
  and trust is high
  and modesty low
  and anxiety is low
  then judicial is favored
```

R_2

```
if achievementstriving is high
  and deliberation is med
  and activity is high
  and trust is med
  and modesty is low
  and anxiety is low
  then suspicious is favored
```

R_3

```
if achievementstriving is high
  and deliberation is high
  and activity is high
  and trust is low
  and modesty is low
  and anxiety is low
  then persistent is favored
```

These first three reasoning rules determine a preference for the three attitudes *judicial*, *suspicious* and *persistent*. A requirement for preferring each of these attitudes is a high value for *achievement striving*, since by challenging an opponent's claim, the agent strives the achieve its personal goal by attacking the opponent's move. *Deliberation* is matched on a range from a low value to a high value, since each of the three attitudes respectively requires better motivation of the opponent, where for the *judicial* attitude, the agent is allowed to provide for a supporting argument without any further requirements, whilst the *persistent* attitude requires the agent to provide for a justified supporting argument. When moving a *why* move, the agent is active, therefore a high value indicates a preference for each of the three attitudes. Similar to the *deliberation* facet, the *trust* facet is matched on a range from a high value to a low value for the three attitudes *judicial*, *suspicious* and *persistent* respectively. The *judicial* attitude allows for a less trusting agent, since the requirements for the agent to not challenge are less than those of not challenging in case of the *persistent* attitude. The *suspicious* attitude lies in between the two mentioned attitudes. For each of the three attitudes it is required for the agent to be immodest, since the interpretation of the facet *modesty* in the personality model specifies that modesty indicates an agent's preference towards not challenging. In case the agent is anxious, the agent is less likely to challenge an opponent's claim, therefore this facet is matched at a low value in the agent's personality configuration.

R_4

```
if deliberation is not low
  and trust is not high
  then judicial is disfavored
```

In case the personality configuration does not match the values for *deliberation* and *trust*, the *judicial* attitude should not be preferred.

R_5

```
if deliberation is not med
  and trust is not med
  then suspicious is disfavored
```

R_6

```
if deliberation is not high
  and trust is not low
  then persistent is disfavored
```

Similar to reasoning rule R_4, in case the values for *deliberation* and *trust* do not match for the attitudes *suspicious* and *persistent*, the agent should rather prefer a different attitude.

R_7

```
if achievementstriving is high
  and deliberation is low
  and activity is high
  and straightforwardness is low
  and modesty is low
  and anxiety is low
  and angryhostility is high
  then tentative is favored
```

Recall that the *tentative* attitude allows the agent to challenge a proposition regardless of its knowledge. Since this attitude contributes to the agent achieving its personal goal by attacking the claim of an opponent, the *achievement striving* facet is expected to have a high value to prefer this attitude. In contrast with the first three attitudes, this attitude indicates low *deliberation*. The agent is not at all required to provide for support for the challenge move, indicating that the *deliberation* facet should have a low value for the agent to prefer this attitude. By moving a *why* move, the agent is active, resulting in a high value for *activity* indicating that the agent prefers this attitude. Since the agent is allowed to challenge the claim, even though the agent knows that the claim by the agent is justified. According to the specification of the *straightforwardness* facet in the personality model, the facet should have a low value for the agent to prefer this attitude. Similar to reasoning rules R_1, R_2 and R_3, the facets *modesty* and *anxiety* should have a low value. In contrast with these first three rules, being angry can be an additional reason for the agent to prefer this attitude, since the agent can obstruct its opponent by verbosely moving challenge moves, even though the agent knows the opponent's claims are justified.

R_8

```
if achievementstriving is not high
  and deliberation is not low
  and straightforwardness is not low
  then tentative is disfavored
```

R_9

```
if achievementstriving is low
  and deliberation is low
  and activity is low
```

```
and straightforwardness is low
and anxiety is high
and angryhostility is high
then indifferent is favored
```

The *indifferent* attitude disallows the agent to make challenge moves. Since this attitude does not contribute to the agent achieving its personal goals, the agent is expected to have a low value for *achievement striving* to prefer this attitude. In addition, since the agent is disallowed to make a challenge move, regardless if the agent can provide for a supporting argument, the agent is expected to be non-deliberate. By not making challenge moves, the agent is not showing activity. For this reason the *activity* facet is expected to have a low value in the personality configuration of the agent. Similar to the motivation for reasoning rule R_7 the facet *straightforwardness* is expected to be low. The interpretation of the facet *anxiety* in the personality model indicates that a high value for this facet indicates the agent to disfavor making challenge moves. For this reason, the value for this facet is expected to be high for the agent to prefer the *indifferent* attitude. Similar to R_7, the *indifferent* attitude allows the agent to obstruct its opponent. Being angry may be a motivation for the agent to obstruct its opponent.

R_{10}

```
if achievementstriving is not low
 and deliberation is not low
 and straightforwardness is not low
 then indifferent is disfavored
```

6.5.4 Reasoning Rules for Retraction

The next set of reasoning rules that will be defined are the reasoning rules for retraction, associated with the *retract* speech act. Recall that the personality model introduced the five attitudes *regretful*, *sensible*, *retentive*, *incongruous* and *determined*. The first three attitudes respectively allow the agent to retract its claim whenever the agent can construct an argument for the contrary of the claimed proposition, construct a stronger argument for the contrary of the claimed proposition and can construct a justified argument for the contrary of the claimed proposition. The *incongruous* attitude allows the agent to retract its claim regardless if the agent can construct an argument for the contrary of the claimed proposition. Lastly, the *determined* attitude disallows the agent to retract its own proposition.

R_1

```
if achievementstriving is low
```

```
  and deliberation is low
  and activity is high
  and modesty is not low
  then regretful is favored
```

R_2
```
if achievementstriving is med
  and deliberation is med
  and activity is high
  and modesty is not low
  then sensible is favored
```

R_3
```
if achievementstriving is high
  and deliberation is high
  and activity is high
  and modesty is not low
  then retentive is favored
```

The first three attitudes *regretful*, *sensible* and *retentive* constrain the argument the agent must be able to construct before the agent is allowed to retract its own claim. These three attitudes are increasingly more strict in terms of these constraints. For this reason the facets *achievement striving* and *deliberation* are matched from low to high for each of the three attitudes. By more easily retracting its own claims, the *achievement striving* facet is expected to be low to prefer the *regretful* attitude, while the *retentive* attitude requires a highly *achievement striving* agent. The *sensible* attitude lies in between, since the *sensible* and *retentive* attitudes require the agent to be capable of providing for more deliberate arguments for the contrary of the claimed proposition than the *regretful* attitude requires. For this reason, the *retentive* attitude is matched as high, the *sensible* attitude is matches as medium and the *regretful* attitude is matched as low for the *deliberation* facet. For the agent to retract its own claim, it is expected that the agent is not immodest. Moreover, by retracting its claim the agent is active.

R_4
```
if achievementstriving is not low
  and deliberation is not low
  then regretful is disfavored
```

R_5
```
if achievementstriving is not med
```

```
  and deliberation is not med
  then sensible is disfavored
```

R_6
```
if achievementstriving is not high
  and deliberation is not high
  then retentive is disfavored
```

Like we have seen with earlier definitions of reasoning rules, in case the personality configuration does not match the expected values, prefer a different attitude.

R_7
```
if achievementstriving is low
  and deliberation is low
  and activity is high
  and modesty is high
  and anxiety is high
  and depression is high
  then incongruous is favored
```

The *incongruous* attitude allows an agent to retract its claim, even if the agent cannot construct an argument for the contrary of a claimed proposition. Since this allows for the agent to retract, while the agent has no reason for doing so, the agent is expected to have a low value for the facet *achievement striving*. Again, since the agent's motivation for retraction may be absent, the agent is expected to be not deliberative. By making *retract* moves, the agent is active, so a high value for the facet *activity* is expected. The description of the facet *modesty* in the personality configuration indicates a preference for retracting own claims. For this reason, *modesty* is expected to be high. In addition, a motivation for the agent to unfoundedly retract its claims may be that the agent is anxious. For this reason, a high value for the facet *anxiety* matches a preference for this attitude. Lastly, according to the description of *depression* in the personality model, a high value for this facet indicates a preference for the *incongruous* attitude.

R_8
```
if achievementstriving is not low
  and deliberation is not low
  and modesty is not high
  then incongruous is disfavored
```

R_9

```
if achievementstriving is high
  and deliberation is low
  and activity is low
  and straightforwardness is low
  and modesty is low
  and angryhostility is high
  then determined is favored
```

The *determined* attitude disallows the agent to move a *retract* move. An agent could be motivated to never retract its claims due to being achievement striving. In this case, the agent never retracts at the expense of its opponent. For this attitude to be preferred, it is expected that the *deliberation* facet in the personality configuration has a low value, since the agent does not require any support for not making a *retract* move. In addition, the agent is not active by not moving *retract* moves, resulting in an expected low value for the *activity* facet. By never retracting, the agent is non-straightforward and immodest, since it is never going to retract its claims. A reason for preferring this attitude may be due to the agent being angry, in which case the agent will obstruct its opponent by never retracting its claims.

R_{10}

```
if achievementstriving is not high
 and deliberation is not low
 and modesty is not low
 then determined is disfavored
```

6.6 Reasoning Rules for Argumentation

The last set of reasoning rules that will be defined are the reasoning rules for argumentation, related to the *argue* speech act type. Recall from chapter 4 that the set of argumentation attitudes consists of the attitudes *hopeful*, *dubious*, *thorough*, *misleading*, *fallacious* and *devious*. The first three attitudes are related to the level of deliberation in the supporting argument for moving a *argue* move. The *misleading* attitude allows the agent to move an *argue* move while the agent does not have a supporting argument for the move. In contrast, the *fallacious* attitude allows the agent to move an *argue* move while the agent knows that the proposition is false. Lastly, the *devious* attitude disallows the agent to move an *argue* move at all.

R_1

```
if achievementstriving is high
  and selfdiscipline is high
```

```
  and deliberation is low
  and activity is high
  and straightforwardness is high
  and anxiety is low
  then hopeful is favored
```

R_2
```
if achievementstriving is high
  and selfdiscipline is high
  and deliberation is med
  and activity is high
  and straightforwardness is high
  and anxiety is low
  then dubious is favored
```

R_3
```
if achievementstriving is high
  and selfdiscipline is high
  and deliberation is high
  and activity is high
  and straightforwardness is high
  and anxiety is low
  then thorough is favored
```

The facet *achievement striving* for the first three attitudes should be high, since these attitudes make the agent move *argue* moves that support the agent in achieving its goal by defending itself. The agent in addition is expected to be *self-disciplined*, *straightforward*, not *anxious* and *active*, since, respectively, the agent is adding to lines of dispute by moving *argue* moves, by selecting these attitudes is direct and sticks to the rules and contributes to the dialogue. The facet *deliberation* is the differentiating factor between these three attitudes, since the agent will prefer increasingly well-motivated moves from preferring a *hopeful* to preferring a *thorough* attitude.

R_4
```
if deliberation is not low
  then hopeful is disfavored
```

R_5
```
if deliberation is not med
  then dubious is disfavored
```

R_6
```
if deliberation is not high
 then thorough is disfavored
```

Rules R_4, R_5 and R_6 support this, by indicating that a different strength value for *deliberation* indicates that another attitude is more preferred.

R_7
```
if achievementstriving is high
 and selfdiscipline is med
 and deliberation is low
 and activity is high
 and straightforwardness is low
 then misleading is favored
```

R_8
```
if deliberation is not low
 and straightforwardness is not low
 then misleading is disfavored
```

R_9
```
if achievementstriving is high
 and selfdiscipline is med
 and deliberation is not low
 and activity is high
 and straightforwardness is low
 and angryhostility is high
 then fallacious is favored
```

R_{10}
```
if deliberation is low
 then fallacious is disfavored
```

For rules R_7 to R_9, the agent is expected to be *achievement striving* since the agent by selecting one of two attitudes *misleading* or *fallacious* is allowed to contribute a move to the dialogue that allows the agent to achieve its personal goal by defending its proposition. Since the agent is allowed to respectively bullshit and lie using these attitudes, the facet *self-discipline* is mildly associated with these attitudes. *Deliberation* is a differentiating facet between these two attitudes. For the first attitude, the agent is expected to be not *deliberative* since the agent is allowed to move an *argue* move that is entirely unsupportable by the agent. When selecting the *fallacious* attitude, the agent must be able to provide an argument, either for the proposition or for the contrary, which

allows the agent to be deliberative. For both attitudes, the agent is expected to not be *straightforward*, since the agent is allowed to not stick to the rules. By moving an *argue* move, the agent is active. For the *fallacious* attitude, an additional reason for selecting this attitude might be for the agent to be angry and obstruct its opponent.

R_{11}

```
if achievementstriving is low
 and selfdiscipline is low
 and deliberation is low
 and activity is low
 and anxiety is high
 and modesty is high
 then devious is favored
```

R_{12}

```
if selfdiscipline is not low
 and deliberation is not low
 and activity is not low
 then devious is disfavored
```

The *devious* attitude disallows the agent to support its proposition. Therefore, the agent is expected to not be *achievement striving*, since the agent does not defend itself. Its *self-discipline* is expected to be low, since the agent does not add to lines of dispute and its *deliberation* is low since it never moves well-motivated moves. In addition, the agent is expected to be *inactive*, as the agent is not contributing moves to the dialogue. Two facets that might explain the agent's preference for the *devious* attitude are the agent being *anxious* or *modest*.

6.7 Reasoning Algorithm

The last sections introduced and explained the reasoning rules that are included in the agent's reasoning system. For the agent to reason according to these reasoning rules, the reasoning system includes what will be called the *reasoning algorithm* of the agent. This reasoning algorithm uses the reasoning rules and the fuzzy inference system to compute preference orderings over locutions available in the argumentation framework in case of *action selection* and preference orderings over attitudes in case of *action revision*. These preference orderings combined with an algorithm that uses these preference orderings to generate moves for the agent allows the agent to reason based on its personality configuration.

The reasoning algorithm makes use of *reasoning engines* which for *action revision* exist for every type of locution available in the argumentation framework, as well as one for *action selection*. Thus, for every locution in the framework, one corresponding reasoning engine exists that contains the reasoning rules for that locution. In addition, one reasoning engine exists for *action selection* that contains the corresponding reasoning rules. A reasoning engine takes a personality vector, $\mathcal{PV}_{\mathcal{AR}}$ in case of *action revision* and $\mathcal{PV}_{\mathcal{AS}}$ in case of *action selection* and maps the strengths of facets in the personality vectors to input variables in the reasoning rules. The fuzzy inference system is then used to compute a preference ordering over attitudes in case of *action revision* and locutions in case of *action selection*.

These reasoning engines are used to compute preference orderings in the reasoning algorithm. After computing these orderings, these are used to let the agent generate moves and contribute generated moves to the dialogue. The attitudes are used to determine whether an agent is allowed to generate a move with a locution at a point in the dialogue as illustrated by algorithm 4.

Algorithm 4 Action revision

```
1:  function ACTIONREVISION(O)
2:      i ← 0
3:      M ← ∅
4:      U ← ∅
5:      for l ∈ O do
6:          if ¬(l ∈ U) then
7:              a ← getAttitude(l, i)
8:              if allowedBy(a) then
9:                  M ← M ∪ generateMoves(a)
10:                 U ← U ∪ {l}
11:             end if
12:         end if
13:         i ← i + 1
14:     end for
15:     removeDoubleTargets(M)
16:     return M
17: end function
```

This algorithm takes an ordering of locutions O and returns a set of generated moves M. For every locution in the ordering the ith attitude is looked up, tested is whether a move with this locution is allowed by the attitude. If so, the attitude is used to generate a set of moves which is added to the overall set of moves generated by the algorithm. A test is included that ensures that every locution distinctly generates moves. The function *removeDoubleTargets* makes sure that no two generated moves have the same target.

The function *generateMoves* uses the reasoning engine associated with the locution to generate a preference ordering over attitudes belonging to the locution. Note that in algorithm 4 *generateMoves* with a single parameter, the attitude, wraps an underlying *generateMoves* function specific to the attitude.

The function *generateMoves* is implemented for each of the attitudes and returns a set of moves allowed for the particular attitude. An example of an implementation can be given for an attitude like the *thoughtful* attitude as illustrated in the next algorithm:

Algorithm 5 Generate moves function for the *thoughtful* attitude

```
1:  function GENERATEMOVES(dialogue, agent)
2:      newMoves ← ∅
3:      attackers ← getActiveAttackers(dialogue)
4:      for attacker ∈ attackers do
5:          newArgument ← ∅
6:          if attacker is why move then
7:              newArgument ← arg(beliefs(agent), premise(attacker))
8:          else if attacker is claim move then
9:              newArgument ← arg(beliefs(agent), ¬proposition(attacker))
10:         end if
11:         if newArgument ≠ ∅ then
12:             if isJustified(newArgument) then
13:                 newMoves ← newMoves ∪ {claimMove(newArgument)}
14:             end if
15:         end if
16:     end for
17:     return newMoves
18: end function
```

The functions *beliefs* which returns the knowledge base of the agent, *premise* which returns the attacked premise, *proposition* which returns the claimed proposition, *arg* which generates an argument if possible and *claimMove* which creates a *claim* move for a given argument are based on functions present in the BAIPD framework, based on Erik Kok's framework functions.

6.8 Implementation

BAIPD contains an implementation of the argumentative agent with personality. This agent has a personality configuration which is subdivided into action section and action revision personality facets. The BAIPD framework is responsible for hosting the persuasion dialogue, which is governed by a persuasion platform after the deliberation platform in the BAIDD framework of Kok [Kok13]. The persuasion platform makes sure that turn taking, adding generated moves and additional things

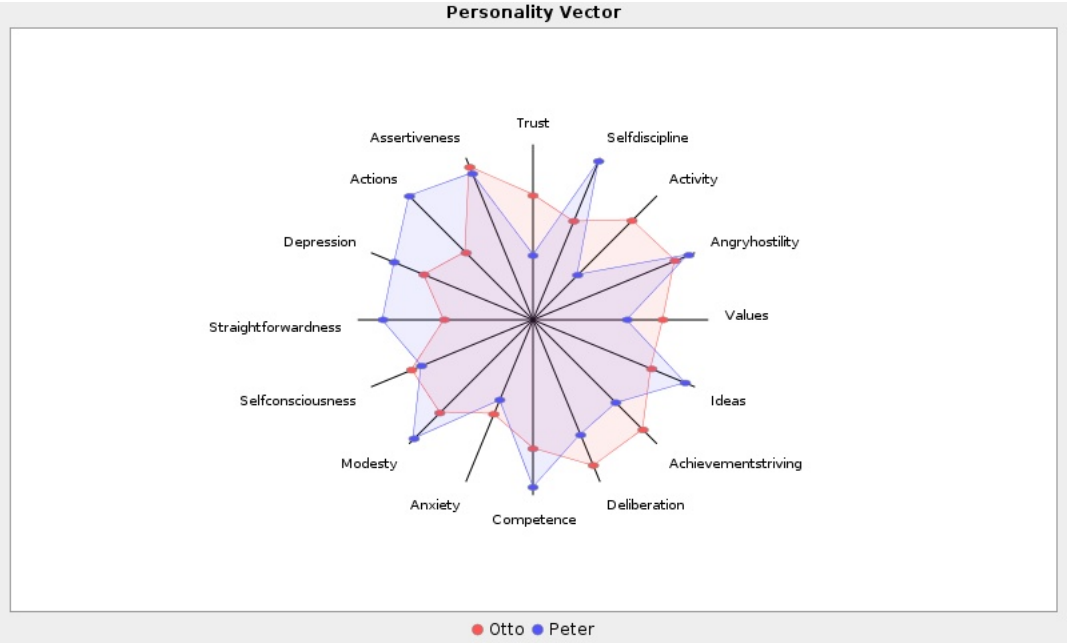


Fig. 6.4: A plot showing the personality vectors of two agents in the BAIPD framework.

related to the dialogue process like protocol rules is handled. When turns shift to a participant in the dialogue, the persuasion platform requests the agent to generate moves. Since the agent has a description of its personality by means of its personality vectors, the agent is able to generate moves by utilizing the reasoning engines to compute corresponding orderings over speech acts types and attitudes. Figure 6.4 shows the personality vectors of two agents Peter and Otto in the BAIPD framework. The graph shows the facets included in the personality vectors of the agents, where the inside of the graph yields a strength value of 0 and the outside of the graph represents a strength value of 1. The personality configurations of the agents in this example are chosen randomly. The andness value for this example is set as 0.9.

As we have seen, multiple reasoning engines exist. As always, a distinction is made between action selection and action revision, resulting in an ordering over speech act types, computed for agents Peter and Otto as seen in figure 6.5. In addition, multiple reasoning engines for action revision are used to compute an ordering over attitudes belonging to these speech act types as seen in figure 6.6.

Two prominently favored speech act types are *why* and *argue* for Peter. By looking at the personality vector of Peter, this preference can be explained by high strengths of the personality facets *ideas* and *competence*. These orderings are used by the reasoning algorithm to generate moves using the *generateMoves* function we have seen earlier in this chapter. This function generates, based on the description of the individual attitudes, the moves that given the context of the dialogue are allowed and should be contributed to the dialogue by the agent. The outcome of such a dialogue can be seen in figure 6.7. The attitude names along the edges in the graph

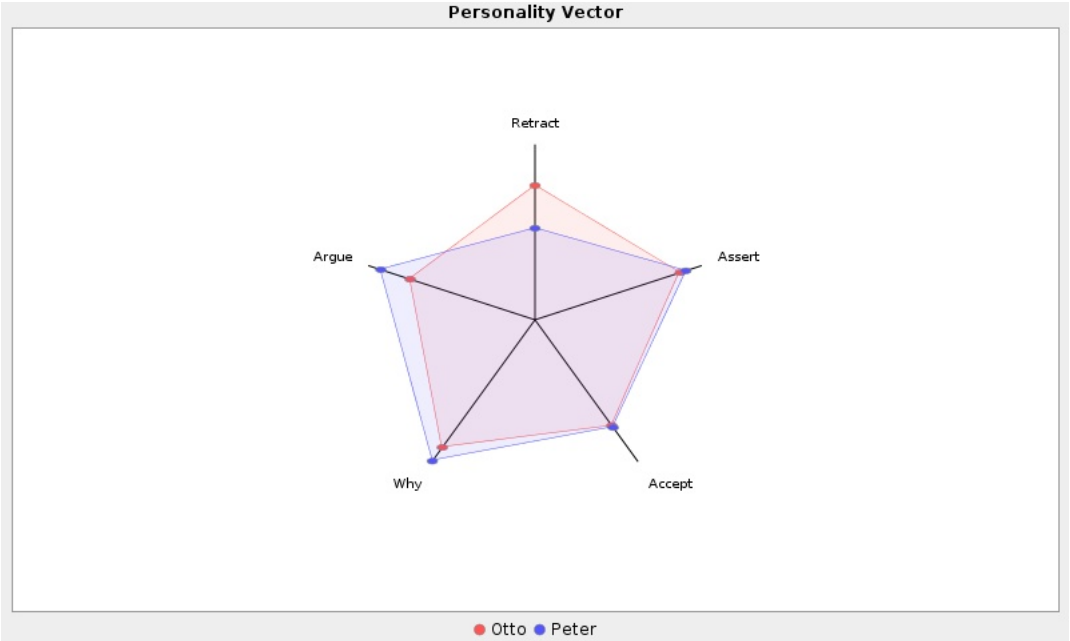


Fig. 6.5: A plot showing the action selection ordering of two agents in the BAIPD framework.

denote the attitude that was used to generate the move. As an example, the *tentative* attitude generated Otto's first move *why endangersHumanity*. In addition to the reasoning algorithm described earlier in this chapter, additional rules are added to make sure the agent generated moves that are consistent and do not validate protocol rules:

- If an agent generates a move that claims some proposition p, supports an argument with conclusion p or moves a *why* move challenging p, the agent is not allowed to retract p or concede p.

- If the agent surrendered to a move, the agent is not allowed to generate a move with the surrendered to move as a target.

These rules make sure the agent conforms to consistency and validity.

6.9 Conclusion

This chapter introduced the reasoning system of the argumentative agent. The reasoning system consists of reasoning rules that guide the reasoning process of the agent by taking the personality vectors as input and outputting preference values for either locutions in the argumentation framework or attitudes associated with these locutions. These reasoning rules are based on fuzzy logic rules and as such offer a possibility to easily modify these rules, extend the set of personality facets or adjust the matching of strength values in the personality vectors. In addition to

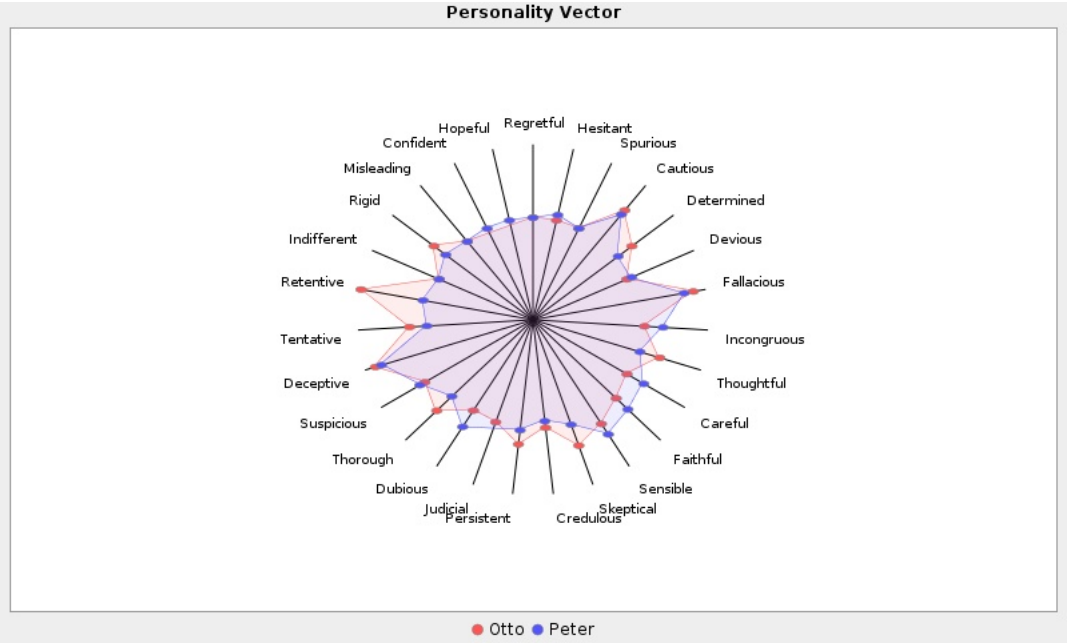


Fig. 6.6: A plot showing the action revision ordering of two agents in the BAIPD framework.

these reasoning rules, the reasoning system contains a reasoning algorithm. This reasoning algorithm uses the reasoning rules to determine preference orderings and generate moves to contribute to the dialogue accordingly. The reasoning algorithm makes use of a reasoning engine, computing the preference orderings using a fuzzy inference system. Emphasized was that the reasoning rules introduced in this chapter are easily adjustable according to the desired behavior of the agent. Moreover, the reasoning system including the reasoning rules should be regarded as a system for introducing personality in argumentative agents, instead of as an exhaustive description of the personality of such agents.

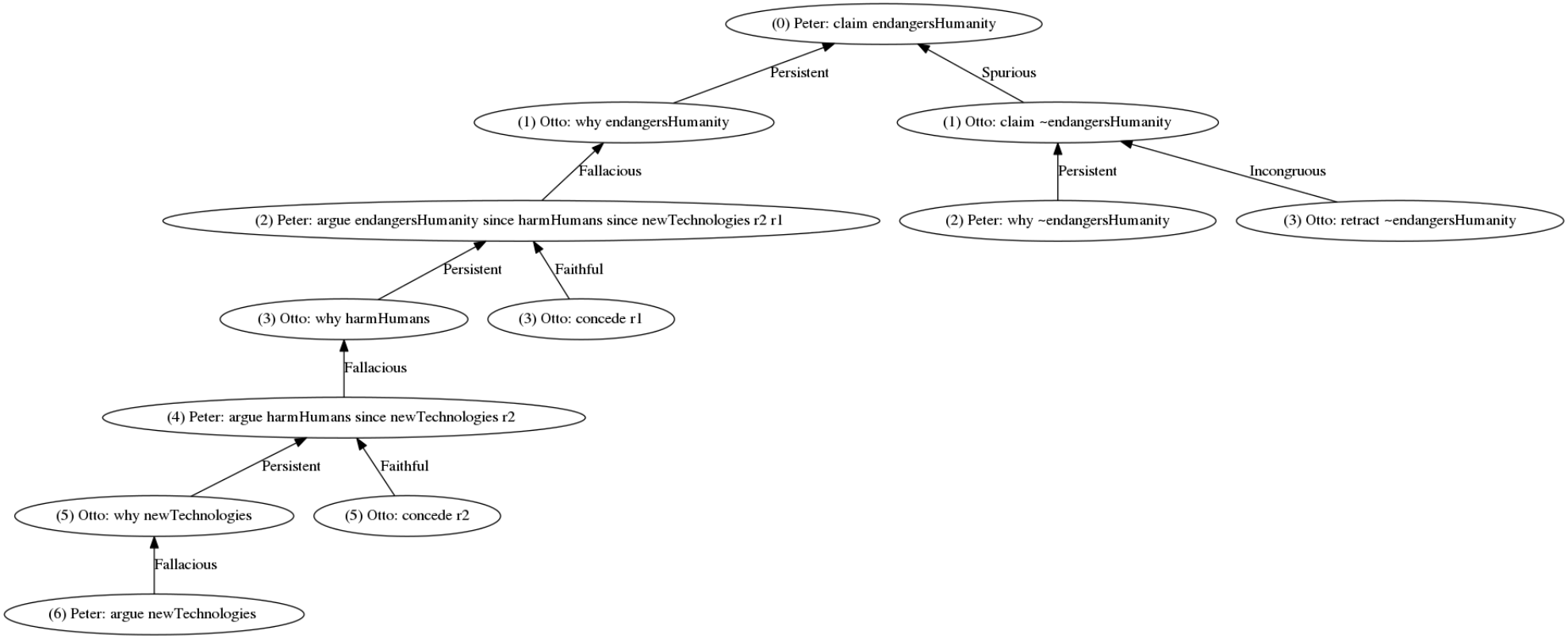


Fig. 6.7: The sample dialogue as a dialogue tree.

7 Modelling the personality of opponents

The previous chapter discussed the reasoning system of the agent. The agent with personality allows for the configuration of an agent according to some preferred behavior in a specific application.

In addition to introducing personality in the agent, introducing a way of modelling the opponent allows for the agent to adjust its strategy based on the personality of its opponent. For example, let us imagine an agent that is configured to prefer a *faithful* attitude, indicating that the agent will easily accept an opponent's claim. Suppose this agent is faced with an agent that prefers one of the *spurious* or *deceptive* attitudes, indicating that the agent is able to bullshit or lie according to Hommes' [Hom15] definition. Now suppose the agent would have some method of continuously modelling the opponent based on the moves played by the opponent. Using this opponent modelling method, the agent models the opponent as preferring his opponent's preference for one of the *spurious* or *deceptive* attitudes. If our agent is rational we would expect the agent to decrease its preference for the *faithful* attitude, since preferring the *faithful* attitude makes the agent vulnerable to the opponent's lies or bullshit.

This chapter will discuss the modelling of agents opponents based on the personality model introduced in this thesis. Among other things, this will lay the foundation for optimizing an agent's dialogue strategies based on opponent modelling. However, research on this is outside the scope of the current research. Optimizing strategies based on opponent models is beyond the scope of this research and will be left open as a topic for future research. Throughout this chapter the familiar sample dialogue as re-presented in figure 7.1 will be used to exemplify our method of modelling the opponent's personality.

7.1 Attitude Status

Up until this point the knowledge base of the agent has always been used for generating moves allowed by attitudes preferred by the agents. This knowledge, however, is private and can not be accessed by another agent. This makes it impossible to model the personality of an agent's opponent directly. By playing

moves an agent does, however, share information about its internal state in terms of commitments. Based on these commitments we can analyze the attitudes that have possibly triggered the agent to play a move. By application of modus tollens on the information provided by the agent's move in the dialogue, we are able to preclude attitudes as belonging to moves that are played in the dialogue. Subsequently, we are able to perform abduction on the set of possible attitudes for a move in the dialogue by continuously eliminating possible attitudes for moves in the dialogue once new moves are played in the dialogue. Here, *possible* attitude refers to possible attitudes according to the definition of the attitude in the personality model. As a simple example:

> An agent with a *hesitant* attitude cannot assert any proposition.
>
> The agent has asserted a proposition (claim).
>
> ---
>
> The *hesitant* attitude is not a possible attitude for the claim move.

This reasoning allows us to assign, based on the moves that are added to a dialogue, the possible attitudes that can be associated with a move in the dialogue. Likewise, we can apply this reasoning pattern to the other attitudes associated with the *claim* speech act.

> An agent with a *confident* attitude can assert any proposition for which he can construct an argument.
>
> The agent has asserted a proposition (claim), but cannot construct an argument for the claim.
>
> ---
>
> The *confident* attitude is not a possible attitude for the claim move.

An agent could observe that the opponent cannot construct an argument in this case by challenging the opponent's claim. If the agent provides for support for the claim, under the assumption that the provided support is not lied or bullshit, the agent can construct an argument for the claim. This can be observed by the commitments of the opponent based on an *argue* speech act. Similar to the pattern above, we can apply the same reasoning to the *careful* and *thoughtful* attitudes.

> An agent with a *careful* attitude can assert any proposition for which he can construct an argument and no stronger counterargument.
>
> The agent can construct a stronger counterargument.
>
> ---
>
> The *careful* attitude is not a possible attitude for the claim move.

An agent could observe that the opponent has a stronger counterargument when the opponent provides for an argument for the contrary of the claim and this counterargument is stronger. This could be by the opponent being inconsistent or incoherent. In addition, the introduction of a *question* locution as in [WK95] would allow for the agent to ask about the opponent's opinion about the contrary of its claimed proposition.

> An agent with a *thoughtful* attitude can assert any proposition for which he can construct a justified argument.
>
> The agent cannot construct a justified argument.
>
> ---
>
> The *thoughtful* attitude is not a possible attitude for the claim move.

An agent could observe whether or not the opponent can construct a justified argument by observing the opponent's commitments. This reasoning pattern can be applied to remaining attitudes as defined in the personality model. The *attitude status* determines for the attitudes in the personality model, which attitude is *possible* for that move.

Example 7.1.1. In the dialogue between Peter and Otto, the attitude associated with Peter's move P_1 cannot determined on its own without access to Peter's knowledge base.

P_1: *claim* endangersHumanity

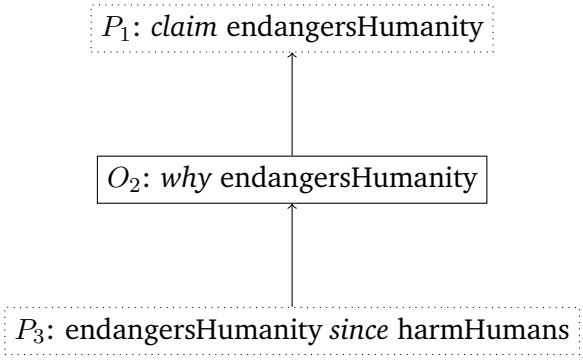
Peter's claim can be the result of any of the attitudes; *confident*, *careful*, *thoughtful*, *spurious* and *deceptive*. The attitude *hesitant* is ruled out, since Peter in fact made the claim.

To keep track of the possible attitudes associated with a move in a dialogue, the *attitude status function* is introduced.

Definition 7.1.1. An *attitude status function* is a function $A : M^{\leq\infty} \times \{P, O\} \longrightarrow \wp(\mathcal{AT})$, where $\mathcal{AT}$ denotes the set of attitudes.

This function defines for a move in the dialogue the set of possible attitudes based on the commitments of a player in the dialogue. As new moves are added to the dialogue, this function is used to redetermine the set of possible attitudes associated with the move. As the set of commitments increases, more information becomes available to the agent to determine the attitude status of move which helps the agent prune the attitude status of the moves in the dialogue.

Example 7.1.2. Based on the support provided by player Peter denoted by p in the dialogue d



$A(d, P_1)$ will contain the attitude names; *confident*, *careful* and *thoughtful* based on the commitments of the agent at this point in the dialogue

$C_d(p) \subseteq \{\text{endangersHumanity}, \text{harmHumans}, \text{harmHumans} \supset \text{endangersHumanity}\}$

These three attitudes allow Peter to generate the move P_1 based on the commitments after P_3. The remainder of assertion attitudes is not allowed:

- By the *spurious* and *deceptive* attitude, Peter is not allowed to generate a claim for *endangersHumanity*, since Peter can construct an argument for *endangersHumanity*,
- the *hesitant* attitude can be ruled out since Peter moved a *claim* move.

What can be observed from this example is the pruning of the attitude status. This, however, depends on the moves by the agent itself. Had the agent not moved move O_2, the agent had not been able to rule out the *spurious* and *deceptive* attitudes.

7.2 Modelling Algorithm

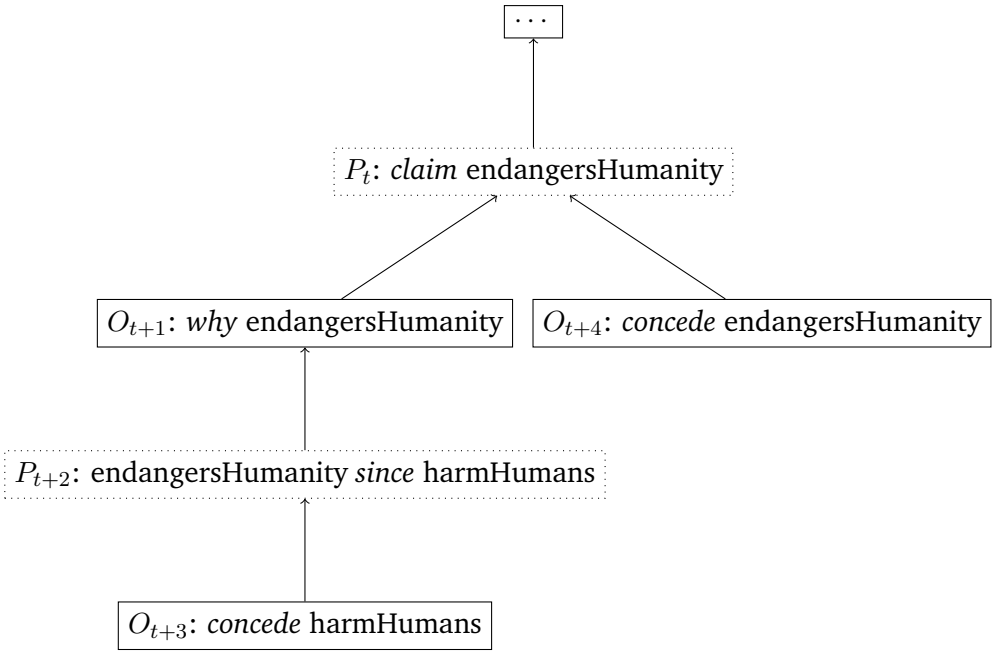
Let the following algorithm concretize the attitude status function. The algorithm depends on the commitment function as present in Prakken's framework. In addition, the algorithm makes use of the earlier defined *generateMoves* function. In this case the knowledge base that was left implicit earlier has been replaced by the set of commitments for player pl.

Algorithm 6 Attitude status function

```
1: function ATTITUDESTATUS(d, m)
2:     C ← C_pl(d, m)
3:     A ← ∅
4:     for attitude ∈ AT do
5:         if m ∈ generateMoves(d, pl, C) then
6:             A ← A ∪ {attitude}
7:         end if
8:     end for
9:     return A
10: end function
```

For every new move, the attitude status of moves in the dialogue is determined based on the current set of commitments. As can be observed from this algorithm, the attitude status is constructed from the empty set. As the attitude status narrows down based on new information in terms of moves that are added to the dialogue, it is not the case that the attitude status necessarily results in a singleton attitude. An example of such a case will be illustrated by the following example.

Example 7.2.1. Consider the following example continuation of a dialogue with $t-1$ moves.



As an example, let us say that the opponent has a preference for the *cautious* acceptance attitude and a preference for the *judicial* challenge attitude. Moreover, the opponent prefers the *accept* speech act type over the *why* speech act type and

the opponent prefers both these speech act types over all other speech act types. When the proponent claims *endangersHumanity*, the opponent first tests whether its acceptance attitude allows for the acceptance of *endangersHumanity*. The opponent can construct an argument for ¬*endangersHumanity*, but not for *endangersHumanity*, the *cautious* attitude disallows the opponent to accept the claim. Next, the opponent tests whether the challenge attitude allows the opponent to challenge the claim. Since the opponent cannot construct an argument for *endangersHumanity*, the opponent is allowed to challenge the claim. Next, the proponent provides support for the claim. Now, the opponent tests whether he is allowed to accept *harmHumans*. Suppose the agent is able to provide an argument for *harmHumans* and no stronger argument for ¬*harmHumans*; then the opponent is allowed to accept *harmHumans*. Next, based on the newly provided information, the opponent is able to construct an argument for *endangersHumanity*. This argument is stronger than its argument for ¬*endangersHumanity* and the opponent accepts the proponent's initial claim.

Since the proponent has provided support for the claim, the opponent can rule out the *spurious* and *deceptive* attitudes for the proponent's move P_t. This is, however, under the assumption that the provided support by the proponent is not lied or bullshit, which cannot be observed in this example. The opponent can also rule out the *hesitant* attitude, since the proponent made the claim.

After the proponent provided support for its claim, the opponent concedes the supporting argument and subsequently concedes the proponent's claim. The remaining attitudes *confident*, *careful* and *thoughtful* are still possible attitudes for move P_t. By conceding, the opponent did not receive enough information to further narrow down the set of possible attitudes.

As we can observe from this example, the personality configuration and the moves that are generated by the agent accordingly strongly determine the outcome of the dialogue. This makes that it is not necessarily the case that the attitude status of a move converges to a singleton attitude. As the length of the dialogue increases, more information will become available based on new commitments that possibly affect the attitude status assignment of moves in that dialogue.

7.3 Using the Attitude Status

The attitude status allows for observing the personality of the opponent. Even though the attitude status does not necessarily result in a singleton attitude as the dialogue increases in the number of moves, as this is highly dependent on the behavior of the agent itself, the attitude status describes for every move in the dialogue the possible attitudes. When combining the attitude statuses of every move by the

opponent in a dialogue, the agent is able to construct an estimation of its opponent's configuration. As an example, a histogram of attitudes can be constructed that is updated whenever the attitude statuses in the dialogue are recomputed. In this case, the frequencies of attitudes in the histogram would be an estimation of the configuration of the opponent. As more moves are added to the dialogue, this estimation is likely to increase in precision. Moreover, the histogram can be updated over multiple dialogues to analyze the behavior of an opponent over more than one dialogue to increase precision.

This estimation can serve as input to learning algorithms like an artificial neural network, Bayesian networks or genetic algorithms or mathematical optimization methods. Based on this input and information about the configuration of the agent, the algorithm allows for the optimization of the agent's strategy based on the analysis of its opponent in relation to its own personality.

7.4 Conclusion

This chapter discussed a method of modelling the personality of the opponent. Using a modus tollens style reasoning, an attitude status is continuously updated by means of abduction. This attitude status defines what attitudes are possibly associated with a move in the dialogue. Using the attitude status, a model of the opponent. The model of the opponent can be used in future research to optimize the behavior of the agent according to its knowledge of its opponent.

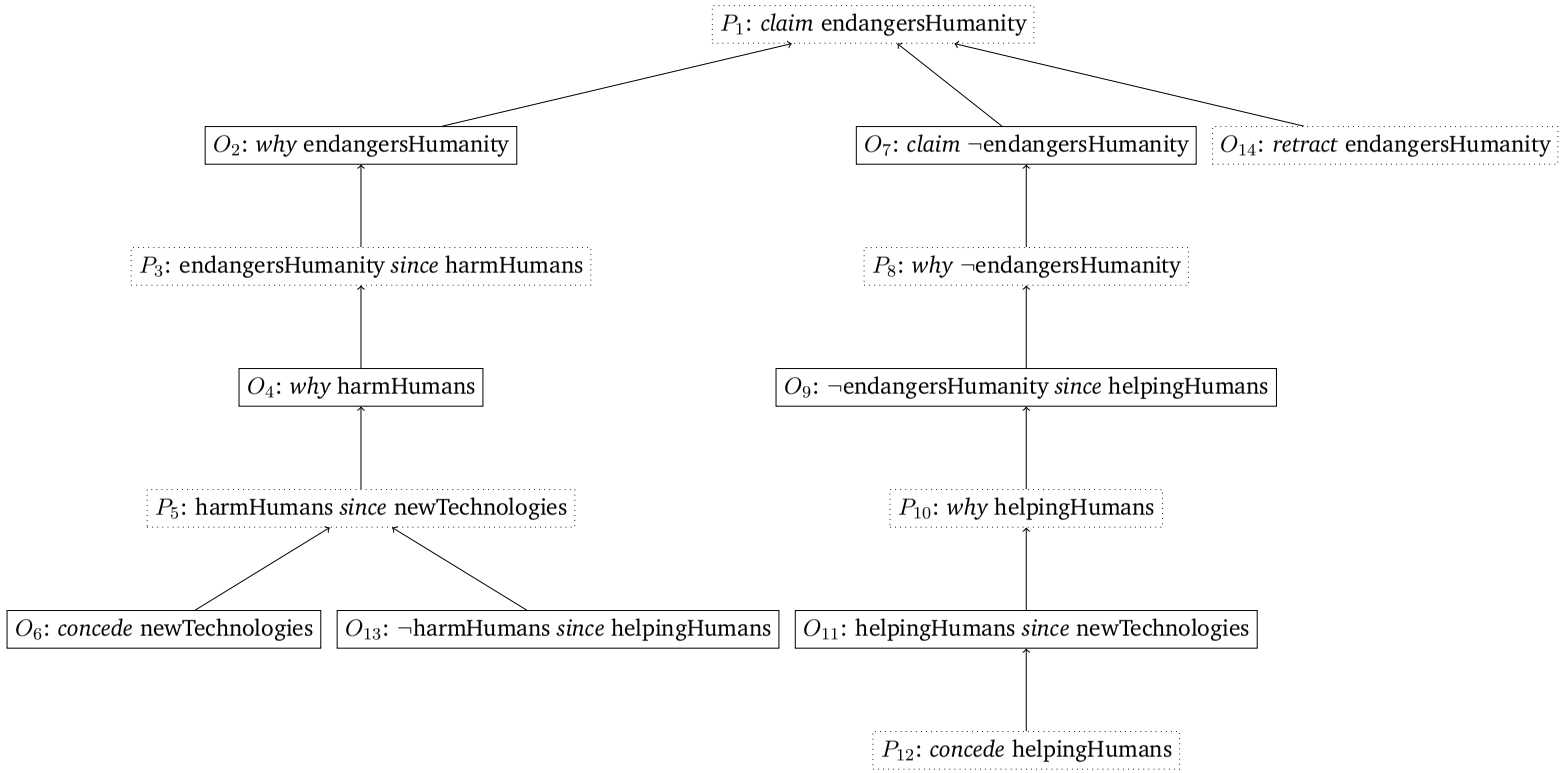


Fig. 7.1: The sample dialogue as a dialogue tree.

8 Conclusion

In this thesis, personality in argumentative agents was investigated. To this end a personality model was introduced based on personality theories originating from the field of personality psychology. This personality model consists of (i) an agent's personality vectors, describing the personality configuration of the agent in terms of strengths of personality facets in the personality model and (ii) agent attitudes, indicating under what circumstances an agent is allowed to move a specific locution in the dialogue. In addition to a formalization of the concept of personality captured by the personality model, the implementation of such an agent was introduced by adjusting Erik Kok's testbed for experimentation with argumentative agents. Moreover, the inner workings of the agent's reasoning according to its personality was introduced in terms of the agent's reasoning system. Lastly, a method for modelling the opponent was investigated, allowing for an agent to model the personality of its opponent.

Four research questions were put forward in the introduction of this thesis, which can be answered as follows:

On how a personality can be introduced to argumentative agents for persuasion dialogues (research question 1) and how a model for personality in argumentative agents can be devised that allows argumentative agents to reason according to a personality configuration (research question 2) Chapter 3 first introduced the personality theory that forms the basis for the personality model that was introduced in chapter 4. The personality model of the agent is (i) a description of the agent's personality in terms of 15 personality facets that each have their interpretation in the personality model in the context of argumentation and (ii) a personality vector, describing for each of these personality facets the strength, such that the personality vector of the agent is a configuration of the agent's personality. The personality model serves as the agent's description of its personality and is used to allow the agent to reason according to its personality.

On how an argumentative agent featuring personality can be implemented (research question 3) Chapter 5 discussed how Erik Kok's testbed for experimentation with argumentative agents BAIDD can be adjusted such that this testbed is usable for persuasion dialogues, as the testbed was originally created for deliberation dialogues. BAIPD, a variation on BAIDD was introduced that allows for the implementation

and experimentation with persuasion dialogues. Chapter 6 discussed the reasoning system of the agent. This reasoning system consists of (i) a description of reasoning rules that determine the preferences over locutions available in the argumentation framework and agent attitudes depending on the strengths of personality facets in the agent's personality vectors and (ii) a description of the agent's reasoning algorithm, used to, based on the reasoning rules in the reasoning system and the agent's personality configuration generates moves according to the configuration of the agent's personality. The agent was implemented in the BAIPD framework [@Eth16].

On how an argumentative agent featuring personality can model the personality of its opponent (research question 4) Chapter 7 introduced an opponent modelling method that uses a modus tollens style reasoning scheme to eliminate possible attitudes given information shared by the participants in an argumentation dialogue. The elimination of possible attitudes abductively narrows down the attitude status of moves in the argumentation dialogue. Maintaining the distributions of possible attitudes used by the opponent yields a model of the opponent's personality. It was mentioned that future research in the optimization of agent strategies in argumentation could use this information to automatically optimize agent strategies by means of for instance artificial neural networks.

This research has contributed to research in artificial intelligence in the following ways; (i) this research combined the fields of agent technology and personality psychology and as such contributes to the field of multi-agent systems and computational argumentation by extending the agent with a description of personality, (ii) this research introduced fuzzy reasoning in the implementation of the agent's reasoning process and offers and alternative to game-theoretical approaches, (iii) this research introduced a new way of steering an agent's behavior according to a description of its personality, (iv) this research extended the agent attitudes of Parsons et al., (v) this research introduced a method of modelling the behavior of opponents.

As for topics for future research, optimizing the strategies of agents could be investigated. The opponent modelling approach can serve as the input to an optimization approach that uses the agent's knowledge of its opponent to optimize its strategy not only based on what moves would be optimal, but what moves would be optimal given the personality of the agent and the personality of its opponent. As an example, suppose that the agent is trusting and models its opponent as being deceptive. The agent would prefer, even though the agent is trusting by nature, to not trust its opponent and be reluctant to accept propositions by its opponent. Another topic of future research is the extension of the current research outside of the field of argumentation theory. Introducing personalities to agents in different domains could

benefit for instance the interaction between software agents and humans as, especially when the agent is able to model the personality of a human, agents are able to adjust their behavior according to humans. This could for instance benefit training situations. Many theorems about the properties of different personalities of agents can be formulated such as "a dialogue with two agents that are achievement-striving increase the length of the dialogue" or "a deliberate and and straightforward agent contributes to a truthful outcome of the dialogue", researching such theorems will benefit in proving certain desirable or undesirable properties of the personalities of agents. Lastly, the current research could be extended to other types of dialogues or could be extended with additional locutions or more personality facets allowing for more fine-grained specification of the agent's personality.

Bibliography

[Amg+06] L. Amgoud, L. Bodenstaff, M. Caminada, et al. *Final review and report on formal argumentation system. Deliverable D2. 6*. Tech. rep. ASPIC IST-FP6-002307, 2006 (cit. on p. 7).

[App+12] A. Applebaum, Z. Li, A. R. Syed, et al. „Firewall configuration: An application of multiagent metalevel argumentation“. In: *Proceedings of the 9th Workshop on Argumentation in Multiagent Systems*. 2012 (cit. on p. 1).

[Ash+04] M. C. Ashton, K. Lee, M. Perugini, et al. „A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages.“ In: *Journal of Personality and Social Psychology* 86.2 (2004), pp. 356–366 (cit. on p. 24).

[Boy95] G. J. Boyle. „Myers-Briggs Type Indicator (MBTI): Some Psychometric Limitations“. In: *Australian Psychologist* 30.1 (1995), pp. 71–74 (cit. on p. 24).

[CM92] P. TT. Costa and R. R. McCrae. *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO FFI): Professional manual*. Psychological Assessment Resources, 1992 (cit. on pp. 24, 28).

[Dun95] P. M. Dung. „On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games“. In: *Artificial Intelligence* 77.2 (1995), pp. 321–357 (cit. on p. 7).

[EE65] H. J. Eysenck and S. G. B. Eysenck. „The Eysenck Personality Inventory“. In: *British Journal of Educational Studies* 14.1 (1965), pp. 140–140 (cit. on p. 24).

[Had+12] C. Hadjinikolis, S. Modgil, E. Black, and P. McBurney. „Mechanisms for Opponent Modelling“. In: *2012 Imperial College Computing Student Workshop*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2012 (cit. on p. 1).

[Hom15] R. M. Hommes. „Improper behaviour in argumentation based persuasion dialogues“. MA thesis. Utrecht University, 2015 (cit. on pp. 38, 85).

[JS99] O.P. John and S. Srivastava. „The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives“. In: *Handbook of Personality: Theory and research (Vol 2)*. Ed. by L. A. Pervin and O. P. John. Guilford Press, 1999, pp. 102–138 (cit. on p. 23).

[Jun71] C.G. Jung. *The collected works of C.G. Jung: Psychological Types*. Routledge and K. Paul, 1971 (cit. on p. 24).

[Kok13] E. M. Kok. „Exploring the practical benefits of argumentation in multi-agent deliberation“. PhD thesis. Utrecht University, 2013 (cit. on pp. 4, 43, 79).

[MA75] E. H. Mamdani and S. Assilian. „An experiment in linguistic synthesis with a fuzzy logic controller“. In: *International Journal of Man-machine Studies* 7.1 (1975), pp. 1–13 (cit. on p. 50).

[MC08] R. R. McCrae and P. T. Costa. „The five-factor theory of personality“. In: *Handbook of personality: Theory and research*. Ed. by O. P. John, R. W. Robins, and L. A. Pervin. Guilford Press, 2008. Chap. 5, pp. 159–181 (cit. on p. 23).

[McC+02] R. R. McCrae, P. T. Costa, A. Terracciano, et al. „Personality trait development from age 12 to age 18: Longitudinal, cross-sectional and cross-cultural analyses“. In: *Journal of Personality and Social Psychology* 83.6 (2002), pp. 1456–1468 (cit. on p. 23).

[O'H87] M. O'Hagan. „Fuzzy decision aids“. In: *Proceedings of the 21st Asilomar Conference on Signal, Systems and Computers*. IEEE, 1987 (cit. on p. 57).

[Par+03] S. Parsons, M. Wooldridge, and L. Amgoud. „Properties and complexity of some formal inter-agent dialogues“. In: *Journal of Logic and Computation* 13.3 (2003), pp. 347–376 (cit. on pp. 1, 37).

[Pit93] D. J. Pittenger. „Measuring the MBTI and coming up short“. In: *Journal of Career Planning and Employment* 54.1 (1993), pp. 48–52 (cit. on p. 24).

[Pra05] H. Prakken. „Coherence and flexibility in dialogue games for argumentation“. In: *Journal of Logic and Computation* 15.6 (2005), pp. 1009–1040 (cit. on pp. 1, 14).

[Pra06] H. Prakken. „Formal systems for persuasion dialogue“. In: *The Knowledge Engineering Review* 21.2 (2006), pp. 163–188 (cit. on pp. 1, 47).

[Pra10] H. Prakken. „An abstract framework for argumentation with structured arguments“. In: *Argument and Computation* 1.2 (2010), pp. 93–124 (cit. on pp. 7, 10).

[Rie+13] T. Rienstra, M. Thimm, and N. Oren. „Opponent Models with Uncertainty for Strategic Argumentation“. In: *23rd International Joint Conference on Artificial Intelligence*. 2013 (cit. on p. 1).

[WB99] D. G. Winter and N. B. Barenbaum. „History of Modern Personality Theory and Research“. In: *Handbook of personality: Theory and research*. Ed. by L. A. Pervin and O. P. John. Elsevier, 1999. Chap. 1, pp. 3–27 (cit. on p. 23).

[Wei11] T. L. Van der Weide. „Arguing to motivate decisions“. PhD thesis. Utrecht University, 2011 (cit. on p. 1).

[Wig96] J. S. Wiggins. *The five-factor model of personality: Theoretical perspectives*. Guilford Press, 1996 (cit. on p. 23).

[WK95] D. N. Walton and E. C. W. Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. State University of New York Press, 1995 (cit. on pp. 1, 87).

[Yag88] R. Yager. „On ordered weighted averaging aggregation operators in multicriteria decisionmaking“. In: *Systems, Man and Cybernetics* 18.1 (1988), pp. 183–190 (cit. on p. 56).

[Zad73] L. A. Zadeh. „Outline of a new approach to the analysis of complex systems and decision processes“. In: *Systems, Man, and Cybernetics* 3.1 (1973), pp. 28–44 (cit. on p. 50).

[Zem92] R. Zemke. „Second Thoughts about the MBTI“. In: *Training* 29.4 (1992), pp. 43–47 (cit. on p. 24).

[Zho+14] Q. Zhong, X. Fan, F. Toni, and X. Luo. „Explaining Best Decisions via Argumentation“. In: *Proceedings of the European Conference on Social Intelligence*. 2014, pp. 224–237 (cit. on p. 1).

Websites

[@Eth16] M. E. Etheredge. *BAIPD*. 2016. URL: `https://bitbucket.org/metheredge/baipd` (cit. on p. 94).

[@Gol99] L. R. Goldberg. *International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences*. 1999. URL: `http://ipip.ori.org/` (cit. on p. 28).

[@SV09] M. South and G. A. W. Vreeswijk. *ASPIC Java Components*. 2009. URL: `http://aspic.cossac.org/` (cit. on p. 43).