Agents with a Human Touch:
Modeling of Human Rationality in Agent Systems

Thesis submitted in accordance with the
requirements of the University of Liverpool
for the degree of Doctor in Philosophy

by

Fahd Saud Nawwab

June, 2010

To my parents, wife and daughters,
with love and gratitude

Abstract

Will it be possible to create a self-aware and reasoning entity that has the capacity for decision making similar to that we ascribe to human beings?

Modern agent systems, although used today in various applications wherever intelligence is required, are not ready for applications where human rationalities are usually the only option in making important decisions in critical or sensitive situations.

This thesis is a contribution to this area: a decision-making methodology is introduced to address the different characteristics that an agent should have in order to be better trusted with such critical decisions.

The work begins with a study of philosophy in the literature (Chapter 2), which reveals that trust is based on emotions and faith in performance. The study concludes that a trustworthy decision has five main elements: it considers options and their likely effects; it predicts how the environment and other agents will react to decisions; it accounts for short- and long-term goals through planning; it accounts for uncertainties and working with incomplete information; and, finally, it considers emotional factors and their effects. The first four elements address decision making as a product of "beliefs"; the last addresses it as a product of "emotions". A complete discussion of these elements is provided in Section 2.1.

This thesis is divided into two main parts: the first treats trust as a product of beliefs and the second treats trust as a product of emotions.

The first part builds the decision-making methodology based on argumentation through a five-step approach where first the problem situation representing the actions available to the agent and their likely consequences is formulated. Next, arguments to perform these actions are constructed by instantiating an argumen-

tation scheme designed to justify actions in terms of the values and goals they promote. These arguments are then subjected to a series of critical questions to identify possible counter arguments so that all the options and their weaknesses have been identified. Preferences are accommodated by organising the resulting arguments into an Argumentation Framework (we use Value-Based Argumentation [VAF] for this approach). Arguments acceptable to the agents will be identified through the ranking of the agent's values, which may differ from agent to agent. In the second part (Chapters 5 and 6), this methodology is extended to account for emotions. Emotions are generated based on whether other agents relevant to the situation support or frustrate the agent's goals and values; the emotional attitude toward the other agents then influences the ranking of the agent's values and, hence, influences the decision.

In Chapters 4 and 6, the methodology is illustrated through an example study. This example has been implemented and tested on a software program. The experimental data and some screen shots are also given in the appendix.

Acknowledgments

Firstly, people who have shaped my personality and made the difference in my life that got me to reach this stage.

My mentor and teacher, Dr. Abdulhafeez Ameen who has planted in me the seed of the love for research and the love for education. Hatem El-Kady, my ex-manager in IBM, as he believed in my capabilities and taught me the very important lesson of believing in myself and my potential. He always showed me what I am capable of and always pushed me to be the best I could.

Secondly, the support I received from my family. My mother, Huda Muhammad Alam, has always been on my side with prayers, encouragement and love. My father, Saud Yusuf Nawwab, has always believed in the pursuit of academic excellence. He has always pushed me to excel and be distinguished and showed me how to dream big and how to achieve those dreams. I owe thanks to my siblings Nada, Faisal and Shada for their encouragement and prayers. I am also thankful to Ismail Ibrahim Nawwab and Wasimah Ahmad; I could not have wished for better in-laws. I am grateful for their support during my studies.

And I could have never achieved or completed this thesis without the love, understanding and support of my wife, Rahmah Ismail Nawwab, as she shared with me the complete journey and gave me nothing but love, care and support. This was also a joyful experience because of my daughters Hibah and Huda who were very understanding and patient despite the long periods they did not get to see their father.

Thirdly, I was privileged to be part of a great academic institution the University of Liverpool. My research has been an extremely smooth and rewarding experience because of the great facilities, tools and resources offered by the university. The help, guidance and support I received from the staff as well have

made my travel through this academic landscape even more enriching.

I was also honoured to be part of the Computer Science Department. They have helped fund my participation in several conferences during my study, provided me with a decent work environment and access to resources. I was happy to learn how esteemed the department is; whenever I collaborated with fellow researchers outside Liverpool, I have always been proud introducing myself as part of this department.

I am thankful to Ramzi El-Eid, my manager, colleague and friend who has always been supportive and understanding.

Finally, I am heartily thankful to have benefited from my supervisors, Paul Dunne and Trevor Bench-Capon. When I look back at the beginning of my doctoral studies and compare myself then with the person that I am today, I immediately realise the effects of the great shift and transformation that their supervision had on me. I am privileged to have been supervised by such well-esteemed and recognised individuals. As I am pleased that my PhD is coming to an end, I am also determined that my journey with my supervisors will not. Once my supervisors, I am looking forward to having them as my post-doctoral mentors.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis identifies the characteristics that raise the level of Trust between entities and then uses these characteristics to develop a methodology of decision making to allow the delegation of higher level tasks to software agents.

Trust in decision making is a product of beliefs and also emotions (as discussed in depth in Section 2.1). This methodology involves building a decision-making methodology based on beliefs and extending this to incorporate emotions. The thesis also examines an extensive case study.

This chapter provides an overview of the topic and a general overview of the work.

First, Section 1.1 describes the main motivation. Section 1.2 outlines the different capabilities that will form the basis of this research. Section 1.3 gives an overview of principles of Multi-Agent Systems that are relevant to our work. Section 1.4 then outlines the main contribution of this thesis and gives an overview of how the principles and elements of 1.2 and 1.3 are used to build toward the goal. Section 1.5 will outline the thesis structure and finally, Section 1.6 summarises the introduction.

1.1 Research Motivation

The main motivation behind this research is to support the delegation of high-level decisions to software agents by building a methodology of decision making with the capabilities needed to allow those agents to be trusted with important decisions.

The kind of decisions we mean by “High Level” are critical and/or sensitive. Critical decisions involve matters of high risk and strategic importance, such as air traffic control and budget allocations. Sensitive decisions are matters of personal preference, such as decorating a room.

From a high-level perspective, our aim is to help answer the following question:

> “What aspects can raise our confidence level in agents so as to allow them to take over decision making and how can these aspects be implemented?”

We provide a discussion on Trust in order to identify the capabilities that helped us answer this question. Providing these capabilities is then the main focus of the methodology offered by this thesis.

1.2 The Basic Elements of Trust

This section discusses the main capabilities that a decision-making methodology needs to have for the agent to be trustworthy. The methodology presented by this thesis is built on this discussion.

Trust is a combination of beliefs and emotions; either we have sufficient evidence to believe that someone can be trusted, or we have the instinct and emotional inclination to trust that person. As a belief, when Trust is given, more control is handed to the trustee, allowing more power over more decisions; also the truster is hopeful that he would be better off trusting the trustee than otherwise. That is the truster believes that the trustee’s decisions will be at least as good as his own. Thus the trustee’s ability to make decisions means also giving the power of using resources. So, the trustee should have the ability to commit appropriate resources relevant to the goals he is trying to achieve, giving the trustee the ability to make decisions means giving him also the power of using resources.

We trust someone as we believe that he has what it takes to be trustworthy, so, we trust because of belief, but philosophers argue that this “belief” is not enough and Trust is sometimes based on “emotions” and not just beliefs. Emotions play a big role in Trust in two capabilities: first, they help to resolve conflicts when there are overall equally good choices as far as belief is concerned. Second, emotions are an important element in decisions that have social and/or behavioural aspects combined.

Other than Trust, considerations of emotions can be beneficial in decision making in many different aspects other than the two mentioned above. Among these benefits are triggering resequencing in response to changes in priorities, fostering cooperation among different agents by rewarding cooperation and punishing defection and in applications where decision making involves expression of those decisions.

Philosophy (Section 2.1) suggests that for a decision-making methodology to yield trustworthy decisions, it should have five main capabilities:

- Practical Reasoning: This is the basic element in decision making whereby the agent needs to understand the environment it is situated in and consider all its available options for different decisions and the likely effects of making these decisions.

- Social Interactions: When taking an action, results are not solely dependent on the agent's actions, but also on how the environment will react to this decision. For example, when getting into a plane to travel, trust is placed in the pilot to actually arrive safely to the destination and this decision to take the plane is based on the assumption that the other agent, the pilot, will fly the plane safely.

- Planning: This considers not only the direct effects of actions but also the ability to align with the different goals and aspirations of the agent in the short and long term. A decision that is less suitable in a particular situation might be better given the long-term plans of the agent and might help other future decisions to achieve more goals.

- Uncertainties: Decisions are always taken using incomplete and uncertain information. Thus, when making a decision, the agent should consider uncertainties and manage risk appropriately. The agent must understand and consider the surrounding environment with regard to the completeness of information related to the specific decision.

- Emotions: An agent must have the ability to evaluate the importance of emotional factors as they affect the decision, as well as the ability to integrate those factors into the decision itself.

1.3 Multi-Agent Perspective

The previous section discussed the main motivations behind this study and explored the notion of Trust from a philosophical perspective. We have argued that Trust is a product of belief and of emotion. Overall, these lead to a mapping of the elements of Trust to five main capabilities.

The class of decisions considered here are of two types: the critical type (such as allocation of budgets, hiring, procurement, general strategic directions and promotions). For these we currently make use of intelligent systems for suggestions, but not actual decisions; hence, automating those processes is not yet a possibility. The other type of decisions we address are sensitive decisions. For example, agents may be used in a personal-assistant scenario to choose books or movies for an individual based on historical preferences. Amazon.com runs intelligent computations that provide suggestions built from your profile, but these computations cannot be trusted to make actual decisions on our behalf (I will never let Amazon buy a book for me to read without consulting me first). These decisions also, then, are not unthinkingly accepted.

The term agent in this thesis means autonomous intelligent agents which can be defined as:

> "An agent is a computer system that is situated in some environment,
> and that is capable of autonomous action in this environment in order
> to meet its design objectives". [73][Page 4]

Aristotle [11] took Practical Reasoning to be reasoning that concludes in an action. This can be translated as the ability to reason intelligently within the environment about what actions are best.
Another view is given by Bratman [38] where Practical Reasoning is all about choosing a response as a result of interactions.

Moreover, this response is selected based on the agent's goals, beliefs and values. For Searle [116], values are the principles, standards, and aspirations that guide human actions. Actions are not selected only by facts; subjective values also play a role. Individuals do not always make decisions based solely on facts and figures. Morals, integrity, cultural values and other elements play a role, particularly with respect to how people prioritize their goals and choose between options.

Another point to consider is that values are not built-in; rather, they evolve from circumstances within the external world and dynamically change with time and experience. Thus, one's values and their importance can also be seen as a memory of expertise or experience. For example, babies have no regard for the value of friendship, but as they grow older, they start to interact with their parents and give more importance to the value of family. They feel hunger, adding the value of food. When meeting other children, the value of friendship is introduced. With this mix of values, those young children sometimes face situations where they need to decide whether to eat a meal or go play with their friends. In such situations, they begin to order their values. If they chose to play with their friends, the value of friendship is promoted to be of more importance.

The order of values differs among individuals. While the youngster in the previous example preferred friendship over food, another one might prefer the opposite, thus, explaining the rational disagreement of Searle and Perelman that different audiences will accept different arguments and make different choices [97][116].

Although this work is not directed toward the multi part in Multi-Agent Systems, there is a very nice slogan by Wooldridge [130]: "There is no such thing as a single agent system". Wooldridge suggests by this slogan that self-interested agent systems must address the question of adapting to the world and to other agents; their reasoning and actions alone are never sufficient to make any conclusions. A methodology of decision making for a self-interested agent cannot be considered sufficient unless it addresses how other agents will react in the environment. Thus while we do not address collective decision making or negotation, we do take account of the behaviour of other agents in determining the consequence of decisions.

Thus, it is very important to remember that consequences do not depend merely on actions, but also on how other agents react. For example, repeating an action does not guarantee the same results. There are a number of key elements to consider. Wooldridge mentions the following:

1. Agents are designed by different individuals with different goals and interests. Therefore, the interactions between these agents are considered

negotiations where agents are self-interested.

2. Agents are assumed to be acting autonomously. Decision making is not fixed in agents, but rather computed and reasoned about dynamically as they respond to situations.

Therefore, self-interested agents need to behave robustly in the presence of inconsistency or, as in the terminology of Lesser and Corkill [79], be functionally accurate. Lesser and Corkill [79] further identify the following characteristics:

1. Problem solving is not confined to a sequence of actions; rather, it progresses opportunistically.

2. Agents communicate higher level plans and results rather than raw data.

3. Uncertainty and inconsistency are resolved during the reasoning process with partial information.

4. The solutions are not tied to a single route, but should have many ways to achieve the same result. Agents require strategies (in the same theoretical sense) rather than single plans.

An agent works autonomously and decides what goals are to be pursued and how they can best be achieved. Hence, trusting an agent will not only involve believing that he has the capacity to solve problems, but also that he has the intention/motivation to then act as expected. To be trusted, the agent must be able to choose the goals, plan to achieve them, consider alternatives, choose the best alternative if needed and make sure that this process is acceptable to the audience. And if things do not go as expected, the agent should be able to resequence and set a different course of action.

There are different approaches to decision making. The approach we adopt in our methodology is based on argumentation. In reasoning for the best decision, the agent will consider different routes to achieve the goal and build arguments justifying each route. Those arguments are associated with the different values that accepting them will promote. Ordering of those values will then give us a criteria to set preferences among different arguments. This ordering will take account of how important the values are and the emotional aspects of the situation. The reasoning process should also account for the behaviour of other agents as the outcome of any action is dependent on the joint action of the agent and other agents situated in the environment. Hence, other agents behaviours and uncertainties must be accounted for.

1.4 Thesis Contribution

This thesis will first introduce a decision-making methodology that considers the belief aspects of decision making. Our contribution will:

1. Use argumentation as a way to model Practical Reasoning through an instantiation of an argumentation scheme as a presumptive reasoning for action.

2. Consider joint actions to address how the environment and other agents would react in conjunction with the agent's own actions.

3. Address uncertainties in the methodology and allow it to build decisions when information is incomplete about the environment or unexpectedness in actions outcomes.

4. Exploit the role of emotions in the process of decision making.

5. Examine the effect emotions would have on the decision-making process compared to social aspects by providing a mechanism to balance between the influence of emotions and beliefs.

6. Present a detailed example study.

We first introduced a methodology of decision making from beliefs perspective in [86] addressing the first three points above. This was then complemented by another paper [87] that added emotional aspects to the methodology, addressing the fourth and fifth points above.

1.4.1 Areas of Further Development

This thesis provides a methodology where emotions can be embedded and considered part of the decision-making process. It does not explain those emotions or discuss their meanings.

The principles given by this methodology considers short- and long-term visions of the agent in the decision-making process from the perspective of goals and values. It does not cover planning aspects in depth.

The value order of the audience in the decision-making framework is very dynamic in the sense that it changes according to events occurring in the environment. Thus, this value order is basically a product of many past events, allowing us to consider it as a store of historical information. Mechanisms to allow the agent to look into historical information and consider it in this process are beyond the scope of this study.

1.5 Thesis Structure

This thesis is structured into seven chapters as follows:

Chapter 1 presents the research motivation and a general overview of the topic. **Chapter 2** provides the literature survey where relevant work is discussed. **Chapter 3** introduces the proposed decision-making methodology.

Chapter 4 is a case study of the aspects introduced so far. **Chapter 5** extends the decision-making methodology to include emotions. **Chapter 6** is an extension of the case study that incorporates emotional aspects. **Chapter 7** concludes the study and summarises the main outcomes, discusses the findings of the examples, and aligns them with the main objectives of the work. **Appendix A** offers snapshots for the implementation of the case study.

1.6 Summary

To summarise, this thesis introduces a decision-making methodology by which agents can consider rational, social and emotional aspects relevant to the decision. This chapter has introduced the main capabilities addressed by this methodology and gave an overview of the relevancy with the field of Multi-Agents. The next chapter reviews the literature relevant to this work.

Chapter 2

Literature Review

This chapter discusses the literature relevant to this thesis. The chapter details references used to build up the methodology and also discusses work that has approached the same issues differently.

Section 2.1 provides the philosophical motivation. Section 2.2 provides the essential background to the field of Multi-Agent Systems. Section 2.3 discusses approaches to Trust in Multi-Agent systems. Section 2.4 gives an overview on Planning in Artificial Intelligence. Section 2.5 focuses on argumentation in Artificial Intelligence (AI). Section 2.6 reviews the literature on emotions. Section 2.7 considers two theories in the decision-making field. Finally, Section 2.8 concludes with a summary of this chapter. And finally Section 2.9 introduces Part I of the thesis (Chapters 3 and 4).

2.1 Trust

This section explores Trust and trustworthiness in order to establish an understanding of what it means to trust a person, or an agent. This is done by translating the concept of Trust from the philosophical and psychological worlds to AI. We can thus identify the five capabilities of the decision-maker that contribute to trustworthiness.

Bailey [24][Page 1] gives an example from Plato's Republic:

> *"Plato recounts a dialogue between Socrates and Glaucon, Plato's older brother. In it, Glaucon argues that only the fear of detection and punishment prevents a human being from breaking the law and doing evil for the sake of his own self-interest. Glaucon thinks that this natural fact is demonstrated by the shepherd Gyges, who found a*

> *gold ring which made him invisible whenever he twisted it on his finger. On realising the ring's power, Gyges used it to seduce the queen, murder the king, and take the throne. Glaucon's claim then, is that every one of us, however law-abiding and good we might seem, would do as Gyges did, or something else in our self-interest, if we could avoid detection and punishment. And, Glaucon claims, we would be right to do so, since each human being's only interest is their own self-interest, and we have no interest in justice and morality for their own sakes".*

This idea of self-interest that Plato presents raises the question of when we can trust someone else. Is fear of detection the only motivation to do so and why should an agent fear detection?

A trustworthy person is someone we can place our Trust in and rest assured that he will not betray us. Bailey [24] also argues that another important attribute of Trust is that when we trust others, we rely on them to take care of important matters; thus, they have the capacity to help us, as to harm us. Placing trust means that we rely on the trustee to take care of resources which in turn they could misuse.

Faulkner [56] defines Trust as a prediction of reliance on an action, based on what a party knows about the other party. Also, Trust is always about the unknown. If we are guaranteed a certain result, trust is of no importance. On the other hand, the more uncertain we are about the outcome of an action, the more important trust is in the relationship.

A more detailed view is offered by Coleman [44] who suggests that Trust involves the following principles:

1. *"Trusting an entity allows for actions that cannot be done otherwise".*
   If we trust an agent, it may identify options that we do not have the capacity or time to find for ourselves.

2. *"The truster is better off than if he had not trusted the trustee".*
   This suggests that the agent will save us time, or find better options.

3. *"Trust involves transfer of resources".*
   If we are to profit from the agent, we must devolve some control to it, and we must trust the agent to be trustworthy.

2.1.1 The Capabilities Needed for Trusted Decisions

The work of Ramchurn et al.[101] examines the specific roles of Trust in Multi-Agent Systems, in particular with respect to agent to agent interaction. Ramchurn et al. defines reasons for Trust as the belief in competence, willingness, persistence and motivation. Frijda [62] also mentions the importance of emotions in establishing the satisfaction of the course of action chosen by the trustee, the influential factors on emotions, and the influences of emotions on intentions to act.

Now, to bridge the gap between philosophy and multi-agent systems' we capture the understandings of Trust and offer the points below as the main capabilities for a trusted decision in agent systems.

1. Practical Reasoning/Competence:
   This is the first and basic prerequisite of trust; to be trusted, the agent must be able to evaluate options and their likely effects to be able to evaluate and decide on the effects of the decisions on the trustee.

   A trustee will always have many possibilities for a course of action to choose from. So, the very first basic capability of our methodology is that it should consider options and their likely effects.

   For example, if I know that a computer shop has a single model of computers, there is no point in giving weight to how trustworthy the salesman is, as I can only buy that model. However, if I know that this shop has many varieties and I decided to give the salesman full authority when choosing a computer for me, it becomes more important to trust him.

2. Social Interaction:
   There is no point in trusting an entity unless we know that with this power of trust it will perform better than we could do ourselves. We need to know that the trustee will actually perform the actions that would result in the desired goals which the truster sees himself better off delegating. However, we need to understand that it is not our actions alone that bring about desired results. In order for anything to happen, we expect other behaviours

from the entities around us which we hope will yield the expected results. For example, if I want a piece of information from the university archives, I can go to the store and spend hours looking for the information, but I will be better off trusting the database system to provide the information I need. This means that our trustee should understand the environment and be able to predict how it will react (whether other agents or objects in the environment) to his actions to achieve the desired goals. In summary, Trust requires *Social Interaction.*

The agent should be able to interact with the environment around him to understand and also persuade or influence. It is the understanding that our actions are never enough for anything to happen and the result of any action will depend on the result of other agents interacting within the same environment.

3. Planning:

   A trustworthy agent controls resources; hence, this agent must not be short-sighted, but should work purposively and with a long-term strategy. Planning also allows for complex and difficult goals to be considered.

   Trust becomes crucial when we transfer resources. These resources may be physical, financial, intellectual or temporal. Transferring the decision-making ability -the focus of this study- means placing the power of commitment with the trustee and that the trustee has full power to commit resources. The trustee should have full understanding of the truster's resources and his willingness to commit these resources. Of course, this allocation of resources depends on their importance in relation to the goals. The trustee should understand how strategic and important the goal is compared to the importance of the resources he is about to commit. The trustee should also be able to look at the overall picture when making a commitment. While the goal itself may not justify the commitment, achieving this goal may make a more important goal possible in future planning. Consideration of whether there should be doubts if the goal to be achieved merits the expenditure of resources required must also be given. Achieving the goal facilitates the realisation of future goals.

4. Uncertainties/The Unknown:

   Agents often work with incomplete information, inconsistent beliefs, and uncertain outcomes. A methodology of action selection must address un-

certainty.
With perfect information, an agent can be automated. However, in a world of incompleteness and uncertainties, trust becomes important. In other words, trust can be placed in situations where the agent will not be able to achieve the goal without trusting other agents. For example, I will not be able to fly to a destination unless I trust the pilot. This can also be seen as when we are in a state where we have uncertainties; an action cannot be made without a certain degree of trust in the trustee. This is another capability our trusted agent should have.

5. Emotions:
   We have considered trust based on beliefs where we trust someone when we only believe he has the capabilities to deserve our trust. We cannot, however, always see Trust as based only on belief and facts: trust is sometimes irrational and a product of emotions.

   So, is trust a belief, i.e, a belief in someone's trustworthiness? Many philosophers presume that it is not [36, 69]. Holton [69] gives the example of trusting a friend to be sincere without believing the friend will be sincere. Arguably, if one already believes that to be the case, one would have no need to trust the friend. It is also possible to believe that someone is trustworthy without trusting that person, which suggests that trust could not merely be a belief in someone's trustworthiness.

2.2 An Introduction to Multi-Agents

The field of Multi-Agent Systems emerged from Computer Science [113, 125] to address the continuing need for intelligence [128, 130]. From a high-level perspective, agents are computer systems that have the capacity to take actions without any prior instructions and in response to the environment. Agents are able to act in unfamiliar environments and deal with unclear situations that they might not have clear instructions on how to handle. All of this is done by providing those agents with a reasoning power to enable them to figure out what to do in each and every situation rather than prescribing it. According to Wooldridge [130][Page 15]:

> "An agent is a computer system that is capable of independent actions on behalf of its user or owner to satisfy its design objectives. A

> Multi-Agent System consists of multiple agents that interact with each other. These agents will thus require the ability to cooperate, coordinate and negotiate".

Agents are autonomous entities that do not work on hard-coded programs but are rather able to make decisions independently. This autonomy gives an agent the power to adapt to unexpected events in its environment and where situations require a decision and information to take that decision are incomplete in the light of unknowns in the environment. This autonomy also allows for the social and emotional aspects of Trust to be modeled in an agent.

Meyer [80] summarises agents' properties into four parts:

1. Agents are situated in an environment.

2. Agents are able to react to unexpected situations arising in the environment.

3. Agents can proactively set and pursue their own goals on behalf of a user.

4. Agents are able to cooperate and socialise with other agents to achieve their own goals.

We use a combination of Meyer's [80] and Wooldridge's [130] views as our definition of an agent. So, an agent is a computer system that is capable of independent actions on behalf of its owner to satisfy its design objectives. An agent would have the ability to react to unexpected situations, sets and pursues its own goals and can cooperate with other agents in the system to achieve its own goals.

2.2.1 Belief-Desire-Intention (BDI)

Although not adopted in this thesis, BDI is a popular example of a reasoning architecture in rational agents. Bratman [38] introduced intentions, which is a state of affairs that an agent has chosen to commit to realising. This is commonly referred to in Multi-Agent systems as the BDI model. The BDI model of human practical reasoning was developed as a way of explaining future-directed intentions.

One use of BDI was presented by Georgeff and Lansky [65], who proposed a system, called Procedural Reasoning System (PRS), for reasoning about complex tasks in dynamic environments.

PRS allows agents to consider not only goals but also beliefs and intentions; PRS can also reason about its own internal state giving it the power to change its own beliefs, desires and intentions. The main components of PRS are a database that stores beliefs, a set of goals (desires and intentions), a set of procedures on how to achieve goals (plan library), namely "Knowledge Area", and, finally, an intention stack, which is basically all the agent's current intentions (the desires to which it is currently committed or in other words its current plan). This system has been implemented as a robot system in the role of an astronaut's assistant. For example, if this robot was asked to fetch a wrench, it would calculate the procedure and perform it. If in the case of fetching the wrench a more important event happened, such as an accident in the spaceship, it would reason that it must abandon its current mission and react to the accident. Moreover, if it went to the place where the wrench should be and failed to find it, it would also reason as to where it might have gone and would be able to either return and report it was missing or keep looking in more places where the wrench is likely to be [128].

Another use of BDI was presented by Wooldridge [127] where he introduced a logic of BDI for describing and executing plans $\mathcal{L}$. Rao and Georgeff [102] introduced a modeling system to be used within BDI agents, which is basically a formalisation of intentions based on a branching-time of possible worlds.

Rao and Georgeff [102] model the world using a time tree, which is a temporal structure with a branching-time future. Events then transform from a state true at a time point to a state true at the next time point. Branches represent the different choices the agent has at each point of time. Thus, if the agent at one point has two branches labeled e1 and e2 respectively, this means that if the agent chose the action presented by e1, he will move along the branch to the time point at the end of e1.

Other agents, however, might lead to an unexpected state giving rise to indeterminacy.

Rao and Georgeff summarise their approach:

> "We enforce this notion of compatibility by requiring that, for each

> belief-accessible world w at a given moment in time t, there must be a goal accessible world that is a sub-world of w at time t". [102][Page 474]

This is referred to as "Strong Realism" and builds upon a notion of realism of Cohen and Levesque [43] who consider how the prevention of agents from adopting a goal that is believed to be unachievable.

There are three main elements that Rao and Georgeff address in their formalisation:

1. Intentions are treated as first-class objects on a par with beliefs and goals and cannot be reduced.

2. They make a distinction between the choices available to agents and the possibilities of outcomes from those actions.

3. They defined an interrelationship between beliefs, goals and intentions. This helps in situations so that aspects of the outcomes which are not desired need not to be seen as commitments of the agents.

2.3 Approaches to Trust in Multi-Agent Systems

In this thesis, the concept of Trust means the generic trust that we place as users of the system in the complete agent system to take over more responsibilities. Nonetheless, the concept of Trust appears also in multi-agent systems as establishing the trust among agents in a distributed system.

2.3.1 Trust in Multi-Agent Systems

The concept of Trust in multi-agent systems concerns the need for Trust among several autonomous agents. The need for trust among agents is mainly linked to uncertainty: In the absence of information, Trust becomes more significant to achieve the agent's design objectives.

The need for Trust in multi-agent systems connects directly to its definition and mainly the concepts of autonomy, flexibility, uncertainty and the need to react in a dynamic environment [117]. Two challenges in specific were identified by Ramchurn et al [101]. First, agents are likely to represent different stakeholders with different goals and aspirations. Second, given that such a system is open, agents can join and leave at any given time avoiding consequences of their actions.

There are two ways to look at Trust: first, individual-level trust, where agents can trust each other by reasoning about the reliability and honesty of the other agents. This mainly occurs as the agent gets into several iterations of actions with the other agent and can learn from the results of those iterations. Also, individual-level trust can be acquired by reputation (the agent asking other agents in the environment about the trustworthiness of the trustee agent). Moreover, the agent could reason not about the trustee's honesty but could also reason about its motivations and based on which decide on its trustworthiness [101].

Second, system-level trust where the rules of the system are sufficient to ensure Trust in interaction shifting the Trust from the individual to the system. Such systems can enforce Trust by eliminating any gain that results by lying, the reputation of the lying agent can be spread by the system or agents can be audited as they enter the system for reliability and perhaps asked for a third-party reference [101].

2.3.2 Delegation

Like Trust, delegation can be defined as entrusting a representative to act on your behalf [91]. Agents act autonomously and with a certain degree of flexibility, nonetheless, most probably an agent would rely on other agents to achieve some goals or in other scenarios those goals can be achieved easily and with less cost if the agent had the ability to delegate some of its tasks.

Delegation can be done to a single agent or a group of agents and in the case of issuing an imperative to a group of agents, this can either be distributively, for example "all of you stand up" or collectively, for example "One of you shut the door" [90]. Those imperatives might not necessarily be actions to be performed or propositions to be achieved, they can be delegations of responsibilities for actions in progress or a state of affairs to be achieved.

Norman et al. [91] gives a number of possible distinct imperatives that a logic of delegation can consider:

I don't care who achieves (state of affairs) A, but it must be achieved
I don't care who does (action) α, but it must be done
You must, through your own, direct, intervention achieve A
You must, through your own, direct, intervention do α
You must ensure that A is achieved by someone other than yourself
You must ensure that α is done by someone other than yourself

The above imperatives revolve around either the distinction between actions and state of affairs and whether or not to permit/forbid/enforce further delegation. The work of Hamblin [67] introduces and discusses Imperatives extensively. Hamblin indicates five features for an imperative: a time scale, a distinction between actions and states, physical and mental causation, agency and agent reduction and intensionality.

Norman et al. adapts the concepts of Hamblin and develops a model of delegation [91]. This model considers the distinction between actions and state of affairs and monitors agent's performance in terms of delegating tasks. This was then extended by Norman et al. [90] to account for group delegations for the difference between distributive and collective groups. The work in [92] then gives a detailed logic of delegation with a detailed analysis of the logical, normative and inferential aspects of agents.

2.4 Planning

Planning is a key consideration in achieving complex goals and it goes beyond just prerequisites of those goals to considerations of joined planning to achieve multiple goals and aligning among the different ways of achieving those different goals. An important aspect of planning is that a plan specifies the steps to be executed to reach a goal, but does not necessarily specify the full order of execution. So, a Plan is a partial order on steps mostly referred to as "nonlinear planning" [114].

Chapman introduced a new planner "TWEAK" [40]. TWEAK has three layers: a plan representation, a way to make a plan achieve its goal and a control structure. One concept in this model is incompleteness where the model starts with an incomplete plan and builds toward its completion. This incompleteness is either from a temporal perspective where the steps are known but how they are ordered is not or codesignation constraint where steps themselves are not specified.

Some of the common terminologies in AI planning are: situations that are prerequisite to other situations are *establishers*, disallowed steps in the systems are *clobbers* and steps that would make clobbers legal are *white knights*. In planning situations (or what we call states) are a set of propositions. A *problem* is basically the initial situation and the final situation. A *goal* is the set of propositions of the final situation of the problem.

Most of the work that is taking place nowadays on planning can be routed back to either: means-end analysis (GPS) of Newell et al. [88] and the action model (STRIPS) of Fikes et al. [57]. The domain-independent approach of planning though started with HACKER of Sussman [120] and nonlinear planners with NOAH of Sacerdoti [114].

2.5 Argumentation in Artificial Intelligence

Argumentation is a method which can be used to perform practical reasoning. We use argumentation when negotiating with another agent to promote our ideas, and we also use argumentation to justify our own actions; where we weigh different arguments, evaluate risks and see potential outcomes and then come up with a justification good enough to make a choice. Agent communication through argumentation is an approach widely adopted in AI [104, 33, 31, 100, 41, 126, 42]. We will now review the literature behind formalising argumentation.

2.5.1 What are Arguments?

Argumentation can be defined as: Argumentation is a set of assumptions where we can derive conclusions [33].

Arguments are useful in handling conflict and uncertainty in the knowledge base and can also be used for justification in decision making. In general, they provide a way to approach non-monotonic reasoning [50].

A functional model of the evaluation of arguments, taking seriously the procedural and dialogical aspects of argumentation was addressed by Gordon et al. [66] which provides a mathematical structure to determine the acceptability of statements. Arguments in practical reasoning to perform an action can be accepted when providing premises on which a goal can be achieved through an action. For example, by doing X, I will achieve Y then I should perform X [123].

Argumentation depends on argumentation schemes which are stereotypical patterns of reasons that presumptively justify a conclusion, subject to satisfactorily addressing the critical questions of the scheme.

The methodology used in this thesis to address argumentation in practical reasoning is by regarding practical reasoning as a species of presumptive argument based on a particular argumentation scheme: given an argument, we have a presumptive reason for performing the action. This presumption however can always be challenged and withdrawn [12] in the light of the critical questions associated with the scheme. Challenging those presumptions is the way we can identify alternatives, weigh differences, identify risks, and address inconsistencies.

The argument scheme forms the presumptive reason to justify an action. This presumption will include premises and conclusions and could be challenged through critical questions in five different ways: denial of premises, existence of better alternatives, existence of side effects, interference with other actions and finally impossibilities relating to the conclusion.

2.5.2 Abstract Argumentation

An established method of abstract argumentation was proposed by Dung [50] in which an argument is accepted and assessed according to the other arguments which attack it and their attacker.

An argumentation framework is a finite set of arguments X, and a binary relation between pairs of those arguments called an attack. Argument A is set to be acceptable with respect to a set of arguments S if any attack by any other argument is defeated (i.e., some argument in S attacks the attacker). A set S is conflict-free if it does not contain any pair of arguments that attack each other. Dung then introduces the notion of admissible Preferred Extensions, as maximal admissible sets of arguments in a framework. An extension to Dung's framework, which is used here as the basis of the proposed methodology, is Value-Based Argumentation Framework (VAF) of Bench-Capon [28]. The basic principle in VAF is that it allows arguments to be evaluated according to the social values they promote by considering those values in the analysis of the evaluation.

2.5.3 Argumentation and Practical Reasoning

Aristotle [11] gives an example of practical syllogism:

Dry food suits any human.
Such-and-such is a dry food.
I am a human.
This is a bit of such-and-such food.
This concludes to:
This food suits me. [103](Page 33)

Practical Reasoning is reasoning about what should be done according to some set of criteria. These criteria often compete with each other and their importance is intrinsically subjective; thus, what should be done is always relative to a particular agent in a particular situation from a particular perspective. Hare emphasises the importance of motive:

> "To get people to think morally it is not sufficient to tell them how to do it; it is necessary also to induce them to wish to do it".[68][Page 224]

One motive behind reasoning and deciding is suggested by Damasio, who says:

> "It is perhaps accurate to say that the purpose of reasoning is deciding and that the essence of deciding is selecting a response".[45][Page 165]

This can be translated as the ability to reason intelligently in the context of a particular environment about what actions are best. A more formal definition is the one given by Bratman [38][Page 17]:

> "Practical reasoning is a matter of weighing conflicting considerations for and against competing options, where the relevant considerations are provided by what the agent desires/values/cares about and what the agent believes".

So, Practical Reasoning is about intelligently selecting a response by interacting with the environment. Moreover, this response is being selected based on the agent's goals, beliefs and values. Searle mentions:

> "Assume universally valid and accepted standards of rationality, assume perfectly natural agents operating with perfect information, and you will find that rational disagreement will still occur; because, for

> example, the rational agents are likely to have different and inconsistent values and interests, each of which may be rationally acceptable". [116][Page xv]

Different approaches took place to model argumentation in logic and BDI, one of which is by Bench-Capon et al. [32] where the positive and negative side effects are captured by accruing the applications of the practical syllogism. Another approach is by Amgoud and Prade [8] where they build a methodology to logically instantiate arguments from a knowledge base proposing characteristics of the attack relation among arguments addressing possible conflicts. Amgoud and Prade [7] incorporate the notion of strategy in modeling dialogues between autonomous agents based on the beliefs and goals of the agent. Rahwan and Amgoud [99] provide an argumentation framework for BDI agents that generate consistent desires and plans for achieving those desires through arguing about beliefs, desires and plans to evaluate the worth of desires and cost of resources. We will however base our approach on the use of argumentation embodied in [14].

Audiences

Following on Searle's view of rational disagreements, it is clear that differences would sometimes occur between rational self-interested agents because each agent would have a different view of the world around it and different values and interests. Each point of view might be rationally acceptable. When an action is chosen, whether the argument justifying the choice of action is acceptable or not really depends on the aspirations and values of the agent to which it is addressed, what Perelman[97] calls the *audience*.

> "Each [party] refers in its argumentation to different values...the judge will allow himself to be guided, in his reasoning, by the spirit of the system i.e., by the values which the legislative authority seeks to protect and advance".[97][Page 152]

This view of the audience has the significance of shifting the focus from the beliefs of the speaker to those of the audience accounting for differences in opinions. [12, 27]

Hunter [71, 72] has considered the notion of audiences in AI from the basic idea that different audiences will have a different understanding of the environment. He uses an argumentation framework to evaluate the believability of arguments and build a preference model where arguments for an action can be compared to others according to the audience's view. The work in [47, 48, 89] also uses audiences to determine preferences.

Values

In sociology and philosophy, values are the principles, standards, or qualities that guide human actions. Actions are not selected only by facts (not all rational people make the same choices). Independent values always play a role. We do not always make our decisions and reason based on facts and figures; rather, morals, integrity, cultural values and other factors play a role.

Values are not built-in; rather, they evolve from dealing with the external world and are dynamic and change with time and experience. So, values and how one rates their importance can also be seen as a memory of expertise; or the conclusion of a life-time experience.

Argument Scheme

Arguments are directed to a specific person or persons at a specific time. This implies that arguments can be evaluated differently in different situations/states.

Walton [123] provides an argumentation scheme justifying actions in terms of their consequence. This was elaborated in [12] in the following scheme:

In the current State R
Perform Action A
To get to State S
Achieving Goal G
Promoting Value V [12]

Arguments in practical reasoning suggest that an action should be performed when providing premises at a certain state on which values can best be promoted when achieving a goal through an action [12, 31].

We live with uncertainties and incomplete information where we cannot have the facts necessary to determine an action. In such a case, argumentation can be treated as a mean to reach a conclusion. Moreover, even if we assume that we live in a perfect world where information is complete and available, we will still have rational disagreement between agents with different interests where argumentation can play the role of mediation to reach the best acceptable agreement, given the subjective perspective of the agents concerned.

Critical Questions (CQs)

In the previous subsection, we identified our argumentation scheme on the basis of presumptive justification of action, where we regarded practical reasoning as species of presumptive argument. By this approach we have the ability to challenge and withdraw the presumptive justification. An agent who does not accept an argument can challenge and provide counter arguments to show that it is faulty. The attack on practical arguments can come from four critical questions in Walton [124]. An attack can promote alternative directions, question the truthfulness of the premises upon which the argument was built, argue the importance of the goal the argument is trying to achieve and mention destructive side effects. Atkinson expanded these concerns into sixteen Critical Questions (CQs) reflecting a more nuanced view of the scheme, which can be seen as the guidelines to define attacks later in the Value-Based Argumentation Framework (VAF)[12].

- CQ1: Are the believed circumstances true?

- CQ2: Assuming the circumstances, does the action have the stated consequences?

- CQ3: Assuming the circumstances and that the action has the stated consequences, will the action bring about the desired goal?

- CQ4: Does the goal realise the value stated?

- CQ5: Are there alternative ways of realising the same consequences?

- CQ6: Are there alternative ways of realising the same goal?

- CQ7: Are there alternative ways of promoting the same value?

- CQ8: Does doing the action have a side effect which demotes the value?
- CQ9: Does doing the action have a side effect which demotes some other value?
- CQ10: Does doing the action promote some other value?
- CQ11: Does doing the action preclude some other action which would promote some other value?
- CQ12: Are the circumstances as described possible?
- CQ13: Is the action possible?
- CQ14: Are the consequences as described possible?
- CQ15: Can the desired goal be realised?
- CQ16: Is the value indeed a legitimate value?

CQ1-CQ4 concern the denial of premises while CQ5-CQ7 explore alternatives.

CQ8-CQ10 address side effects. CQ11 deals with interference with other actions. CQ12-CQ16 are disagreements relating to impossibility.

Action-Based Alternating Transition System (AATS)

A commonly used logic in Multi-Agents Systems is Alternate-Time Temporal Logic (ATL) [1], which is a logic of cooperative ability intended to support reasoning about the powers of agents and the coalition of agents in game-like Multi-Agent Systems. This was extended by Wooldridge and Van der Hoek [132] to Normative Alternate-Time Temporal Logic (NATL), which allows representation of normative aspects. A normative system is a set of constraints on the actions that may be performed in any given state and defines whether or not the action is legal for every possible system state and action. The semantic structure underpinning ATL is known as Action Based Alternating Transition Systems (AATS).

AATS was introduced in Wooldridge and Van der Hoek [132] as a foundation to formally describe a system in which several agents combine to determine the transition between states. In Wooldridge and Van der Hoek [132], an AATS with n agents is an (n+7) tuple. This was then extended by Atkinson and Bench-Capon in [14] to include the notion of values so that each agent has a set, Av, of values drawn from an underlying set of values V, and every transition from the set Q may either promote, demote, or be neutral with respect to those values. The states of the AATS represent possible states of affairs and the transitions between them through joint actions, that is, actions composed from the individual actions of the relevant agents. Thus, if two agents can each independently choose one of three actions, there will be nine possible joint actions. Atkinson and Bench-Capon [14] extended the AATS model to include and describe values. A full description of AATS is given in Chapter 3.

Value-Based Argumentation Framework (VAF)

VAF was introduced by Bench-Capon [28] to represent rational disagreement by extending Dung's Argumentation Framework with the inclusion of values. Many studies took place on VAFs in the past few years [31, 32, 29, 82].

A VAF is defined by a triple $\langle H(X, A), \nu, \eta \rangle$, where $H(X, A)$ is an argumentation framework, $\nu = v_1, v_2, ..., v_k$ a set of k *values*, and $\eta : X \rightarrow \nu$ a mapping that associates a value $\eta(x) \in \nu$ with each argument $x \in X$. A *specific audience*, α, for a VAF $\langle H, \nu, \eta \rangle$, is a total ordering of ν. We say that v_i is preferred to v_j in the audience α, denoted $v_i \succ_\alpha v_j$, if v_i is ranked higher than v_j in the total ordering defined by α [54].

One significant benefit of VAFs is that they allow differentiation between audiences. By associating arguments with values, we are able to give weight to different arguments, thus allowing us to choose whether an attack is successful or not by comparing the importance of the value promoted by each argument.

The Preferred Extension of a VAF is then the maximal conflict-free subset S of arguments so that no argument defeats any other argument in S for that audience.

2.6 Emotions

> "The influence of emotions upon beliefs can be viewed as the port through which emotions exert their influence upon human life. Beliefs fueled by emotions stimulate people to action, or allow them to approve of the actions of others in political context". Frijda [61][Page 1]

Emotions play a big role in our everyday decision making. One can argue that emotions should not be an important part of a decision-making system. However, we do not live in a perfect world where everything about the environment is known and all behaviours from others are predictable. Rather, because we live in a world that is full of uncertainties, we cannot be sure about our actions given only the rational aspect of evaluating them. Damasio suggests:

> "Emotions and feeling, along with the covert physiological machinery underlying them, assist us with the daunting task of predicting an uncertain future and planning our actions accordingly".[45][Page xiii].

Moreover, the combination of all emotional aspects makes us distinctive in who we are and hence how we Trust each other. So, emotional values can help in the confidence issues that we raised earlier. Building an agent with emotional features gives it an identity that one can trust.

> "Emotions and feelings can cause havoc in the process of reasoning under certain circumstances. Traditional wisdom has told us that they can, and recent investigation of the normal reasoning process also reveal the potentially harmful influence of emotional biases. It is thus even more surprising and novel that the absence of emotion and feeling is no less damaging, no less capable of compromising the rationality that makes us distinctively human and allows us to decide in consonance with a sense of personal future, social convention, and moral principle". Damasio [45][Page xii]

Ultimately, people do use emotions in decision making, where purely rational decision making is mistrusted as 'cold' or 'inhuman' is evidence that emotions must have some beneficial effect. If emotions had no evolutionary value, we would all be completely rational within our limitations.

Several approaches have been proposed to model emotions. The Emotional Belief-Desire-Intention (EBDI) model built by Jiang and Vidal [74] is based on the BDI model of Bratman [37] by incorporating an emotional function. Reilly's [107] decision-making model is dependent on emotions and targeted to expressing them rather than influencing a decision. Padgham and Taylor [94] present a system built to treat emotions with goal-oriented behaviours trying to build the personality aspects of agents. The work of other researchers [119, 46, 119] gives us a formalisation aimed at reducing the nondeterminism in an agent's decision making.

The methodology this thesis offers builds the emotional system within an argument-based model, giving it the ability to evaluate the arguments from both rational as well as emotional perspectives and, moreover, determine the degree to which decisions are influenced by emotions.

2.6.1 Frijda's Account of Emotions

Frijda [58, 60] claims the existence of six basic emotions (desire, happiness, interest, surprise, wonder and sorrow). Another important aspect of emotions is the level of their effect to the decision-making process. Frijda says:

> "As to the law of change itself: the greater the change, the stronger the subsequent emotion. Pleasure after suspense is considerably stronger than what the same event produces without prior uncertainty". [60][Page 11]

Frijda [58] gives a definition for the emotions process, beginning with the stimulus to influencing a decision. Emotions are also preconditions for ensuring the satisfaction of the system's major goals [62].

> "Emotions alert us to unexpected threats, interruptions and opportunities" [59][Page 506-507]

Another important aspect that Frijda [58] mentions is that the degree of difference in prior expectancy and stimulus affects the reaction to the event.

2.6.2 The Ortony, Clore and Collins (OCC) Model

Another factor to consider in decision making is the personality of the agent or simply the different emotions the agent possesses and how they are related.

Ortony, Clore and Collins (OCC) [93] addresses the structure of emotions and how they are related and identifies twenty-two emotions organised into a hierarchy. The OCC model also provides a specification of the conditions which give rise to each emotion in terms of the notions of objects, events and actions. The OCC model includes notions of intensity of emotions, and identifies a number of variables that influence the intensity of the emotions experienced.

The OCC model is well suited for computational implementations. Other researchers (e.g., Reilly [110, 109, 106, 107]; and Steunbrink et al. [80, 81]) have implemented versions of these models. Reilly used OCC specifically to build agents that are *Believable*. His main aim was to have agents that can be useful for artistic nature as actors in a play for example. Steunbrink et al. purposes of modeling emotions are similar to ours in that they were aiming at improving decision making and they use OCC to implement a deliberation and decision language for agents to use. In this thesis, I model emotions and use OCC to basically support decision making through influencing the agents preference. The OCC model specifies a clear hierarchy of emotions where twenty-two basic emotion types are organised into groups. Each emotion is elicited through a defined set of conditions on the basis of actions, goals and values. This model has attracted computer scientists as it provides flexibility in implementation and expansion. It provides generic views of basic emotions that then can be extended to include complex emotions or shortened to a subset of those emotions. Detailed definitions of emotion types are given in Appendix C.

2.6.3 Reilly's Model of Emotions

Reilly [107] has introduced a methodology to implement the OCC model where emotion generators are the set of rules that take the input of the agent and environment and produce emotion structures. Emotion storage takes the emotion structures and defines how they will be combined and decayed, (i.e., how the agent will deal with more than one emotion structure at a time and how long it will take until this emotion fades away). Emotion expression is mapping emotions to behavioural features. Emotion actions are how emotions affect goals, plans and actions.

Reilly's objective in his model was to serve the art community by creating emotional characters for artistic modeling where the objective in this thesis is to use emotions to support decision making. Reilly identifies how emotions take place:

1. Cognitive-Appraisal Emotions: emotions happen as we evaluate the state.
2. Reflex Emotions: emotions as a direct reflex to a sudden change in the environment.
3. Social Contagion: emotions as a result of a common emotion by most of the surrounding agents.

In his model of emotions, Reilly introduces the notion of *threshold*, where different emotions have different weights that change according to events taking place in the environment. This change in emotions weights is different from one agent to another. This notion of threshold is important. Although we would like an agent to be accountable for the emotional aspects of decisions, it is important to be able to control how much the particular emotional response of an agent affects these decisions. We need the emotions to have an effect, but not for the agent to be so volatile as to make it entirely inappropriate.

Stuenbrink et al [119] presented a formalisation of OCC and formally defined the twenty-two emotions in terms appropriate to BDI agents. We use their definition (after adapting it to AATS) later on in Chapter 5 (definition 5.1).

2.7 Decision Theory, Game Theory, and Value-Based Argumentation

Other popular approaches to decision making include game theory and decision theory.

Game theory has been adopted as a mathematical approach to analyse interactions between self-interested agents [34, 131]. The basic concept behind game theory is that in order to find the best decision to be taken we must first evaluate the outcome of other actions performed by other agents in the environment, only then, we can guarantee a best decision [95]. One of the early adopters of game theory to solve coordination problems was Rosenchein [111] and in negotiation [77, 112, 115].

Decision theory is a set of mathematical techniques for making decisions about what action to take when the outcomes of the various actions are not known [96]. It can also be viewed as a one person game (for example, Bayesian networks in [10], the influence diagram in [70]).

Both decision and game theories describe the notion of the best action as the action that maximises the utility of the agent.

The notion of Value-Based Argumentation Frameworks (VAFs) is an extension of the standard Argumentation Frameworks originally proposed by Dung, which is able to show how rational decisions are possible in cases where arguments derive their force from the social values their acceptance would promote instead of maximising the utilities as in decision and game theories.

1. Acceptability
   Game and decision theories aim at finding the set of actions that will maximise the utility of the decision maker which can be viewed as the computation of probabilities and values for these actions; where VAF is aimed at finding the set of actions that is not necessarily the best, but rather accepted by the audience, hence, promoting the agent's trustworthiness.

2. Positions can be justified
   "Negotiating using a particular game-theoretic technique might make it very hard to understand how an agreement was reached" [130][Page 148]. The main aim of this study is to increase confidence in the decisions an agent makes. Hence, a proper justification of each direction is a necessity. The kind of decisions we will usually refer to are the decisions we allow agents to take on our behalf. Argumentation-based approaches give us the ability to relate to the decisions agents make.

3. Positions can be changed
   Utilities in game theory are usually fixed and assumed to stay the same during the whole process. From a human-rationale perspective, this is not true as we tend to change our values and preferences along with the negotiation process. These preferences are dependent on how secure they are and what kind of side effects they may have. Moreover, such changes need to be explicable as appropriate responses to events.

2.8 Summary

This chapter started by explaining the main motivation behind this study as we are trying to find a way to make us (humans) trust agents' actions and decisions more so we can assign more critical and/or sensitive tasks to them.

Intelligent autonomous agents are entities that have the capability to interact with other agents. Those agents do not work on predefined plans, but rather reason about their action given certain inputs as they go along. They have a sense of freedom where actions and plans are made by the agent for the agent to satisfy his design needs.

Trust based on beliefs, simply assumes that the trustee will never betray us and it includes transferring authorities and also understanding that this Trust we are placing on the agent is crucial to the decision process and our agent cannot do a good job without it. Moreover, when trusting someone we should be better off than not trusting him.

And then we argued that Trust is not always gained through beliefs and sometimes it is an emotional state. So, Trust can either be based on beliefs (proving that someone is trustworthy) or based on emotions (someone is just a friend of mine, or his smile and attitude show that he is trustworthy), or as we believe a combination of both. As even if we believed someone is trustworthy, we would not trust him with important decisions that might affect us if we have feelings of hatred toward him.

As a conclusion to the Trust discussion, we said that in order to build an agent which is trustworthy, the five elements below need to be incorporated in the decision-making process in this agent:

1. Practical Reasoning:
   The ability to evaluate options and their likely effects. We defined Practical Reasoning both in the eyes of philosophers and computer scientists. The importance and logic behind argumentation was then introduced along with the work of Dung in Argumentation Frameworks (AF) and then the extension of Bench-Capon in VAF. All of this was then formalised by Atkinson who also added the concept of Critical Questions (CQs). The formalisations were introduced based on AATS and the main basic definitions that are relevant to this study were also mentioned.

2. Social Interaction:
   The agent should be able to interact with the environment around him to understand and also persuade or influence. It is the understanding that our actions are never enough for anything to happen, and everything we do, we expect something else to happen for our desired result to happen. For example, I can decide to drive to work today, but this depends not only on me but also on the car actually moving as I expect it to.

3. Planning:
   Bratman believes that the ability to plan is what makes us human. A trustworthy agent will have control over my resources and hence we want to make sure that this agent is not shortsighted, but rather works purposively and with a long-term strategy that matches mine. Planning also allows for complex and difficult goals to happen.

4. Uncertainty and Incompleteness:
   If I know that the agent is working with perfect information and clear predefined actions with clear and assured results we will have no problem trusting that agent. Agents however always work with incomplete information, inconsistent beliefs and moreover are not always sure about the outcome of his actions. A methodology of action selection must address these problems and so address uncertainties. One aspect here is the importance of adapting to the world through learning. Another aspect is Atkinson's Critical Questions (CQs), which address uncertainties in argumentation by a full consideration of the possibilities.

5. Emotions:
   Emotions also are very important. This work will show how the methodology of decision making introduced can sufficiently accommodate emotions and how it can apply emotional theories and practices in VAF.

2.9 Introducing Part I: Trust as a product of Beliefs

Philosophy suggests that Trust is based on two considerations: the basis of beliefs is where our beliefs in the abilities, experience, ...etc of the other person to achieve a certain goal makes us trust him. The second is emotions where a person with emotional feelings and with the ability of weighing matters with emotional consideration is better trusted (for example, we trust our friends perhaps in some matters more than anyone else only on the basis of the friendship and emotions we have toward them). Beliefs are about whether someone can perform an action, where emotions relate to whether they will choose to do so.

The work from now on will be divided accordingly into two parts: Part I, which comprises the following two chapters, will introduce a methodology of action selection that would account for agents' beliefs (Chapter 3) and then gives a detailed experimental study (Chapter 4) where the methodology is tested and evaluated.

Part II, which will follow in Chapters 5 and 6 will extend this methodology to incorporate emotions and extend the experimental study to explore the effects of emotional considerations on the decision-making process.

Part I

Trust as a Product of Beliefs

Chapter 3

The Decision-Making Methodology

In Section 2.1, a trustworthy decision was described as a product for both beliefs and emotions. Five elements were identified for a trustworthy decision: the first four relate to decisions as a product of beliefs, and the last is related to emotions. This chapter addresses the first four.

A methodology of action selection using presumptive argumentation is introduced here. The methodology aims at reaching a decision that meets the agent's goals and aspirations based on its beliefs and value preferences.

The methodology has five steps, four of which are based on the work of Bench-capon and Atkinson [29], [22] and [12]; the fifth is introduced here to complement and address the aspects that this thesis considers.

As the work builds on Value-Based Argumentation Framework (VAF), Section 3.1 starts by proposing an extension to the argumentation scheme and various related issues. Section 3.2 describes the five-step approach used in the methodology of action selection, as well as the basic formalisations and a brief example for illustration. Section 3.3 summarises the work and its relation to the basic elements of Trust. The summary also connects the work of this chapter to the next.

3.1 Development of Value-Based Argumentation and Critical Questions

This section explains the contribution of this methodology to the existing research on VAFs. First, an overview of VAFs with an explanation of the major points of concern and how these are then addressed in this methodology is presented, followed by an introduction to the methodology.

3.1.1 Overview of VAF Elements

In practical reasoning, argumentation is used to justify proposals for action. Arguments for action often face counter arguments for different proposals that defeat them. On the other hand, weak arguments might be supported by other stronger arguments. So, due to these stronger arguments, weaker arguments become justified. For example, when a researcher writes a paper, this action may not yield any beneficial result unless it is supported by another argument: this paper should then be published, which makes the argument to write the paper much stronger.

The notion of Argumentation Framework (AF) was introduced by Dung [50] to allow for the evaluation of arguments in the context of a set of arguments by building a relationship model between them (Chapter 2.5.2). A number of developments of Dung's model, extending the interpretation of argument relation, have been proposed (e.g., Preference-Based Argumentation Frameworks (PAF) [5], bipolar [6], ...etc). We use the development proposed by Bench-Capon [28] in which a relation between arguments and the various social values they promote is presented. Those values are ordered to allow each argument in the relationship model to be preferred over other arguments according to an audience (defined as an ordering of values), which determines whether attacks succeed or fail. The use of VAFs requires values to be associated with arguments. This can be done using the argument scheme for practical reasoning developed in [12], the formal definition of VAFs was given in Chapter 2.5.3.

In [12], Atkinson builds on Walton's [123] sufficient condition scheme for practical reasoning and extends it to the following argument scheme:

In the current circumstances R,
We should perform action A,
To achieve circumstance S,
Achieving goal G, which will promote the value V.

At times, however, an action does not necessarily promote any values or achieve any goals, but simply enables another action that would promote a more preferred value with respect to the audience. Saying that, it is not enough to consider the performance of the action itself, but the sequence of actions that might follow performing this action. The argumentation scheme proposed then slightly changes to:

In the current circumstances R,
We should perform action A,
To achieve/enable achievement of circumstance S,
Achieving/enabling achievement of Goal G,
which will promote/enable the promotion of the value V.

Arguments are formed by instantiating this scheme. Those arguments are then subjected to sixteen Critical Questions (Section 2.5.3) which give rise to counter arguments, and Action-Based Alternating Transition System (AATS) is used as a foundation to model the domain and allow the generation of both arguments and counter arguments using the techniques of [14], also explained earlier in Section 2.5.3. AATS formalisations are given in Definition 3.1.

When discussing the consequences of actions, it should be noted that the result of any action performed also depends on what the other entities in the environment will do. The performance of one action might have different outcomes depending on what joint action it will be part of.

The methodology defines the argumentation problem and formalises it as an AATS, which is used to generate an argumentation framework which can be evaluated as a VAF. A further step is introduced here (step number 5) to account for uncertainties, future actions, and calculations of probabilities in consideration of single actions, where the result is a set of accepted arguments that presents acceptable actions at each stage of the decision-making sequence. These possibilities are sequenced using the criteria of safety, opportunity and threat to determine the best order in which to execute the actions.

3.1.2 A Methodology for Decision Making

The methodology passes through five steps:

1. Formulating the Problem: Produce a formal description of the problem scenario to give all relevant possible actions, values and related factors that may influence the decision. This is accomplished by building an AATS appropriate to the problem.

2. Determining the Arguments: On the basis of the AATS, arguments providing justifications of the various available actions are identified through the argument scheme. Counter arguments are identified using a subset of the Critical Questions as interpreted in the terms of an AATS.

3. Building the Argumentation Framework: in this step, the arguments and attacks between them are organised into an Argumentation Framework. Because the argument scheme associates arguments with the values they promote or demote, arguments can be annotated with these values, yielding a VAF.

4. Evaluating the Argumentation Framework: The arguments within VAFs are evaluated for acceptability with respect to the specific audience representing the decision maker, characterised by the ordering of values subscribed to by the agent making the choice.

5. Sequencing the Actions: The set of actions acceptable to the agent in the previous stage is now put into a suitable order for execution.

3.2 The Decision-Making Model

This section gives details of the five-step methodology described briefly in Section 3.1.2. Each subsection presents a step. The previous work is mentioned and then the contribution made by this thesis is detailed and formalised.

3.2.1 Formulating the Problem

In this first step, the problem is formulated as an Action-Based Alternating Transition System (AATS), as described in [15]. AATS was introduced as a foundation formally to describe a system in which several agents combine to determine the transition between states (see Section 2.5.3). An AATS is defined by Wooldridge and Van der Hoek [132] as an (n+7) tuple where n is the number of agents. Atkinson and Bench-Capon [15] expand upon this idea to include the notion of values, so that each agent has a set of values, Av, drawn from an underlying set of values V, and every transition between members of the set Q of different possible states may either promote, demote, or be neutral with respect to those values.

To present a particular problem, we first describe relevant situations using a set of propositional variables. The variables which are true at a given stage correspond to different possible states of the system. Only valid combinations are included in states. Thus, given P,Q and the knowledge that P → Q, only those states will be used. Using knowledge in this way is useful to control the proliferation of others. Next, we identify the relevant agents, different possible actions the agents can perform, and how these move between states with each transition representing a joint action of the agents involved. Finally, we identify the values and relate them to the transitions between states.

A tree of possible future states can be derived from AATS. Given an initial state denoted q_0, we can construct a tree representing possible future states with q_0 as the root. The next level (level 1) of states below q_0 are all the possible states that would occur given the performance of the different joint actions identified earlier and possible for the agent at q_0.

Definition 3.1. *[14]*

An AATS is a $(2n + 8)$ *element tuple:*
$S{=}\langle Q, q_0, Ag, Ac_1, Ac_n, Av_1, Av_n, \rho, \tau, \Phi, \pi, \delta\rangle$

Q *is a finite, non-empty set of states*

$q_0 \ = q_x \in Q$ *is the initial state*

$Ag \ = \{1, ..., n\}$ *is a finite, non-empty set of agents*

Ac_i *is a finite, non-empty set of actions, for each* $i \in Ag$ *(where* $Ac_i \bigcap Ac_j = \emptyset$ *whenever* $i \neq j$*)*

Av_i *is a finite, non-empty set of values* $Av_i \subseteq V$*, for each* $i \in Ag$

ρ : $Ac_{Ag} \rightarrow 2^Q$ *is an action precondition function, which for each joint action* $\alpha \in Ac_{Ag}$ *defines the set of states* $\rho(\alpha)$ *from which* α *may be executed*

τ : $Q \times J_{Ag} \rightarrow Q$ *is a partial system transition function, which defines the state* $\tau_(q, j)$ *that would result by performing* j *from state* q *- note that, as this function is partial, not all joint actions are possible in all states (cf. the precondition function above)*

Φ *is a finite, non-empty set of atomic propositions*

π : $Q \rightarrow 2^{\Phi}$ *is an interpretation function, which gives the set of propositions satisfied in each state: if* $p \in \pi(q)$*, then this means that the propositional variable* p *is satisfied (equivalently, true) in state* q

δ : $Q \times Q \times Av_{Ag} \rightarrow \{+, -, =\}$ *is a valuation function which defines the status (promoted (+), demoted (−) or neutral (=)) of a value* $v_u \in Av_{Ag}$ *ascribed by the agent for the transition between two states:* $\delta(q_x, q_y, v_u)$ *labels the transition between* q_x *and* q_y *with one of* $\{+, -, =\}$ *with respect to the value* $v_u \in Av_{Ag}$

In our approach, this modeling is step one in the methodology of action selection. The output of this step can best be shown as a graph (see Figure 3.1).

As an extension to the basic AATS modeling, this methodology considers not only the direct successors of the state where the decision is being made, but also more remote future states. Such consideration allows for the accommodation of the concepts of *Establishers* in AI Planning. Establishers are the prerequisites for a state of affairs to occur. So, a state that promotes publication has usually an establisher of another state that would first have a paper written and a successful registration at a conference (for example, in figure 3.1 q3 is an establisher for q7).

Example: The Head of the Department (HoD) Dilemma

A department head is about to make a decision to send a student to a conference, send him for training, or ask him to write a paper that could then be published.

Sending the student to a conference would promote the value ‘Attendance’ (v1); sending the student to training would promote the value ‘Knowledge’ (v2). Asking the student to write a paper and then sending him to a conference would promote ‘Publication’ (v3).
In Figure 3.1, the states below level 1 are future states that then form possibilities from all different states and help the agent reason not about his direct actions, but also about the possible effects of those actions, and the actions they will enable. For simplicity, we make the assumption in this example that a student will not have time to write a paper if he attended training, so, after a training course the only possible action is to go directly to a conference without preparing a paper. We also assume that when a student is registered for a conference, he can only write a paper afterwards in preparation for that conference and will not

have time to attend training. Nonetheless, when the student is firstly asked to write a paper, he will be able to choose among the two different options (Training or Conference).

The different states are then linked together with their parent states by the joint actions that would cause them to occur. For example, q0 and q1 are linked with an arrow pointing at q1 indicating that a transition is possible from q0 to q1, this arrow is labeled (Train) indicating that the joint action that would instantiate such a transition is when the HoD asks a student to go to training and the student obliges. The arrow is also labeled v2, which is the value that such a transition would cause (experience). Each box represents a state and has the state name (q0, q1, ...etc) and also the propositions that are true at that state (has a paper written, is registered for a conference, attended a training course, ...etc). In our example, although labeled differently, q5 and q7 are the same state where the HoD has realised the same propositions but in a different order. In our approach so far we differentiate among states only by the propositions that are true at that state not by how the state was reached or when.

In AI Planning *situations* are a set of propositions that denote a state of affair which we refer to here as a state. The concept of a *problem* in Planning is basically the initial situation and the final situation where the goal is realised. In our formulation the initial situation is basically $q0$ and we differ in concept from Planning in that we do not specify final situations in our definition of a problem but we rather look at value important to the agent being realised regardless of what final situation (state) they could be realised in.

3.2.2 Determining the Arguments

Now, the problem is formulated into an AATS, and all possibilities are identified. This subsection introduces step two of the methodology, where the arguments are built and formalised.

The method of justifying actions is built in terms of presumptive justification through an argument scheme, followed by a process of critical questioning to see whether the presumption can be maintained, as described by Walton[123]. We use the argument scheme presented by Atkinson et al[22], which specifically extends the sufficient condition scheme of Walton [123] to enable discrimination between the effects of an action (the consequences), the desired effects (the goal),

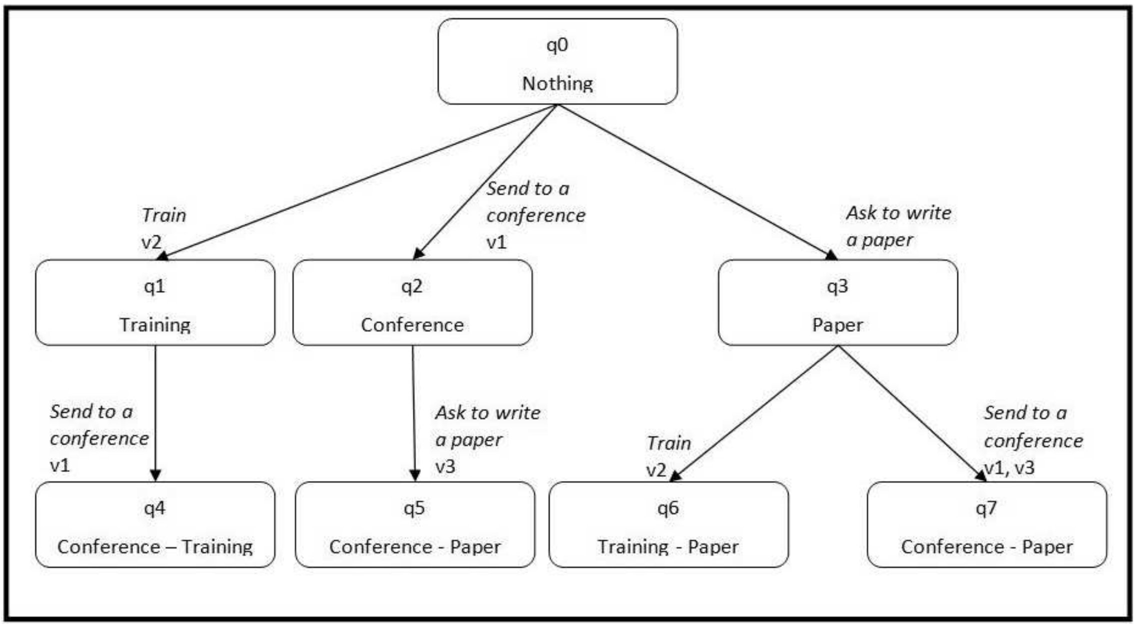


Figure 3.1: The AATS Model of the example

and the reason why these effects are desired (the value).

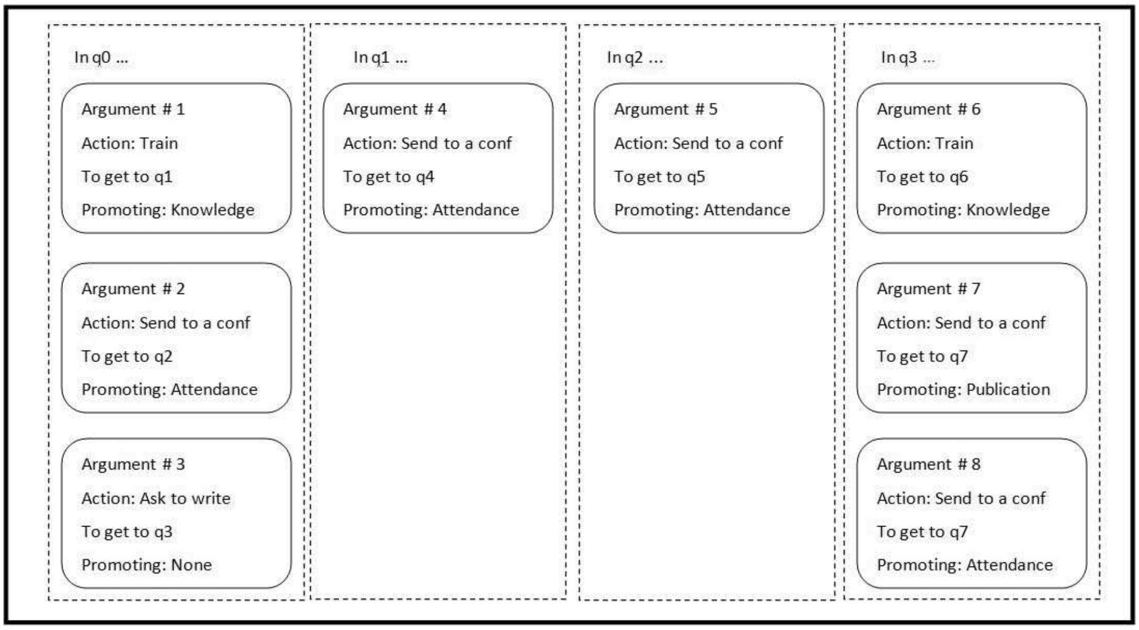
Thus, the argument scheme is as follows: in the current state, the agent should perform action A to reach a new state in which goal G is true, promoting value V in terms of an AATS, this can be defined as:

Definition 3.2. *[22]*

The initial state q_0 *is* $q_x \in Q$*;*
Agent $i \in Ag$ *should perform* α_i*, consistent with the joint action* $j_{Ag} \in J_{Ag}$ *where* $j_{Ag_i} = \alpha_i$*,*
so that $\tau(q_x, j_{Ag}) = q_y$*,*
$p_a \in \pi(q_y) \setminus \pi(q_x)$ *and*
for some $v_u \in Av_i$*,* $\delta(q_x, q_y, v_u) = +$*.*

The methodology presented uses this as the second step after formulating the problem scenario in an AATS model.

Arguments can be created from the AATS model (previous step) easily. Each link between connected states presents a transition; each transition which promotes a value provides an argument justifying the joint action presented by this transition. If the link promotes more than one value, each value is used to produce a separate argument (see Figure 3.2). Some values would require more than one transition to be realised as they would require multiple actions. In our example v3 (Publication) would only be promoted if the student had registered to attend a conference and at the same time has written a paper for that conference, it is not sufficient to just write a paper in order to promote Publication (in figure 3.1 Publication (v3) is promoted either by the transition to q2 then q5 or q3 then q7).



In q0 ...

Argument # 1
Action: Train
To get to q1
Promoting: Knowledge

Argument # 2
Action: Send to a conf
To get to q2
Promoting: Attendance

Argument # 3
Action: Ask to write
To get to q3
Promoting: None

In q1 ...

Argument # 4
Action: Send to a conf
To get to q4
Promoting: Attendance

In q2 ...

Argument # 5
Action: Send to a conf
To get to q5
Promoting: Attendance

In q3 ...

Argument # 6
Action: Train
To get to q6
Promoting: Knowledge

Argument # 7
Action: Send to a conf
To get to q7
Promoting: Publication

Argument # 8
Action: Send to a conf
To get to q7
Promoting: Attendance

Figure 3.2: The different arguments of the example

An aspect we add in this methodology to the work of Atkinson is the notion of Uncertainty where consequences of actions are not always entirely predictable. In this case, actions are executed aiming at a certain result, which often will not come about because it has dependencies on other actions performed by other agents.

When an agent performs an action where the results solely depend on itself, it is very easy to assume the resulting state. However, for example, when a teacher asks a student to write a paper, the student may or may not succeed, and so the teacher cannot be certain which state will be reached. There may also be uncertainty about the initial state: if an action is performed in a state other than the one assumed, the state reached may be different. So, when a joint action is performed at any state, the result is known and predicted, but when we are sure about only a subset of this joint action, for example, an individual action might have different possible states that result depending on what the other agents in the joint action do.

To address uncertainty properly, we add a new formulation in Definition 3.3, where we define $\underline{\tau}$ a function that yields the different possible states that might result by performing a single action.

Definition 3.3. *Given a sequence* $\underline{J} = j_1, j_2, ..., j_n$ *of joint actions involving agent* A_i
the action sequence for agent A_i *is* $\underline{\alpha} = \alpha_1, \alpha_2, ..., \alpha_n$
where $\alpha_j \in AC_i$ *is the action performed by* A_i *at step* J_j
$\underline{\tau} : Q \times AC_i \rightarrow Q$ *where:*

$\underline{\tau}(q, \alpha_x) = q\prime$: $\exists$ *a joint action* j *with* α_x = *action performed by* A_i *in* j *such that* $\tau(q, j) = q\prime$

This extension to the argumentation model allows:

1. Considering possibilities when deciding on the course of action to be taken. If an action is desirable but might have a potential of not reaching the desirable state, this action might be replaced with another one that has a more reliable outcome. For example, if Agent A (Ag_A) would like to go to London to see a play and can take his car or catch the train. We would say that there are two arguments (Arg_{Train} and Arg_{Car}). Going with the car will be faster in the best case: Ag_A knows for sure that if he chooses Arg_{Train}, he will reach London on time, while Arg_{Car} might have two possible results as he is not sure about the directions or the traffic conditions. Even if going by car is faster, it might not result in the desired goal of attending the play. Thus, Ag_A might choose the train if unwilling to accept that risk.

2. Estimating probabilities of whether the desired goals are achieved.

As this methodology aims at finding the best sequence of actions possible, the account of uncertainty will help us choose the best possible sequence of actions. For example, if the methodology results in the actions a,b,c,d being justified, the sequence adcb might yield a higher probability of achieving the result than abcd.

Example: The Head of the Department (HoD) Dilemma

Figure 3.2 shows the different arguments available to the HoD at various stages. The HoD has a choice of values to promote by either sending the student on a training course or sending him to a conference. Asking him to write a paper will not itself promote any values, as he has also to go to a conference afterwards. Our joint action here is represented by the HoD asking the student, who can comply or refuse to do something. If the joint action is (HoD: Send a student to training; Student: Succeeds), we then reach q_1, but if it is (HoD: Send a student to training; Student: Fails), we remain in q_0.

3.2.3 Building the Argumentation Framework

In this step, we link the different arguments together, using the work of Bench-Capon and Atkinson et al.[29] and [14].

The previous step gave us a number of arguments associated with values and a set of attack relations. These can be organised into a Value-Based Argumentation Framework (VAF). Figure 3.3 extends the view of Figure 3.2 and turns it into a VAF. Looking back at the Critical Questions (CQs), the relationship with different arguments can then be identified.

Two concepts are critical in this step: VAF and CQs.

Value-Based Argumentation Framework (VAF): VAF is an extension of the Argumentation Framework that allows for values to be associated with arguments. This allows arguments to be given different strengths according to preference ordering of values for a particular audience. This allows conflicts to be resolved differently by different audiences.

Definition 3.4. *[31]*

A VAF is a triple $< H(X, A), \nu, \eta >$, where H(X,A) is an argumentation framework, $\nu = v_1, v_2, ..., v_k$ a set of k values, and $\eta : X \rightarrow \nu$ a mapping that associates a value $\eta(x) \in \nu$ with each argument $x \in X$. A specific audience, α, for a VAF $< H, \nu, \eta >$, is a total ordering of the values ν. We say that v_i is preferred to v_j to the audience α, denoted $v_i \succ_\alpha v_j$, if v_i is ranked higher than v_j in the total ordering defined by α.

Critical Questions (CQs): CQs were introduced by Walton [123]. Atkinson then [14] identify a set of sixteen CQs relevant to the argument scheme. CQs challenge the presumptive conclusion of an argument scheme. Each scheme is associated with a characteristic set of CQs. These questions can identify five kinds of attacks:

1. Denial of Premises: Questioning the validity of the information given.
2. Alternatives: Considering other possibilities that would have better or equal results.
3. Side Effects: Asking how this action would affect other elements in the environment.
4. Interference: Asking if this action would disturb other actions in any way.
5. Impossibility: Questioning the possibility of stated actions, values, goals or states.

CQs are now used to address the factors that may lead to the presumptive justification being overturned. In Atkinson et al.[22], sixteen CQs were identified, but in any given scenario, not all of them will be relevant. For our purposes, we need to consider only six as some of the questions concern problems of language or epistemic matters that do not apply to a single agent. These are defined in terms of an AATS in Definition 3.5.

Definition 3.5. *[15]*

CQ1: *Are the believed circumstances true?*
$q_0 \neq q_x$ *and* $q_0 \notin p(\alpha_i)$.

CQ11: *Does doing the action preclude some other action that would promote some other value?*
In the initial state $q_x \in Q$*, if agent* $i \in Ag$ *participates in joint action* $j_n \in J_{Ag}$*, then* $\tau(q_x, j_n)$ *is* q_y *and* $\delta(q_x, q_y, v_u)$ *is* $+$*. There is some other joint action* $j_m \in J_{Ag}$*, where* $j_n \neq j_m$*, such that* $\tau(q_x, j_m)$ *is* q_z*, such that* $\delta(q_x, q_z, v_w)$ *is* $+$*, where* $v_u \neq v_w$.

CQ2: *Assuming the circumstances, does the action have the stated consequences?*
$\tau(q_x, j_n)$ *is not* q_y.

CQ7: *Are there alternative ways of promoting the same value?*
Agent $i \in Ag$ *can participate in joint action* $j_m \in J_{Ag}$*, where* $j_n \neq j_m$*, such that* $\tau(q_x, j_m) is q_z$*, such that* $\delta(q_x, q_z, v_u)$ *is* $+$.

CQ8: *Does doing the action have a side effect which demotes the value?*
In the initial state $q_x \in Q$*, if agent* $i \in Ag$ *participates in joint action* $j_n \in J_{Ag}$*, then* $\tau(q_x, j_n)$ *is* q_y*, such that* $p_b \in \pi(q_y)$*, where* $p_a \neq p_b$*, such that* $\delta(q_x, q_y, v_u)$ *is* $-$.

CQ9: *Does doing the action have a side effect which demotes some other value? In the initial state $q_x \in Q$, if agent $i \in Ag$ participates in joint action $j_n \in J_{Ag}$, then $\tau(q_x, j_n)$ is q_y, such that, $\delta(q_x, q_y, v_w)$ is $-$, where $v_u \neq v_w$.*

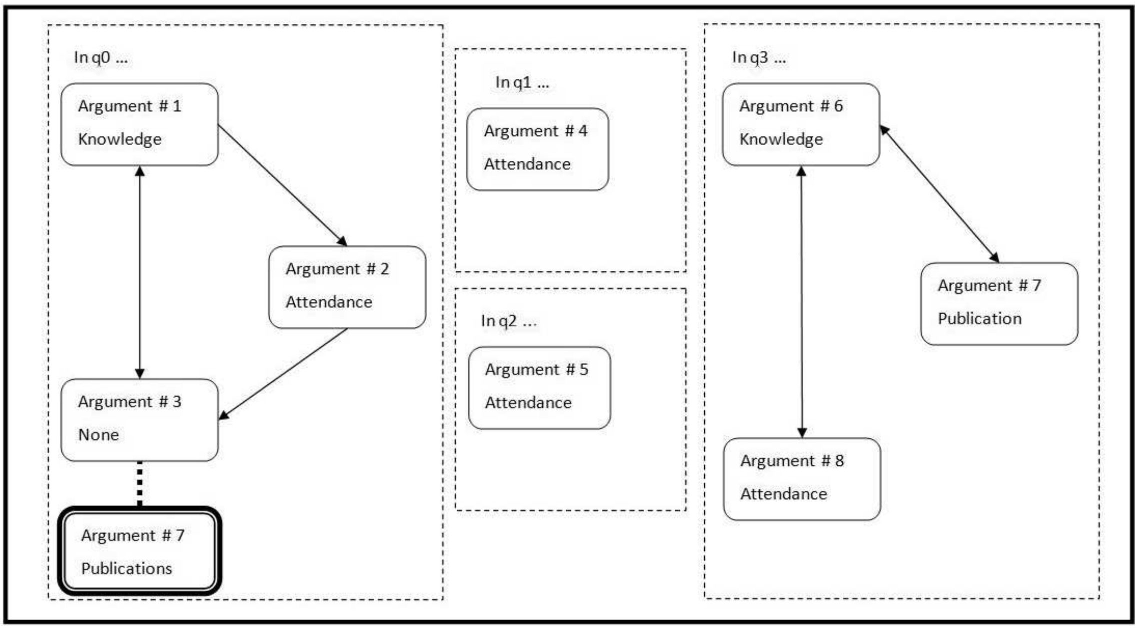


Figure 3.3: The VAF Model

We use only six of the sixteen Critical Questions. Some CQs are related to problem formulation, which is addressed in building the AATS model (CQs 2,3,4,12,13,14,15 and 16). CQs 5 and 6 are also not considered because they relate to possibilities of realising the same goal or the same consequences; in our model, these possibilities form different branches in the AATS tree. CQ10 states whether this action promotes any other value. In our model, this is unnecessary; if an action promotes other values, this will result in separate arguments for each value promoted.

The modeling of the relationship using VAF and CQs gives the following benefits:

1. Strength related to social values: This allows values to give strength for arguments, allowing persuasion from social rationality perspectives, rather than only pointing to information or logical connections.

2. Individual preference between values: This introduction of values' strengths allows an individual audience to evaluate these arguments by stating preferences between them.

3. Systematic identification of attacks: The introduction of CQs through their different attack possibilities provides a systematic and formal methodology to build the relationship between arguments.
4. Exhaustive identification of the attacks: In this step, uncertainties are considered as all arguments are subjected to CQs. The list developed in [12] accounts for possibilities of argumentation schemes that can be used in persuasion over action.

The next subsection now sets a preference system that becomes the basis of evaluating the different arguments to eventually select the desirable ones and dismiss the rest.

Example: The Head of the Department (HoD) Dilemma

In Figure 3.3, The Value-based Argumentation Framework (VAF) for four states (q0, q1, q2 and q3) is presented. Each box represents an argument and contains the argument label along with the social value this argument promotes. The arrow between the boxes defines the attack relationship. So, in q0 argument 1 attacks argument 2 and 3. The highlight around argument 7 in q0 indicates that it is an argument that belongs to a different state of affairs and the dashed line to argument 3 indicates that it is an argument that defends argument 3. Arguments to write a paper and attending a training course attack each other as choosing one excludes the other and both claim they promote better values (arguments 1 and 2). The argument for writing a paper is supported by argument 7 (Figure 3.3), which promotes the value of Publication as writing a paper will allow the value of Publication to be promoted through argument 7 (Figure 3.3).

3.2.4 Evaluating the Model

Having constructed a VAF, the next step is to evaluate the attacks and determine which arguments will be acceptable to the agent. The strength of each argument is determined by associated values. Given the ordering of values desired by the agent, we can determine what arguments will be acceptable, and determine the Preferred Extension (PE) with respect to the audience endorsed by the agent. This PE, which will be unique and non-empty, provided every cycle of arguments involves at least two distinct values [53], represents the maximal set of acceptable arguments for that agent.

Definition 3.6. *[30]*

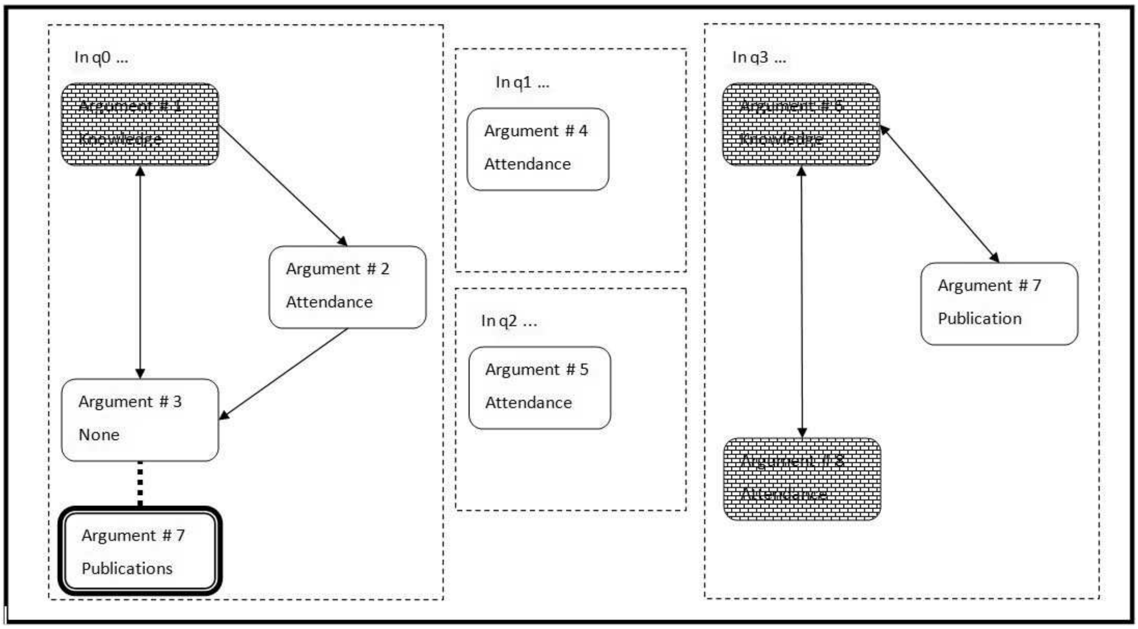


Figure 3.4: The result of the evaluation

Let $< H(X, A), V, \eta >$ be a VAF and α an audience.

1. *For arguments x, y in X, x is a successful attack on y (or x defeats y) with respect to the audience α if: $< x, y > \in A$ and it is not the case that $\eta(y) \succ_{\alpha} \eta(x)$.*

2. *An argument x is acceptable to the subset S with respect to an audience α if: for every $y \in X$ that successfully attacks x with respect to α, there is some $z \in S$ that successfully attacks y with respect to α.*

3. *A subset R of X is conflict-free with respect to the audience α if: for each $< x, y > \in RR$, either$< x, y > \notin A$ or $\eta(y) \succ_{\alpha} \eta(x)$.*

4. *A subset R of X is admissible with respect to the audience α if: R is conflict-free with respect to α and every $x \in R$ is acceptable to R with respect to α.*

5. *A subset R is a Preferred Extension (PE) for the audience α if it is a maximal admissible set with respect to α.*

An Audience is an ordering of values; the associated PE represents the acceptable arguments to the audience. If there are no cycles in the same value, the PE for a given audience will be non-empty and unique. It is noted that this also accounts for arguments indirectly defending each other, and we saw through the model how an attack was dismissed not because of the argument itself, but because of an argument defending the attacking arguments.

Now, the set of acceptable arguments has been identified, but what if those arguments are not prerequisites of each other? Which action should be performed first? Does the order in which the actions are performed have any effect on the overall state? The next section considers these questions.

Example: The Head of the Department (HoD) Dilemma

Figure 3.4 assumes the following value order ($V3 > V2 > V1$) preferring Publication over Training and Training over Attendance, which concludes that the HoD would either send the student to the conference and then ask him to write a paper that can be published there or he would ask the student to first write the paper and then attend the conference. Both sequences are acceptable to the HoD. The next subsection determines the best sequence.

3.2.5 Sequencing the Actions

An entirely novel aspect of our methodology is the sequencing of actions, where we formally consider the effect of performing a sequence of joint actions and see their effect on the AATS model. This is useful to:

1. Recognise actions that might not have a direct impact on the initial state or might not promote any values immediately, but would enable future joint actions that promote better values. We can see this as also allowing arguments in future states to support current arguments in their current state.

2. Allow future arguments against actions to be considered. Some arguments might not have a direct impact on possibilities in the current state, but would prevent future actions from being performed.

The above accounts for what is called in AI Planning *clobbers* and *white-knighting.* Clobbers are states that are not allowed in the system. In our example states that promote Publication are not allowed in the initial state as the student has not yet written a paper. White-knights are situations that would make a clobber legal. In our example, a state that would realise a paper written is a white-knight to a state that would then promote Publication.

We extend the AATS model of [14] to also account for action sequencing by introducing a $\underline{J}$ as a sequence of joint actions and τ_s as the function to yield the state when performing $\underline{J}$.

Definition 3.7. *The state transition function (for joint actions) $\tau_s : q \times J_{AC}$ is defined as:*

$$\tau_s(q, \underline{J}) = \begin{cases} \tau(q, j) & if \underline{J} = j \\ \tau(\tau_s(q, \underline{J\prime}), j_n) & if \underline{J} = J' j_n \end{cases}$$

"$J\prime j_n$ is a sequence composition where j_n follows J'"

This definition is an extension of Definition 3.1 and in particular builds on the transition system function τ, where $\tau(q, j)$ gives us the resulting state from q by performing the joint action j; $\tau_s(q, \underline{J})$ gives us the resulting state from q by performing the sequence of joint actions $\underline{J}$.

The result at the end of this step is a graph that has a full picture of all possible sequences of joint actions and their effects in terms of goals, values and state. This can also be seen as a modeling of all possibilities according to the agent's actions and the environment's reaction.

This is the final step of the methodology. The prior four steps generate a set of actions acceptable to the agent. We now organise those actions into a sequence of actions. This is important as sequencing the actions differently might have different results. So, as we might instruct our agent to perform action α_A hoping for the joint action (α_A, β_A), we might end up in (α_A, β_C), which, in such cases an alternative sequence might need to be computed. In summary, joint actions will always lead to a definite known state, but individual actions might lead to different possible states depending on what other agents in the joint action would do. This, in turn, affects the actions the agent can perform, as well as their desirability.

All the actions in our sequence are acceptable in the sense that they have survived the critique by the posing of CQs and have no attackers preferred to them. Often, this set contains several actions, any of which could be beneficial to perform in the current state. These should be sequenced in terms of *safety*, so that unexpected consequences will not prevent the other actions in the set from being performed; *opportunity*, because the performance of an action may enable some desirable action not available in the current state to be executed subsequently;

and *threat*, where potential side effects are considered.

Next, is a detailed description of the three principles mentioned in action sequencing.

Safety: A joint action will lead to a definite state, but an individual action from this joint action can lead to multiple possibilities. If we are hoping to execute a series of actions in a sequence, one of those actions might lead to an undesirable state where the rest of the sequence of actions might be prevented. In considering safety, we would choose a sequence of actions that no matter what state any possible joint action might lead to, it would be possible to execute the rest of the sequence. A safe action can be defined as an action that participates in different joint actions, all of which will lead to states where other actions in the sequence can be performed.

Definition 3.8. *A sequence of actions $\underline{\alpha}$ is considered safe if each $\alpha_x \in \underline{\alpha}$ can be performed in the particular sequence in all possible states when performing the joint action.*
In order to formalise this idea we introduce the following notation. Given $\underline{\alpha} = \alpha_1\alpha_2 \dots \alpha_n$ and i with $0 \leq i \leq n$, we use β_i to denote the initial sequence of i actions, i.e. $\beta_i = \alpha_1 \dots \alpha_i$ where $\beta_0 = \varepsilon$
We now wish to define the set of possible states that can arise in terms of these initial actions thereby allowing us to indicate whether the given the sequence is safe or otherwise.
Given β_i we denote this set of sttaes using $\sigma(\beta_i)$ so that
*$\sigma(\beta_0) = q_0$ (i.e before *any* action is performed we are in the initial state)*

$$\sigma(\beta_i) = \begin{cases} \emptyset & \text{\textit{if} } (\sigma(\beta_{i-1}) = \emptyset) \textit{or} \neg(\sigma(\beta_{i-1}) \subseteq \rho(\alpha_i)) \\ \bigcup_{q\in\sigma(\beta i-1)} \tau(q, \alpha_i) & \textit{otherwise} \end{cases}$$

Opportunity: In the performance of a sequence of actions, we would say that the different possibilities of sequencing those actions are paths. There are two issues to consider here: first, different paths of those actions might promote different values. When choosing a path, we should consider if different ways of executing our actions would promote different values.

The second aspect is probability. Although the different paths would lead to the same goal, some paths have a higher probability of achieving the desired goal and promoting the values.

Definition 3.9. *An opportunity exists if either:*

(1) A different path of joint actions $\underline{j}_2 \neq \underline{j}_1$ *exists where* $\underline{j}_2$ *promotes different values.*

$\tau_s(q_0, \underline{j}_1) = q_1$, $\tau_s(q_0, \underline{j}_2) = q_2$ *and* $q_1 \neq q_2$

An opportunity is when $\delta(q_0, q_1, v_1)$ *is* $+$ *and* $\delta(q_0, q_2, v_2)$*is* $+$ *and* $v_1 \neq v_2$

Threat: When different possibilities arise when performing a sequence of actions, some are not desirable and might lead to negative values. In considering whether we should perform an action, we should consider the extent to which it is possible for things to go wrong. For example, we can set a threshold, and if the probability of failure were to exceed the threshold, we might be better off not following this sequence.

Definition 3.10. *A threat is when the performance of a certain sequence of actions* $\underline{\alpha}_x \in \underline{j}_x$ *the probability of demoting a value* v_u *will be high.*

$Prob(\underline{j}_x | \underline{\alpha}_x, \delta(q_i, \underline{\tau}(q_i, \alpha_i), v_u)$ *is* $\rightarrow X$ *where* X *is a predefined threshold*

Example: The Head of the Department (HoD) Dilemma

A student might not be motivated to write a paper if he knows that he will attend a conference regardless of his choice or ability to complete a paper. If the teacher in this case asks the student to write the paper before registering him for the conference, this would be a *safer* sequence because whatever happens with the first action it will be possible to then register him for the conference.

If the HoD is also thinking of sending this same student to a training course, perhaps sending him to the training course before actually asking him to write the paper might have the *opportunity* of a better written document rather than if he went to training after writing the paper.

In our example, we have two sequences (Paper-Conf-Training) and (Training-Paper-Conf). If we emphasise the value of promoting the publication rate of the department, the second sequence has a higher probability in the *opportunity* of promoting this value rather than the first one.

If the student has a health condition and sending him to that specific conference might seriously affect his health, the HoD would weigh the *threat* of this student's health actually being affected. If the probability is below a certain threshold, he might feel comfortable sending him. Otherwise, he would be better off not promoting the value of the department's publication rate rather than demoting his student's health.

3.3 Summary

As mentioned in Section 2.1, a trustworthy decision is a decision with the following capabilities: it identifies the options for actions and their likely effects (Practical Reasoning); it considers the surroundings and how relevant objects would react (Social Interaction); it makes plans that serve short- and long-term visions (Planning); it addresses side effects, completeness, or correctness of knowledge and unexpectedness (Uncertainty); and it addresses emotional and behavioural aspects in the decision process.

In this chapter, we presented a methodology that provides a foundation to address the first four of these capabilities.

The first four steps are an extension of the work of Atkinson et al.[14]. This work has been extended to account for uncertainties, considerations of future actions attacking or supporting current actions and enabling the calculations of probabilities.

The fifth step was introduced in this thesis to account for variations in sequencing a series of accepted arguments.

The methodology has five stages:

1. Formulating the Problem: Produce a formal description of the problem scenario to include all possible actions, values and all related factors that may influence the decision. This addresses capability one in the capabilities mentioned above (**Practical Reasoning**).

2. Determining the Arguments: Arguments providing justifications of the various available actions are provided.

3. Building the Argumentation Framework: In this step, the arguments and attacks between them identified in the previous step are organised into an

Argumentation Framework. This addresses the second capability (**Social Interaction**), allowing the consideration of social interaction between different agents and the environment and shows how it can affect us.

4. Evaluating the Argumentation Framework: The arguments within VAFs are determined to be acceptable or not with respect to a specific audience, characterised by the ordering of values subscribed to by the agent making the choice. This allows for the subjective elements of decision making.

5. Sequencing the Actions: The set of actions deemed to be acceptable to the agent in the previous stage must now be put into a suitable order in which they should be performed. Allowing the consideration of **uncertainties** through calculations of possibilities and probabilities. Moreover, this step provides an overview of future actions and how they can be planned into different paths, addressing the **planning** capability.

Our work in action sequencing takes into consideration some concepts of AI Planning such as clobbering and white-knighting where we consider prerequisites and states that should be realised in order to achieve a certain value (clobbers and white-knights were discussed earlier in Section 2.4). We also share the same concept of state of affairs with AI Planning where we call them states it is called situations in planning. We model the concept of establishers as well in constructing the argumentation framework where arguments could defend other arguments in other states as they form a prerequisite for them. Our work nonetheless provides a different view to the concept of the *goal* where we look at goals from a value perspective (values being realised), AI Planners look at goals as the state of affairs to be reached.

The next chapter is a detailed implementation of this methodology on a specific case study, the HoD dilemma is used as an illustration throughout this chapter.

Chapter 4

Experimental Application

Chapter 3 proposed a methodology of action selection that is based on presumptive justification of actions through Value-Based Argumentation. The presented methodology addresses four out of five elements that philosophy suggests can elicit Trust in any decision-making mechanism.

This chapter presents a detailed case study to demonstrate this methodology. This example has been implemented and tested. The application was chosen to show a critical situation where important decisions, which we would not usually trust to software agents, need to be made, linking clearly to the issue of trust. The example also provides the necessary settings to show convincing evidence of the validity of the methodology with reference to the basic elements of trust. Finally, the example provides evidence of scalability to different scenarios.

This chapter states the aims of this experiment and details the approach that was used. It next describes the example in full detail and goes through the five different steps of the methodology. Finally, it presents and evaluates the results.

Section 4.1 provides an introduction where the properties are identified and the aims of this experiment stated. Section 4.2 introduces the study, and provides the background to the problem and its set up. Section 4.3 then applies the five steps to the methodology on the settings previously identified. Section 4.4 presents the results, evaluates the different scenarios, and finally discusses how the main aim of this thesis is addressed, i.e, creating a decision-making methodology that can be better trusted to allow the assignment of more critical/sensitive tasks to agents. Section 4.5 summarises the chapter. And finally Section 4.6 introduces Part II of the thesis (Chapters 5 and 6).

4.1 Introduction

> "If you have a beast that has the capacity of forming beliefs on the basis of its perception, and has the capacity for forming desires in addition to beliefs, and also has the capacity to express all this in a language, then it already has the constraints of rationality built into those structures". Searle [116][Page 22]

In this section, an informal introduction to the example used in this chapter is given. This includes three subsections where the main aims and goals that we hope to achieve through this experiment are detailed, along with the measurement criteria that are used. We then explain the approach along with the resulting application and its evaluation.

4.1.1 Main Aims of this Experiment

The example used and discussed throughout this chapter was chosen to address:

- An application of relevance to most environments that can be easily translated to other settings where the difficulties of the choices are common. This is important to establish the applicability of the methodology to a wider range of applications.

- A critical situation where important decisions, with which we do not usually trust an agent, need to be made. This is perhaps the most important element in this application as the main aim is eventually to show how the methodology could enhance the level of confidence we have in agents to undertake such tasks.

- Short and long term goals and aspirations. These are needed to show how an action can be taken into consideration to direct achievement of goals and future actions.

- Evidence of the validity of the methodology and to link the results to the main elements of Trust identified in Chapter 2.

Recall, from Chapter 2, philosophical studies have proposed that when one entity trusts another, Trust is a component of rational aspects or beliefs and also a component of emotions between these two entities. Five different elements of trust suggest that a decision-making methodology addressing these elements can then be trusted. The first four (which relate to rationality/beliefs) were addressed and discussed in the methodology of Chapter 3 and are applied through the example in this chapter. The fifth element, relating to emotional aspects is the subject of Chapters 5 and 6.

Below is a list of the main aims that the study in this chapter hopes to achieve.

1. Give a better understanding of the methodology discussed in Chapter 3 through a detailed analysis of all the steps and definitions mentioned in a real-world example.
2. Show the applicability of the methodology being implemented.
3. Help show the linkage to elements of trust from an experimental perspective.
4. Build different setups and scenarios to demonstrate how behaviours of the agents in the environment respond to different settings.
5. Finally, evaluate all the different results.

4.1.2 The Approach

To conduct this experiment, a program was built using Microsoft Visual Basic on the .Net environment. A complete set of screen shots of the experiment is given in Appendix A.

The approach of this chapter is as follows:

The problem we are trying to solve is presented; then, interactively, the five steps of the methodology are applied to this problem showing us how each step can be implemented and the respective results of each step.

In each step, the single view from the state where the action being made, as well as a wider perspective of the overall problem with respect to all different possible states is shown. We also provide a brief description of how this was then implemented and converted into the experiment. Figures are used to illustrate the different results.

The results of this experiment are sequences of actions available for the agent to choose from.

4.2 Introducing the Case Study

The next four subsections introduce the case study where the problem and the related assumptions are explained.

4.2.1 The Problem

Our agent is the head of an academic department (HoD) in a university, and he is faced with a dilemma of how to appropriately allocate the department's budget where he needs to balance costs and consider departmental and individual interests of his staff. Our agent (HoD) receives requests relating to travel funding to attend two specific conferences. He receives requests from three different students and needs to decide which of them to send. Students 1 (S1) and 2 (S2) are new students. S1 asks to go to a nearby conference, which will be less expensive; S2 asks for a different conference which will cost more. However, S2 has prepared a good paper that might help the department's publication rate. Student 3 (S3) is an experienced student asking to be sent to a local conference and, although no paper has been prepared, she is an excellent networker who is likely to impress other delegates and so promote the reputation of the department. The conferences are on different topics, so S2's paper would not be suitable for the local conference, but both conferences are of equal standing. The budget will only allow two students to be sent.

4.2.2 Assumptions

To accurately build the example study, several assumptions must be made.

- The total number of agents mentioned in this example is four (the HoD and three students (S1, S2 and S3). In this experiment, the actions and behaviours of only the HoD will be considered throughout, although various possible reactions from S1-3 will need to be taken into account. The model is built for the HoD to take his decision. This means that the different values, goals and states considered are relevant to the HoD only.
- An arbitrary value of 3 is assigned as a budget available to the HoD.
- Emotional considerations are not yet considered in this example; the example will be extended in Chapter 6 to include them.

These assumptions were not critical to the experiment set-up but made simply for the clarity of the presentation.

4.2.3 Settings of the Example

In this subsection, the tables given below are intended to summarise and clarify the problem as explained above. Table 4.1 introduces the agents present in the environment; Table 4.2 details the relevant possible actions from those agents;

Table 4.3 combines the actions in the previous table to produce all possible joint actions; Table 4.4 summarises the relative values to the HoD and how they can be promoted. Table 4.5 is the cost of doing any action.

Agent	Description
HoD	Head of the Department
S1	First student
S2	Second student
S3	Third student

Table 4.1: All agents in the environment

Agent	Action	Reference
HoD	Send Sn to a conference	$\alpha1(n)$
HoD	Asks Sn to write a paper	$\alpha2(n)$
Sn	Student n goes to the conference	β_n
Sn	Student n does not go to the conference	$\neg\beta_n$
Sn	Student n writes a paper	γ_n
Sn	Student n does not write a paper	$\neg\gamma_n$
Sn	Student n does nothing	$\neg\delta_n$

Table 4.2: All Possible Actions

Joint Ac	Combination	Description
$J1_n$	$\alpha1(n), \beta_n$	HoD sends Sn to a conference and she attends
$J2_n$	$\alpha1(n), \neg\beta_n$	HoD sends Sn to a conference and she does not attend
$J3_n$	$\alpha2(n), \gamma_n$	HoD asks Sn to write a paper and she does
$J4_n$	$\alpha2(n), \neg\gamma_n$	HoD asks Sn to write a paper and she does not

Table 4.3: All Possible Joint Actions (Remaining agents are assumed to do nothing of relevance to the scenario)

Table 4.2 shows that the HoD can either ask the student to write a paper or send them to a conference; the student, on the other hand, has four possible actions. A student can either write/or not write the paper, attend/or not attend the conference. Table 4.3 then shows how these different actions can be combined in joint actions. There are four possible joint actions for each student in our example: the HoD sends a student to a conference and student attends, the HoD sends a student to a conference and the student fails to attend. The HoD asks a student to write a paper and the student does; the HoD asks a student to write a paper and the student does not. Note that some combinations are not possible: a student cannot attend a conference to which he has not been sent. In this example we focus on the HoD as our agent and we assume that changes would occur only when the HoD takes actions. Therefore, we do not consider pos-

Ref	Value	Condition for promotion
H(n)	Happiness of Sn	When Sn is sent to a conference
E(n)	Experience of Sn	When Sn has attended a conference
P	Department's Publication Rate	When a student writes and attends
R	Reputation of the Department	When an experienced student writes and attends

Table 4.4: HoD values

sibilities where the students would perform anything without being told. Hence, table 4.2 does not include the negation of the HoD's actions (for example $\neg\alpha 1(n)$).

As shown in Table 4.4, there are four different values of importance to the HoD. The happiness of each student (which is promoted every time a student is sent to a conference); the experience of the student (promoted when a student is sent for the first time to a conference); the publication rate of the department which can be promoted by sending a student who has written a paper to a conference; the reputation of the department (which is promoted when an experienced student attends and publishes in a conference).

Item	Cost
Initial Budget	3
Writing a Paper	0
Sending S1 to a conference	1
Sending S2 to a conference	2
Sending S3 to a conference	1

Table 4.5: Cost Breakdown

4.3 Applying the Methodology

This section is divided into five subsections, each representing one of the five steps of the methodology of action selection that was presented in Chapter 3.

Note that in the screen shots of the implementation the joint action are referenced differently from Table 4.3. (J0, J1 and J2) are sending students 1, 2 and 3 to conferences. (J4, J5 and J6) are asking students 1, 2 and 3 to write a paper.

4.3.1 Formulating the Problem

We first consider the different propositions that the HoD would consider: whether there are funds currently available in the budget (Budget); whether the students can be sent to attend (Attendance S(1-3)), whether the students have written a paper (Paper S(1-3)) and, finally, whether the students have attended a conference before (Previous S(1-3)).

Now, we define all possible actions that the HoD can take in all circumstances. Those are defined in table 4.2 but for better presentation in this chapter we refer to them as either to ask anyone of the three students to write a paper (Write(S1), Write(S2), Write(S3)) which corresponds with $\alpha 2(1)$ $\alpha 2(2)$ or $\alpha 2(3)$ in table 4.2. Or to agree to send a student to the requested conference (Send(S1), Send(S2), Send(S3)) which corresponds to $\alpha 1(1)$ $\alpha 1(2)$ or $\alpha 1(3)$ in table 4.2. These actions may change the state of Paper(Si) or Attendance(Si) respectively (see tables 4.2 and 4.3).

The values significant for the HoD are listed. Those values are then associated with the various conditions according to Table 4.4.

By esteem, we mean the general enhancement of the reputation of the department that comes from an impressive individual making an impact at a conference and raising the profile of the department's research, the research links established, and such like. Note that happiness and experience are relative to individual students, whereas the other values are relative to the department, although realised through properties of the individual students.

The states are designated by a code comprising four groups separated by a hyphen as follows: Budget, Attendance, Paper, and Previous (B-XXX-XXX-XXX), where each X represents a student and is either 1 or 0 depending on whether or not the corresponding proposition is true in that state. The first digit is the available budget at that state; the second, third and fourth digits are the attendance indicator of students 1,2, and 3. So, 000 means no students have been sent to any conference; 001 means the third student has been sent; 110 means the first and second student have been sent. If we look at the state as a whole, (3-110-000-000) would mean that the HoD has 3 points as available budget, students 1 and 2 have been sent to conferences, none of the students has written a paper or has previously attended a conference. The second group of 3 Xs (5th, 6th, and 7th digits) show whether any of the three students has written a paper; e.g., 001 indicates that the third student wrote a paper. The third group of 3 Xs (8th, 9th and 10th digits) shows if any of the students has previous experience attending conferences; 101 would mean that the first and third students have had previous

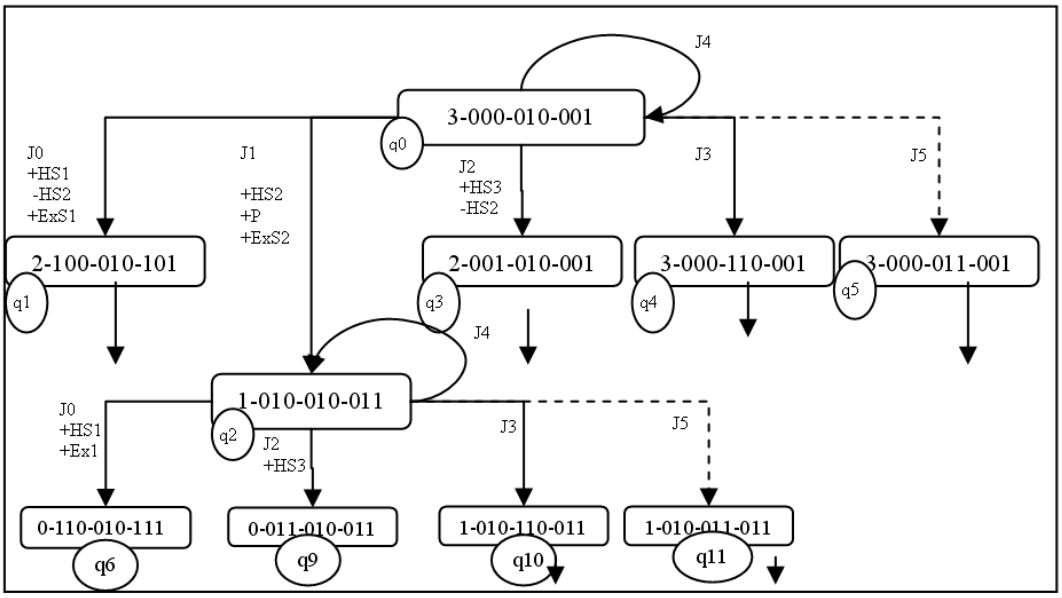


Figure 4.1: AATS of the Example

experience. Another example of a full state is (2-101-001-100). This means that the HoD has two points left in his budget, students 1 and 3 have been sent to conferences where student 3 had a paper written at his conference. Student 1 has been to conferences before. Compared to student 3, student 1 is a bit more experienced. Before we move into modeling the state transitions, let us look at the initial state q0. Budget is set to 3: the cheaper conference costs 1, and the expensive conference 2, so that we can send at most two students. S1 and S3 consume 1 point from the budget whenever chosen whereas S2 consumes 2 points. S3 has already attended a previous conference and S2 has a paper ready. Thus,

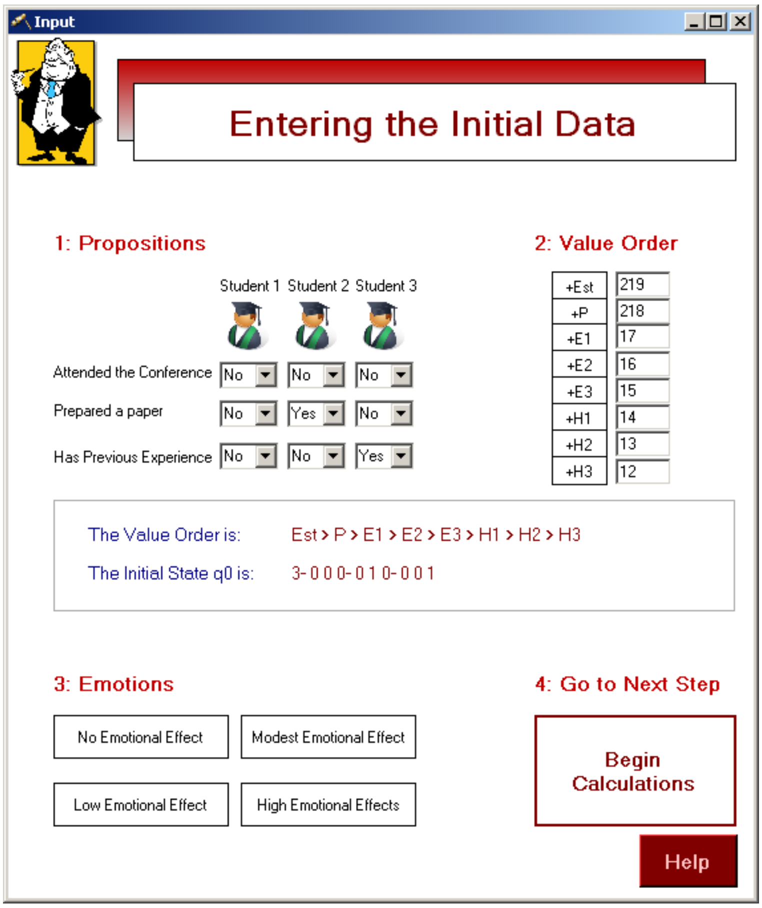
q0 = (3-000-010-001)

Figure 4.1 shows the initial state and some example transitions from that state. A complete view of the AATS is given in Appendix B. Action J0 is Send(S1); J1 is Send(S2); J2 is Send(S3) (it is assumed that the student will attend if sent to a conference; that is why we do not consider joint actions of students not attending). Where a paper is requested and written, we have J3 for S1 and J5 for S3, while J4 represents a request that does not result in a paper. So, the result of requesting a paper depends on the student's action to determine its effect. The transitions are also labeled with the values they promote or demote. Budget = 0 represents a terminal state since no further actions are possible.

For example, $q2$ from the figure represents a state where the second student has been sent to a conference and successfully attended. While looking into the move from $q0$ to $q2$, we see that the first digit on the state (which records the budget) has been reduced from 3 to 1, as the conference costs 2 points. Moreover, the third digit has changed from 0 to 1 indicating that the student has attended the conference. The digit before the last has changed from 0 to 1 indicating that the student has also successfully published a paper at that conference since this specific student had a paper written in the previous state (i.e, $q0$). Along the line leading to $q2$, we see $J1$, which is the joint action of the HoD sending the student to the conference and the student attending the conference. Below that we see $+HS2, +P, +ExS2$ indicating that this action also promotes three different values of importance to the HoD (the happiness of the student sent, the publication rate of the department, and the experience of the student). Below $q2$ we see five different possibilities depending on what joint action is performed. If the HoD decided to send the first student to a conference and he attends this will lead to $q6$ where the budget will become 0, and the second digit will turn into 1 indicating the successful attendance of S1 and the same applies to the third digit from the end indicating the experience of S1. Asking the second student to write a paper from $q2$ will yield the same state $q2$ as the second student has already written a paper.

Recall from Section 3.2 the concept of joint actions where consequences of actions are not always entirely predictable and actions are executed in the hope for a certain result, which often does not come about because it has dependencies on the actions performed by other agents. Hence, any executed action might have different results. When an agent performs an action where the results solely depend on itself, as with sending a student to a conference, it is very easy to assume the resulting state. When, however, the HoD asks a student to write a paper, the student may or may not succeed. And so he cannot be certain which state will be reached. This is shown in Figure 4.1 with a dashed line.

Figures 4.2 and 4.3 show the first step of methodology. In Figure 4.2, the user is asked to enter the different propositions of the environment. It is divided into three sections where the first section is setting the initial state $q0$ by indicating the current status of the students. In the second section the user sets the ordering of values. In the third section the users set the degree to which emotions will affect the decision-making process (will be relevant in the discussions in Chapters 5 and 6).



1: Propositions

	Student 1	Student 2	Student 3
Attended the Conference	No	No	No
Prepared a paper	No	Yes	No
Has Previous Experience	No	No	Yes

2: Value Order

+Est	219
+P	218
+E1	17
+E2	16
+E3	15
+H1	14
+H2	13
+H3	12

The Value Order is: Est > P > E1 > E2 > E3 > H1 > H2 > H3

The Initial State q0 is: 3-0 0 0-0 1 0-0 0 1

Figure 4.2: Inserting the Information

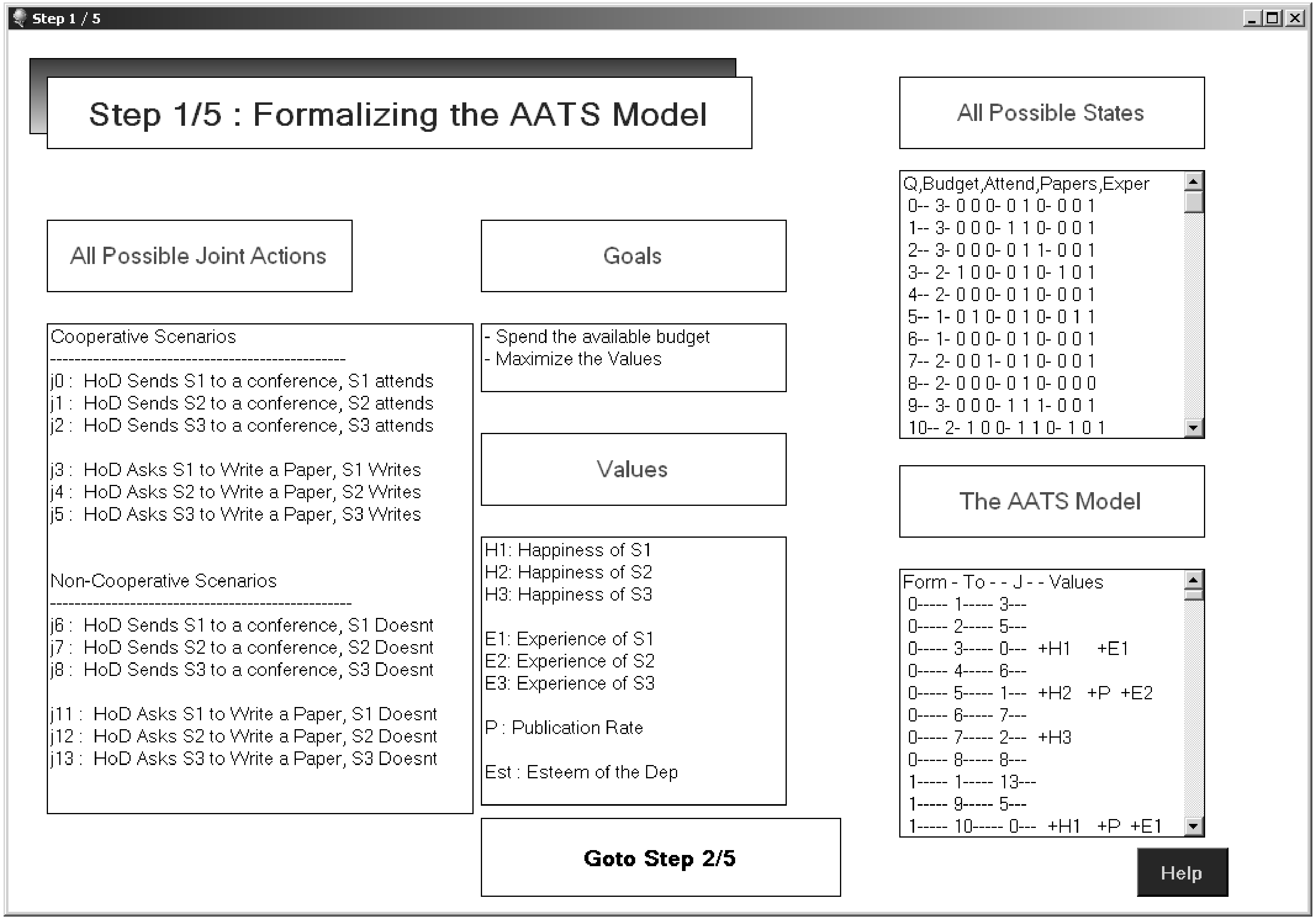
Figure 4.3 is the first step of the methodology where the variables of the system are shown (on the left) and the AATS model is calculated and shown (on the right). The AATS is shown in two different tables where the upper table shows all the possible states of the system and the lower one all possible transitions.

For clarity, it was assumed in this experiment that students do not collaborate on writing papers, and if any student was asked to write a paper, he will do it on his own.

We now move to the second step where we start building arguments for and against performing the various actions.

4.3.2 Determining the Arguments

The AATS allows the evaluation of each action at every state and relates the actions to propositions and the values they promote. Figure 4.4 shows the arguments that can be made for performing the actions available in the initial state.



Step 1/5 : Formalizing the AATS Model

```
Cooperative Scenarios
------------------------------------------------
j0 :  HoD Sends S1 to a conference, S1 attends
j1 :  HoD Sends S2 to a conference, S2 attends
j2 :  HoD Sends S3 to a conference, S3 attends

j3 :  HoD Asks S1 to Write a Paper, S1 Writes
j4 :  HoD Asks S2 to Write a Paper, S2 Writes
j5 :  HoD Asks S3 to Write a Paper, S3 Writes


Non-Cooperative Scenarios
------------------------------------------------
j6 :  HoD Sends S1 to a conference, S1 Doesnt
j7 :  HoD Sends S2 to a conference, S2 Doesnt
j8 :  HoD Sends S3 to a conference, S3 Doesnt

j11 :  HoD Asks S1 to Write a Paper, S1 Doesnt
j12 :  HoD Asks S2 to Write a Paper, S2 Doesnt
j13 :  HoD Asks S3 to Write a Paper, S3 Doesnt
```

```
- Spend the available budget
- Maximize the Values
```

```
H1: Happiness of S1
H2: Happiness of S2
H3: Happiness of S3

E1: Experience of S1
E2: Experience of S2
E3: Experience of S3

P : Publication Rate

Est : Esteem of the Dep
```

```
Q,Budget,Attend,Papers,Exper
0-- 3- 0 0 0- 0 1 0- 0 0 1
1-- 3- 0 0 0- 1 1 0- 0 0 1
2-- 3- 0 0 0- 0 1 1- 0 0 1
3-- 2- 1 0 0- 0 1 0- 1 0 1
4-- 2- 0 0 0- 0 1 0- 0 0 1
5-- 1- 0 1 0- 0 1 0- 0 1 1
6-- 1- 0 0 0- 0 1 0- 0 0 1
7-- 2- 0 0 1- 0 1 0- 0 0 1
8-- 2- 0 0 0- 0 1 0- 0 0 0
9-- 3- 0 0 0- 1 1 1- 0 0 1
10-- 2- 1 0 0- 1 1 0- 1 0 1
```

```
Form - To - - J - - Values
0----- 1----- 3---
0----- 2----- 5---
0----- 3----- 0---  +H1     +E1
0----- 4----- 6---
0----- 5----- 1---  +H2   +P  +E2
0----- 6----- 7---
0----- 7----- 2---  +H3
0----- 8----- 8---
1----- 1----- 13---
1----- 9----- 5---
1----- 10----- 0---  +H1   +P  +E1
```

Figure 4.3: Step 1 : Formulating the problem

We can see from Figure 4.4 how these arguments differ with respect to the values promoted. The next step is to use Critical Questions to identify which arguments are open to attack.

The arguments are determined as follows: looking into the AATS, any movement from one state to another state yields an argument which can justify this move. If the move promotes more than one value (e.g., sending a student to a conference might promote P and H), then each one represents a separate argument justifying this action (e.g., Arg1 and Arg2 both propose J0, but Arg1 promotes H, and Arg2 promotes E).

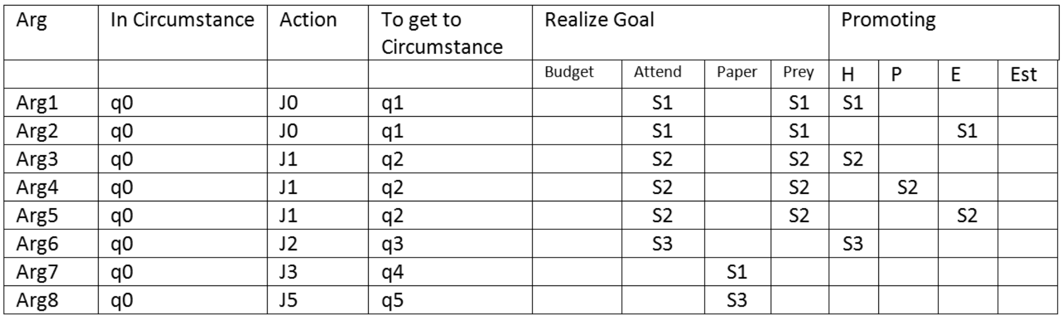
These arguments are structured into the argumentation scheme presented in Chapter 3, and are represented in the table shown in Figure 4.4 and Figure 4.5.

Arg1 and *Arg2* recommend sending the first student to a conference. Both options would realise the goals of the student attending a conference and obtaining experience by moving from $q0$ to $q1$. The first option argues that this will promote the student's happiness (which is an important value to the agent), while the other option suggests the same action for a different reason (the promotion of the student's experience).

Arg3, *Arg4*, and *Arg5* recommend sending the second student to a conference, which would result in moving from $q0$ to $q2$. Where *Arg3 and Arg5* use the same arguments as *Arg1 and Arg2* but relate to student 2, *Arg4* argues that sending the second student would promote *Publication* as this student has a paper ready. *Arg6* is an argument to send the third student to the conference, which would result in reaching $q3$ and promotes the happiness of the third student.

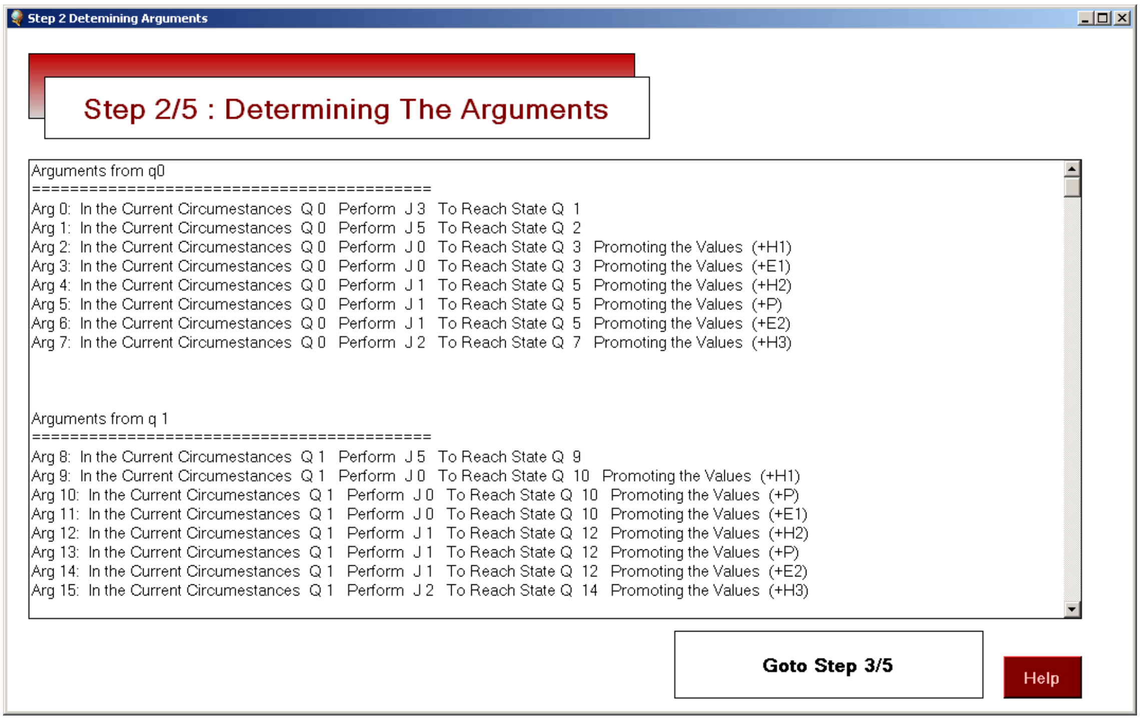
Arg7 and *Arg8* recommend asking the students (first and third) to write a paper first to reach ($q4$ and $q5$) where the goal of writing a paper is achieved. On their own, *Arg7 and Arg8* do not promote any value of relevance to the HoD, but if they were followed by an action of sending those students to conferences, they would promote *Publication* in addition to any values promoted by attendance.

Figure 4.5 is a snapshot from the implementation that shows a list of all arguments, not only from the initial state, but from all states of the system. This is shown in accordance with the argumentation scheme considering future states



Arg	In Circumstance	Action	To get to Circumstance	Realize Goal				Promoting			
				Budget	Attend	Paper	Prey	H	P	E	Est
Arg1	q0	J0	q1		S1		S1	S1			
Arg2	q0	J0	q1		S1		S1			S1	
Arg3	q0	J1	q2		S2		S2	S2			
Arg4	q0	J1	q2		S2		S2		S2		
Arg5	q0	J1	q2		S2		S2			S2	
Arg6	q0	J2	q3		S3			S3			
Arg7	q0	J3	q4			S1					
Arg8	q0	J5	q5			S3					

Figure 4.4: Different Arguments



Step 2/5 : Determining The Arguments

```
Arguments from q0
==========================================
Arg 0:  In the Current Circumestances  Q 0   Perform  J 3   To Reach State Q  1
Arg 1:  In the Current Circumestances  Q 0   Perform  J 5   To Reach State Q  2
Arg 2:  In the Current Circumestances  Q 0   Perform  J 0   To Reach State Q  3   Promoting the Values  (+H1)
Arg 3:  In the Current Circumestances  Q 0   Perform  J 0   To Reach State Q  3   Promoting the Values  (+E1)
Arg 4:  In the Current Circumestances  Q 0   Perform  J 1   To Reach State Q  5   Promoting the Values  (+H2)
Arg 5:  In the Current Circumestances  Q 0   Perform  J 1   To Reach State Q  5   Promoting the Values  (+P)
Arg 6:  In the Current Circumestances  Q 0   Perform  J 1   To Reach State Q  5   Promoting the Values  (+E2)
Arg 7:  In the Current Circumestances  Q 0   Perform  J 2   To Reach State Q  7   Promoting the Values  (+H3)


Arguments from q 1
==========================================
Arg 8:  In the Current Circumestances  Q 1   Perform  J 5   To Reach State Q  9
Arg 9:  In the Current Circumestances  Q 1   Perform  J 0   To Reach State Q  10   Promoting the Values  (+H1)
Arg 10:  In the Current Circumestances  Q 1   Perform  J 0   To Reach State Q  10   Promoting the Values  (+P)
Arg 11:  In the Current Circumestances  Q 1   Perform  J 0   To Reach State Q  10   Promoting the Values  (+E1)
Arg 12:  In the Current Circumestances  Q 1   Perform  J 1   To Reach State Q  12   Promoting the Values  (+H2)
Arg 13:  In the Current Circumestances  Q 1   Perform  J 1   To Reach State Q  12   Promoting the Values  (+P)
Arg 14:  In the Current Circumestances  Q 1   Perform  J 1   To Reach State Q  12   Promoting the Values  (+E2)
Arg 15:  In the Current Circumestances  Q 1   Perform  J 2   To Reach State Q  14   Promoting the Values  (+H3)
```

Figure 4.5: Step 2: The different arguments in the system

discussed in Chapter 3.

4.3.3 Building the Argumentation Framework

In this step, we take the arguments built in the previous step and subject them to the critical questions explained in Chapter 3, which would yield a relationship model between those arguments where arguments attack each other, which will allow us to eliminate weaker arguments, with respect to the HoD's preferences.

Arguments in the argumentation framework represent a justification to perform an action and hence move to a different state of affairs. This justification is linked with social values that are important to the agent. This was introduced in [14] where every transition from the set Q may either promote, demote, or be neutral with respect to those values. The critique offered by the Critical Questions may question the arguments claim of promoting a value, or might even bring attention to other values that can be demoted.

Below is a sample explanation of applying the CQs to the arguments from q0. Figure 4.6 presents the VAF of the model.

CQ1: Are the stated circumstances true?

This question arises in this example from the fact that although the HoD believes that the initial state is 3-000-010-001 (q0) where S2 has written a paper, he cannot actually be sure that the other students have not also written papers or that S2 has in fact written a paper. This results in eight different possible initial states. So, in this case, all the arguments are open to this attack. For simplicity and better presentation of the example, we only consider this type of uncertainties (not having written a paper). Nonetheless, in a more extended version of the example study, we might have uncertainties also about the budget, actually attending the conference and the experience of the students.

The agent could assume that all states are possible and build up the argumentation model with all the possible states in mind. This would result in multiple Preferred Extensions (PEs), one for each possible initial state. The common elements in all PEs would then represent justifications of actions, which are unaffected by the uncertainties with respect to what is true in the initial state.

CQ11: Does the action preclude some other action which would promote some other value?

Any use of a budget for a purpose might preclude its use for other purposes. If S1 or S3 are sent to the conference without being asked to write a paper, the chance to promote Publication is lost. Moreover, this will also lose the chance to promote Esteem, which requires S3 to be sent with a paper written. Thus $Arg1$, $Arg2$, and $Arg6$ are all attacked by an argument, $A1a$, in that they prevent the promotion of Publication. $Arg6$ is also attacked by an argument, $Arg6a$, that it precludes the promotion of Esteem.

CQ17: Does the action have the stated consequences?

This question occurs when we need to consider joint actions or cases in which the agent is not in sole control of the state reached. In our example, this is represented by the possibility of the request to write a paper not being met. Thus, $Arg7$ and $Arg8$ are attacked by $Arg7a$, so that the joint action might turn out to be $J4$.

CQ8: Does the action have side effects which demote the value?

Sending any of the students other than S2, will demote the happiness of S2 since he has already written a paper. So, this will give an argument, $Arg1b$, against $Arg1$ and $Arg6$: while these arguments promote happiness of S1 and S3, the actions they justify also demote the happiness of S2.

CQ9: Does the action have side effects which demote some other value?

If a student who has written a paper is not sent, the happiness of that student will be demoted. This provides the basis for an argument against doing any action which involves not sending S2 for any reason other than the promotion of happiness. Thus $Arg2$ is subject to an attack from $Arg2a$ since it would demote S2's happiness.

CQ7: Are there other ways to promote the same value?

Both $Arg2$ and $Arg5$ are based on the promotion of experience. This question indicates that they attack one another. Similarly, happiness can be promoted by any of $Arg1$, $Arg3$ and $Arg6$. These also mutually attack.

Now, we have identified all possible attacks from the different arguments we are able to form the Value-Based Argumentation Framework. Figure 4.6 is a graphical representation of the framework in the example.

Argument	CQ
Arg1b	CQ8
Arg2a	CQ9
Arg1a	CQ11
Arg7a	CQ17
Arg6a	CQ11

Table 4.6: Arguments rising from the CQs

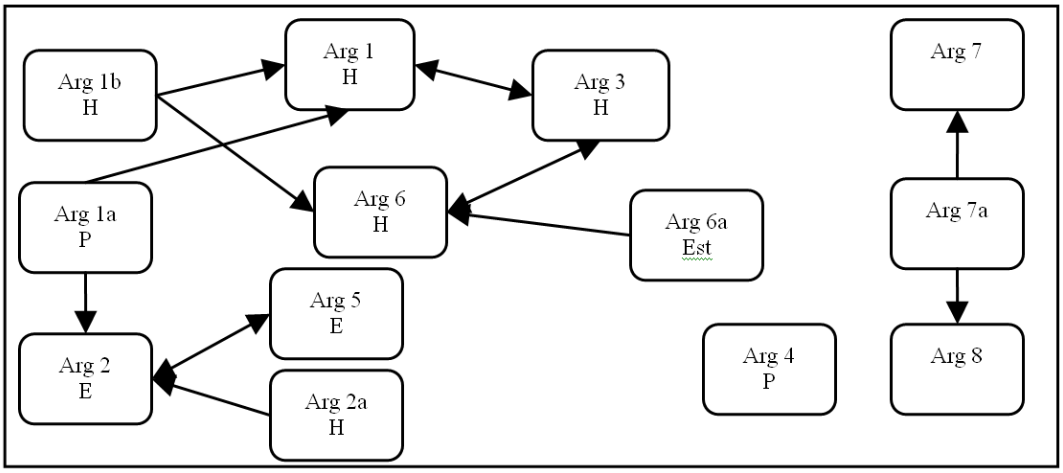


Figure 4.6: VAF of the model

4.3.4 Evaluating the Argumentation Framework

From the VAF in Figure 4.6, we can see that $Arg4$ has no attacking arguments. Thus, it appears in every Preferred Extension, irrespective of the ranking of values. The status of $Arg7$ and $Arg8$ depend on whether the HoD is confident that the papers will be written if requested. Suppose his confidence is sufficient, and so $Arg7$ and $Arg8$ are acceptable. In order to determine which of the remaining arguments are acceptable, the values that the HoD wishes to promote at the particular time need to be ordered. Suppose that the value ordering is as follows: Esteem > Publication > Experience > Happiness.

This gives us the ability to resolve the conflicts that we have in the model by eliminating unsuccessful attacks. $Arg1b$ defeats $Arg1$ and $Arg6$, leaving $Arg3$ for the Preferred Extension. Although $Arg2$ is not defeated by $Arg2a$, it is defeated by $Arg1a$, and so $Arg5$ survives. Thus our Preferred Extension is $Arg1a$, $Arg1b$, $Arg2a$, $Arg3$, $Arg4$, $Arg5$, $Arg6a$, $Arg7$, $Arg8$. In terms of actions, sending S2 can be justified, as can be requesting a paper from S1 and S3. See Figure 4.7.

A full view of the Preferred Extension with consideration of all states is given

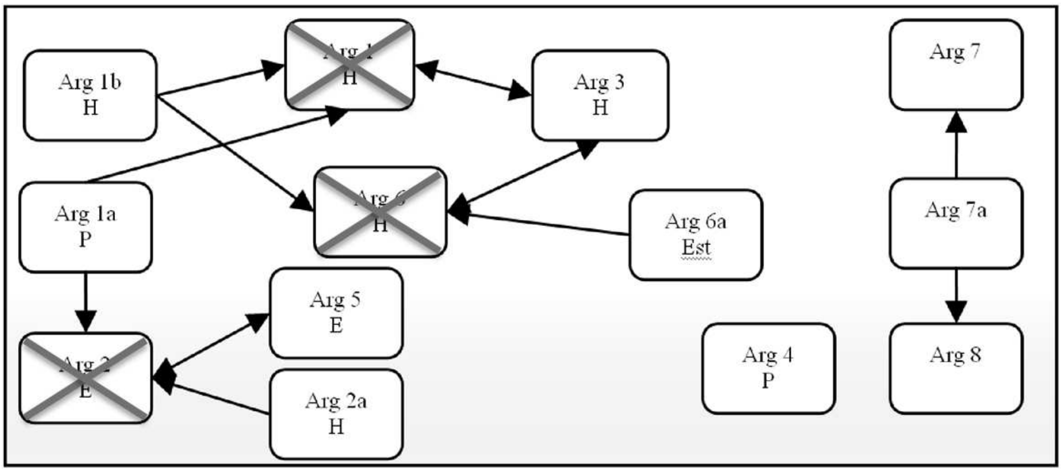


Figure 4.7: Evaluation of Framework

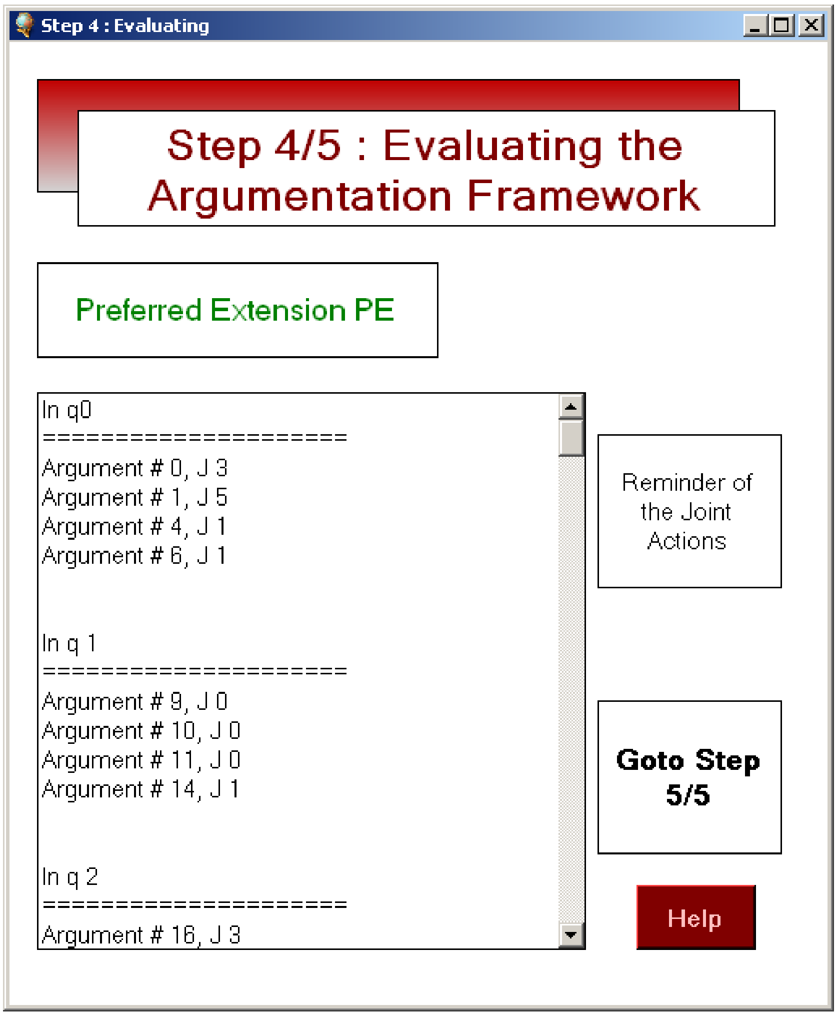
in Figure 4.8, where we can see the PE in each state within our AATS. This snapshot covers arguments from all different states and in consideration of all possible CQs.

We can see that the HoD has three options in $q0$: he can either send the second student to a conference and promote Publication, ask the first student to write a paper, or ask the third student to write a paper. If the agent chooses any of those options, the student will either fail, which would return the system to the same state (if he was asked to write a paper) or a different state where the budget is reduced and the student does not attend a conference (if he was asked to attend a conference) or the student might adhere and be successful in achieving the expectations which might lead to the expected state where successful arguments are also calculated (See Figure 4.8), then the HoD will have another set of possible actions to perform. For example, in $q1$, the HoD can only send the first student or send the second student to a conference; no other action passes the evaluation of the VAF in $q1$ (See Figure 4.8).

4.3.5 Sequencing the Actions

The result of the successful arguments from all states can now be modeled as a graph (Figure 4.9), where as we have different arguments in the PE of every state we see that there are different possibilities at each state. For example, depending on what action the HoD takes in $q0$, he can reach $q2$, $q4$, or $q5$.

The last step identified three actions, which would move to q2, q4, or q5. However, if the confidence in the student's ability to produce a paper was misplaced, the state would remain q0. Some sensible sequence for these actions


Step 4 : Evaluating
Step 4/5 : Evaluating the Argumentation Framework
Preferred Extension PE
Reminder of the Joint Actions
Goto Step 5/5
Help

```
In q0
======================
Argument # 0, J 3
Argument # 1, J 5
Argument # 4, J 1
Argument # 6, J 1


In q 1
======================
Argument # 9, J 0
Argument # 10, J 0
Argument # 11, J 0
Argument # 14, J 1


In q 2
======================
Argument # 16, J 3
```

Figure 4.8: Step 5: Evaluation of the argumentation framework

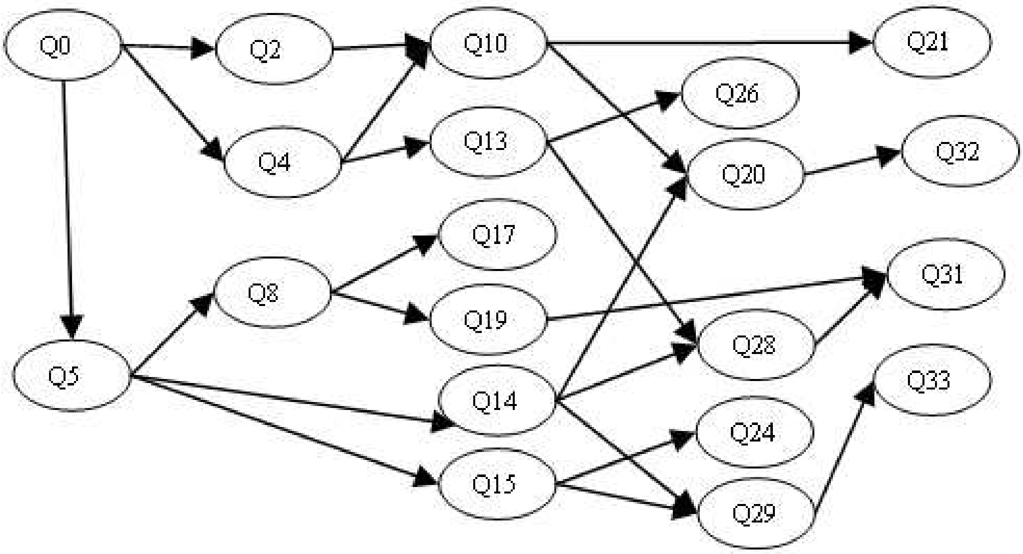


Figure 4.9: Possible Sequences

must be chosen. This choice needs to consider uncertainty about the outcome to eventually reach the state best for the agent given the action of the other agent. This requires looking ahead to consider what is possible in the states that would result from our action. In our work we consider sequencing of action by ordering their execution one after the other, we do not consider parallel actions, choices of actions, iteration and interleaving.

There are three issues we should consider here. First, we need to consider whether the action is safe, in the sense that if it fails, we do not move to a state where our other desirable actions are no longer possible. Next, we must consider opportunities: what additional values can be promoted in the next state? If we ask S1 to write a paper, we have the possibility of promoting Publication (the chance to promote S1's Experience already exists as $Arg2$), and, although we can already promote this by sending S2, S1's Publication would be an additional benefit. If we ask S3 to write, we can create the possibility to promote Esteem, as well as the additional Publication. There are also threats: if S1 and S3 write papers and are not sent, they will be unhappy. Since we have said that we prefer Esteem to Experience, we prefer the opportunities created by requesting a paper from S3, and so prioritize this action over asking S1. Sending S2 demotivates S1 and so reduces the likelihood of her producing a paper. Suppose, however, we make this judgment: then we should request a paper from S3 before sending S2. If S3 does write the paper, we move to another state in which the recommended actions are to send S2 and S3. If, however, S3 does not produce a paper, we have no possibility of sending S3, and no threat of making S3 unhappy. We should then ask S3 first, if S3 does not write a paper we should request a paper from S1, in the hope of making an opportunity to promote Publication. If S1 does write, we should then send S1 and S2: and even if S1 does not write a paper he should be sent as this will still promote Experience with respect to S1.

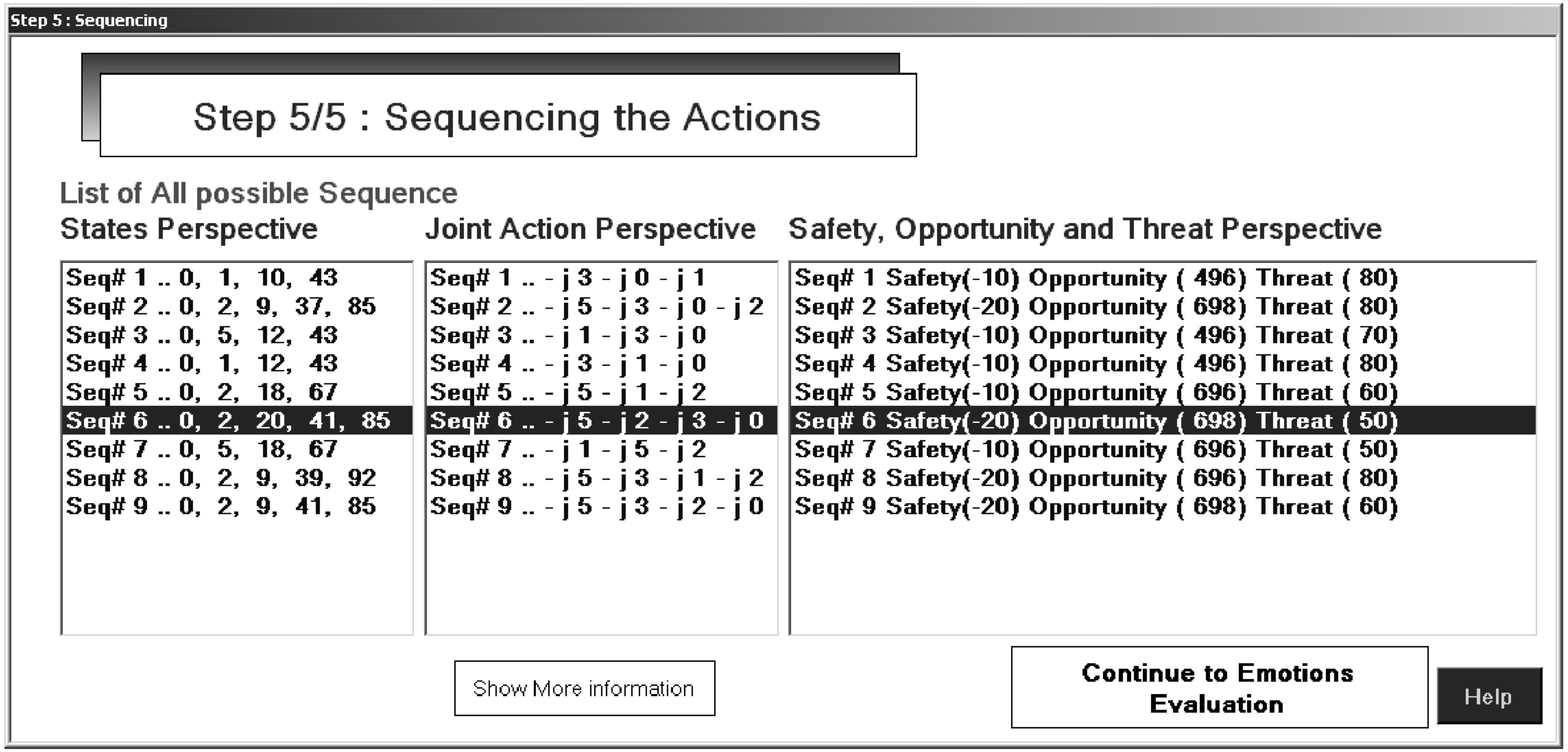
The Final Result

Figure 4.10 is a snapshot of the calculation of Safety, Opportunity and Threat in the experiment. Sequence number 6 is the best choice, as it has the lowest threat and highest opportunity in the system. The sequence is as follows (Ask S3 to write, Send S3, Ask S1 to write a paper, then send S1). This is because asking the third student first where the budget is still high would give him the comfort that he would actually be sent if successful and thus reduce the threat of S3 not writing the paper.

Note that in the implementation (Figure 4.10) we quantify Safety, Opportunity and Threat with numerical values instead of treating them in a binary fashion. This was found to give more accurate analysis in this implementation. It is more convenient to distinguish the level of Safety of different sequences rather than just flagging it with either Safe or Unsafe. The same goes for Opportunity and Threat as it was more useful in the implementation to exactly define how big or small the Opportunity is and in Threat how serious the threat is. As for the numbers themselves, those are arbitrarily chosen.

At the initial stage $q0$, sending S2 to a conference promotes Publication where sending S3 would only promote Happiness. The HoD nevertheless decides to send S3 rather than S2 in the second step after asking S3 to write a paper. The reason behind that is when S3 has successfully written the paper, sending him to a conference promotes better value than sending S2 as S3 will not only promote Publication, but would also promote Esteem.

After that the HoD has a choice of either sending S2 or S1. He chooses to ask S1 to write a paper and then send him because it is cheaper than sending S2 although S2 has a paper written.



Step 5 : Sequencing

Step 5/5 : Sequencing the Actions

List of All possible Sequence

States Perspective	Joint Action Perspective	Safety, Opportunity and Threat Perspective
Seq# 1 .. 0, 1, 10, 43	Seq# 1 .. - j 3 - j 0 - j 1	Seq# 1 Safety(-10) Opportunity (496) Threat (80)
Seq# 2 .. 0, 2, 9, 37, 85	Seq# 2 .. - j 5 - j 3 - j 0 - j 2	Seq# 2 Safety(-20) Opportunity (698) Threat (80)
Seq# 3 .. 0, 5, 12, 43	Seq# 3 .. - j 1 - j 3 - j 0	Seq# 3 Safety(-10) Opportunity (496) Threat (70)
Seq# 4 .. 0, 1, 12, 43	Seq# 4 .. - j 3 - j 1 - j 0	Seq# 4 Safety(-10) Opportunity (496) Threat (80)
Seq# 5 .. 0, 2, 18, 67	Seq# 5 .. - j 5 - j 1 - j 2	Seq# 5 Safety(-10) Opportunity (696) Threat (60)
Seq# 6 .. 0, 2, 20, 41, 85	Seq# 6 .. - j 5 - j 2 - j 3 - j 0	Seq# 6 Safety(-20) Opportunity (698) Threat (50)
Seq# 7 .. 0, 5, 18, 67	Seq# 7 .. - j 1 - j 5 - j 2	Seq# 7 Safety(-10) Opportunity (696) Threat (50)
Seq# 8 .. 0, 2, 9, 39, 92	Seq# 8 .. - j 5 - j 3 - j 1 - j 2	Seq# 8 Safety(-20) Opportunity (696) Threat (80)
Seq# 9 .. 0, 2, 9, 41, 85	Seq# 9 .. - j 5 - j 3 - j 2 - j 0	Seq# 9 Safety(-20) Opportunity (698) Threat (60)

Figure 4.10: Action Sequencing

4.3.6 Observations and Results

Table 4.7 shows the plan of action chosen by the HoD with respect to all possible audiences in the system (we have 24 possible audiences). We have four main values in the system: (The Department's Esteem (Est), Publication rate (P), Student's Experience (E) and student's Happiness (H)). For simplicity, we consider variances here in the values ordering from a general level not from the student level and we assume that all three students are treated equally. So, $H1 = H2 = H3$ which is represented in Table 4.7 with only H and $E1 = E2 = E3$ which is represented in the table with E. $j0 - 2$ are joint actions presenting the HoD asking students 1,2 or 3 to attend a conference and they do. $j3 - j5$ are the HoD asking students 1, 2 or 3 to write a paper and they do.

No	Audience	Course of Action
1	Est > P > E > H	j5 → j2 → j3 → j0
2	Est > P > H > E	j5 → j2 → j3 → j0
3	Est > E > P > H	j5 → j2 → j1
4	Est > E > H > P	j5 → j2 → j1
5	Est > H > E > P	j5 → j2 → j1
6	Est > H > P > E	j5 → j2 → j1
7	H > P > Est > E	j1 → j3 → j0
8	H > P > E > Est	j1 → j3 → j0
9	H > E > Est > P	j1 → j3 → j0
10	H > E > P > Est	j1 → j3 → j0
11	H > Est > E > P	j1 → j3 → j0
12	H > Est > P > E	j1 → j3 → j0
13	E > P > Est > H	j3 → j1 → j0
14	E > P > H > Est	j3 → j1 → j0
15	E > H > Est > P	j3 → j1 → j0
16	E > H > P > Est	j3 → j1 → j0
17	E > Est > H > P	j3 → j1 → j0
18	E > Est > P > H	j3 → j1 → j0
19	P > H > Est > E	j1 → j3 → j0
20	P > H > E > Est	j1 → j3 → j0
21	P > E > Est > H	j1 → j3 → j0
22	P > E > H > Est	j1 → j3 → j0
23	P > Est > H > E	j3 → j5 → j2 → j0
24	P > Est > E > H	j1 → j3 → j0

Table 4.7: All Possible sequences in all possible audiences

We note that whatever the value order chosen, there are only five possible courses of actions possible to the HoD from q0 (Table 4.8).

No	Abbreviated Sequence	The Sequence Explained
1	j5 → j2 → j3 → j0	Ask S3 to write → Send S3 → Ask S1 to write → send S1
2	j5 → j2 → j1	Ask S3 to write → Send S3 → Send S2
3	j1 → j3 → j0	Send S2 → Ask S1 to write → Send S1
4	j3 → j1 → j0	Ask S1 to write → Send S2 → Send S1
5	j3 → j5 → j2 → j0	Ask S1 to write → Ask S3 to write → Send S3 → Send S1

Table 4.8: An overview of all possible sequences

Observations

- We note that regardless of value order, whenever S3 is involved in the plan (Table 4.8, numbers 1, 2 and 5) the HoD always sends S3 before sending the other students to their respective conferences. The reason behind it is that S3 is known to be cautious when other students are sent and might not act as expected when such a thing happens as he would be afraid that he will not be sent and as a result might not write a paper or register for the conference. Calculations of threat in step five of the methodology prompts the HoD to always send S3 first.

- From Table 4.8 we also observe that in all eventualities, the HoD does not send S1 or S3 unless he asks them to write a paper first. No matter what value order we have it is always better to ask S1 and S3 to write a paper before actually sending them to any conference as this might promote Publication.

- Ordering of values affects not only the choice of action, but also the ordering of execution. In the second table, (numbers 1 and 5) and (numbers 3 and 4) suggest the same actions be executed but in different orders depending on the audience.

- We note from Table 4.7 that audiences 3 to 6 agree on the same sequence of actions; the case is the same with (7 to 12 and 19 to 22 and 24) and with (13 to 18). We can draw two different conclusions from that: first, having different audiences in the system might not necessarily result in a conflict since different audiences who disagree on value orders might eventually agree on the plan of action. Second, in cases where the audience is only partially determined, such as 13 to 18, where the only available information we have on the audience is that the value of Experience (E) is the most important value, we actually do not require any additional information to decide on the course of action most suitable to that audience. No matter what value order the audience has if his most important value is Experience, the plan of actions would be the same.

- Because the HoD is confident about S3 where he fears that he would not actually adhere to the HoD's instruction, we note that the HoD chooses to send S3 to a conference only when Esteem (Est) is the most important value in the value order.

- In numbers 13 to 18 in Table 4.7, we show j3 → j1 → j0 as the chosen sequence where in the program it was a tie with j3 → j0 → j1 as they both scored the same when it comes to Safety, Opportunity and Threat. When Experience is the most important value to HoD, it does not matter if S1 or S2 is sent first, and so the program makes an arbitrary choice.

4.4 Elicitation of Trust: Did it Happen?

The experiment is evaluated from two perspectives: one, by considering how the application and the results relate and address the four out of five elements of Trust; two, using measurement of "Values" and "Goals", where the evaluation is based on assessing their differences in terms of changing environment/behaviours.

The four elements of Trust we are trying to address in this chapter are: Practical Reasoning, Social Interaction, Planning and Uncertainties. This section will discuss the results of the experiment in light of the different elements of Trust defined in Chapter 1.

Practical Reasoning (options and their likely effects)

> This corresponds to the first step of the methodology "Formulating the problem" where after the basic elements of the problem were defined a complete tree of possibilities was then built to draw on the different actions the HoD can take and what their effects are from a state, goal or values perspectives (Figure 4.1).
>
> As explained; the HoD will perform one action and then the other agent (Student) will respond with another forming a joint action that can lead

to a certain state promoting or demoting some values. It was shown that when the HoD performs an action, the response from the student may vary giving the possibility of more than one joint action. The AATS model will show these different possibilities and also their impact on future actions.

When the HoD asks the third student to write a paper we saw how this might result in actually writing the paper. This would then allow the promotion of Publication and Esteem once the student attends a conference or the student might choose not to do so, thus losing the opportunity to promote Publication at least. Whereas if a different choice took place (asking Student 1 to write a paper), we might have had the chance to at least promote Publication. Thus, it is important to take all possibilities into consideration when deciding on an action and not just consider beneficial outcomes.

Therefore, considerations of pros and cons were successfully captured in the AATS model.

Social Interaction (other agents' actions and how they affect ours)

Going back to Table 4.2, different possible actions by different agents were defined. Table 4.3 then combined those actions into joint actions to form all the possibilities that could arise in situations where the effects of the agents actions are determined not only by his but also by other agents actions.

This is why consideration from that step onwards was only given to joint actions. When building the argumentation model it was shown how arguments were raised based on joint actions that the HoD was hopeful to achieve and then argued in step three using critical questions to generate counter arguments based on the possibility of other undesirable joint actions actually happening in these circumstances.

In the last step, where actions are sequenced into a path, "Safety" covers possibilities where unwanted joint actions occur, as well as the likelihood of other agents performing unexpected actions. "Opportunity" covers aspects where desirable joint actions might have better chances of yielding a particular order.

Looking at Figure 4.10, we notice that although some sequences suggest the same set of actions, they are ordered differently; thus, the measurements in Safety and Opportunity vary.

Planning (short-term and long-term considerations) Although planning is not part of this thesis (see Section 1.4), this methodology provides a basic consideration of planning where agents consider future possibilities when performing any action, allowing a long-term perspective in each and every decision being made.

This affects all the steps in the methodology, starting from step one, where an AATS model was built not only from the initial state where the decision needs to take place but also covering all possible future states (see Figure 4.1). As we saw, this allowed the HoD to consider the value of Publication when asking a student to write a paper; although writing a paper would not actually promote this value a subsequent action would.

The planning concept might appear more in the last step, where actions are sequenced into different paths and then future possibilities are considered and given their weight.

While sequences 2 and 6 in Figure 4.10 suggest the same set of actions, sequence 6 is preferred as it has less threat than sequence 2.

Uncertainties (side effects and working with incomplete information) Uncertainties are presented in this methodology by subjecting each argument to perform an action to the sixteen different critical questions (Section 2.5.3).

Uncertainties are addressed once again in the sequencing of actions where we have different uncertainties according to different possible sequences and side effects of whether there are possibilities of demoting actions while promoting others. Moreover, we consider the possibility of performing further remaining action when current action fails to reach the expected state.

4.5 Summary

In summary, this chapter provided an experimental study where the methodology presented in Chapter 3 was modeled and tested. This model was implemented in Visual Basic, and has been explained in detail throughout the chapter. A case study was used to represent a dilemma where the head of an academic department wants to send two out of three students to conferences and had to choose between them.

As Chapter 3 provided a theoretical formalisation of the methodology, this example gave a practical explanation through an example study.

Our main motivation is to build a decision-making methodology that can be trusted and eventually allow the assignment of critical decisions to agents instead of humans. The elements of Trust were identified and extracted from philosophy in Chapter 2. We do not claim to have achieved the aspiration of a trusted decision; but rather, a contribution where a foundation is built that can accommodate the basic elements of Trust. A discussion of this contribution was given in Section 4.4.

This is the end of Part I of the thesis in which Trust was considered as a product of belief. Next, Part II extends this methodology in Chapter 5 to accommodate the element of emotions and further extends this example study in Chapter 6.

4.6 Introducing Part II: Trust as a product of emotions

> The influence of emotions upon beliefs can be viewed as the port through which emotions exert their influence upon human life. Beliefs fueled by emotions stimulate people to action, or allow them to approve of the actions of others in political context. Frijad et al. [61][Page 1].

Part I of this thesis covered a methodology based on Trust from the perspective of beliefs and built on rationality using argumentation schemes. Chapter 3 gave a formalised theory, and Chapter 4 provided an experimental study.

Next, is Part II of this thesis in which Trust is looked at from emotional perspectives. The methodology presented in Part I will now be extended by allowing emotions to have an effect. Chapter 5 complements the methodology of decision-making to account for emotions, and Chapter 6 extends the experimental study to examine their effects.

Part II

Trust as a Product of Emotions

Chapter 5

Emotions

In Section 2.1, a trustworthy decision was shown to be a product of both beliefs and emotions, after which five elements were identified as the principal factors influencing Trust. The first four elements related to decisions as a product of belief; the last one related to emotions. Chapter 3 introduced a methodology of action selection that addressed the first four elements of Trust. A detailed experimental study in Chapter 4 implemented these four elements.

This chapter extends this methodology to address the fifth element of a trustworthy decision: emotions. Our focus here is to position emotional factors as a complement to rationality in decision making. The overall decision-making methodology should be built on a combination of beliefs and consideration of emotional aspects rather than relying on only one of them. It is well understood that emotional influence might vary with the type of decision to be made and the type of goal to be achieved.

In the decision-making methodology in Chapter 3 the problem is first formulated, and then a state transition model is built identifying all possibilities. An argumentation model is created, linked, and evaluated to yield a set of acceptable arguments that are then sequenced into a path of actions. The arguments that are acceptable to an agent depend on the way in which that agent orders its values. All of these steps were performed by agents based solely on their beliefs; and emotions do not play any role. We now focus on the effect that emotions can have. The key idea is that emotions influence the ordering of values especially those that relate to particular individuals and this influence on the value order will in turn affect which arguments the agent finds acceptable. In so far as emotions change, the value order that will impact on the action performed. At the time of sequencing actions, emotions will then influence the sequence based on the success rate of this sequence and how it affects the emotional state. If the changes in emotions are strong enough to influence the preference model of the agent (Value Order), resequencing can occur and the new value order is considered. The success or failure and the reasons for the success or failure elicit the emotional responses.

Section 5.1 introduces the subject and the main motivations behind this study. A brief background of the philosophical as well as technical aspects of emotions is given. Section 5.2 explains how emotions are positioned within the decision-making framework. Section 5.3 provides a detailed explanation of the emotional model in rational decision making and links this work to the methodology described earlier. Section 5.4 explains and formalises how this emotional model influences the decision-making methodology. Section 5.5 provides a summary of the main highlights of this chapter.

5.1 Introduction

> "Emotions and feelings can cause havoc in the process of reasoning under certain circumstances. Traditional wisdom has told us that they can, and recent investigation of the normal reasoning process also reveals the potentially harmful influence of emotional biases. It is thus even more surprising and novel that the absence of emotion and feeling is no less damaging, no less capable of compromising the rationality that makes us distinctively human and allows us to decide in consonance with a sense of personal future, social convention, and moral principle." Damasio [45][Page xii]

Researchers in AI have taken several routes when modeling emotions and this study was influenced by some of those approaches.

One important approach, developed by Reilly [107], is targeted at expressing emotions rather than influencing decisions. The objectives of Reilly were completely different from ours as he is focused on expressing emotions for dramatic effect whereas this study is focused on the influences on decision making. Nonetheless, this study is influenced by Reilly's work as it provides a clear and efficient approach to storing, decaying and weighing emotions.

The work of Steunebrink et al [119], which takes the OCC model of [93] as its basis, presented a formalisation of a number of emotions in terms of agents using the BDI framework. Their work provided a concise formalisation of emotions and their effects, which this study will use, suitably adapted from the BDI to AATS settings as its underpinning formal basis.

5.1.1 Motivation

> "The influence of emotions upon beliefs can be viewed as the port through which emotions exert their influence upon human life. Beliefs fueled by emotions stimulate people to action, or allow them to approve of the actions of others in political context". Frijda [61][Page 1]

The inclusion of emotions in rational decision making has been motivated by different aspects:

- While performing a sequence of actions, determine whether resequencing is necessary, or whether the current plan can still be followed. As noted in Steunbrink et al. [119] emotions can be a trigger for resequencing.
- Emotions play an important role in social interaction. Humans adopt emotional attitudes towards one another, and this seems to play an essential role in developing and maintaining cooperation, as well as consistently appropriate behaviour.
- These emotions also seem to act as tie-breakers to enable a reasoned choice between two alternatives that are equally acceptable on purely rational grounds.

.

5.1.2 Emotions: AI and Philosophy

The emotions model of Ortony, Clore, and Collins (OCC) [93] identifies twenty-two emotions organised into a hierarchy. OCC does not claim that it covers all possible human emotions; rather, it identifies emotional types where every type refers to all similar emotions in the same category. The OCC model is attractive to AI researchers, as it provides a specification of the conditions that give rise to each distinct emotion in terms of the computationally familiar notions of objects, events and actions [119]. Also, the OCC model includes notions of intensity of emotions, and identifies a number of variables that influence the intensity of emotions.

Reilly [107] used the OCC model to build a methodology of implementing an emotional system. His aim was to construct believable emotional agents for artists to use to create a dramatic environment. Thus, he was more focused on expressing emotions rather than decision making, it is however useful to us as it gives a fully implemented model based on OCC. The work of this chapter uses some of the concepts and simplifications of OCC Reilly identified especially in its quantitative aspects. This basically covers the mechanisms used for storing emotions, combining them and decaying their intensities. Reilly's work also gives a simplified method of determining intensity, using only a subset of the variables from the OCC model, most importantly the *importance* and *unexpectedness* of the triggering event, which is of particular importance in emotion generation and intensity. As well as storage, this work is inspired by Reilly's mechanisms for combining emotions of a similar type, and enabling them to decay over time.

The OCC model uses the familiar notions of Events, Objects and Goals which map easily to our model where Agents, Goals and Values are the basis.

Events are things that happen, including the actions of agents, therefore, whether values are promoted or demoted and goals are achieved or not maps to the AATS model transitions, where every transition from one state to another can be seen as an event. We recall that the transitions in AATS use the joint actions that are performed to make these transitions to determine the effects on values and which goals are achieved. These events are judged to be either pleasing or displeasing according to the agent's goals. Goals represent anything that the agent wants, so they may be actively pursued. For example, the event of eating dinner when there is a goal to eat would be judged as being pleasing.

Objects (including agents) can be liked or disliked according to an agent's attitudes and behaviour. Attitudes represent personal tastes and preferences.

The work done by the Agents Group at Utrecht e.g., [81, 119] provides a formalisation of the emotions of the OCC model and has shown how *hope* and *fear* in particular can play a role in decision making by triggering resequencing. [119] defines the twenty-two emotional fluents of the OCC model. In Definition 5.1 their emotional fluents formalisation is given; these are subscript states of affairs with the name of an agent where appropriate. Some of the names have been changed where appropriate. For example, *Love* has become *Like* where the emotion of Love is too extreme for our purpose.

Definition 5.1. ***Emotional Fluents;****[119] The set emotions is the set of emotional fluents, which is defined as follows:*

$$
\begin{array}{ll}
\multicolumn{2}{l}{emotions =} \\
\quad joy_i(k_i), & distress_i(\neg k_i), \\
\quad hope_i(\pi, k_i), & fear_i(\pi, \neg k_i), \\
\quad satisfaction_i(\pi, k_i), & disappointment_i(\pi, \neg k_i), \\
\quad relief_i(\pi, k_i), & fears - confirmed_i(\pi, \neg k_i), \\
\quad happy - for_i(j, k_j), & resentment_i(j, k_j), \\
\quad gloating_i(j, \neg k_j), & pity_i(j, \neg k_j), \\
\quad pride_i(\alpha_i), & shame_i(\alpha_i), \\
\quad admiration_i(j, \alpha_j), & reproach_i(j, \alpha_j), \\
\quad like_i(j), & dislike_i(j), \\
\quad gratification_i(\alpha_i, k_i), & remorse_i(\alpha_i, \neg k_i), \\
\quad gratitude_i(j, \alpha_j, k_i), & displeasure_i(j, \alpha_j, \neg k_i),
\end{array}
$$

In abstract terms an emotion e is determined for some agent i that may involve attributes towards another agent j, and the actions of both agents (α_i, α_j) in respect of a plan π proposed by agent i and how far towards its goal (k_i) this plan has progressed.

We can further refine this view of distinct emotions by grouping those with relevant characteristics. The twenty-two emotions can be divided into eleven pairs: for example, *distress* is the negative correlate of *joy*, and so on down the list. We can group these pairs:

1. Happy-For, Resentment, Gloating, Pity, Admiration, Reproach, Like, Dislike, Gratitude and Displeasure: Emotions directed towards other agents either generically as an overall attitude such as Like and Dislike or because they realise a certain state of affairs (i.e. Happy-for or Gloating) or they performed a recognised action (i.e., Admiration) or performed an action that realises an important goal (i.e., Gratitude).

2. Pride, Shame, Gratification and Remorse: Emotions central to the agent itself and are related to actions attempted and either failed (Shame), succeeded (Pride), or which resulted in a desirable or undesirable events (Gratification and Remorse).

3. Hope, Fear, Relief and Fears-Confirmed: Probabilistic emotions, where whenever the probability of achieving a desirable goal increases or decreases

emotions of (Hope and Fear) occurs then as this hoped/feared goal succeeds or fails, this causes (Relief or Fears-Confirmed). One interesting point here is that Fear is a prerequisite to Fears-Confirmed, whereas Relief only occurs when a goal succeeded and there was some fear initially that it would not.

4. Joy, Distress, Satisfaction and Disappointment: Similar to item 2 above, these are central to the agents themselves and are related to goals of the agent succeeding or failing (Joy and Distress); or succeeding/failing as the agent performs a sequence of actions (Satisfaction and Remorse).

5.1.3 Emotion Types

This subsection describes each emotion type of the OCC model in more detail. The emotions are presented formally (as in Definition 5.1) to show their direction and how they are influenced. The variables that influence the generation and intensity of these emotions are highlighted and then aligned to the decision-making aspects.

Below, the subscript on the emotion itself indicates the agent having the emotion. This is then followed by the variables influencing this emotion whether this is an action (α), other agents (i or j), goal (k), or value (v). The subscript with each variable is the agent to which this variable refers, whether the same agent (i) or another agent (j) or a set of agents (C).

These emotions do not relate to events, but to actions of the agent itself which each form part of the family of events represented by joint actions. They also relate to the expectedness of these actions succeeding or failing. These emotions are related to the "Values" in VAF, where a desirable action is the one that promotes a value, and an undesirable action is the one that demotes a value.

$pride_i(\alpha_i)$, **and** $shame_i(\alpha_i)$,

Occurrence:

When an action α_i promotes a value, the emotion of pride of the same agent i is generated with an intensity that involves components of the expectedness and importance of this state of affairs.

When the action demotes a value, the emotion of shame is generated with an intensity that involves components of the expectedness and importance of this state of affairs.

For example, a student in a university prepares a paper for a high quality conference that is known to have a very low acceptance rate. Moreover, having a paper published in this respected conference would enhance the department's publication rate. Now, when the conference committee accepts this paper, the student will be *proud* with an intensity in line with the importance and unexpectedness of the event.

$admiration_i(j, \alpha_j)$, **and** $reproach_i(j, \alpha_j)$,

These emotions are similar to (Pride and Shame) but relate to a different agent's action, i.e., $i \neq j$.

Occurrence:
When an action α_j (of another agent j) promotes a value important to agent i, the emotion of "Admiration" is generated with an intensity from agent i that involves components of expectedness and importance of this state of affair.
When an action (of another agent) demotes a value important to agent i, the emotion of "Reproach" will be generated with an intensity that has components of expectation and importance of this state of affair.

For example, when the student succeeds in publishing his paper in the respected conference, his colleagues will admire him as he has promoted a value of importance to all of them (the department's publication rate).

$gratification_i(\alpha_i, V_i)$, **and** $remorse_i(\alpha_i, \neg V_i)$,

These are similar to (Pride and Shame): whereas pride and shame are influenced by the action itself, these emotions also consider the events involved whether these are desirable or undesirable, not only outcomes. This relates in VAF to both Values and Goals.

Occurrence:
Gratification and Remorse have two influencing variables: the action the agent performs and the value promoted or demoted in doing so (α_i,V_i). If the agent performs an action that resulted in promoting a value of importance to the agent himself, Gratification occurs; on the other hand if he fails, he experiences Remorse.

For example, the student feels proud because he succeeds in his action and gets his paper published. He also feels gratification because, in addition, his action managed to promote an important value 'the department publication rate'.

$gratitude_i(j, \alpha_j, V_i)$, **and** $displeasure_i(j, \alpha_j, \neg V_i)$,

These are similar to (Admiration and Reproach) with the difference that the events involved are also desirable or undesirable not only the outcome. In VAF, this relates to both Values and Goals. We can see from the variables above that they are similar to admiration and reproach but also relate to the agent's own goals and values.

Occurrence:
If the action performed by another agent is combined with the agent's own action to promote a value, gratitude is experienced. If, however, the other agent's action produces a joint action which fails to promote that value, the agent experiences displeasure.

For example, the Head of the Department (HoD) admires the successful student for his achievement and he also feels gratitude as the student co-operated with his request and together they had a successful joint action that resulted in promoting the value the HoD originally anticipated.

$joy_i(V_i)$, **and** $distress_i(\neg V_i)$,

These emotions relate to an important state of affairs that is either desirable and would cause "Joy", or undesirable and would cause "Distress". This is mapped to the agent's "Values" in the VAF system.

Occurrence:
Whenever a value that is of importance to the agent is promoted or demoted, one of these emotions occurs with an intensity equal to the level of importance of that value, regardless of how or why it occurs.

For example, the department manages to publish three different papers in the esteemed conference; exceeding the HoD's expectations, he will be generally joyful that the value of interest to him has been promoted and this emotion reflects on all the agents and actions that were involved.

$hope_i(C, V_i)$, **and** $fear_i(C, \neg V_i)$,

These are emotions associated with the probabilities of success of a "Goal" increasing or decreasing.

Whenever the probability of achieving a goal increases, the emotion "Hope" is generated with an intensity that is relative to the change in probability and the desirability of the state of affairs.

Whenever the probability of achieving a goal decreases, the emotion "Fear" is generated with an intensity that is relevant to the change in probability and the desirability of the state of affairs.

Occurrence:
When parts of the plan succeed, then Hope is generated; on the other hand, if some steps of the plan fail (where joint actions of agents in C lead to an undesirable state), then Fear occurs. This applies to values that require more than one step to be promoted.

For example, before asking for a promotion, the teacher plans to publish three papers in three different major conferences. As his first paper gets approved, he feels hope; his plan is moving as expected when he publishes the second paper.

$satisfaction_i(\underline{j}, V_i)$, **and** $fears - confirmed_i(\underline{j}, \neg V_i)$,

These relate to a state of affair happening similar to (Joy/Distress) with the difference that they have a precondition of the existence of (Hope or fear) to occur.

Whenever an agent achieves a goal that he was initially hopeful of achieving, the emotion "Satisfaction" occurs. On the other hand, whenever an agent fails to achieve a goal that he was initially fearful of achieving, the emotion "Fears-Confirmed" occurs.

Occurrence:
Whenever a sequence where hope exists is completed successfully, "Satisfaction" occurs. When a sequence fails where fear existed, "Fears-Confirmed" occurs.

For example, as hope exists and the teacher continues with his sequence to publish all three papers, he asks for a promotion; once his sequence finishes and he manages to get his promotion, the emotion Satisfaction would take place. On the other hand, if the sequence does not progress well (he fails to publish one of the papers), so the emotion fear was in place, if eventually he failed to achieve his goal ‘promotion’ he feels ‘Fears-Confirmed’.

$relief_i(\underline{j}, V_i)$, **and** $disappointment_i(\underline{j}, \neg V_i)$,

These are similar to (Satisfaction and Fears-Confirmed), with the difference that they relate to undesirable events.

Whenever a goal is achieved of which the agent was initially fearful the emotion “Relief” occurs. Whenever a goal fails that the agent was initially hopeful of achieving, the emotion ”Disappointment” occurs.

Occurrence:
Whenever a sequence is completed successfully where fear existed, “Relief” would occur and when a sequence fails where hope existed, “Disappointment” will take place. If an agent decides on a sequence and starts executing it, when an action fails during execution, the agent will resequence in order to find an alternative to achieving a goal. At the same time, the agent will experience fear that his goal might not now be achieved. If he does indeed achieve his goal he now experiences “Relief”. Alternatively, when a sequence is going on course and being followed as expected and then the agent fails to achieve his goal, he experiences “Disappointment”.

For example, if the sequence was not moving well and the teacher fails to publish his first paper, he would feel fear, but if his ultimate goal ‘promotion’ is achieved anyway although fear existed, he would then be ‘relieved’. On the other hand, if the sequence was moving well and the teacher was at all times hopeful that he is on the right track for his promotion if eventually he fails to achieve that goal although the sequence was moving right, he would then be disappointed.

$happy - for_i(j, V_j)$, **and** $pity_i(j, \neg V_j)$,

These emotions relate to a state of affairs that took place regardless of the actions that caused them. This is mapped to “Goals” in the decision-making methodology. These emotions also have a precondition that the emotion “Like” must also exist toward the other agent.

Occurrence:

Whenever a value of importance to the other agent is promoted and Like toward this agent is larger than zero, the emotion of Happy-for is generated and the intensity of this emotion depends on how well the other agent is liked(the intensity of Like) and how desirable is this goal(the importance of the goal).

Whenever a value of importance that is relevant to another agent is demoted and Like toward this agent is larger than zero, the emotion of Pity is generated. The intensity of this emotion depends on how well the other agent is liked(the intensity of Like) and how desirable is this goal(the importance of the goal).

For example, suppose Student A likes Student B, and Student B has done well on her exams, thus promoting a value important to her. Although this value is of no importance to Student A because there is a Like emotion between them, Student A will now be happy for Student B.

$gloating_i(j, \neg V_j)$, **and** $resentment_i(j, V_j)$,

These emotions are the opposites of (Happy-For/Pity) where a state of affair has occurred to someone else but on the other hand this someone is not liked. These emotions have a precondition of another emotion "Dislike" of being present.

Occurrence:
If a disliked agent fails to achieve a certain Value, the emotion "Gloating" occurs with an intensity that is equal to how disliked the agent is and the importance of this state of affairs.

If a disliked agent succeeds in achieving a certain Value, the emotion "Resentment" occurs with an intensity that is equal to how disliked the agent is and the importance of this state of affairs.

For example, the same student in the example above will gloat over the other student if the emotion dislike existed and the other student does badly in the exam.

$like_i(j)$, **and** $dislike_i(j)$,

In this study, we assume that the appeal of an agent increases and decreases along with the values being promoted and demoted by its own actions and joint actions, so that reflects the extent to which other agents cooperate with or frustrate it.

Occurrence:
If another agent performs an action that leads to a joint action which promotes some value, the intensity of “Like” increases. When the joint action demotes a value, “Dislike” increases. The intensity is dependent on the importance of the state of affairs and the expectedness.

For example, when the researcher succeeds in promoting his department’s publication rate, he will be liked by the HoD because it is an important value to him. If the researcher on the other hand fails to publish as originally expected by the HoD and therefore causes the value ‘Publication Rate’ to be demoted, he will be disliked by the HoD.

Modeling Emotions

Gratitude and gratification are central because they relate to all these elements (actions, values and goals), the difference being that gratitude is key for other directed emotions while gratification is key for self-directed emotions.

Now, Gratitude and Displeasure can be used as the basis to model other agents influences, but they cannot be used to model the agent’s own action. We use Gratification and Remorse the same way as Gratitude and Displeasure to model the agent’s own action. If a value is promoted by the agent itself, its gratification increases; if by another agent, its Gratitude increases and if by combination of both its Gratitude and Gratification both increase. But how can we know who to blame or who to congratulate? How can we know which agent in the joint action caused the effect? This can directly be applied to our model by looking into the joint action. If out agent was hoping for joint action (a,b) to promote V1, and what actually happens is (a,c) and V1 is not promoted, then the hopeful joint action did not occur because the other agent did not perform as originally expected (we expected b, but it did c). Thus, Displeasure is generated toward it. If the expected joint action took place and the value is indeed promoted, then both agents have reacted as expected and there is gratification (to the agent himself) and gratitude (toward the other agents in the joint action).

With this view, all other emotions can be derived from these two pairs (Gratitude and Displeasure, Gratification and Remorse). Pride, for example, is the sum of all gratification emotions caused by a certain action. Pity is the sum of all remorse feelings another agent has toward a certain value with a condition that the other agent is liked (intensity of Like emotion is over zero); then intensity can be seen as combining remorse of the other agent with my liking toward it.

The same method can be applied to all emotions. The 11 pairs of emotions have 11 positive emotions and 11 negative emotions. The positive emotions, which are a combination of gratitude/displeasure and gratification/remorse dependent on who caused the state of affairs, occur whenever a value is promoted. The cause of positive emotions is adherence to the intended sequence; negative emotions occur whenever a value is not promoted where expected or a value is demoted where not expected. If both agents do as expected in the joint action, then both gratitude and gratification occur, as do other related emotions. If both fail “we expect (a,b) and we get (c,d)” then remorse and displeasure occur together with the other negative emotions. If the intended joint action (a,b) does not occur and (a,c) happened instead but the value is also promoted, then gratification occurs toward the agent itself, but not gratitude toward the other one as it has made no contribution to its success.

5.2 The Role of Emotions

This section links emotions to the decision-making model by first identifying the structure to address the different aspects of emotions and showing how these can then be mapped into the decision-making methodology.

5.2.1 The Structure

The different steps forming and maintaining the emotions structure are the following:

Emotion Generation: Represents sets of rules that take inputs from the environment relative to the expectation of the agent in order to instantiate the emotion. This will produce emotion structures describing the emotion type, intensity, cause and direction.

Emotion Storage: Takes the emotion structure of various emotions and stores them with two considerations: intensity and the set of emotions influencing an agent. The intensity of an emotion decays over time according to a number of factors. In representing combinations of emotions, the intensity of actions and degree are considered.

Behavioural Features Mapping: The behavioural features are the intermediate step between the emotions computation and the decision-making process. These are user-created features that map different stored emotions to the values of the agent in the value ordering of the decision-making methodology. This may affect the value order, which is the mechanism by which the emotional state of the agent influences the choices it makes.

5.2.2 The Role in Decision Making

Decision making in the methodology explained in Chapter 3 critically depends on the ordering of the agent's values. Chapter 3 explained that although the plan of action/sequence of action is defined for future action, this might change according to how well or poorly the actual effects of the actions conform to the expected actions. Emotions influence the process in the same manner where the ordering of values is affected by the different emotions that the agent might have. This could cause resequencing whenever an ordering of values changes and will certainly impact on subsequent decisions. This happens through the influence of emotions on behavioural features. Emotion structures are first built identifying their type, intensities and direction. Those emotions are stored with a certain decay function and then combined and mapped to behavioural features.

5.3 Emotions and Rational Decision Making

A distinctive feature of our approach is that the decision to be taken is not based entirely on emotions, but on beliefs and rationality when what is rational is relative to the emotional state of the agent. Emotions then may or may not actually change the decision depending on the degree of their influence on the decision-making process. This section covers the main aspects of emotions and how they can be realised and linked to rational decision making.

5.3.1 Generation

The OCC model identifies twenty-two emotions as the basic emotions. These were formalised by Steunbrink et al. [119] in terms of the common notions of Plans, Goals and Actions. In the OCC model, each emotion is classified in terms of its generation into first the cause of this emotion as a description and then variables affecting the generation of this emotion.

An example of this is the emotion “Hope”, which is defined in OCC as “pleased about the prospect of a desirable goal”. This emotion can also be seen from an agent’s perspective as increasing the expectedness of a desirable joint action occurring. OCC defines the variables influencing intensity as:

1. The importance of the goal
2. The expectedness of this goal to happen

The emotion “Hope” occurs every time the expectedness of a desirable joint action increases; this normally occurs if the prerequisites to this state of affairs are brought about. Looking back at the example in the previous chapter, the publication goal is achieved if first a student writes a paper and then successfully registered for and attended a conference. The HoD will be more hopeful to achieve the goal of publication if the student succeeded in writing a decent paper as this increases the probability of the goal of publication being achieved.

Emotions are generated within our model whenever transitions occur in the AATS model. Every transition moves to a different state of affairs where values are either promoted or demoted and actions to reach subsequent states become possible or not possible. Goals and preconditions are achieved or not and the probabilities of goals are increased or decreased. Accordingly, these notions of Value, Goals and Probabilities reflect the changes that happen in the state of affairs and instantiate the emotions in this model.

We store the emotions generated in a structure with the following attributes:

Emotion type (Anger, Joy, ...etc)
Cause (Goal failing,Value demoted, ...etc)
Direction (is this a general emotion or directed toward an agent)
Intensity (a scale of how strong it is) See 5.3.2
Decay (an equation of how the emotion will decay) See 5.3.3

5.3.2 Intensity

The intensity of emotions depends on the importance of reaching a goal and the unexpectedness of an action. It is now necessary to quantify importance and unexpectedness.

Definition 5.2. ***Importance and Unexpectedness***

Let $Audience_i$ be a partial order given to a set V of n values agent $i = \langle v_n > v_{n-1} > ... > v_1 \rangle$. Let q be a state such that moving from q_0 to q promotes the set of values $V_v \subseteq V$.
$Rank_i(V)$ is the ranking of values $v \in V$ for $Audience_i$, i.e., the position at which the value v occurs in $Audience_i$
$V_n(q)$=$V : V$ is promoted at q
$Importance_i(q)$=$\sum_{w \in V_v(q)} Rank_i(w)$
Let VAF_0 be the VAF formed by agent i in q_0 and m the number of preferred extensions of VAF_0 with respect to $Audience_i$. The unexpectedness of the action of agent i, $unexpectedness_{q_0}(i)$ in q_0 will be $1 - 1 \div m$.

The calculation of emotion intensities occurs after the fifth step of the methodology of decision making presented earlier in Chapter 3, where an action is selected and executed, the results are known, and a decision is made as to whether the effects of emotions and sequencing are sufficient to change the value order; if not, resequencing might be necessary.

Suppose agent i in q_0 desires to reach q_d and, given the action chosen by agent j, α_j, q_a is the state actually reached. The intensity of any emotion is $importance_i(q_d) \times unexpectedness_j(q_a)$ given that the conditions of that emotion occurs.

Definition 5.3. ***Emotion Intensities***

Let i and $j \in Ag$ and values:value v_i and v_j. Suppose agent i seeks a state q_1 from the current state q_0 to promote value v_i ,i.e., $\delta(q_0, q_1, v_i) = +$. Let $J = \langle a_0, a_1, \ldots, a_n \rangle$ be an intended *joint action whose effect achieves this value, and $J' = \langle a_0, b_1, \ldots, b_n \rangle$ be the* actual *joint action performed. Agent j is* cooperative *w.r.t the joint action J if $b_j = a_j$; otherwise agent j is said to* frustrate *joint action J.*

This means that cooperation and frustration are linked to whether the agent performing the joint action performs as expected or not, if the value we hope to achieve is promoted by an unexpected joint action it would frustrate the agent. We can now model emotions felt by agents i and j towards other agents in terms of the outcome of a joint action J' relative to the value promoted. We have the following possibilities where performing J' results in the state q' and intensity refers to $importance_i(q_d) \times unexpectedness_i(q_a)$ and α is the action performed by i and α_s is the planned sequence of actions.

1. If $\delta(q_0, q', v) = +$ then $gratitude_i(j, b_j, v)$ for all cooperative agents and $gratification(b_j, v)$ if $b_i = a_i$.

2. If $\delta(q_0, q', v) = -$ then $displeasure(i, b_j, v)$ for all frustrating agents and $remorse(b_j, v)$ if $b_i \neq a_i$

The above four emotions capture any action performed in the environment either by the agent himself or another agent, discussed earlier in 5.1.3. Moreover, it captures emotions that are directed toward other agents and emotions toward the agent himself. The rest of the emotions can now be seen as a combination of these.

Calculation of Emotions Intensity

In the decision-making model presented in this thesis, an agent i decides on a sequence of actions to be performed $\underline{\alpha} = (\alpha_1, \alpha_2, ..\alpha_n)$ part of a sequence of joint actions he is hopeful will happen $\underline{J} = (J_1, J_2, ..J_n)$ and promote some values at each step $v_i \in V$. As the agent starts performing his sequence of actions, three possible reactions can occur:

Possible Reactions in the system:

1. *Joint Actions:* Values would be promoted or demoted as expected in the original sequence.

2. *Side Effects:* Other values important to the agent would be promoted/demoted unexpectedly.

3. *Sequence of Actions:* The desired sequence of joint actions would be affected or changed.

Depending on the importance and unexpectedness of each of the three factors above, emotions are generated and might eventually change the decision-making process. This is in line with the main definitions of occurrence of every emotion discussed in 5.1.3. Below is a more detailed discussion of the various emotions and how they are generated and calculated in our approach.

1. Let $G_i(j, v, t)$ be the intensity of gratitude felt by agent i toward agent j at time t and $D_i(j, v, t)$ be the intensity of displeasure felt by agent i toward agent j at t both with respect to v. Let $EM = (G_i(j, v, t)) - (D_i(j, v, t))$.

This captures the basic emotion at every transition.

2. Let $\sum_{J=2}^{J=n} G_i(J, v, t)$ be the intensity of gratitude felt by agent i toward the other agents at t in respect of v and $\sum_{j=2}^{j=n} D_1(j, v, t)$ be the intensity of displeasure felt by agent i towards the other agents at t with respect to v. Let $EM_J = (G_i(J, v, t)) - (D_i(J, v, t))$. This captures the overall Gratitude emotions i feel towards all agents in the system.

3. Let $\sum_{V=1}^{V=n} G_i(j, V, t)$ be the intensity of gratitude felt by agent i toward j at t in respect of all values and $\sum_{V=1}^{V=n} D_1(j, V, t)$ be the intensity of displeasure felt by agent i toward the agent j at t with respect to all values. Let $EM_V = (G_i(j, V, t)) - (D_i(j, V, t))$. This captures the overall Gratitude emotions i feels toward agent j in the system relating to all values.

Like and Dislike Looking back at the list of possible reactions of the system, like and dislike can occur as a direct result from *joint actions* and whatever value they promote or demote (item 1) or can also be caused by *side effects* whenever an unexpected value changes (item 2). Now:
$Like_i(j, v, t + 1) = Like_i(j, v, t + 1) + EM$.

Joy and Distress Looking back at the list of possible reactions of the system, Joy and Distress are caused from the results of the *joint action* performed and whether it promoted value as originally intended. Differently from Like and Dislike above side effects do not have an effect here, also, Joy and Distress relate to all agents in the environment.
$Joy_i(v, t + 1) = Joy_i(v, t) + EM_J$.

Pride and Shame Pride and Shame are emotions similar to Joy and Distress where they are caused from the results of *joint actions*, the only difference from it is that it relates to the success caused by the individual agent and how he made the difference in the promotion or demotion of values. Our methodology presented here considers actions only from a joint actions perspective where a group of agents participates and results are then driven. Therefore, results of Pride and Shame would be identical to Joy and Distress.

Admiration and Reproach Those are emotions directed towards a specific agent but capture all relevant values that agent promotes and is caused by the result of joint action and side effects.
$Admiration_i(j_i, t+1) = Admiration(j_i, t+1) + EM_V$.

Happy-For and Gloating In our methodology, this is redundant as the only difference between those emotions and Admiration/Reproach is that it relates to values that are not important to the agent i. Considerations of values relating to other agents have no effect on our current agent and so are out of the scope of this study.

Hope and Fear This is related the sequence of action being executed as planned with no changes. Let $\underline{J}_0$ be the the original sequence generated by the agent and $\underline{J}_t$ be the current sequence of joint actions. Where $J_t \neq J_0$,
$Hope_i(\underline{J}_0, t+1) = Hope_i(\underline{J}_0, t) + EM_J$ if $\underline{J}_{t+1} = \underline{J}t$ otherwise, $Fear_i(\underline{J}_0, t+1) = Fear_i(\underline{J}_0, t) + EM_J$.

Satisfaction and Disappointment Those emotions require a precondition which is the existence of the emotions Hope and they relate to the sequences of actions.
Where $J_{t+1} = J_n$ and $Hope_i(\underline{J}_0, t) > 0$

$Satisfaction_i(\underline{J}_{t+1}, t+1) = Satisfaction_i(\underline{J}_t, t) + EM_J$ if $\underline{J}_{t+1} = \underline{J}_0$ otherwise, $Disappointment_i(\underline{J}_{t+1}, t+1) = Disappointment_i(\underline{J}_t, t) + EM_J$.

Relief and Fears-Confirmed Those emotions require a precondition which is the existence of the emotions Fear and they relate to the sequences of actions.

Where $J_{t+1} = J_n$ and $Hope_i(\underline{J}_0, t) < 0$

$Relief_i(\underline{J}_{t+1}, t+1) = Relief_i(\underline{J}_t, t) + EM_J$ if $\underline{J}_{t+1} = \underline{J}_0$ otherwise, $Fears - Confirmed_i(\underline{J}_{t+1}, t+1) = Fears - Confirmed_i(\underline{J}_t, t) + EM_J$.

Gloating and Pity Those emotions relate to reaction towards other agents and their perspective value (which are of no importance to our agent). This will not have any effect on decision making and so will not be discussed further.

5.3.3 Decay

The concept of decay in emotions is very important. Emotions do not stay the same at all times. When we are angry, anger starts at a certain intensity and then dissolves over time. Nonetheless, time is not the only factor in decaying emotions. Sometimes other factors might play a role (for example, the reason that caused the emotion in the first place disappears). In this thesis, we do not focus on decay but it is still necessary to model it, otherwise, emotions would only increment unrealistically. We adopt the approach of Reilly [107] in modeling decay without change.

Emotions start with a certain intensity and then decay over time and they differ in how this happens. It is plausible to consider that if an agent is Joyful at achieving a certain Goal, this emotion will decrease and then dissolve overtime.

A simple decay system might just assume that intensity of every emotion will decay by 1 with every time tick (e.g., Like(j,t) = Like(j,t-1) - 1). Another system might have different rates of decay for different emotions.

In designing a decay function, three scenarios can be considered:

Stays The Same Where the emotions intensity will not decay with any factor and can only increase as events occur. This might be useful in applications where other external entities are involved in performing similar tasks on an irregular basis where if a contractor failed to deliver on his project the emotion of dislike might always stay the same ($Emotion_t = Emotion_{t-1}$).

Over Time Where a simple function will be attached to time and the emotion will decay consistently over time. This might be the most common way to implement decay ($Emotion_t = Emotion_{t-1} - Parameter$). This would be seen clearly in emotions like Hope where if the goal went a long time without being achieved Hope will decrease.

By Condition If the condition that caused the emotion was not true anymore the emotion will then decay. An example is when the HoD is fearful that the student will not publish his paper, this emotion will disappear as the student publishes his paper. ($Emotion_t = Emotion_{t-1}$ if Condition = True and $Emotion_t = 0$ if Condition = False).

Complex Structures Where decays can have rules causing them to decay over time and also have a condition that might alter the decay process. Moreover, the generation of new emotions might result in decaying others. For example, an increase in Like might lead to a decay in Resentment.

A suitable decay function with either time constraints or condition constraints will be combined with the emotional structure and stored in the emotional system. As our methodology of using AATS does not explicitly account for time as a basis, we consider each movement from one state to the other as a time tick to account for decay over time.

5.3.4 Combination and Storage

In a given event, several instances of the same emotion may take place. In our example of Chapter 4, a student might succeed in writing a paper and then attend a conference which would give rise to two joy inducing events, the overall effect on Joy can be dealt with in several ways.

Let us assume that an agent has three joy structures with intensities of x, y and z: how can we determine the intensity of joy the agent actually has. Three models are defined here and it is possible to use any of them [107]:

Winner takes all Joy in the above example will be $max(x, y, z)$; the advantage of this mechanism is that small emotions will not suddenly cause a big emotional reaction as they are added, the disadvantage though is that small emotions will now not be considered and have no impact on the decision-making process.

Additive Joy will be x+y+z, this has the advantage of full and fair considerations of all emotions, but can also cause havoc in the system as minor emotions may cause a strong reaction.

Logarithmic Joy will be (Log(x+y+z)). This is a middle way where all emotions are realised but at the same time the emotion will not have a big change as emotions are occurring.

The emotional considerations take place after step five of the methodology where an action has been performed and calculations are being made as to whether resequencing is necessary or not.

The next section will discuss how the different emotional structures can influence the decision-making process.

5.4 Influence on Decision Making

This section details how the emotional model now influences the decision-making methodology explained in Chapter 3. First, an intermediate step will be created between emotions and rational decision-making called “Behavioural Features”, which will assist the mapping to values in the value order. Some changes are then applied to the current decision-making methodology, in particular, Values in the value order and States in the AATS, to cater for the emotional influence. The concept of thresholding is then introduced to control the effect of emotions on decision making, and the final subsection shows the mechanism of applying the overall picture.

5.4.1 Behavioural Features

This is a level between emotions and values in the decision-making process. Introducing this level has the benefit of simplifying the changes in value orders.

Repression of Emotions It might not be desirable that all emotions would have an effect on the value order, e.g., an agent might be influenced by Anger but at the same time the emotion of Dislike might not make any difference with respect to particular decisions. This can get more complicated if it is relevant to condition in the environment: i.e., a teacher might be affected by Joy and Like throughout the year but when it comes to grading his student those emotions should be repressed.

Complexity in Emotion Mapping The value order may be influenced by more than one emotion. If the value V1 is influenced by emotions Em1, Em2 and Em3, we can either build a relationship such as V1= V1 + Em1 + Em2 + Em3 to indicate how the value would change with those emotions. The emotions involved however might have different influences on the values so that we should use weights to make the relationship V1 = V1 + (0.2 X Em1) + (0.5 X Em2) + (0.3 X Em3). Behavioural features will provide a more straightforward way to map these weighted emotions, where (B1= (0.2 X Em1) + (0.5 X Em2) + (0.3 X Em3)) will then make it easier to map to values.

Redirection of Emotions An agent might have another level of complexity where attitudes are built toward other agents that are unrelated to the emotion itself, e.g., the HoD might spend more budget on training (value of experience) when the publication rate is high).

The behavioural features differ with every person and perhaps are the elements that determine the person's attitude as they link values and goals to behaviours and potentially actions. Accordingly, the setup that takes place here would determine how well emotions would eventually influence the choice of action. And because this relationship is not unique and it is different among individuals it is then quantified in Reilly's model [107] and therefore ours. This step represents the potential mismatch between modeling human reasoning and the need for quantification in computer systems.

Behavioural features can also influence Aggressiveness toward another agent, which involves a mix of components such as shame and dislike. Another possible behavioural feature can be a good mood, a component of joy and pride.

This indirection can be eliminated and emotions can directly be mapped to values in the value order, but it would become extremely difficult to formulate rules as assigning emotions to values becomes more complex. Behavioural features then act as intermediate factors between the raw emotions and their effects on behaviour.

5.4.2 Weight Assignment

The mechanism by which emotions will influence decisions is by their impact on the value ordering of the emotional agent.

Values in the Value Order (VO) are ranked with either a ">" to indicate a value higher than another or "=" to indicate that both values have the same preference. An example of a VO is "$V1 > (V2 = V3)$" indicating the preference of V1 over both V2 and V3 and also that both V2 and V3 have equal preference for the agent which means that if the agent wants to choose between them, the agent would randomly choose either: he does not have any information that would lead to preferring one over the other from a rational perspective. In addressing emotions we add a subscript to values in the VO giving a weight to each value "$V1_3 > (V2_2 = V3_2)$". V2 and V1 have the same importance since they are equal: V1 has a greater importance. This change will now allow emotions to change the VO by altering these weights. If an emotion affected V1 and lowered its importance, it will be ranked below V2 and V3 resulting in "$(V2_2 = V3_2) > V1_1$".

In the emotional model we introduced two new concepts to evaluate the models from an emotional perspective (*Importance* and *Expectedness*; see Section 5.3.2). If there were different possible states that would occur moving from the initial state, the states that would be desirable (G) will have higher importance than the ones that do not. On the other hand, states that promote more important values (VS_+) will in consequence be more important, similarly, when values are demoted (VS_-). To quantify importance we can use a formula:

$Importance(q0) = (VS_+) + (G) - (VS_-)$

One aspect not considered here is that this calculation of importance limits the view to one move in the AATS model. A transition in AATS might not promote many values or achieve a goal, but it would allow for a future transition. We can call this a strategic move. To fully consider strategic moves, importance should be calculated on paths of actions (paths of actions were discussed in Chapter 3), and then the importance of a transition is defined, with the possible opportunities that it might offer in the future as well as what it will promote immediately. To simplify, we do not consider this aspect of opportunity further.

The last aspect is "Expectedness". This is indicated in the AATS model initially and assigned to every state accordingly as

If $Expectedness(q0)$ then $Unexpectedness(q0) = 1 - Expectedness(q0)$

A trivial way to look at expectedness, which can also be very useful when an agent is uncertain or has no knowledge about expectedness, is to assume equal expectedness of all possibilities. If we assume that there are five possible joint actions from q0 that would lead to anything between (q1 to q5), then $Expectedness(q1) = 1/5 = 0.20$ and the same value is given to the remaining states. This can be used in cases of uncertainty and then in emotion evaluation causing Expectedness to be neutral. If there is a formula expressing how likely the other agents are to perform the actions available to them, a better calculation of expectedness can be given. This would, however, depend on domain knowledge and models of other agents which are outside the scope of this thesis.

5.4.3 Thresholds

The decision that the agent is trying to make is one based on his own beliefs, as well as influenced by emotions. Some agents are more volatile than others. To reflect this, we identify the degree of influence that emotions have on the decision-making process. Threshold is the mechanism used to control the effect of emotions whenever it is desirable to vary the emotional influence and to ensure that things will not get out of control with high emotional influence.

Emotions vs. Beliefs :

It might be desirable at certain scenarios to have an agent with higher emotional influence while in other scenarios it is undesirable to have an agent easily influenced by emotions. For example, it might not be desirable to have a HoD in a university that is highly influenced by emotions in his budgeting decisions, where for more social situations, emotions can have greater effects.

Recall that values in the VO are assigned weights. The threshold is the difference between those values. In the example above, we considered the VO: $V1_3 > (V2_2 = V1_2) > V0_1$. An agent with this VO is said to be a volatile agent, where any small change in emotions would result in a value change (the difference is 1). A more stable agent might rank his values as such $V1_{30} > (V2_{20} = V1_{20}) > V0_{10}$, which now has a difference of 10 points, making it hard for any minor emotion to influence the decision-making process. An even stricter agent who would prefer to be minimally influenced by emotions would choose: $V1_{300} > (V2_{200} = V1_{200}) > V0_{100}$, which would mean that emotions play a very minimal part in his decisions.

Values vs. Other Values :

Another aspect of thresholding is differentiating among values when a particular group of values should be relatively stable with respect to any emotional influences. In the example, a fixed factor was added as the difference between all values to control the generic thresholding of emotions.

5.4.4 The Methodology of Decision Making

A relationship model is built to relate behavioural features to the VO, where the weights subscripted to the VO are affected by the intensities of the behavioural features, i.e., V1 = V1 - (0.5 × Aggressiveness(i)), which means that V1 will be affected by how aggressive the agent is toward i and the level of this effect is 50%.

Chapter 3 discussed a five-step methodology where a transition model is built and arguments of transition are constructed and then connected together in a VAF indicating a relationship of attack between them. Successful attacks are then identified using a value preference order, and the remaining surviving arguments are placed in a Preferred Extension (PE), which is sequenced in the fifth step, identifying differences in possible paths of actions to identify the best possible path that the agent should follow.

After the execution of the first action and identifying the resulting state, the intensity factor (See 5.3 Def 5.2) is calculated with the input of the initial state and the resulting state. Intensity has the components of Importance and Expectedness (Section 5.3.2).

Accordingly, different emotions are generated whenever the condition for instantiating this emotion occurs (Section 5.3.3). Existing emotions are decayed according to the rules defined (Section 5.3.4).

The behavioural features map is updated and the different values in the VO are affected according to the rule set earlier (Section 5.4.1).

If the weights of values mean that the ordering of values does not change, the agent will continue the execution of his initial plan. This means that emotions did not reach the threshold to actually affect the decision-making process (Sections 5.4.2 and 5.4.3).

If the weights of values changed the ordering of values, the agent will take the new VO and go through a resequencing process, where steps 4 and 5 will be revisited, and then step 5 will produce a new sequence of actions, which may result in a new plan of action.

Emotional Attitude

Behavioural features are elements like aggression, defensiveness and generosity, which are formulated as a combination of emotions. These behavioural features act as an intermediate step between emotions and decision making as they are mapped to values in the VO. The structuring of the behavioural feature partially determines the attitude of the agent. If dislike (Emotion) is mapped to aggression

(Behaviour), which is then directed towards a value in the decision system, the result will be an aggressive agent that can easily get angry; where if the mapping of dislike to behaviour is something like Aggression=(0.5) × Dislike, we would have a less aggressive agent.

The concept of thresholding then controls the effects of emotions on the decision-making process (Section 5.4.3). The combination of both thresholds and behavioural features can now create the "Attitude" of the agent.

In other words, the overall attitude of the agent in our system can differentiate between how much his decisions are affected by emotions (Thresholds) and what kind of mapping is done between what the agent feels (Emotions) and what he would actually do (Behavioural Features).

5.5 Summary

This chapter showed how emotions can be integrated into a rational decision-making process by extending the methodology to address emotional aspects.

The basis for our work is OCC [93] where we use the foundation and definitions of emotions. The work of Reilly [107] builds the mechanisms of how emotions can be translated and treated from philosophy and mainly the concepts of emotions: Generation, Decay and Behavioural Features were taken. The work of Steunbrink et al. [119] provides the basic formalisation of the OCC emotions (Definition 5.1) which then this work extends to cater for different usages of emotions.

Our interests are not in the study of emotions, but rather in determining the foundation that can be used to implement and accommodate emotional aspects in decision making.

Another necessary aspect which our mechanism offers is the combination of rationality and beliefs in decision making with the irrationality of emotions and provides a mechanism where this relationship can be controlled.

The main points this chapter addressed were:

1. Emotion Generation, where the rules of instantiation of emotions were linked to the elements of the AATS model of this methodology with the common aspects of Values and Goals. It was also shown how emotions

can be a component of other emotions when they are considered from a decision-making perspective. An emotion-generation model was built and formalised.

2. Intensity, which is the degree of strength each emotion has. Intensity is a component of both unexpectedness and importance of the state of affairs and this links to the state of the AATS, the expected outcome and the actual outcome. We can then see what different joint actions were performed and so identify cooperative and non-cooperative agents.

3. Decay, are rules that control how emotions will decrease and then diminish.

4. Influence on the Decision-Making process, emotions will be combined in the behavioural features' structure which will affect the value order as the intended sequence of actions is performed so that the values will be assigned weights that change according to the intensities of the emotions that are mapped to them. When the change is strong enough, the VO will change and the plan of action will change accordingly.

5. Thresholding, is a method of controlling the degree of effect emotions will have on the decision-making process. This is basically the difference in weights between the different values in the VO. Whenever the difference is larger, the decision will be more resistant to emotions, so that we can represent more or less volatile agents.

The next chapter extends the example shown in Chapter 4 by implementing the ideas of this chapter so as to include the effects of emotions on decision making.

Emotion	Description	Occurrence
$pride_i(\alpha_i)$, and $shame_i(\alpha_i)$	Relates to the agent's own actions	When an action α_i promotes/demotes a value
$admiration_i(j,\alpha_j)$, and $reproach_i(j,\alpha_j)$	Relate to different agent's action	When an action α_j (of another agent j) promotes/demotes a value important to agent i
$gratification_i(\alpha_i,V_i)$, and $remorse_i(\alpha_i,\neg V_i)$	Consider actions and events	The action the agent performs and the value promoted or demoted in doing so (α_i,V_i)
$gratitude_i(j,\alpha_j,V_i)$, and $displeasure_i(j,\alpha_j,\neg V_i)$	Consider actions and events toward other agents	If the action performed by another agent is combined with the agent's own action to promote a value.
$joy_i(V_i)$, and $distress_i(\neg V_i)$	Relate to an important state of affairs	Whenever a value that is of importance to the agent is promoted or demoted
$hope_i(C,V_i)$, and $fear_i(C,\neg V_i)$	Associated with the probabilities of success of a "Goal" increasing or decreasing	When parts of the plan succeed or fail
$satisfaction_i(\underline{j},V_i)$, and $fears-confirmed_i(\underline{j},\neg V_i)$	Relate to a state of affair happening with a precondition of (Hope or Fear)	Whenever a sequence where Hope exists is completed successfully
$relief_i(\underline{j},V_i)$, and $disappointment_i(\underline{j},\neg V_i)$	Those relate to undesirable events	A sequence is completed successfully where Fear existed, "Relief" and when a sequence fails where Hope existed, "Disappointment" will take place
$happy - for_i(j,V_j)$, and $pity_i(j,\neg V_j)$	These emotions relate to a state of affairs that took place regardless of the actions that caused them	A value of importance to the other agent change and Like toward this agent is larger than zero
$gloating_i(j,\neg V_j)$, and $resentment_i(j,V_j)$	A state of affair has occurred to someone else but on the other hand this someone is not liked	A disliked agent succeeds or fails
$like_i(j)$, and $dislike_i(j)$	The appeal of an agent increases and decreases along with the values being promoted and demoted	An agent performs an action that leads to a joint action which change some value.

Table 5.1: Summary of Emotion Types presented in 5.1.3

Chapter 6

Extension of the Experimental Study

This thesis aims at finding a methodology of decision making that can improve the level of Trust we can place in agents to eventually be able to delegate more critical tasks to them. Philosophy suggests that Trust is a combination of beliefs and emotions; thus, the thesis has been divided into two parts where trust is handled from the perspectives of rationality and beliefs of the agents about their environment and their capabilities (Part I) and which is extended to accommodate emotions (Part II).

Part I introduced the theoretical work (Chapter 3), where a methodology of decision making was constructed based on argumentation and empirically investigated through a case study (Chapter 4). Part II started with the theoretical work in Chapter 5, introducing the basic elements of emotions and how they can be integrated into the decision-making framework. This chapter now explores those elements of emotions by extending the experiment in Chapter 4 to account for them.

Chapter 5 introduced the OCC model of basic emotional types and how they relate to each other. This was then formalised and linked with issues of intensity of those emotions, decay and combination. Emotions were then linked to Value Orders (VO) in the decision-making methodology through behavioural features that govern and control emotions. The concept of thresholds was also introduced to control the extent of emotional influences on the decision-making process and particularly the level of influence that emotions might have in comparison with the beliefs of the agent.

This chapter extends the experiment presented in Chapter 4 to include those concepts of emotions.

Section 6.1 introduces the main aims of this experiment and presents an overview of the final results. Section 6.2 introduces the example study and gives a detailed explanation of the mechanisms with which the methodology has been implemented. Section 6.3 takes the implementation through a couple of scenarios and drives a discussion with observations on the methodology as it has been implemented. Section 6.4 studies the behaviour of the system on a larger scale, considering multiple scenarios and analysing them. Section 6.5 offers a high-level discussion of the benefits gained by the introduction of emotional aspects in the methodology. Section 6.6 concludes the work of this chapter.

6.1 Introduction

This chapter extends the experiment presented in Chapter 4 to account for emotional aspects. The experiment so far showed how a Head of a Department (HoD) in a university can make a decision on how he can appropriately spend the department's budget based on his beliefs of what would be best for the department and beliefs in the knowledge of the environment and how other agents can react according to any action. This chapter extends this idea and enables the HoD to factor emotions into his decisions.

It is important to show how the decision process can be balanced between beliefs and emotions, how this effect of emotions can be controlled, and when it is appropriate to allow for emotional influence and when a decision should not be influenced by them. The experiment needs to show also in practice the beneficial effect of emotional consideration in decision making and how the decisions of the HoD can actually be improved if he considers emotional aspects.

6.1.1 Main Aims of this Experiment

A methodology addressing the first four elements of Trust was built in Chapter 3 and experimented in Chapter 4. This chapter now extends the experiment to address the fifth element.

Below is a list of the main aims of this experiment and what this chapter hopes to achieve.

1. Give a better understanding of the work on emotions and the formalisations presented in Chapter 5 through an example relevant to the topic.

2. Show the applicability of this work by means of a substantial implementation.

3. Build different setups and scenarios where conditions and reactions from the environment differ, allowing an analysis of the degree of the effects of emotions in order to improve the decision-making process and analyse the effect of different setups on the process.

4. Show the different reactions the agent would have with different levels of emotional influence by changing the level of threshold for the emotions in the decision-making process.

5. Compare the results of having an agent with emotional influence with the results of having an agent using a pure rational decision-making process.

6. Link the final view to the basic elements of trust and show the full picture of the methodology in accordance with the elicitation of trust.

6.1.2 An Overview of the Empirical Observations

The case study this chapter offers starts by providing settings for all the emotional aspects of the agent, which in this case is a department head trying to send several students to conferences. Section 6.3 and 6.4 then present some scenarios using those settings.

One observation is that emotional aspects affect not only short-term actions but also future and even unrelated actions. Emotions of Dislike, for example, caused by an action at the beginning of a scenario can affect a decision unrelated to the first decision.

Emotions also have higher influence when things are not moving as expected (whether better or worse), and almost no influence when everything is moving as anticipated. Emotions thus re-enforce successful plans and trigger resequencing in response to failure or unexpected success.

Emotions sometimes influence the agent's preferences, but that does not mean they will also influence the actual decision. In many cases the plan remains best even when the preferences as expressed by its values order change.

6.2 The Case Study

As mentioned in Chapter 4, the experiment was implemented and tested in the .Net environment using Visual Basic. Screen shots of the actual experiment, along with a brief description of how it has been implemented are provided in Appendix A.

As this chapter extends the work done previously in Chapter 4, the next subsection briefly recalls the approach and example mentioned there.

The experiment started by identifying the problem scenario and the different settings of the problem and ended with the sets of actions the agent can take. Each choice represents a sequence of actions that can be executed one after the other.

The example study presented here focuses on emotional considerations. Calculations are made between the execution of actions in the sequence and will either affirm the initial sequence or propose a different one according to the progress that has been made.

6.2.1 A Review of the Initial Experiment

We recall that our agent is a head of an academic department (HoD) in a university, and he is faced with a dilemma of choosing what to do to appropriately allocate the department's budget where he needs to balance costs and consider departmental and individual interests. Our agent (HoD) has requests relating to travel funding to attend two specific conferences. He received requests from three different students and needs to decide which of them to send. Students 1 (S1) and 2 (S2) are new students. S1 is asking to go to a nearby conference, which will be cheaper; S2 is asking for a different conference, which will cost more, but has prepared a good paper that might help the department's publication rate. Student 3 (S3) is an experienced student asking to be sent to the local conference. Although she did not prepare a paper for submission, she is an excellent networker who is likely to impress other delegates and so promotes the reputation of the department. The conferences are on different topics, so S2's paper would not be suitable for the local conference, but both conferences are of equal standing. The budget only allows two students to be sent.

Table 6.1 introduces all related agents in the environment; Table 6.2 details all possible actions from those agents; Table 6.3 combines the actions in the previous table to produce all possible joint actions; Table 6.4 summarises the relative value to the HoD and how it can be promoted; Table 6.5 is the cost of doing any action.

Agent	Description
HoD	Head of the Department
S1	First student
S2	Second student
S3	Third student

Table 6.1: All agents in the environment

Agent	Action	Reference
HoD	Send Sn to a conference	$\alpha 1(n)$
HoD	Asks Sn to write a paper	$\alpha 2(n)$
Sn	Student n does well at the conference	β_n
Sn	Student n does poorly at the conference	$\neg\beta_n$
Sn	Student n writes a paper	γ_n
Sn	Student n does not write a paper	$\neg\gamma_n$

Table 6.2: All Possible Actions

Joint Ac	Combination	Description
$J1_n$	$\alpha 1(n), \beta_n$	HoD sends Sn to a conference and she attends
$J2_n$	$\alpha 1(n), \neg\beta_n$	HoD sends Sn to a conference and she does not attend
$J3_n$	$\alpha 2(n), \gamma_n$	HoD asks Sn to write a paper and she does
$J4_n$	$\alpha 2(n), \neg\gamma_n$	HoD asks Sn to write a paper and she does not

Table 6.3: All Possible Joint Actions

Ref	Value	Condition
H(n)	Happiness of Sn	When Sn is sent to a conference
E(n)	Experience of Sn	When Sn has attended a conference
P	Department's Publication Rate	When a student writes and attends
R	Reputation of the Department	When an experienced student writes and attends

Table 6.4: HoD values

In building the model of emotions and specifically in implementing it into a computer program, several assumptions were made in addition to the assumptions mentioned earlier in 4.2.2. Numeric values are assigned to different variables throughout this example. This assignment is done arbitrarily where the numbers are arbitrarily chosen with certain relative values to serve the purpose of evaluation. The numbers are not meant to express any "truth", however, they are

Item	Cost
Initial Budget	3
Writing a Paper	0
Sending S1 to a conference	1
Sending S2 to a conference	2
Sending S3 to a conference	1

Table 6.5: Cost Analysis

chosen so that their relationships are consistent with the characterisation of the agent.

6.2.2 Settings in the Example

Before going into the main model it is necessary to fix the initial settings of the example study. For simplicity, the experiment does not consider all emotions in the OCC emotional model. In particular emotions relating to goals of other agents are not considered.

Thresholds

As mentioned in Section 5.4.3, this is the point where emotional influence on decision making can be controlled. This is achieved through weights expressing the worth of the various values. Two components will be needed, one to express the volatility of the agent, and one to express the intrinsic importance of values.

Emotions vs. Beliefs: Our aim is that decisions should be based on beliefs and influenced by emotional factors, and so it is very important now to quantify this influence. In our case, a Department Head should not be over influenced by emotions in his decisions, particularly budget-related decisions. Since emotions impact on decisions by changing the value order, the *relative* difference in weights between values will determine whether an emotional experience of a particular intensity will alter this order. We, therefore, use a threshold TR to set the difference between values. The greater the TR, the more resistant the agent is to emotions. When TR is small, the difference between the values is small and the agent is volatile with respect to emotions so that a minor emotional influence can dramatically change the value order. If, however, TR is high, the agent is rather stable and has a high resistance to emotional influence, and intense emotions, or a series of events reinforcing a particular emotion, are required for there to be an effect of the value order. Values are ordered, and for n values, the first will receive the weight $((N-1) \times TR)$ the second $((N-2) \times TR)$ and so on down to the least important which receives no relative weight, $(0 \times TR)$.

Values vs. Other Values: The second aspect in setting preferences is designed to recognise the fact that some relative value orderings never change no matter what happens: a businessman would always give more importance to the profit margins of his company over the social activities the company organises no matter what emotions contributed to raising the importance of social activities. In our example, the HoD should always prefer the values of "Publication Rate" and "Department's Reputation" over a student's experience or happiness. This intrinsic worth of values is represented by a second threshold, TI. Values may be assigned to one of three categories: high worth (TI_H), medium worth (TI_M) and low worth (TI_L). Now while values within the same category change in accordance with TR, only a drastic emotional experience can move a value from one category to another. This is useful in cases where we want to have some values that should always be more important than others (in our example Esteem and Publication). It may still be possible for a value to move categories, but this would require a series of very intense emotions. It is exceptional and typically undesirable: if for example liking for a particular student promoted that student's happiness above the departmental values, we would have a situation where decisions were being taken for quite the wrong reasons.

Weights Assignment - Value Order

The initial VO is set to (Esteem > Publication > Experience > Happiness) or in an abbreviated form (Est > P > (E1=E2=E3) > (H1=H2=H3)). This represents that, although the HoD values the happiness and experience of his students, the reputation and publication of the department is considered more important.

We must now assign numerical values to represent the weights of these values and hence the volatility of this ordering with respect to emotions. Whenever the difference between value weights is high, the agent will be less responsive to emotions.

Following the discussion above we need to consider two aspects: the first is emotions vs. beliefs, where we need to generally define the level of influence that an emotion has on the decision-making process by determining the relative distance between the value weights. In this example, we say that the HoD should not to be too easily influenced by emotions; hence, we arbitrarily set TR to 20. The second aspect is the intrinsic importance of given values. Some values inside the VO need to have a higher resistance to emotional influence than others. In our example, it would make sense that the HoD will always value Esteem (Est) and Publication (P) above Happiness (H) and Experience (E) no matter what happens from an emotional perspective, and these two values should remain protected from any change; thus, Esteem and Publication receive TI_H and Happiness and Experience TI_M. As actual numbers we arbitrarily set TI_H to 1940 and TI_M to 10.

The weight of the values will now be as follows. The least important value is Happiness. This has only intrinsic worth and no relative worth. It will therefore be set to TI_M. Experience has the same intrinsic worth, but some relative worth: $TI_M + TR$. Publication has both high intrinsic worth and more relative worth: $TI_H + 2 \times TR$. Finally Esteem has high intrinsic worth and the most relative worth: $TI_H + 3 \times TR$.

The VO with weights using the numbers identified above for in this experiment:

$VO_0 = Esteem_{2000} > Publication_{1980} > (E1_{30} = E2_{30} = E3_{30}) > (H1_{10} = H2_{10} = H3_{10})$.

If an emotional influence occurs that increases the weight of the value of Experience, it would now be extremely hard for it to become worth more than Publication, as the difference is 1950. If, in contrast, some students fail to publish papers at a conference as expected, an experience of Distress would occur, and the importance of Publication will rise and become more valued than Esteem as the difference in weights between Esteem and Publication is not high enough to resist it. Initially, all the students are treated equally in terms of the importance of their happiness and experience, but if student 1 has made an unexpected achievement raising the emotions of Joy and Like for the HoD, the value of H1 becomes more important than H2 and H3. Note that particular values held with respect to different agents, such as happiness, will always be more responsive to emotional influence than distinct values.

Weight Assignment - Importance and Unexpectedness

The next assignment we need to make is to the different transitions in the AATS model. From Chapter 5 the intensity of any emotion is calculated by multiplying the importance of the state reached by the unexpectedness of this transition to occur. Here we need to assign weights to both importance and unexpectedness in the model.

Condition	Effect on Importance
A paper written	+15
Happiness Promoted	+17
Experience Promoted	+20
Publication Promoted	+28
Esteem Promoted	+35
Two Students Sent	+30

Table 6.6: Importance of each transition

Condition	Agent	Unexpectedness
A paper written	S1	0.2
	S2	0.2
	S3	0.6
Attended the Conference	S1	0.1
	S2	0.1
	S3	0.4

Table 6.7: Unexpectedness of each transition

Importance corresponds to the transitions that would be made between the different states (See Table 6.6). For example, if a student writes a paper and attends a conference and thus moves from q_x to q_y, he promotes Happiness and Publication, and so the Importance in this case is 17 + 28 = 45. If this was an inexperienced student attending a conference for the first time, q_y would also be recognising experience and importance is 45 + 20 = 65.

We note that the importance of every transition when considering emotions is not relative to values only (See table 6.6). We say that sometimes there are events that occur in the system that elicit an emotional response but would not necessarily be a value to the agent. In our example, we have two cases. A transition to a state that causes a paper to be written would elicit an emotional response and would have an importance on 15 (Table 6.6) even though it is not a relative value to our audience. The other case is the complete fulfillment of sending students to a conference (two students sent has an importance of 30), where this is not a value of importance or consideration to the audience but would, nevertheless, elicit an emotional response as it is of emotional importance to the audience. These transitions represent goals: the intermediate goal of a paper having been written has a small importance, and the overall goal, the highest importance.

Actual Publication and the Department's Esteem are the most significant factors, so a higher value is assigned to them. The final goal that the HoD is trying to achieve is sending two students to conferences. States that realise this fact have a high value in importance too.

Now, it is time to give some assumptions to unexpectedness. We apply the percentages in Table 6.7 to calculate the unexpectedness of each transition. We assume that both S1 and S2 have a probability of actually writing a paper when asked by 80% where the third student is less likely to do so and there is a 40% chance only that the student might actually write, the same goes for attendance as we are 90% sure that S1 or S2 would actually attend when asked whereas S3 is 60%.

The value of the unexpectedness of having written a paper by S1 can now be calculated from the table (1 - 0.8 = 0.2). S3 attending is (1 - 0.6 = 0.4).

The intensity to which emotions are experienced will eventually be captured from these percentages. It is quite normal and expected for new students such as S1 or S2 to actually register for and attend conferences when asked to do so. It is then rational to say that the agent (HoD) should not have intense emotions when it happens. On the other hand, if a new student (S1 or S2) does not attend the conference, this is a very unusual state of affairs, and it is rational to say that the agent (HoD) will now experience intense emotions.

6.2.3 Generation of Emotional Structures

As described earlier in Section 5.1.3 Gratitude and Displeasure are used as a basis to model all emotions because of the way they relate to Actions, Values and Goals. Whenever the transition adheres to the original intended plan the emotion of Gratitude will occur and will be the base to calculate the rest of positive emotions. The same holds for Displeasure, whenever the transition frustrates the original plan, Displeasure will occur and will be used as the basis for the rest of negative emotions.

Below is a discussion of how the rest of the emotions can be elicited in our case study.

$Joy(G)_{HoD}$, **and** $Distress(G)_{HoD}$,

Joy occurs whenever a value is promoted as intended and Distress occurs whenever a value was expected to be promoted and it does not. So, Joy occurs when $J1_n$ is expected and happens, and Distress occurs when $J1_n$ is expected and $J2_n$ happens.

$hope_{HoD}(\underline{j}, G)$, **and** $fear_{HoD}(\underline{j}, \neg G)$,

Those emotions occur depending on the probability of a goal being achieved increases (Hope), or decreases (Fear).

A student would usually be hopeful that he would be sent to a conference whenever he is asked to prepare a paper that is suited for a particular conference, but that is not enough. As the HoD might be asking the student to prepare a paper as a backup plan or for future needs. Thus, the student will need to also know that the HoD has a plan in place to actually send him to that conference. In short, the Hope emotion occurs whenever a student S has written a paper and there is a plan to send him to a conference.
As the student knows that the HoD has a limited budget that would allow him to only send few students to a conference, if a student was working on preparing a paper and he then knew while doing so that another student has been sent to a conference he would then be fearful that the budget might not allow him to be sent and he might even start questioning the HoD's intentions on whether or not he is sending him. In short, Fear for S occurs with respect to his own happiness whenever a student T is sent where there is a plan to ask student S to write a paper.

$satisfaction_{HoD}(\underline{j}, V_i)$, **and** $fears - confirmed_{HoD}(\underline{j}, V_i)$,

If Hope exists and the goal related to hope has actually been achieved, Satisfaction occurs. And if fear exists and the goal related to Fear was not achieved, Fears-Confirmed occurs. Hope and Fear are prerequisites to these emotions.

A student S will be satisfied if he was hoping for $J1_S$ to occur and then it actually does happen. So, the emotion Hope is a precondition to Satisfaction. In short, Satisfaction occurs when Joy(G) and Hope($\underline{j}$,G) Exists.
On the other hand if the student was fearful that something would not hap-

pen and it does not he would then have the emotion of Fears-Confirmed. So, the emotion Fear is a precondition to Fears-Confirmed. In short, Fears-Confirmed occurs when Distress(G) and Fear($\underline{j}$,G) exist.

$pride_S(\alpha_S)$, **and** $shame_S(\alpha_S)$,

Pride and Shame are emotions that relate to the agent itself only and its individual action. Whenever a value is promoted or enabled (writing a paper enables publication) the agent will experience Pride and whenever a value is demoted or disabled the agent experiences Shame.

Pride occurs when a student S has successfully realised $J1_S$ or $J3_S$ and $\alpha_S = \beta$ or γ. Shame occurs when a student S realises $J2_S$ or $J4_S$ and $\alpha_S = \neg\beta$ or $\neg\gamma$

$like_{HoD}(S)$, **and** $dislike_{HoD}(S)$,

Like and Dislike are directly related to the promotion or demotion of values.

Like occurs when a student S promotes the value of importance to the HoD (Publication, Esteem, Experience).
Dislike occurs when a student S demotes the value of importance to the HoD (Publication, Esteem, Experience).

$admiration_{HoD}(S, \alpha_S)$, **and** $reproach_{HoD}(S, \alpha_S)$,

Those emotions relate to other agents and their own success or failure. When a Liked agent succeeds in achieving a goal (experiencing Pride as a result) Admiration occurs and when this liked agent fails (experiencing Shame as a result) Reproach occurs.

Admiration occurs when $Like_{HoD}(S)$ and $Pride_S(\alpha_S)$. Reproach occurs when $Like_{HoD}(S)$ and $Shame_S(\alpha_S)$

6.2.4 Decay and Combinations

Decay

Table 6.8 shows the settings used in this experiment to represent Decay. Those are arbitrarily assumptions and are in reference to the example Reilly gave in [107]. The concept of Decay has been described in 5.3.4. Future possible extensions to this work might consider alternative combinations of different decays and analyse the behaviour of the system accordingly.

Emotion	Decay
Joy(G)	50% with each transition
Distress(G)	50% with each transition
Hope($\underline{j}$,G)	Decays to 0 when a student is sent to a conference and attends
Fear($\underline{j}$,G)	Decays to 0 when the other student S writes and attends a conference
Satisfaction($\underline{j}$,G)	Decays to 0 when Joy(G) becomes 0
Fears-Confirmed($\underline{j}$,G)	Decays to 0 when Distress(G) becomes 0
Pride(α)	50% with each transition
Shame(α)	50% with each transition
Like(S)	30% with every transition
Dislike(S)	30% with every transition
Admiration(S,α)	50% when Pride(α) becomes 0
Reproach(S,α)	50% when Shame(α) becomes 0

Table 6.8: Decay functions of the Example study

Combination

Three different methods were explained in 5.3.5. For simplicity, this experiment uses simple addition to combine similar emotions.

As a simple example, if we assume that the initial value of Joy(G) is 5, and then an action occurs which causes another experience of Joy(G) with the intensity 7, the overall emotion Joy(G) is $(\frac{5}{2}) + 7 = 12$. We have multiplied 5 by 50% to account for decay in the transition.

6.2.5 Settings of Behavioural Features

At this stage the connection between the emotional state of the agent (HoD) and the decision-making methodology is made by setting the different behavioural features and linking them to the values in the VO.

Behavioural Features

Mood = (Joy(G) - Distress(G) + Hope(G)) / 2

Friendliness(S) = (Mood + (Like(S) - Dislike(S)))/2 + (Pride(S) - Shame(S))

Cheerful = (Mood + Satisfaction($\underline{j}$,G))/2

Defensive(S) = (Dislike(S) + Reproach(S))/2

Disappointed(G) = (Fears-Confirmed($\underline{j}$,G) + Fear) /2 + Distress

Links to the Values

$H_S = H_S$ + Mood + Friendliness(S)

$E_S = E_S$ - Defensive(S) + Cheerful + (Friendliness(S)/3)

P = P + Disappointed(P)

Est = Est + Disappointed(Est) + Mood

6.3 Scenarios

The previous sections explained the settings used within the case study. This section is now a walk through of actual runs of the system. At the beginning, an assumption that the HoD has a neutral emotional state is made; thus, Tables 6.9 and 6.10 show the initial states of the HoD at the beginning of the experiment.

As shown in Chapters 3 and 4, the methodology of action selection brings about several plans (sequences) of action to be performed by the agent (HoD) and presents them as possibilities that are subjected to a Safety, Threat and Opportunity evaluations (See Section 3.2.5) to determine which sequence of actions can best be chosen. A sequence is the list of ordered actions that the HoD executes one by one. If a path has five actions, the transitions that are made are called stages. So, executing the first action in the sequence takes the agent from stage 1 to stage 2. A path with five actions will have five different stages. The process ends when the budget becomes insufficient to send any additional students to conferences.

In this section a path is chosen and the agent starts executing the actions in this path where the emotions vary from one stage to the other upon the results that are achieved from executing the action. It will be explained how these emotional variations will affect the execution of this plan of actions.

The following scenarios reflect the same settings used in Chapter 4: the HoD has a budget of three and 3 students, where the first and the second are inexperienced students, and the third student is an experienced student. The second

Joy=0	Distress=0
Hope=0	Fear=0
Satisfaction=0	Fears-Confirmed=0
Pride=0	Shame=0
Like=0	Dislike =0
Admiration=0	Reproach=0

Table 6.9: The Emotional State at the beginning

Mood=0	Friendliness=0
Cheerful=0	Defensive=0
Disappointed=0	

Table 6.10: Behavioural State at the beginning

student has prepared a paper for submission but is asking to go to an expensive conference whereas the other students are asking to go to a less expensive conference. The initial state q0 is then 3-000-010-001.

The degree to which emotions influence the decision-making process differs according to the behaviours of agents in the environment. Two extreme scenarios are presented in the next subsections where first cooperative agents will be considered in which students are mainly following the plans of the HoD. The next subsection has a similar setup, but uses non-cooperative agents where the students are consistently behaving contrary to the HoD's expectations. During the discussion of these scenarios other variations will be presented such as changing the threshold or VO or the initial state.

Note that in the screen shots of the implementation the joint actions are referenced differently from Table 6.3. (J0,J1 and J2) are sending students 1, 2 and 3 to conferences. (J4, J5 and J6) are asking students 1,2 and 3 to write a paper.

6.3.1 Scenario No.1 Cooperative Agents

The following paragraphs discuss the findings in the first scenario. The numerical results of each stage are shown in Table 6.11.

The initial Value Order is: $VO_0 = Esteem_{2000} > Publication_{1980} > (E1_{30} = E2_{30} = E3_{30}) > (H1_{10} = H2_{10} = H3_{10})$

When the five-step methodology is executed, the HoD will have six possible sequences to choose from; the optimal choice given his VO is: " Seq = $J3_3$ then

Emotions	Stage 0	Stage 1	Stage 2	Stage 3
Joy	0	0	32	20.75
Hope	0	9	0	0
Satisfaction	0	0	32	32
Pride(S2)	0	0	0	4.75
Pride(S3)	0	9	36.5	18.25
Like(S2)	0	0	0	6.65
Like(S3)	0	0	32	22.4
Admiration(S2)	0	0	0	9.5
Admiration(S3)	0	9	73	70.75
Behavioural Features				
Mood	0	9	41	50.5
Friendliness(S1)	0	4.5	20.5	25.25
Friendliness(S2)	0	4.5	20.5	39.5
Friendliness(S3)	0	13.5	73	77.75
Cheerful	0	4.5	36.5	46
Value Order				
Est	2000	2009	2050	2100.5
P	1980	1980	1980	1980
E1	30	33.75	58	90.25
E2	30	33.75	58	95
E3	30	36.75	79	128.25
H1	10	23.5	85	160.75
H2	10	23.5	85	175
H3	10	32.5	146.5	274.75

Table 6.11: Emotions and Behavioural Features of all stages in Scenario 1

$J1_3$ then $J1_2$ ". That is, he would first ask the third student to write a paper ($J3_3$) and then send him to a conference ($J1_3$) and then send the second student (who has a paper written) to his conference ($J1_2$).

Stage 1

The agreed Sequence as discussed is (Seq = $J3_3 \rightarrow J1_3 \rightarrow J1_2$) (Seq = Ask S3 to Write → Send S3 → Send S2)

As we are presenting a cooperative scenario, the student successfully writes the paper. Intensity is calculated according to Tables 6.6 and 6.7.

The emotional state is now different because the HoD will now have emotions of Admiration toward S3 and also Hope that the Publication value might hopefully be achieved in the future as S3 has successfully written a paper.

Accordingly, the behavioural state also changes where the general Mood and Cheerfulness of the agent will get better and his friendliness toward his students will increase and will increase especially for S3 as he is the one who caused the good state of affairs.

Now, as mentioned in Section 6.2.6, this will affect the VO as follows: the calculations are shown in Table 6.11:

$VO_1 = Esteem_{2009} > Publication_{1980} > E3_{36.75} > (E1_{33.75} = E2_{33.75}) > H3_{32.5} > (H1_{23.5} = H2_{23.5})$

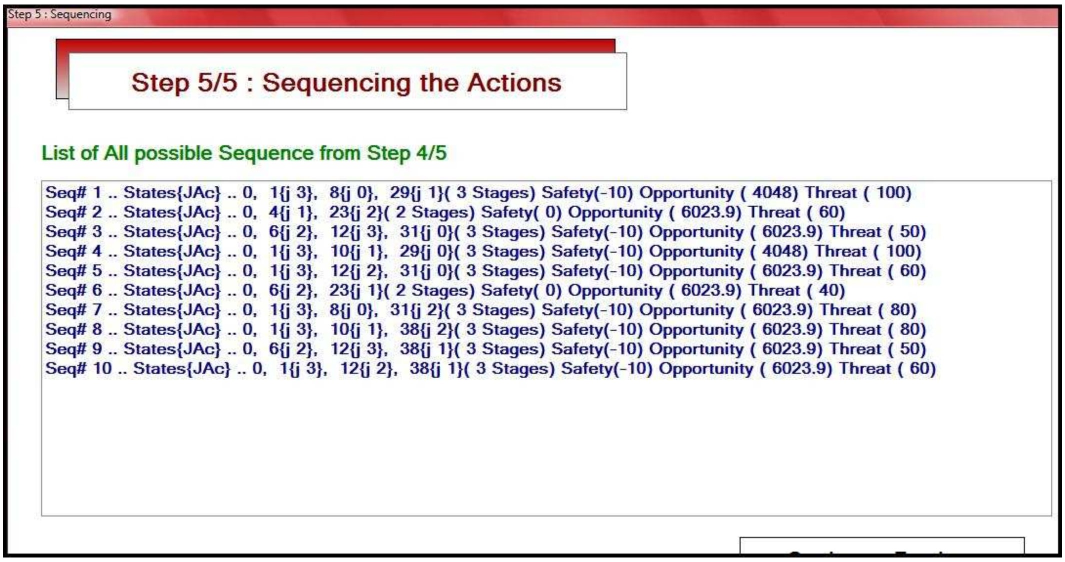
We note the following: although generally the happiness of students has increased in importance and H3 has increased the most, it did not increase enough to be more important than the value of experience. If the threshold had been set to 10, the happiness of the 3rd student would have became more important than the experience value. There is an issue here, in that it might be thought wrongly that a department head will give more importance to a certain student's happiness over the experience of all other students just because he succeeded in writing a paper.

The Happiness/Experience of the third student is of more value now compared to other students. One of the benefits of emotional considerations is to address uncertainties and equalities. (H1=H2=H3) in the initial VO had one of two interpretations: it either meant that the HoD genuinely did not have any preference of one student's happiness over the other (Equality) or the information is basically absent in the system (Uncertainty). Without emotional considerations, such situations would have been dealt with randomly without giving any preference to any particular choice/value, while now the department head has a good enough reason to prefer the Happiness and Experience of S3 over the other two since there is no other information that would help make a choice. This will be further discussed in 6.4.4.

Stage 2

With the change that took place in the VO, the HoD will go through a resequencing process and will be presented with different options one of which is to continue with the original plan (Seq = $J3_3 \rightarrow J1_3 \rightarrow J1_2$) (Seq = Ask S3 to Write → Send S3 → Send S2). Calculations of Safety, Opportunity and Threat

(See figure 6.1) indicate that the original plan remains the best choice which is as it should be as events are occurring as expected when the original plan was made. In general resequencing should not occur if the plan is on track, rather emotional changes should tend to reinforce the plan.


Step 5 : Sequencing
Step 5/5 : Sequencing the Actions
List of All possible Sequence from Step 4/5

```
Seq# 1 .. States{JAc} .. 0,  1{j 3},  8{j 0},  29{j 1}( 3 Stages) Safety(-10) Opportunity ( 4048) Threat ( 100)
Seq# 2 .. States{JAc} .. 0,  4{j 1},  23{j 2}( 2 Stages) Safety( 0) Opportunity ( 6023.9) Threat ( 60)
Seq# 3 .. States{JAc} .. 0,  6{j 2},  12{j 3},  31{j 0}( 3 Stages) Safety(-10) Opportunity ( 6023.9) Threat ( 50)
Seq# 4 .. States{JAc} .. 0,  1{j 3},  10{j 1},  29{j 0}( 3 Stages) Safety(-10) Opportunity ( 4048) Threat ( 100)
Seq# 5 .. States{JAc} .. 0,  1{j 3},  12{j 2},  31{j 0}( 3 Stages) Safety(-10) Opportunity ( 6023.9) Threat ( 60)
Seq# 6 .. States{JAc} .. 0,  6{j 2},  23{j 1}( 2 Stages) Safety( 0) Opportunity ( 6023.9) Threat ( 40)
Seq# 7 .. States{JAc} .. 0,  1{j 3},  8{j 0},  31{j 2}( 3 Stages) Safety(-10) Opportunity ( 6023.9) Threat ( 80)
Seq# 8 .. States{JAc} .. 0,  1{j 3},  10{j 1},  38{j 2}( 3 Stages) Safety(-10) Opportunity ( 6023.9) Threat ( 80)
Seq# 9 .. States{JAc} .. 0,  6{j 2},  12{j 3},  38{j 1}( 3 Stages) Safety(-10) Opportunity ( 6023.9) Threat ( 50)
Seq# 10 .. States{JAc} .. 0,  1{j 3},  12{j 2},  38{j 1}( 3 Stages) Safety(-10) Opportunity ( 6023.9) Threat ( 60)
```

Figure 6.1: Scenario 1 the choices at stage 2

The action $J1_3$ succeeds, and the third student has now attended the conference. First, the existing emotions decay according to Table 6.8. Next, the new emotions are calculated. As the third student is an experienced student who has also written a paper, values of Publication and Esteem are promoted and the new emotional structure has elements of Joy for Publication, Joy(P), and Joy for Esteem, Joy(Est), S3 will have Pride and the HoD will like S3 more and admire him.

The behavioural features now include an increase in the general mood of the HoD and Friendliness (specially toward S3).

The Value Order is now:
$VO_2 = Esteem_{2050} > Publication_{1980} > H3_{146.5} > (H1_{85} = H2_{85}) > E3_{79} > (E1_{58} = E2_{58})$

A dramatic (and perhaps undesirable) change now occurs in the VO where the HoD will now prefer the Happiness of his students over their Experience. This was mainly because S3 has promoted enough values to make the experience of the students less important than their Happiness. In what follows this means that actions that do not promote Experience but Happiness will now be preferred.

In the previous stage, if two students had the chance to be sent to conferences where one of them had attended previous conferences, the attack that represents the importance of Experience will not succeed. Nevertheless, this effect will not last for long, as the reasons that caused happiness to increase will now decay and once a state of affairs occurs that raises the importance of experience they will become more important again in the VO. Psychologically the situation represents a person being more amenable to requests when his mood is good.

Stage 3

As the VO has changed in the previous step, resequencing will occur and the original plan (Seq = Ask S3 to Write → Send S3 → Send S2) is then evaluated to see whether it still represents the best option for the HoD.

Where S2 has a paper written, this means that sending S2 will promote Publication and thus the original plan will still be followed as Publication is still more important than Happiness and Experience despite the changes that occurred in the value order.

The final Value Order of this scenario is:
$VO_3 = Esteem_{2100.5} > Publication_{1980} > H3_{274.75} > H2_{175} > H1_{160.75} > E3_{128.25} > E2_{95} > E1_{90.25}$

If the scenario took place with a different sequence where the first student would be asked to write a paper and then sent instead of sending S2 (which would achieve the Publication goal as well) the plan would now change as not sending S2 would demote Happiness which has become a more important value than the Experience of S1.

Running the same scenario with a slightly higher threshold will keep the VO as is in the previous step and Experience would still be more important than Happiness, and no resequencing will be triggered.

A general observation in this scenario is that although emotions have played a role in changing the VO of the agent, they did not influence the actual decision-making plan. This is an expected behaviour as other agents were acting cooperatively within the scenario. Whenever an agent has a plan and the plan while in execution was delivering the expected results, emotions should have a minimal influence on changing it.

6.3.2 Scenario No.2 Non-Cooperative Agents

The following paragraphs discuss the findings in the scenario where the numerical results are provided in Table 6.12.

The HoD has the following initial VO ($VO_0 = \mathit{Est} > \mathit{Publication} > \mathit{Experience} > \mathit{Happiness}$). The Initial Value Order is:

$VO_0 = \mathit{Esteem}_{2000} > \mathit{Publication}_{1980} > (E1_{30} = E2_{30} = E3_{30}) > (H1_{10} = H2_{10} = H3_{10})$

As the five-step methodology is executed, the HoD is presented with six possible sequences to choose from out of which one presents an optimal choice from a sequencing perspective: " Seq = $J3_3$ then $J1_3$ then $J1_2$" the HoD may first ask the third student to write a paper ($J3_3$) and then send him to a conference ($J1_3$) and then send the second student (who has a paper written) to his conference ($J1_2$).

Stage 1

Say student S3 fails to write the paper. This failure gives rise to the emotions of Dislike and Reproach towards S3: S3 also feels ashamed of his failure. This is then incorporated into the behavioural features, where the HoD is now less friendly and defensive to S3.

As mentioned in Section 6.2.6, this will affect the VO in the following way:

$VO_1 = \mathit{Esteem}_{2000} > \mathit{Publication}_{1980} > (E1_{30} = E2_{30}) > E3_{21} > (H1_{10} = H2_{10}) > H3_1$

Whenever presented with an opportunity to choose among the students, the HoD now always prefers S1 and S2 over S3. However, if S3 can promote Publication or Esteem, he is chosen, as those values are still preferred.

Stage 2

Things have now moved in a different direction from what was initially expected. With the change that took place in the VO, the HoD now goes through a resequencing process and is presented with different options, of which the best is ($J1_2$ - $J3_1$ - $J1_1$) that is a change to the original plan.

Unlike the previous scenario, the resequencing process now suggests a different plan, where S3 is no longer considered and the HoD will now consider S1 instead and give a plan to ask him to write a paper and then send him to a conference hoping that he would achieve the goal of publication.

The action $J1_2$ succeeds and the 2nd student has now attended the conference. First, existing emotions will decay (Table 6.8). Emotions of Joy and Admiration will be experienced and S2 will experience Pride. At the same time, the emotion of Fear will be generated as the HoD is now fearful that by sending the 2nd student, S1 will not agree to write a paper (the next action) as the latter's hopes of being sent are now less.

The Value Order is now:
$VO_2 = Esteem_{2009.75} > Publication_{1983.25} > E2_{35} > E1_{32} > H2_{29.5} > H1_{19.75} > E3_{16} > H3_{5.65}$

The success Student 2 had in attending the conference raised the values of Happiness to a level where they became even more important that the Experience of the 3rd student (who failed to write in the previous stage).

Stage 3

With the change that took place in the VO, the HoD will go through a resequencing process and will be presented with different options out of which the best is still ($J1_2$ - $J3_1$ - $J1_1$) which is the same as the current plan.

The first student now fails to write a paper giving rise to different emotions and behavioural features (see Stage-3 in Table 6.12) and causes the VO to change to:

$VO_3 = Esteem_{2016.25} > Publication_{1986.5} > H2_{39.9} > E2_{39} > E1_{16} > E3_{13} > H3_{7.5} > H1_{6.6}$

Stage 4

The HoD now has a dilemma: whether to choose S1 or S3 to attend the conference where both had failed to write a paper. S1 will promote Experience as he is a new student where S3 will promote Happiness which is more important than S1's Happiness. Although the HoD started the process treating everyone equally and rating everyone's happiness equally as a result of what has happened in Stage 3 (the failure of S1), S1's happiness is no longer a priority to the HoD.

As S3 has failed to write a paper earlier his Happiness is still much lower than the value of S1's Experience. Hence, the HoD will choose to send S1 to the conference even though his Happiness is less important than S3's Happiness.
If this has occurred before S3 was asked to write the paper, the HoD would have chosen S3 as his Happiness would have been more important than S1's Experience.

6.3.3 Non-Cooperative with Higher Emotional Influence

This is another scenario where the volatility of emotional influences is high or in other words the difference in weights between the different values are set low. Figure 6.2 shows the tracker screen of the experiment from the program.

As S2 has written a paper sending him would promote Publication. And as S3 has Experience sending him would promote the department's Esteem.

The HoD starts with the preference of the department's Esteem over Publication and over the student's Experience and Happiness, as in the previous scenarios. The ideal sequence chosen at stage 5 was (Send 2 - Ask 3 - Send 3) (Figure 6.3). This sequence would, if successful, promote the department's Esteem and then Publication when S2 is sent: these are the most important values to the HoD.

Stage 1

When the HoD asked the second student to attend the conference the student failed to do so. This caused emotions of Distress and Fear (because Publication was not promoted as expected) and Dislike toward the second student. The student himself also had the emotion of Shame as he failed to meet the expectations.

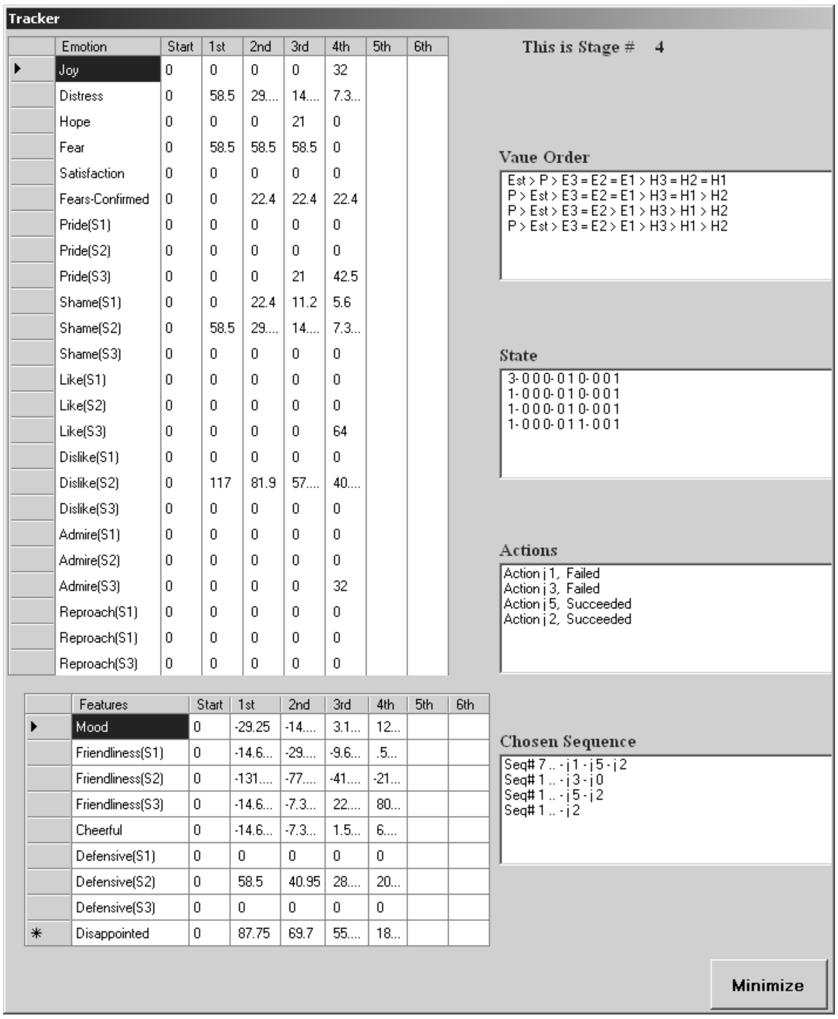
The values of those emotions are considerably higher for two reasons: the unexpectedness of S2 attending was set very low as it was thought very unlikely to happen; and the importance of the next state was set high as the promotion of Publication was expected. Another observation would be the dramatic change in VO. Whereas the HoD failed to promote Publication his Fear caused the change that Publication became more important than the department's Esteem.

Emotions	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
Joy	0	0	6.5	3.25	20.425
Fear	0	0	6.5	6.5	6.5
Pride(S2)	0	0	6.5	3.25	1.625
Pride(S3)	0	0	0	0	18.8
Shame(S1)	0	0	0	12	6
Shame(S3)	0	6	3	1.5	0.75
Like(S2)	0	0	6.5	4.55	3.185
Like(S3)	0	0	0	0	18.8
Dislike(S1)	0	0	0	12	8.4
Dislike(S3)	0	6	4.2	2.94	2.058
Admiration(S2)	0	0	13	18.2	13.91
Admiration(S3)	0	0	0	0	37.6
Reproach (S1)	0	0	0	12	12
Reproach (S3)	0	6	6	5.25	22.175
Behaviours					
Mood	0	0	6.5	3.25	20.4
Friendliness(S1)	0	0	3.25	-16.375	0.01
Friendliness(S2)	0	0	13	7.15	13.43
Friendliness(S3)	0	-9	-1.85	-1.35	36
Cheerful	0	0	3.25	1.625	10.2
Defensive(S1)	0	0	0	12	10.2
Defensive(S3)	0	6	5.1	4.095	12
Disappointed	0	0	3.25	3.25	3.25
Value Order					
Est	2000	2000	2009.75	2016.25	2039
P	1980	1980	1983.25	1986.5	1989
E1	30	30	32	16	10.97
E2	30	30	35	39	48.73
E3	30	21	16	13	18.37
H1	10	10	19.75	6.6	27
H2	10	10	29.5	39.9	73.75
H3	10	1	5.65	7.555	64.6

Table 6.12: Emotions and Behavioural Features of all stages in Scenario 2

Stage 2

The failure to promote Publication in the previous step has caused the HoD to change priorities and place Publication at higher importance in the VO. This has also affected the sequence of actions where the HoD was initially planning to promote Esteem at stage 2 he now has changed the sequence to (Ask Student 1 - Send Student 1) in the hope that the first student would promote the department's Publication.



	Emotion	Start	1st	2nd	3rd	4th	5th	6th
▶	Joy	0	0	0	0	32		
	Distress	0	58.5	29....	14....	7.3...		
	Hope	0	0	0	21	0		
	Fear	0	58.5	58.5	58.5	0		
	Satisfaction	0	0	0	0	0		
	Fears-Confirmed	0	0	22.4	22.4	22.4		
	Pride(S1)	0	0	0	0	0		
	Pride(S2)	0	0	0	0	0		
	Pride(S3)	0	0	0	21	42.5		
	Shame(S1)	0	0	22.4	11.2	5.6		
	Shame(S2)	0	58.5	29....	14....	7.3...		
	Shame(S3)	0	0	0	0	0		
	Like(S1)	0	0	0	0	0		
	Like(S2)	0	0	0	0	0		
	Like(S3)	0	0	0	0	64		
	Dislike(S1)	0	0	0	0	0		
	Dislike(S2)	0	117	81.9	57....	40....		
	Dislike(S3)	0	0	0	0	0		
	Admire(S1)	0	0	0	0	0		
	Admire(S2)	0	0	0	0	0		
	Admire(S3)	0	0	0	0	32		
	Reproach(S1)	0	0	0	0	0		
	Reproach(S1)	0	0	0	0	0		
	Reproach(S3)	0	0	0	0	0		

	Features	Start	1st	2nd	3rd	4th	5th	6th
▶	Mood	0	-29.25	-14....	3.1...	12...		
	Friendliness(S1)	0	-14.6...	-29....	-9.6...	.5...		
	Friendliness(S2)	0	-131....	-77....	-41....	-21...		
	Friendliness(S3)	0	-14.6...	-7.3...	22....	80...		
	Cheerful	0	-14.6...	-7.3...	1.5...	6....		
	Defensive(S1)	0	0	0	0	0		
	Defensive(S2)	0	58.5	40.95	28....	20...		
	Defensive(S3)	0	0	0	0	0		
✳	Disappointed	0	87.75	69.7	55....	18...		

This is Stage # 4

Vaue Order

```
Est > P > E3 = E2 = E1 > H3 = H2 = H1
P > Est > E3 = E2 = E1 > H3 = H1 > H2
P > Est > E3 = E2 > E1 > H3 > H1 > H2
P > Est > E3 = E2 > E1 > H3 > H1 > H2
```

State

```
3-000-010-001
1-000-010-001
1-000-010-001
1-000-011-001
```

Actions

```
Action j 1,  Failed
Action j 3,  Failed
Action j 5,  Succeeded
Action j 2,  Succeeded
```

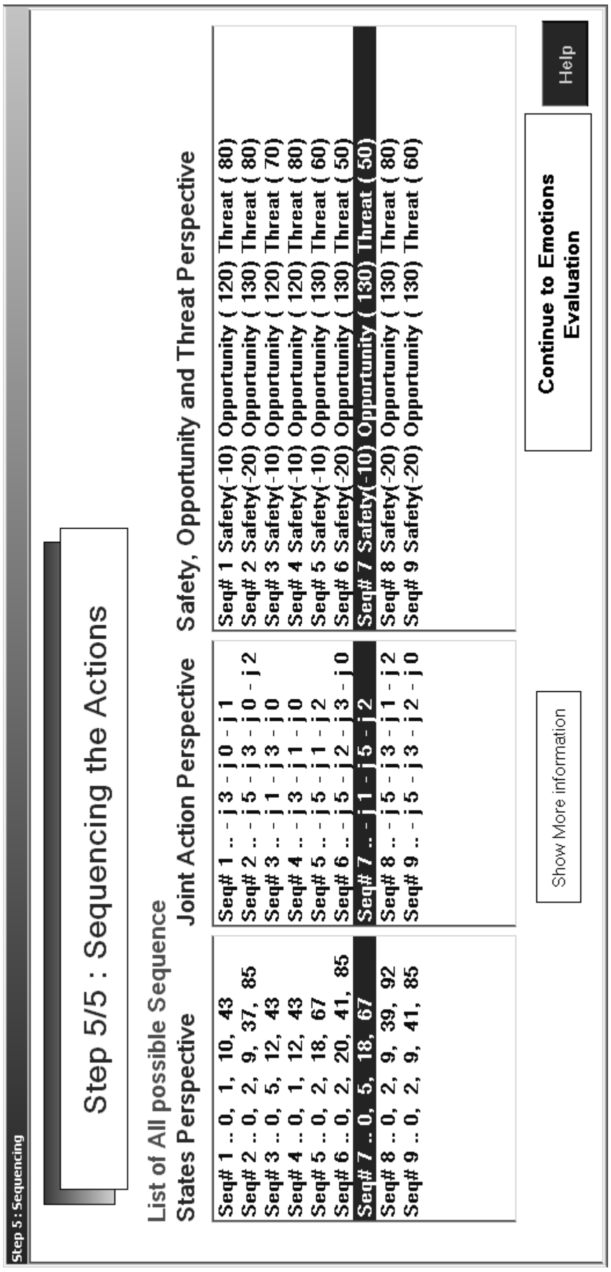
Chosen Sequence

```
Seq# 7 .. - j 1 - j 5 - j 2
Seq# 1 .. - j 3 - j 0
Seq# 1 .. - j 5 - j 2
Seq# 1 .. - j 2
```

Figure 6.2: The Emotional state throughout Scenario 3

We note that where the HoD had initially wanted to consider the third student and ignore the first student, he has now changed priorities with the increased importance placed on Publication and now at this stage changes the plan from promoting Esteem through student 3 to promoting Publication though student 1.

When this action fails, the HoD changes plans again (See stage 3 in Figure 6.2). And although Publication remains his number one priority he chooses to promote Esteem and gives the third student a chance as the first and second student have failed him.



Step 5/5 : Sequencing the Actions

List of All possible Sequence

States Perspective

```
Seq# 1 .. 0,  1,  10,  43
Seq# 2 .. 0,  2,  9,  37,  85
Seq# 3 .. 0,  5,  12,  43
Seq# 4 .. 0,  1,  12,  43
Seq# 5 .. 0,  2,  18,  67
Seq# 6 .. 0,  2,  20,  41,  85
Seq# 7 .. 0,  5,  18,  67
Seq# 8 .. 0,  2,  9,  39,  92
Seq# 9 .. 0,  2,  9,  41,  85
```

Joint Action Perspective

```
Seq# 1 .. - j 3 - j 0 - j 1
Seq# 2 .. - j 5 - j 3 - j 0 - j 2
Seq# 3 .. - j 1 - j 3 - j 0
Seq# 4 .. - j 3 - j 1 - j 0
Seq# 5 .. - j 5 - j 1 - j 2
Seq# 6 .. - j 5 - j 2 - j 3 - j 0
Seq# 7 .. - j 1 - j 5 - j 2
Seq# 8 .. - j 5 - j 3 - j 1 - j 2
Seq# 9 .. - j 5 - j 3 - j 2 - j 0
```

Safety, Opportunity and Threat Perspective

```
Seq# 1 Safety(-10) Opportunity ( 120) Threat ( 80)
Seq# 2 Safety(-20) Opportunity ( 130) Threat ( 80)
Seq# 3 Safety(-10) Opportunity ( 120) Threat ( 70)
Seq# 4 Safety(-10) Opportunity ( 120) Threat ( 80)
Seq# 5 Safety(-10) Opportunity ( 130) Threat ( 60)
Seq# 6 Safety(-20) Opportunity ( 130) Threat ( 50)
Seq# 7 Safety(-10) Opportunity ( 130) Threat ( 50)
Seq# 8 Safety(-20) Opportunity ( 130) Threat ( 80)
Seq# 9 Safety(-20) Opportunity ( 130) Threat ( 60)
```

Figure 6.3: Choices of sequences in Scenario 3

6.4 Analysing the Results

The analysis below is divided into two parts: First, we will discuss the results of applying the methodology to twelve different scenarios and discuss the reactions of the HoD in those different scenarios. Secondly, we will consider all possible value orders in the system and analyse how the agent would react throughout the decision-making stages.

6.4.1 Multiple Scenarios Analysis

We ran the process across different thresholds (Table 6.13) and assuming Cooperative and Non-Cooperative scenarios. In all scenarios, the Starting VO is EST>P>E>H and as a result the HoD will always start the decision-making process choosing the sequence $J1_2 \rightarrow J3_3 \rightarrow J1_3$ (Sending S2 → Asking S3 to write → Sending S3).

Threshold	Final VO	Chosen Sequence
Cooperative Agent		
1	H>E>Est>P	$J1_2 \rightarrow J3_1 \rightarrow J1_1$
5	H>E>Est>P	$J1_2 \rightarrow J3_1 \rightarrow J1_1$
15	Est>H>E>P	$J1_2 \rightarrow J3_3 \rightarrow J1_3$
50	Est>P>E>H	$J1_2 \rightarrow J3_3 \rightarrow J1_3$
1000	Est>P>E>H	$J1_2 \rightarrow J3_3 \rightarrow J1_3$
Mixed	Est>P>H>E	$J1_2 \rightarrow J3_3 \rightarrow J1_3$
Threshold	Final VO	Chosen Sequence
Non-Cooperative Agent		
1	P>Est>E>H	$J1_2 \rightarrow J3_1 \rightarrow J1_1$
5	P>Est>E>H	$J1_2 \rightarrow J3_1 \rightarrow J1_1$
15	P>Est>E>H	$J1_2 \rightarrow J3_1 \rightarrow J1_1$
50	Est>P>E>H	$J1_2 \rightarrow J3_3 \rightarrow J1_3$
1000	Est>P>E>H	$J1_2 \rightarrow J3_3 \rightarrow J1_3$
Mixed	P>Est>E>H	$J1_2 \rightarrow J3_1 \rightarrow J1_1$

Table 6.13: Agents reaction to different emotional thresholds with the following starting VO: $Est > P > E > H$

Observations

1. The original plan of the HoD is always Sending S2 and then S3 after he prepares a paper and S1 would have no role at all in the HoDs original plan. We see that emotional factors change this preference according to the result of Sending S2 at the beginning. If S2 performs badly, the HoD will give Publication more importance and when the emotional threshold is set low, the HoD will quickly change his VO to have P as the most important value. This would then reflect on the action plan where the HoD will now ignore S3 (as S3 in not likely to produce a paper) and consider S1 instead. The HoD will now ask S1 to write a paper and send him to a conference as this better suits the new VO.

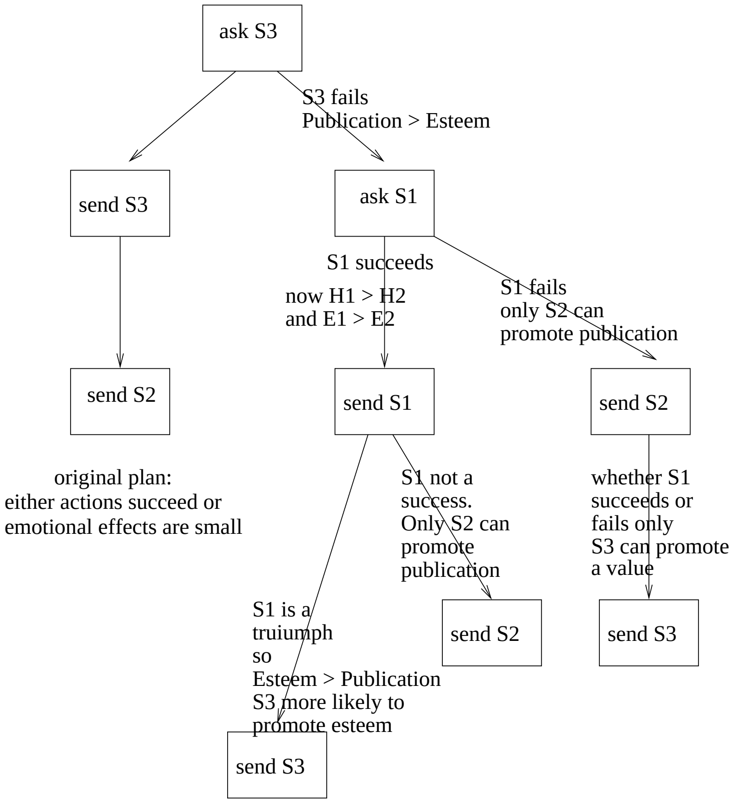


Figure 6.4: Variations arising from emotional influences

2. In cooperative scenarios when emotions have lower effects (higher thresholds) we notice that the HoD would always stick to the original plan and does not deviate from it, meaning that emotions will have little or no influence at all in the decision-making process.

3. In cooperative scenarios when emotions have higher effects (lower thresholds) we notice that the HoD will immediately start preferring students' Happiness and Experience and hence focusing on S1 rather than S3. This happens because the HoD will now become more comfortable and joyful as his plans are moving very well, and when the HoD is volatile and so responsive to emotional changes (as emotions have a low threshold). He will immediately change VO to rank Happiness higher, causing him to prefer sending S1 to a conference instead of S3 (As S1 is more likely to obey).

4. In both cooperative and non-cooperative scenarios we notice that as the Threshold goes higher emotions will have little or no impact on the VO and therefore the chosen sequence. Increasing the threshold, would make the HoD less volatile to changes and, therefore, be more rational in his choices.

When the threshold is set at 1000, we notice that emotions play no role at all in the decision-making process. However, as we have seen, sticking to a failing plan can miss opportunities, and so whenever there is a possibility of an emotional trigger for resequencing, the HoD performs better.

5. As the HoD gets more comfortable when actions succeed we notice that Happiness and Experience get promoted eventually affecting his choice of actions whereby he starts to incline towards S1 rather than S3.

6. In all scenarios, the HoD will always start with Sending S2 as he has a paper ready and would immediately promote Publication. In most cases, the result of this action determines whether he would continue with S3 as planned or switch to S1. The HoD will always keep the original plan of Sending S3 whenever Esteem is at the top of VO, but whenever this changes, the HoD would immediately ignore S3 and focus on S1. This is because the only benefit that the HoD has from S3 is when the value Esteem is of importance, otherwise, S1 would be his choice.

7. We can summarise the tables above in the following: the HoD would rationally choose to Send S2 and then S3 ignoring S1 unless emotional factors were high enough then he would switch to Sending S2 and S1 ignoring S3. This is a good example of the importance of emotions as it shows us how our agent can quickly adapt to changes in the environment and react to results. The HoD has better chances to promote Publication through S1 and S2, but he chooses to go with S3 initially instead of S1 as S3 promotes Esteem. Failure to promote Publication with S2 quickly makes the HoD give more focus to it and he chooses to take the safer path, choosing S1.

A useful overview of the process can be seen in Figure 6.4. This shows what happens with a starting VO of $Est > P > E > H$ for different responses and thresholds.

Emotions can influence the decision making at three points:

1. If S_3 fails to write a paper. Worry about Publication may mean that S_1 is asked to write a paper instead of Esteem being pursued by sending S_3.

2. If S_1 writes a paper. Now liking for S_1 will mean that S_1 is sent before S_2.

3. If S_1 succeeds at the conference. Relief with respect to Publication, together with improved mood after this unexpected bonus may mean that Esteem again becomes the priority and S_3 is sent to pursue this value.

Of particular note is the way in which the emotional response prevented staying with a failing plan and instead sought an alternative way of promoting key values in step 3, and caused a refocus on important departmental goals in response to to the unexpected success of $S1$. It may appear that $S2$ suffers, because he is given no opportunity to impress the HoD. We may think this is appropriate as the aims of the department are furthered. However, if we wish to give the interests of $S2$ more weight, we could increase the liking for $S2$ in the initial position, to reflect that he already had a paper written.

6.4.2 Multiple VO Analysis

In these scenarios (Table 6.14) all possible VO are presented and we analysed the response of the HoD in every step. A couple of assumptions are made here: First, there is a fixed emotional threshold between all values in the VO. Second, we assume that the students will alternate in their response by failing or succeeding to achieve the required action.

Observations

1. In any step during the decision-making process the agent will always preserve the same action plan whenever it is going as expected (stages 2 and 4) and would always consider changing the plan if an action fails (stages 1,3 and 5). This happens because the original VO can change dramatically whenever a step of the plan fails as the HoD responds emotionally to the failure. This change in the VO would immediately affect the plan of the HoD and hence change the next actions he would make. On the other hand, success in the execution of the plan would usually encourage the HoD to to continue on the same course of successful actions.

2. In some cases (such as 7-12 and 19-24) we notice that emotions and resequencing had no influence on the decision plan. It is useful to consider such scenarios to reduce complexity and redundancy in the decision-making process. Looking at the table above we can set a rule that whenever the value H or P are the most highly rated the HoD will skip the resequencing and emotions process during the execution of his actions.

3. A more interesting observation is what we see in 5,6,17 and 18 in stage 3, where even though the HoD succeeds in his plan he still changes it in stage 3 in response to an unexpectedly good outcome. This reflects the fact that priorities are changing and when priorities change the plan also changes.

So, emotions would usually play a big role when things are not moving on track but they can still influence the plan even when it is moving right.

6.5 How Emotions Helped in the Decision-Making Process

This section shows how the work of this chapter contributes to the overall motivation of this thesis. The inclusion of emotional considerations in the HoD's decision-making process helps him to decide when merits are equal over multiple choices. It has also shown how plans might better be changed during the execution of it depending on their results. Emotional consideration was found to be useful as a substitute in the absence of enough information to make a rational decision.

6.5.1 A Tie Breaker in Equally Evaluated Options

Situations where preferences are set equally among different values will mean that the agent might come to situations where a decision will be made on a random basis as rationality will fail to resolve to better decisions. In real life, emotions are usually the judge in situations whenever rationality fails. For example, if the most important values are satisfied, the better liked agent will be chosen.

In the scenarios above, the HoD has no preference among the different students initially (H1 = H2 = H3 and E1 = E2 = E3), but as they started succeeding and failing in bringing about the HoD's values he started to place preferences among them and the VO started to change reflecting this change.

6.5.2 The Necessity of Resequencing

Whenever a decision is needed, considerations of the future are made and the methodology of decision making will not only recommend the next action, but also build a plan of actions that the agent can best follow. This plan is made on certain expectations as to what will happen during its implementation and based on current priorities. Nevertheless, if the actual results were different from the original expectations this might mean that the plan is no longer the best plan to be followed. Using emotions, the action can cause the priorities of the agent to change. If we planned to travel by plane as it is two hours faster than the train and went to buy tickets and found out that the wanted flight has been cancelled and the alternative flight will have a four-hour connection it is wise to resequence and consider if other plans were better in light of the recent results.

That said it also makes sense that if results occurred as expected that the plan is followed as is. This was clearly shown as the main difference between scenario 1 and 2, where the original plan was never affected regardless of changes that occurred in the VO as the students were always delivering the HoD's expectations. The volatility of behaviour shown in table 6.13 for low thresholds is undesirable.

6.5.3 Developing Cooperation

This is one aspect that was not implemented nor studied thoroughly in this work but rather observed as an area for further studies. If agents considered emotional aspects of each other in their evaluation of what actions to take, this might have an influence on their choice of action. When S3 refused to write a paper at the beginning of scenario 2, this affected future considerations that are related to him being sent to a conference (which is not related directly to his failure). If S3 had known that failing to write will have emotional effects on the HoD that might influence future possibilities, S3 might have weighed his options differently and chosen to write the paper.

If this aspect (Cooperation) was considered in the example above, S3 in scenario 2 might choose to write the paper next time based on the motivation to not make the HoD angry.

6.5.4 Absence of Information

Absence of information in real life is the same as equality, e.g., if we want to go home and there are two roads X and Y, one of which is long and the other is short, but we do not know which is which we basically can only assume (X=Y) and choose randomly between them. After experiencing both roads, information will be there as to which is shorter and then the order will be $X > Y$ or $Y > X$. If the experience of S2 is of more importance to the department as he is a promising student with lots of potential but the HoD does not know this information he will initially assume (E1=E2=E3), but as we saw in the scenarios as S2 began to succeed in promoting important values the VO has changed to E2 > (E1=E3).

Such information allowed us to be more specific in setting preferences. Where the HoD used to treat Experience of the students as one value (E) short for (E1=E2=E3) the HoD now can expand on his preference and start setting preferences among the general value of Experience to the specific considerations of each students' experience.

The argument advanced here is that it is better to choose on the basis of emotions when information is absent rather than purely making random choices. Now, of course, this is not always correct as we might for the purpose of fairness choose randomly among actions when information is absent.

6.5.5 Elicitation of Trust

We showed how elicitation of Trust (Chapter 2) is dependent on a combination of beliefs and emotions. Studies of Trust (see Chapter 2) suggest that decisions are trusted better whenever emotional aspects are considered, entities trust each other when they have sufficient knowledge of their abilities and also based on emotional factors that they have towards them. If an agent does not have any emotional considerations in his decisions it will not be possible to build emotional attitudes toward him. The model presented in this thesis was carefully built to integrate both rationality and emotions in the decision-making model and at the same time to provide a mechanism to give flexibility in controlling the balance between these aspects.

It was shown in the scenarios how emotional influences had changed the decision sometimes and also how they were not enough to influence a change at other times.

In particular, we have shown how:

- Emotions can trigger resequencing in response to changes in priorities.
- Emotions can provide a reason based on experience for choosing between rationally equivalent options.
- Emotions foster cooperation.
- Emotions reward cooperation and punish defection.

We believe that all four of the qualities are desirable in a decision maker. Thus, our ability to incorporate them through the mechanism of emotions should foster trust.

6.6 Summary

This chapter provided an example study of the fifth element of a trusted decision-making methodology, “Emotions”. The theoretical bases were explained in Chapter 5. One purpose of providing this example was to give a comprehensive explanation of the mechanisms of the work by using an example relevant to the topic.

To test the applicability of this work, it was implemented and tested in a computer program. Some illustrations of this program have been given in the chapter itself and the rest are detailed in the appendix.

	VO	Stage1	Stage 2	Stage 3	Stage 4	Stage 5
1	$Est > P > E > H$	$J1_2 \rightarrow J3_3 \rightarrow J1_3$	$J3_1 \rightarrow J1_1$	$J1_1$		
2	$Est > P > H > E$	$J1_2 \rightarrow J3_3 \rightarrow J1_3$	$j3 \rightarrow J1_1$	$J1_1$		
3	$Est > E > H > P$	$J3_3 \rightarrow J1_3 \rightarrow J1_2$	$J3 \rightarrow J1_1 \rightarrow J1_2$	$J1_1 \rightarrow J1_2$	$J1_1 \rightarrow J3_3 \rightarrow J1_3$	$J3_3 \rightarrow J1_3$
4	$Est > E > P > H$	$J3_3 \rightarrow J1_3 \rightarrow J1_2$	$J3 \rightarrow J1_1 \rightarrow J1_2$	$J1_1 \rightarrow J1_2$	$J1_1 \rightarrow J3_3 \rightarrow J1_3$	$J3_3 \rightarrow J1_3$
5	$Est > H > P > E$	$J3_3 \rightarrow J1_3 \rightarrow J1_2$	$J3_3 \rightarrow J1_3 \rightarrow J3_1 \rightarrow J1_1$	$J1_2 \rightarrow J1_3$	$J3_1 \rightarrow J1_1$	$J1_1$
6	$Est > H > E > P$	$J3_3 \rightarrow J1_3 \rightarrow J1_2$	$J3_3 \rightarrow J1_3 \rightarrow J3_3 \rightarrow J1_1$	$J1_2 \rightarrow J1_3$	$J3_1 \rightarrow J1_1$	$J1_1$
7	$P > Est > H > E$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		
8	$P > Est > E > H$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		
9	$P > E > Est > H$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		
10	$P > E > H > Est$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		
11	$P > H > Est > E$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		
12	$P > H > E > Est$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		
13	$E > Est > H > P$	$J3_1 \rightarrow J1_2 \rightarrow J1_1$	$J3_3 \rightarrow J1_3 \rightarrow J1_2$	$J1_3 \rightarrow J1_2$	$J3 \rightarrow J1_1 \rightarrow J1_3$	$J1_1 \rightarrow J1_3$
14	$E > Est > P > H$	$J3_1 \rightarrow J1_2 \rightarrow J1_1$	$J3_3 \rightarrow J1_3 \rightarrow J1_2$	$J1_3 \rightarrow J1_2$	$J3 \rightarrow J1_1 \rightarrow J1_3$	$J1_1 \rightarrow J1_3$
15	$E > P > Est > H$	$J3_1 \rightarrow J1_2 \rightarrow J1_1$	$J3_3 \rightarrow J1_3 \rightarrow J1_2$	$J1_3 \rightarrow J1_2$	$J3 \rightarrow J1_1 \rightarrow J1_3$	$J1_1 \rightarrow J1_3$
16	$E > P > H > Est$	$J3_1 \rightarrow J1_2 \rightarrow J1_1$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_3$	
17	$E > H > Est > P$	$J3_1 \rightarrow J1_2 \rightarrow J1_1$	$J3_3 \rightarrow J1_2 \rightarrow J1_3$	$J1_3 \rightarrow J1_2$	$J3_1 \rightarrow J1_1 \rightarrow J1_3$	$J1_1 \rightarrow J1_3$
18	$E > H > P > Est$	$J3_1 \rightarrow J1_2 \rightarrow J1_1$	$J3_3 \rightarrow J1_2 \rightarrow J1_3$	$J1_3 \rightarrow J1_2$	$J3_1 \rightarrow J1_1 \rightarrow J1_3$	$J1_1 \rightarrow J1_3$
19	$H > Est > P > E$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		
20	$H > Est > E > P$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		
21	$H > P > Est > E$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		
22	$H > P > E > Est$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		
23	$H > E > Est > P$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		
24	$H > E > P > Est$	$J1_2 \rightarrow J3_1 \rightarrow J1_1$	$J3_1 \rightarrow J1_1$	$J1_1$		

Table 6.14: Agents reaction to different emotional thresholds

Chapter 7

Summary

The work presented in this thesis produces a methodology of decision making for software agents.

This thesis has been structured with seven chapters and two main parts. The first two chapters introduced the work and presented a literature survey that reviewed related research. From this, we concluded that the decision-making methodology can be seen from two main angles, “Beliefs” and “Emotions”.

Chapters 3 and 4 (grouped under Part I) developed a methodology of decision making on the basis of agent beliefs and rational decision making reflecting subjective preferences. Then, Chapters 5 and 6 (grouped under Part II) extended the methodology to address how decision making could be influenced by emotions.

The work presented has been implemented and tested in a computer program which is referred to within the chapters and also illustrated in the appendix.

This chapter provides a summary of the contribution this thesis offers in Section 7.1 and possible areas of further work in Section 7.2.

7.1 Contribution

The main aim of this work was defined in Chapter 1 as considering the following question:

> “What aspects can raise our confidence level in agents so as to allow them to take over decision making and how can these aspects be implemented?”

The motivation is to develop a decision-making methodology that can be better trusted in critical or sensitive situations which computer softwares are usually not trusted to handle.

The thesis contribution was presented in Section 1.4 as:

1. Use argumentation as a way to model practical reasoning through an instantiation of an argumentation scheme as a presumptive reasoning for action.

2. Consider joint actions to address how the environment and other agents would react in conjunction with the agent's own actions.

3. Address uncertainties in the methodology and allow it to build decisions when information is incomplete about the environment or unexpectedness in actions outcomes.

4. Exploit the role of emotions in the process of decision making.

5. Examine the effect emotions would have on the decision-making process compared to social aspects by providing a mechanism to balance between the influence of emotions and beliefs.

6. Present a detailed example study.

Section 7.1.3 summarises how the above-mentioned points were addressed in the thesis.

7.1.1 A Trusted Decision

To properly address the main thesis question and successfully be able to delegate decision making to agents, it was important to try and understand what we mean by "Trust" and how an agent can be trusted. Section 2.1 offered a review of different aspects of "Trust" and concluded with five main capabilities that a decision making methodology must have for it to be "Trustworthy".

Two main concepts to consider here. First, Trust is a product of both "Beliefs" and "Emotions". An agent can trust another agent if he believes that he is trustworthy, meaning, that the trustee has the capability to take over decision making and the truster would be better off as a result. We also trust others when we know that they consider “Emotions” in their decisions. Although an emotionless trustee might get the calculations right, not all of our decisions are based on calculations. For example, we trust our friends on sensitive matters sometimes only on the basis of emotional considerations. As the study shows, emotions are also an important trigger for resequencing, so that an agent does not stick to a plan when it has become more likely to fail. This idea was used in [119] but here we model more emotions giving a much richer set of resequencing triggers.

Secondly, a trusted agent is defined as an agent that has the capability to: evaluate its options and the likely effects of every option, understand and react to the environment around it, plan ahead and set courses of actions and not just make decisions that would have immediate results. The agent should also be able to make decisions even when not all the information is available to him or if he is not certain about the information he already has. The final capability is emotions. A trusted agent should use emotions to influence decisions, and know how to balance their effects. Noting that it is crucial to account for emotions in decision making, it can also be harmful if the agent’s decision was primarily made based on emotions.

7.1.2 Methodological Approach

In Chapter 3, we presented a methodology of action selection for use by agents. The methodology is intended to raise the trustworthiness we have toward agents, so as to ensure that the demands raised in 7.1.1 are satisfied. The methodology consists of five main steps:

1. Formulating the Problem: produce a formal description of the problem scenario to give all possible actions, values and related factors that may influence the decision.

2. Determining the Arguments: arguments providing justifications of the various available actions are constructed.

3. Building the Argumentation Framework: the arguments and attacks between them identified in the previous step are organised into an Argumentation Framework and relevant values are associated with those attacks

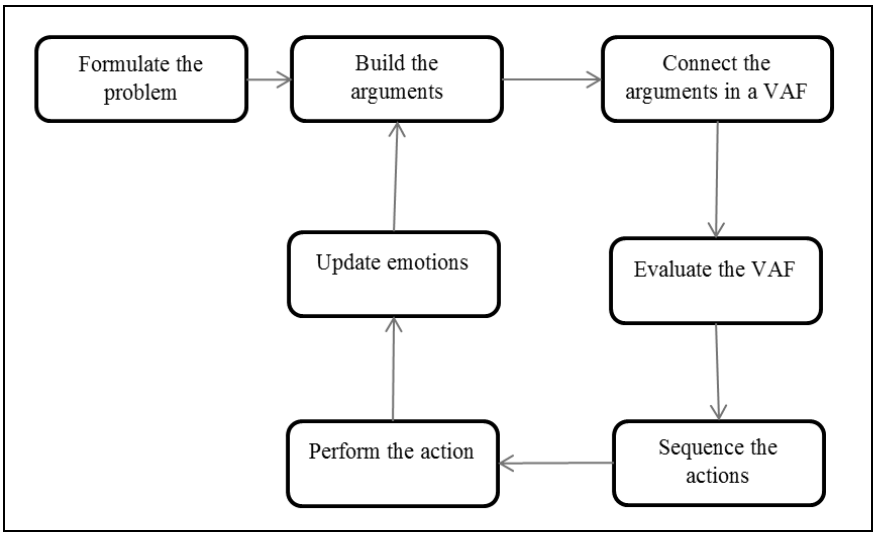


Figure 7.1: Action Life Cycle

resulting in a VAF (Value-Based Argumentation Framework).

4. Evaluating the Argumentation Framework: the arguments within the VAF are determined as acceptable or not with respect to a specific audience, characterised by the ordering of values subscribed to by the decision-making agent.

5. Sequencing the Actions: the set of actions deemed to be acceptable to the agent in the previous stage must now be put into a suitable order in which they should be performed. This allows for uncertainties using estimated probabilities. Moreover, this step enables an overview of future actions and how they can be planned into different paths.

Figure 7.1 gives an overview of the action life cycle where emotional factors will be calculated after seeing the result of an action and before the next action is committed to, which offers the opportunity to resequence.

Emotions are treated as follows: an agent has twenty-two basic emotion types which are tied to the social values the agent originally associates with the decision-making process, the other agents in the scenario and the initial plan. The result of each action would then frustrate or motivate the agent depending on how pleasant and unexpected the actual outcomes were. This would give rise to emotions of various intensities which would affect the behavioural features of the agent. Those emotions are then tied to the ordering of values so that the more extreme shifts in emotions are, the more likely to influence value ordering. This possibility of emotions changing the relevant value order depends on a threshold: if the emotional change was higher than the threshold then the ordering of values will change causing the decision to change.

7.1.3 Main Contributions

We will now address the main contributions of this work (Section 7.1 and Section 1.4):

> *Use argumentation as a way to model practical reasoning through an instantiation of an argumentation scheme as a presumptive reasoning for action:*

The methodology presented in Chapter 3 and then extended in Chapter 5 was built on the proposal of Atkinson et al.[15]. A five-step methodology of action selection defines an argumentation scheme and builds arguments in the second step justifying the execution of any possible action at any state. The reasoning to choose the best action is done in step four where arguments are evaluated based on the merits they provide in terms of values.

> *Consider joint actions to address how other agents would react in conjunction with the agent's own actions:*

The modeling of all possible transitions from the initial state is done in step one of the methodology using AATS. This model accounts for realising all possibilities of other agents' actions (i.e., there might be more than one possible outcome of an agent's action depending on how other agents react). As every possibility forms an argument in the Argumentation Framework, step three of the methodology used critical questions (CQs) to ensure that possible side effects from other agent's actions are accounted for the attack relationship.

> *Address uncertainties and allow for decisions to be made when information is incomplete:*

The critique of arguments in the third step of the methodology through critical questions CQ8-CQ10 addresses possible side effects of any action. These critiques will then be considered in the fourth step where the action will be rejected if the side effect warrants it, according to the ordering of values by the decision maker.

> *Consider the role of emotions in the process of decision making:*

Chapter 5 complements the methodology presented in Chapter 3 by adding a mechanism to integrate emotional aspects into the decision making process. Emotions are generated based on the outcome of any single action and particularly on how important this outcome was and its unexpectedness. Emotions would influence the ranking of values, not directly determining decisions but rather modifying preferences of the decision-making agent.

> *Examine the level of effect emotions would have on the decision-making process compared to social aspects by providing a mechanism to balance between the influence of emotions and beliefs:*

Chapter 5 accounts for the notion of *Threshold* in two ways: an overall threshold that controls how emotional factors would affect the ordering of values; and by addressing different emotional effects on different values in order to account for the fact that when an agent sets his value order some values should be more susceptible to emotions than others.

> *Present a detailed example study:*

Chapter 4 presents an example study of the five-step methodology. Chapter 6 then extends this example by including emotional aspects. This was then experimented in a computer program and snapshots were provided within the chapters and in an appendix.

7.2 Future Direction

7.2.1 Current Limitations

Planning

The current methodology offers very basic functionalities of planning as it considers aspects of side effects. It does not, however, implement a comprehensive planning methodology where an agent can consider in details all possibilities and their side effects, decide on whether a decision actually needs to be planned on the consequences of decisions on other agents and on the environment.

Use of AATS

The AATS does not satisfactorily account for values promoted in virtue of actions rather than in virtue of states reached. This limitation is addressed in [17] which could readily be used as our formal basis.

Emotions

This thesis captures the concepts of emotions in a very limited manner. Emotions in this thesis were taken for granted and as explained in the models of psychology and philosophy. We do not, however, offer an explanation for why and how emotions have the effect they do, but rather assumed this from previous work.

Historical Data

The concept of storage and consideration of historical events is captured in our methodology as we are considering a dynamic value order. The changes in value order are a result of previous actions, so the VO at any point of time is a result of these and their consequences. This method can be improved by implementing a more structured methodology of data storage and how this data can have an influence on the decision.

7.2.2 Possible Extensions

Long term effects of emotions

Emotions change with the outcome of one's actions and in accordance with previous expectations. These changes then affect the value order of the agent and his future decisions.

A possible extension to our work is to study how emotions would behave in the longer term. It would be interesting to see how emotions would look after many iterations and decisions and the aspect of these changes. This study can then propose a recommendation on the effectiveness of this model and how realistic it is.

Further consideration of different scenarios

The scenario chosen in this thesis combines both the needs for sensitivity and criticality with a focus on the implications on one agent (the HoD). A further development would be to consider other different scenarios with different levels of criticality allowing us to provide a concrete evidence of the validity of the model in situations where we would model the decision making of all the agents involved, rather than regarding the actions of other agents as exogenous to the system.

This might also introduce perspectives of multiple agents in the same scenario giving us the ability to compare the reaction of multiple agents to each other's decisions, and for anticipated emotional responses to form part of the agent's decision making. This would help model strategies such as ingratiation.

Optimisation

The implementation was done in a .Net environment using Visual Basic. This has been optimized for accuracy but not speed. A possible development would be to consider other implementation possibilities (Java, C++, ..etc), and also consider possible enhancements to the code to have more efficient calculations in the model.

This would allow more complex scenarios to be tested on the runtime perspective to be assessed.

Considerations of alternatives

The methodology presented sometimes produces multiple acceptable possible actions to the agent. The agent chooses one based on considerations of Safety, Opportunity and Threat.

An interesting development of this work would be to study this aspect in detail, providing a mechanism for the agent to automatically compute the relative importance of these three factors.

Address the current limitations

Further developments to this work would be to address the current limitations in this thesis (Section 7.2.1). Planning methodologies can be integrated to this model offering the ability to differentiate among short and long term goals and decisions. This would also be accompanied with the ability to measure the benefits of planning and how it would improve the quality of decisions.

The study of other accounts of emotions in philosophy and psychology and integrating them with the concepts of Multi-Agent systems is an interesting project and would yield more perspectives on the understanding of trust and the ability to enhance our decision-making methodologies to induce more trust.

The last limitation mentioned above was the consideration of historical information. The current model does not offer a structured method to capture feedback from the environment and have it as an input to the decision-making process.

This can be developed to first study the effectiveness of such an addition and what value it would add to the overall effectiveness of the system and then develop the mechanism where such possibilities can be constructed and tested.

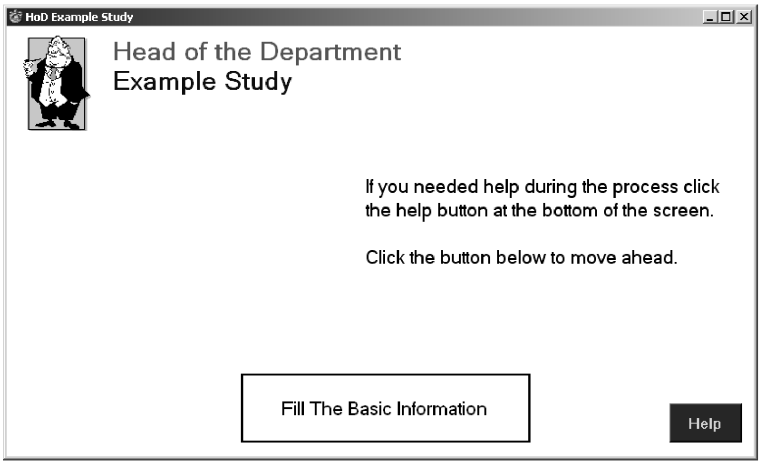
Appendices

Appendix A

Snapshots from the experimental study

This Appendix will give a few snapshots from the software implementation of the methodology. A brief description is given with every snapshot.

Please note that in the screen shots of the implementation the joint actions are referenced differently. (J0,J1 and J2) are sending students 1,2 and 3 to conferences. (J4, J5 and J6) are asking students 1,2 and 3 to write a paper.



Head of the Department
Example Study

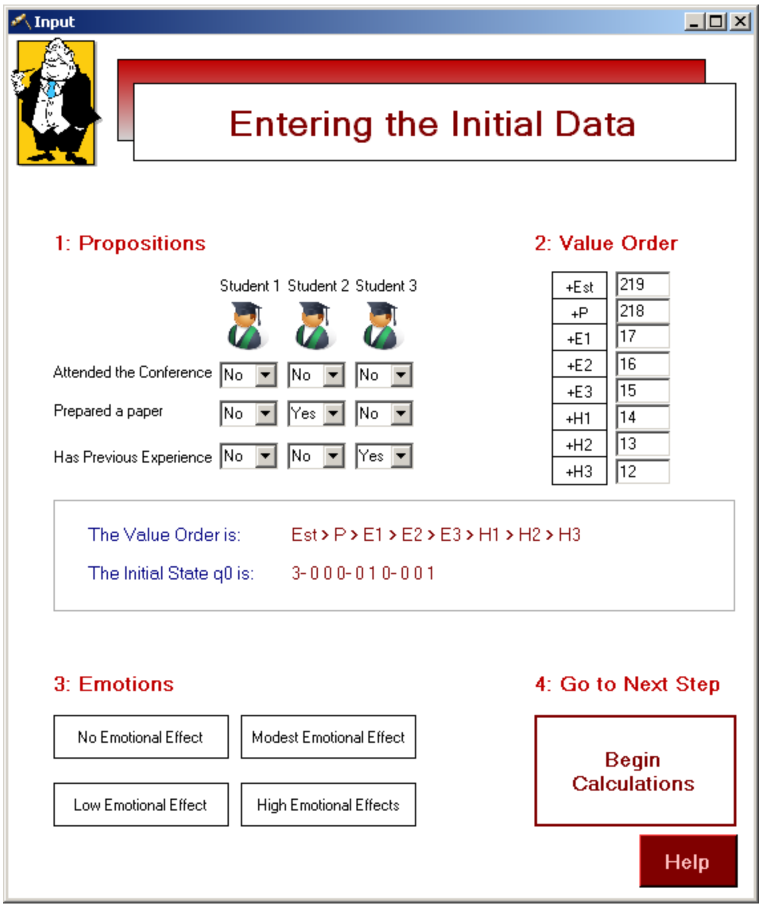
If you needed help during the process click the help button at the bottom of the screen.

Click the button below to move ahead.

Figure A.1: The Main Screen

Figure A.1 is the first screen that appears as the program starts. It gives the title of the case study. A 'Help' button is placed in every screen of the application. When clicked, it gives instructions and guidance on how to interact with the screen.

Figure A.2 is the input screen. Here the user can set the different parameters of the experiment. The Value Order can be set by entering different weights in



	Student 1	Student 2	Student 3
Attended the Conference	No	No	No
Prepared a paper	No	Yes	No
Has Previous Experience	No	No	Yes

+Est	219
+P	218
+E1	17
+E2	16
+E3	15
+H1	14
+H2	13
+H3	12

The Value Order is: Est > P > E1 > E2 > E3 > H1 > H2 > H3

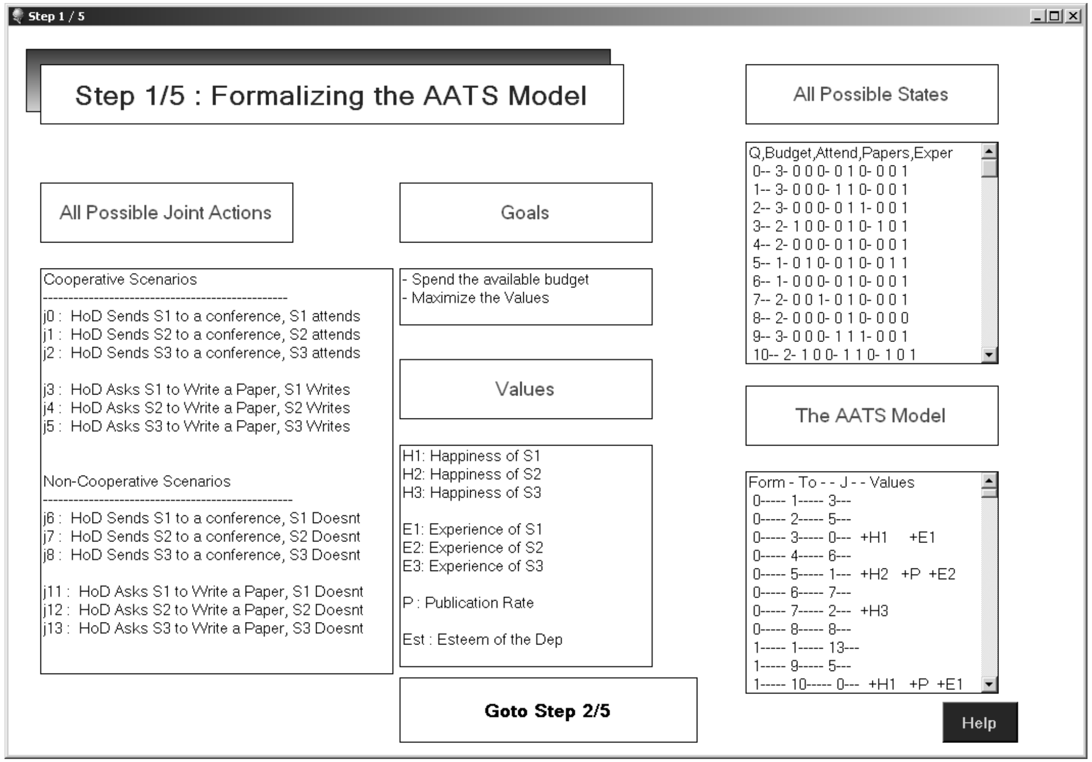
The Initial State q0 is: 3-0 0 0-0 1 0-0 0 1

Figure A.2: Inserting the Information

the input boxes (top right). The initial state can also be altered by changing the drop boxes (top left). The box at the center of the screen summarised the input of the Value Order and the initial state and showed it in the format familiar with the presentation in this thesis.

Another possibility the user has is setting the thresholds of emotions. This places a factor between the different values in the Value Order (bottom right). “No Emotional Effect” would place a factor of 1000 between every value making it hard for emotions to actually play a role in the decision-making process. ‘High Emotional Effects’ would place a factor of 1 between the values making the agent very volatile to emotions.

After the user enters the data, the button ‘Begin Calculations’ can be pressed to start solving the problem. This would take the user then through the 5 different steps of the decision-making methodology.



Step 1/5 : Formalizing the AATS Model

All Possible Joint Actions

Cooperative Scenarios

j0 : HoD Sends S1 to a conference, S1 attends
j1 : HoD Sends S2 to a conference, S2 attends
j2 : HoD Sends S3 to a conference, S3 attends

j3 : HoD Asks S1 to Write a Paper, S1 Writes
j4 : HoD Asks S2 to Write a Paper, S2 Writes
j5 : HoD Asks S3 to Write a Paper, S3 Writes

Non-Cooperative Scenarios

j6 : HoD Sends S1 to a conference, S1 Doesnt
j7 : HoD Sends S2 to a conference, S2 Doesnt
j8 : HoD Sends S3 to a conference, S3 Doesnt

j11 : HoD Asks S1 to Write a Paper, S1 Doesnt
j12 : HoD Asks S2 to Write a Paper, S2 Doesnt
j13 : HoD Asks S3 to Write a Paper, S3 Doesnt

Goals

- Spend the available budget
- Maximize the Values

Values

H1: Happiness of S1
H2: Happiness of S2
H3: Happiness of S3

E1: Experience of S1
E2: Experience of S2
E3: Experience of S3

P : Publication Rate

Est : Esteem of the Dep

Goto Step 2/5

All Possible States

```
Q,Budget,Attend,Papers,Exper
 0-- 3- 0 0 0- 0 1 0- 0 0 1
 1-- 3- 0 0 0- 1 1 0- 0 0 1
 2-- 3- 0 0 0- 0 1 1- 0 0 1
 3-- 2- 1 0 0- 0 1 0- 1 0 1
 4-- 2- 0 0 0- 0 1 0- 0 0 1
 5-- 1- 0 1 0- 0 1 0- 0 1 1
 6-- 1- 0 0 0- 0 1 0- 0 0 1
 7-- 2- 0 0 1- 0 1 0- 0 0 1
 8-- 2- 0 0 0- 0 1 0- 0 0 0
 9-- 3- 0 0 0- 1 1 1- 0 0 1
 10-- 2- 1 0 0- 1 1 0- 1 0 1
```

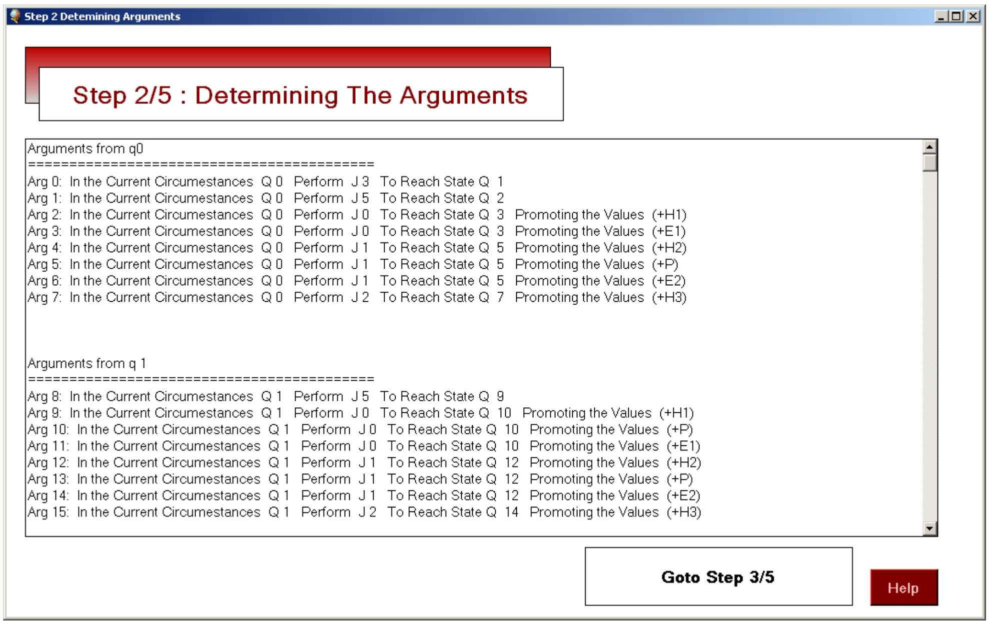
The AATS Model

```
Form - To - - J - - Values
 0----- 1----- 3---
 0----- 2----- 5---
 0----- 3----- 0---  +H1    +E1
 0----- 4----- 6---
 0----- 5----- 1---  +H2   +P  +E2
 0----- 6----- 7---
 0----- 7----- 2---  +H3
 0----- 8----- 8---
 1----- 1----- 13---
 1----- 9----- 5---
 1----- 10----- 0---  +H1   +P  +E1
```

Figure A.3: Step 1 : Formulating the problem

Figure A.3 appears immediately after the user has finished inputting the settings of the example. It basically summarises the results of the calculation that took place in the first step of the methodology. The left and center of the screen are a summary of the basic settings of the example (All the possible joint actions, Values and Goals). On the top right is a calculation of all possible states in the environment given the data that was entered earlier. Bottom right is the AATS of the model giving all the possible transitions of the system and the values each transition might promote.

After reviewing this information, the user clicks go to step 2, which would close this screen and open another one (A.4) that represents step two of the methodology.



Step 2 Detemining Arguments
Step 2/5 : Determining The Arguments
Goto Step 3/5
Help

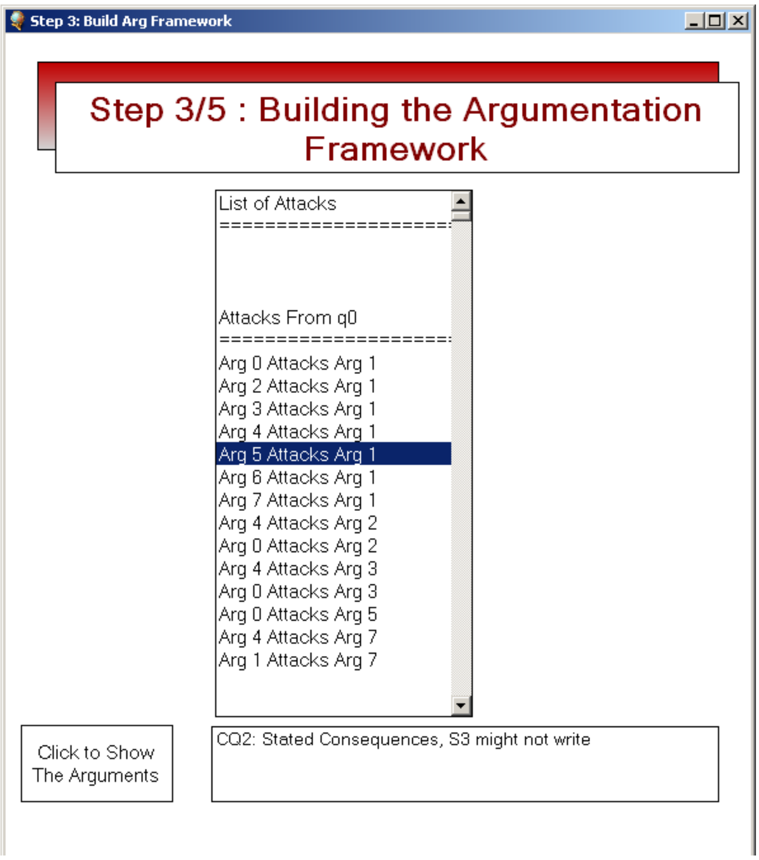
```
Arguments from q0
==========================================
Arg 0:  In the Current Circumestances  Q 0   Perform  J 3   To Reach State Q  1
Arg 1:  In the Current Circumestances  Q 0   Perform  J 5   To Reach State Q  2
Arg 2:  In the Current Circumestances  Q 0   Perform  J 0   To Reach State Q  3   Promoting the Values  (+H1)
Arg 3:  In the Current Circumestances  Q 0   Perform  J 0   To Reach State Q  3   Promoting the Values  (+E1)
Arg 4:  In the Current Circumestances  Q 0   Perform  J 1   To Reach State Q  5   Promoting the Values  (+H2)
Arg 5:  In the Current Circumestances  Q 0   Perform  J 1   To Reach State Q  5   Promoting the Values  (+P)
Arg 6:  In the Current Circumestances  Q 0   Perform  J 1   To Reach State Q  5   Promoting the Values  (+E2)
Arg 7:  In the Current Circumestances  Q 0   Perform  J 2   To Reach State Q  7   Promoting the Values  (+H3)



Arguments from q 1
==========================================
Arg 8:  In the Current Circumestances  Q 1   Perform  J 5   To Reach State Q  9
Arg 9:  In the Current Circumestances  Q 1   Perform  J 0   To Reach State Q  10   Promoting the Values  (+H1)
Arg 10:  In the Current Circumestances  Q 1   Perform  J 0   To Reach State Q  10   Promoting the Values  (+P)
Arg 11:  In the Current Circumestances  Q 1   Perform  J 0   To Reach State Q  10   Promoting the Values  (+E1)
Arg 12:  In the Current Circumestances  Q 1   Perform  J 1   To Reach State Q  12   Promoting the Values  (+H2)
Arg 13:  In the Current Circumestances  Q 1   Perform  J 1   To Reach State Q  12   Promoting the Values  (+P)
Arg 14:  In the Current Circumestances  Q 1   Perform  J 1   To Reach State Q  12   Promoting the Values  (+E2)
Arg 15:  In the Current Circumestances  Q 1   Perform  J 2   To Reach State Q  14   Promoting the Values  (+H3)
```

Figure A.4: Step 2: Determining the arguments

Figure A.4 is step two of the methodology, where all possible arguments are generated and presented. Note that the program would generate all possible arguments from all possible states not just q_0. Arguments are presented here in the same format discussed in this thesis. Pressing the button below will take the user to step three.

Figure A.5 is step three of the methodology where the VAF is generated. The screen now shows all the VAFs in all states of the system. Clicking on any attack would display a text below explaining the reasoning behind this attack and how it is related to Critical Questions. For convenience, a button is added (bottom left) that would expand the screen to show all the possible arguments in the system in case the user wanted to understand the exact components of the attacks (See A.6).

Figure A.7 is the fourth step of the methodology. The attacks have now been resolved and the screen now shows the result of this resolution. The screen would now show the Preferred Extension of the system at every state, showing us all the accepted arguments at every state. We see that in the example captured in A.7, the agent can perform (J3, J5 or J1) only from q_0. In case the user wanted a reminder of what those joint actions are, he can click the button (top right) and it will expand the screen to show joint actions (see figure A.8).



Step 3/5 : Building the Argumentation Framework

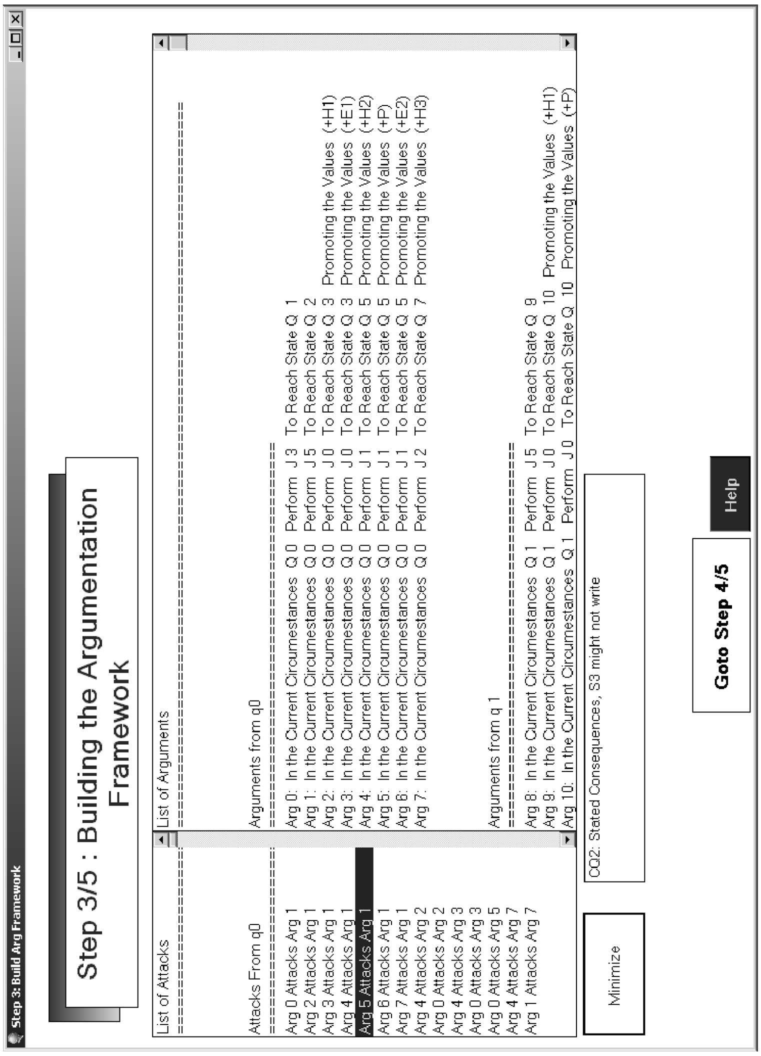
```
List of Attacks
====================

Attacks From q0
====================
Arg 0 Attacks Arg 1
Arg 2 Attacks Arg 1
Arg 3 Attacks Arg 1
Arg 4 Attacks Arg 1
Arg 5 Attacks Arg 1
Arg 6 Attacks Arg 1
Arg 7 Attacks Arg 1
Arg 4 Attacks Arg 2
Arg 0 Attacks Arg 2
Arg 4 Attacks Arg 3
Arg 0 Attacks Arg 3
Arg 0 Attacks Arg 5
Arg 4 Attacks Arg 7
Arg 1 Attacks Arg 7
```

CQ2: Stated Consequences, S3 might not write

Figure A.5: Step 3: Buiding the Argumentation framework

We are now at the final step of the methodology (Figure A.9) where the system would now calculate all possible sequences in the system and show them in term of states (left) and joint actions (center). Moreover, the screen would also show calculations of Safety, Opportunity and Threat (right) to help the user make his decision on which sequence to choose. To help the user choose a button 'More Information' is available and when clicked would expand the screen to show all possible states and joint actions in the system (See figure A.10). The user now chooses the sequence of his choice and clicks 'Continue' to start the emotions evaluation.

Figure A.11 shows the emotional model of the system, it shows (top and left) information on the chosen sequence, number of stages left in that sequence and the next action to be performed according to this plan. The user will now decide what the outcome of this action would be (what the student would actually do in response to the HoD's request) (Succeed or Fail). Based on the user's choice he clicks accept and a screen (bottom) will show what happened. The user then clicks "Replan" which would update the emotions of the HoD and accordingly



Step 3: Build Arg Framework

Step 3/5 : Building the Argumentation Framework

```
List of Attacks
=====================

Attacks From q0
=====================
Arg 0 Attacks Arg 1
Arg 2 Attacks Arg 1
Arg 3 Attacks Arg 1
Arg 4 Attacks Arg 1
Arg 5 Attacks Arg 1
Arg 6 Attacks Arg 1
Arg 7 Attacks Arg 1
Arg 4 Attacks Arg 2
Arg 0 Attacks Arg 2
Arg 4 Attacks Arg 3
Arg 0 Attacks Arg 3
Arg 0 Attacks Arg 5
Arg 4 Attacks Arg 7
Arg 1 Attacks Arg 7
```

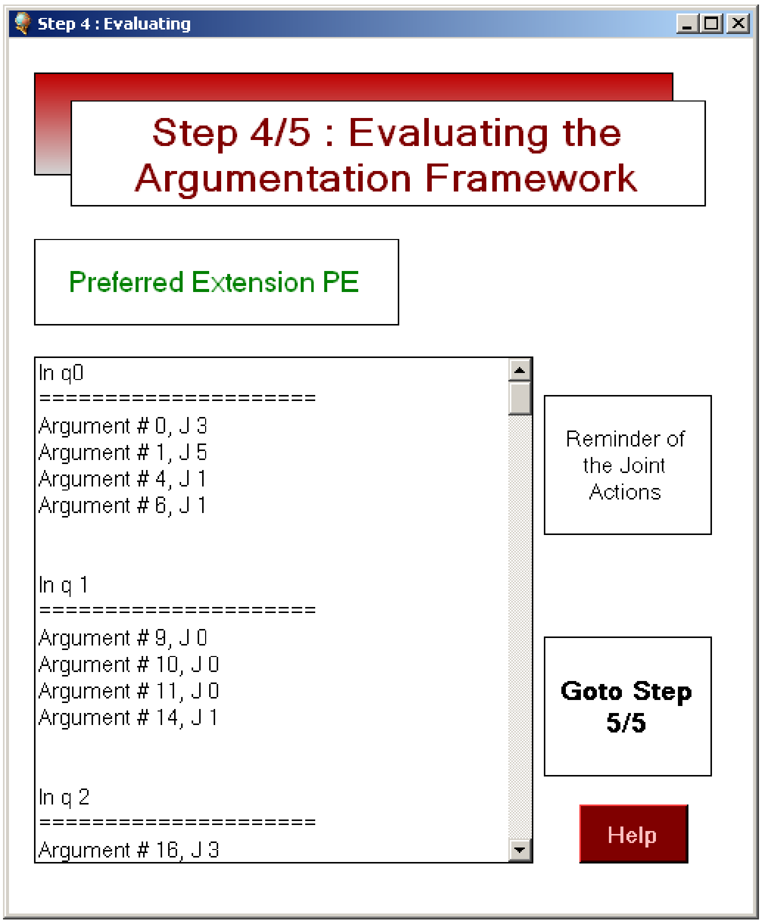
```
List of Arguments
==============================================================================

Arguments from q0
==========================================
Arg 0:  In the Current Circumstances  Q 0   Perform  J 3   To Reach State Q  1
Arg 1:  In the Current Circumstances  Q 0   Perform  J 5   To Reach State Q  2
Arg 2:  In the Current Circumstances  Q 0   Perform  J 0   To Reach State Q  3   Promoting the Values  (+H1)
Arg 3:  In the Current Circumstances  Q 0   Perform  J 0   To Reach State Q  3   Promoting the Values  (+E1)
Arg 4:  In the Current Circumstances  Q 0   Perform  J 1   To Reach State Q  5   Promoting the Values  (+H2)
Arg 5:  In the Current Circumstances  Q 0   Perform  J 1   To Reach State Q  5   Promoting the Values  (+P)
Arg 6:  In the Current Circumstances  Q 0   Perform  J 1   To Reach State Q  5   Promoting the Values  (+E2)
Arg 7:  In the Current Circumstances  Q 0   Perform  J 2   To Reach State Q  7   Promoting the Values  (+H3)


Arguments from q 1
==========================================
Arg 8:  In the Current Circumstances  Q 1   Perform  J 5   To Reach State Q  9
Arg 9:  In the Current Circumstances  Q 1   Perform  J 0   To Reach State Q  10   Promoting the Values  (+H1)
Arg 10:  In the Current Circumstances  Q 1   Perform  J 0   To Reach State Q  10   Promoting the Values  (+P)
```

CQ2: Stated Consequences, S3 might not write

Figure A.6: Step 3: Another view

the Value Order and the state. After that, the system will go back to the input screen which would now reflect the new Value Order and the new state. The system will keep doing this (resequencing) until the HoD exhausts the available budget to him.

To track the progress of the implementation, an icon (see figure A.12) is conveniently placed at the top left of the screen at all times of the experiment 'The Tracker'. When clicked, the tracker would open a larger window (see figure A.13) that would show details on the progress so far in this experiment. This will give details on the changes in Value Order, emotions and behavioural features. The tracker also presents the log of the states transition, action performed and their results and finally the chosen sequence at every step. This can be viewed at any point of time during the execution of the program and can be minimised again



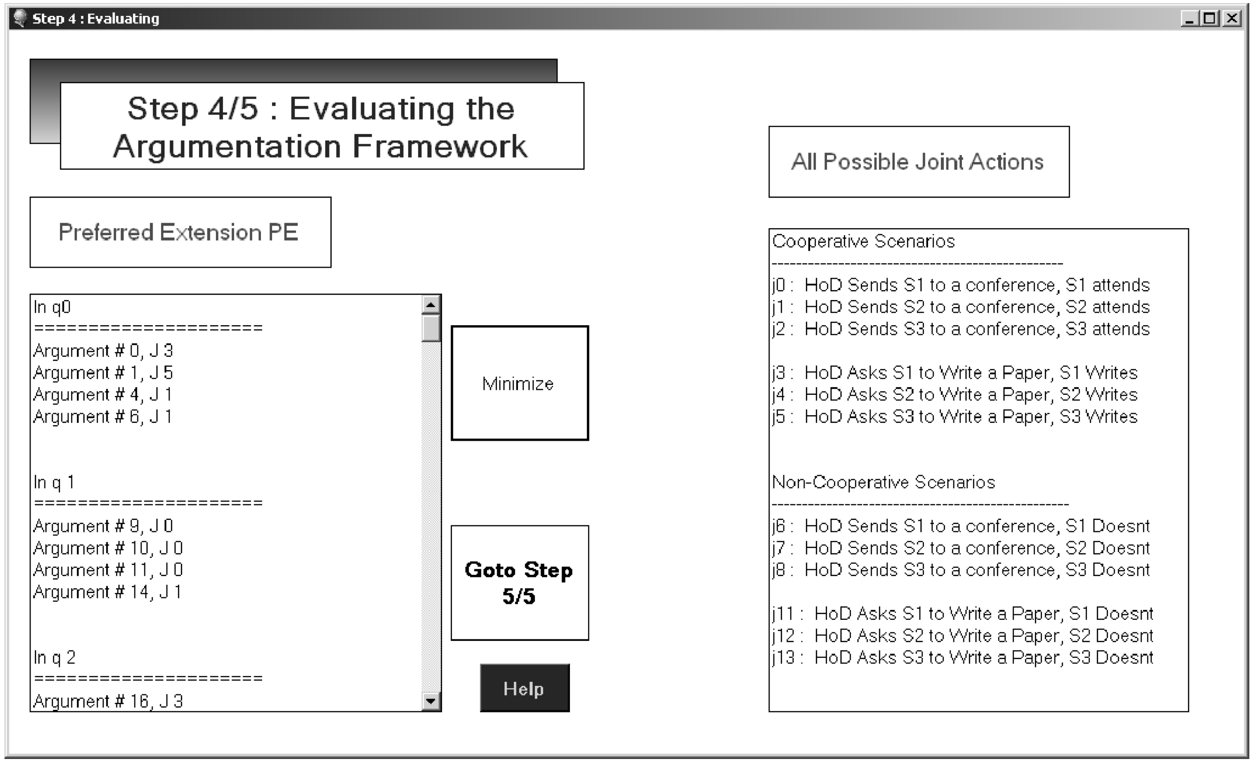
Step 4 : Evaluating
Step 4/5 : Evaluating the Argumentation Framework
Preferred Extension PE
Reminder of the Joint Actions
Goto Step 5/5
Help

```
In q0
=====================
Argument # 0, J 3
Argument # 1, J 5
Argument # 4, J 1
Argument # 6, J 1


In q 1
=====================
Argument # 9, J 0
Argument # 10, J 0
Argument # 11, J 0
Argument # 14, J 1


In q 2
=====================
Argument # 16, J 3
```

Figure A.7: Step 4: Evaluating the argumentation framework

to the small icon of the tracker by clicking “Minimise”.

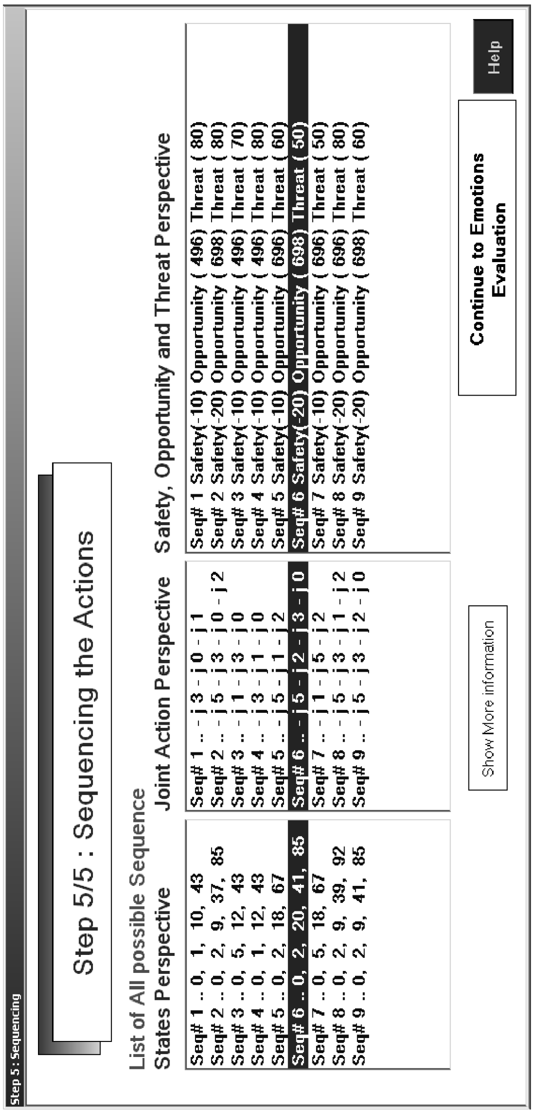
Figure A.14 appears at the very end of the experiment and summarises everything that has happened during the implementation and all the implications that took place. The program closes after this screen.


Step 4 : Evaluating
Step 4/5 : Evaluating the Argumentation Framework
Preferred Extension PE
Minimize
Goto Step 5/5
Help
All Possible Joint Actions

```
In q0
=====================
Argument # 0, J 3
Argument # 1, J 5
Argument # 4, J 1
Argument # 6, J 1


In q 1
=====================
Argument # 9, J 0
Argument # 10, J 0
Argument # 11, J 0
Argument # 14, J 1


In q 2
=====================
Argument # 16, J 3
```

```
Cooperative Scenarios
------------------------------------------------
j0 :  HoD Sends S1 to a conference, S1 attends
j1 :  HoD Sends S2 to a conference, S2 attends
j2 :  HoD Sends S3 to a conference, S3 attends

j3 :  HoD Asks S1 to Write a Paper, S1 Writes
j4 :  HoD Asks S2 to Write a Paper, S2 Writes
j5 :  HoD Asks S3 to Write a Paper, S3 Writes


Non-Cooperative Scenarios
-------------------------------------------------
j6 :  HoD Sends S1 to a conference, S1 Doesnt
j7 :  HoD Sends S2 to a conference, S2 Doesnt
j8 :  HoD Sends S3 to a conference, S3 Doesnt

j11 :  HoD Asks S1 to Write a Paper, S1 Doesnt
j12 :  HoD Asks S2 to Write a Paper, S2 Doesnt
j13 :  HoD Asks S3 to Write a Paper, S3 Doesnt
```

Figure A.8: Another view of the evaluation

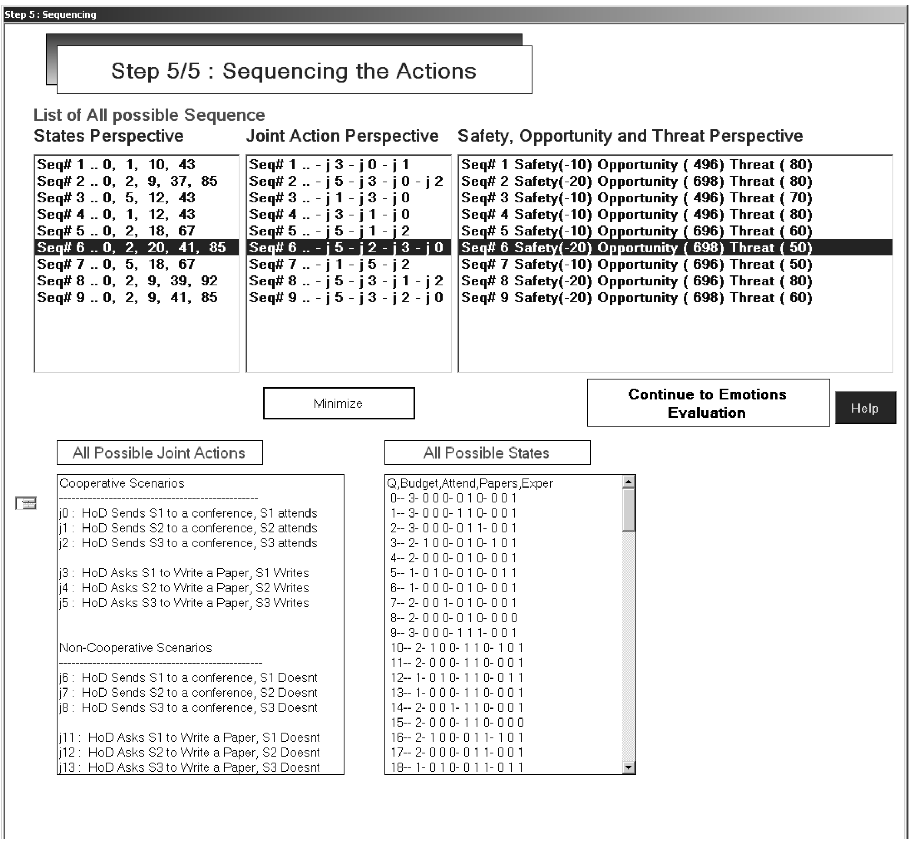


Step 5/5 : Sequencing the Actions

List of All possible Sequence

States Perspective	Joint Action Perspective	Safety, Opportunity and Threat Perspective
Seq# 1 .. 0, 1, 10, 43	Seq# 1 .. - j 3 - j 0 - j 1	Seq# 1 Safety(-10) Opportunity (496) Threat (80)
Seq# 2 .. 0, 2, 9, 37, 85	Seq# 2 .. - j 5 - j 3 - j 0 - j 2	Seq# 2 Safety(-20) Opportunity (698) Threat (80)
Seq# 3 .. 0, 5, 12, 43	Seq# 3 .. - j 1 - j 3 - j 0	Seq# 3 Safety(-10) Opportunity (496) Threat (70)
Seq# 4 .. 0, 1, 12, 43	Seq# 4 .. - j 3 - j 1 - j 0	Seq# 4 Safety(-10) Opportunity (496) Threat (80)
Seq# 5 .. 0, 2, 18, 67	Seq# 5 .. - j 5 - j 1 - j 2	Seq# 5 Safety(-10) Opportunity (696) Threat (60)
Seq# 6 .. 0, 2, 20, 41, 85	Seq# 6 .. - j 5 - j 2 - j 3 - j 0	Seq# 6 Safety(-20) Opportunity (698) Threat (50)
Seq# 7 .. 0, 5, 18, 67	Seq# 7 .. - j 1 - j 5 - j 2	Seq# 7 Safety(-10) Opportunity (696) Threat (50)
Seq# 8 .. 0, 2, 9, 39, 92	Seq# 8 .. - j 5 - j 3 - j 1 - j 2	Seq# 8 Safety(-20) Opportunity (696) Threat (80)
Seq# 9 .. 0, 2, 9, 41, 85	Seq# 9 .. - j 5 - j 3 - j 2 - j 0	Seq# 9 Safety(-20) Opportunity (698) Threat (60)

Figure A.9: Step 5: Sequencing the actions


Step 5 : Sequencing
Step 5/5 : Sequencing the Actions
Minimize
Continue to Emotions Evaluation
Help
All Possible Joint Actions
All Possible States

List of All possible Sequence

States Perspective

```
Seq# 1 .. 0,  1,  10,  43
Seq# 2 .. 0,  2,  9,  37,  85
Seq# 3 .. 0,  5,  12,  43
Seq# 4 .. 0,  1,  12,  43
Seq# 5 .. 0,  2,  18,  67
Seq# 6 .. 0,  2,  20,  41,  85
Seq# 7 .. 0,  5,  18,  67
Seq# 8 .. 0,  2,  9,  39,  92
Seq# 9 .. 0,  2,  9,  41,  85
```

Joint Action Perspective

```
Seq# 1 .. - j 3 - j 0 - j 1
Seq# 2 .. - j 5 - j 3 - j 0 - j 2
Seq# 3 .. - j 1 - j 3 - j 0
Seq# 4 .. - j 3 - j 1 - j 0
Seq# 5 .. - j 5 - j 1 - j 2
Seq# 6 .. - j 5 - j 2 - j 3 - j 0
Seq# 7 .. - j 1 - j 5 - j 2
Seq# 8 .. - j 5 - j 3 - j 1 - j 2
Seq# 9 .. - j 5 - j 3 - j 2 - j 0
```

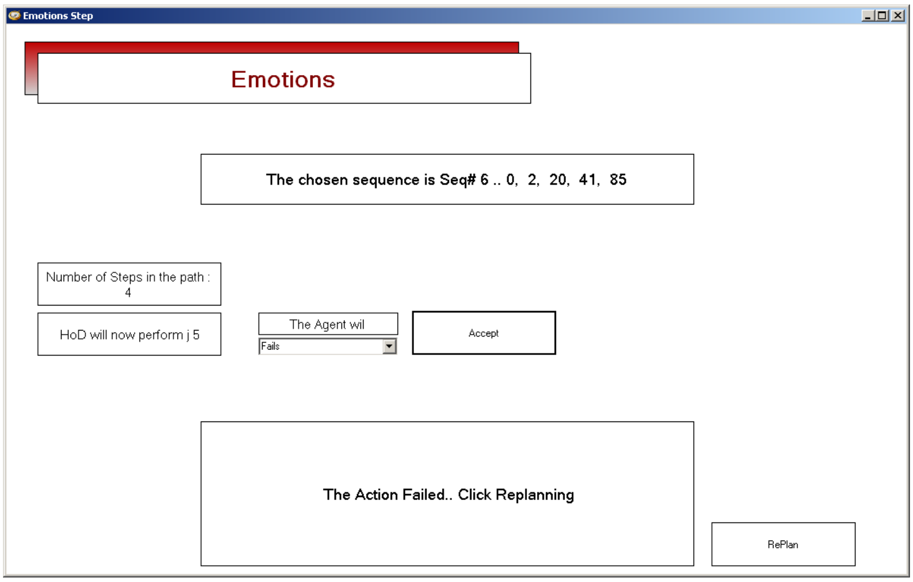
Safety, Opportunity and Threat Perspective

```
Seq# 1 Safety(-10) Opportunity ( 496) Threat ( 80)
Seq# 2 Safety(-20) Opportunity ( 698) Threat ( 80)
Seq# 3 Safety(-10) Opportunity ( 496) Threat ( 70)
Seq# 4 Safety(-10) Opportunity ( 496) Threat ( 80)
Seq# 5 Safety(-10) Opportunity ( 696) Threat ( 60)
Seq# 6 Safety(-20) Opportunity ( 698) Threat ( 50)
Seq# 7 Safety(-10) Opportunity ( 696) Threat ( 50)
Seq# 8 Safety(-20) Opportunity ( 696) Threat ( 80)
Seq# 9 Safety(-20) Opportunity ( 698) Threat ( 60)
```

```
Cooperative Scenarios
-----------------------------------------------
j0 :  HoD Sends S1 to a conference, S1 attends
j1 :  HoD Sends S2 to a conference, S2 attends
j2 :  HoD Sends S3 to a conference, S3 attends

j3 :  HoD Asks S1 to Write a Paper, S1 Writes
j4 :  HoD Asks S2 to Write a Paper, S2 Writes
j5 :  HoD Asks S3 to Write a Paper, S3 Writes


Non-Cooperative Scenarios
-----------------------------------------------
j6 :  HoD Sends S1 to a conference, S1 Doesnt
j7 :  HoD Sends S2 to a conference, S2 Doesnt
j8 :  HoD Sends S3 to a conference, S3 Doesnt

j11 :  HoD Asks S1 to Write a Paper, S1 Doesnt
j12 :  HoD Asks S2 to Write a Paper, S2 Doesnt
j13 :  HoD Asks S3 to Write a Paper, S3 Doesnt
```

```
Q,Budget,Attend,Papers,Exper
 0-- 3- 0 0 0- 0 1 0- 0 0 1
 1-- 3- 0 0 0- 1 1 0- 0 0 1
 2-- 3- 0 0 0- 0 1 1- 0 0 1
 3-- 2- 1 0 0- 0 1 0- 1 0 1
 4-- 2- 0 0 0- 0 1 0- 0 0 1
 5-- 1- 0 1 0- 0 1 0- 0 1 1
 6-- 1- 0 0 0- 0 1 0- 0 0 1
 7-- 2- 0 0 1- 0 1 0- 0 0 1
 8-- 2- 0 0 0- 0 1 0- 0 0 0
 9-- 3- 0 0 0- 1 1 1- 0 0 1
 10-- 2- 1 0 0- 1 1 0- 1 0 1
 11-- 2- 0 0 0- 1 1 0- 0 0 1
 12-- 1- 0 1 0- 1 1 0- 0 1 1
 13-- 1- 0 0 0- 1 1 0- 0 0 1
 14-- 2- 0 0 1- 1 1 0- 0 0 1
 15-- 2- 0 0 0- 1 1 0- 0 0 0
 16-- 2- 1 0 0- 0 1 1- 1 0 1
 17-- 2- 0 0 0- 0 1 1- 0 0 1
 18-- 1- 0 1 0- 0 1 1- 0 1 1
```

Figure A.10: The Final Results



The chosen sequence is Seq# 6 .. 0, 2, 20, 41, 85

Number of Steps in the path : 4

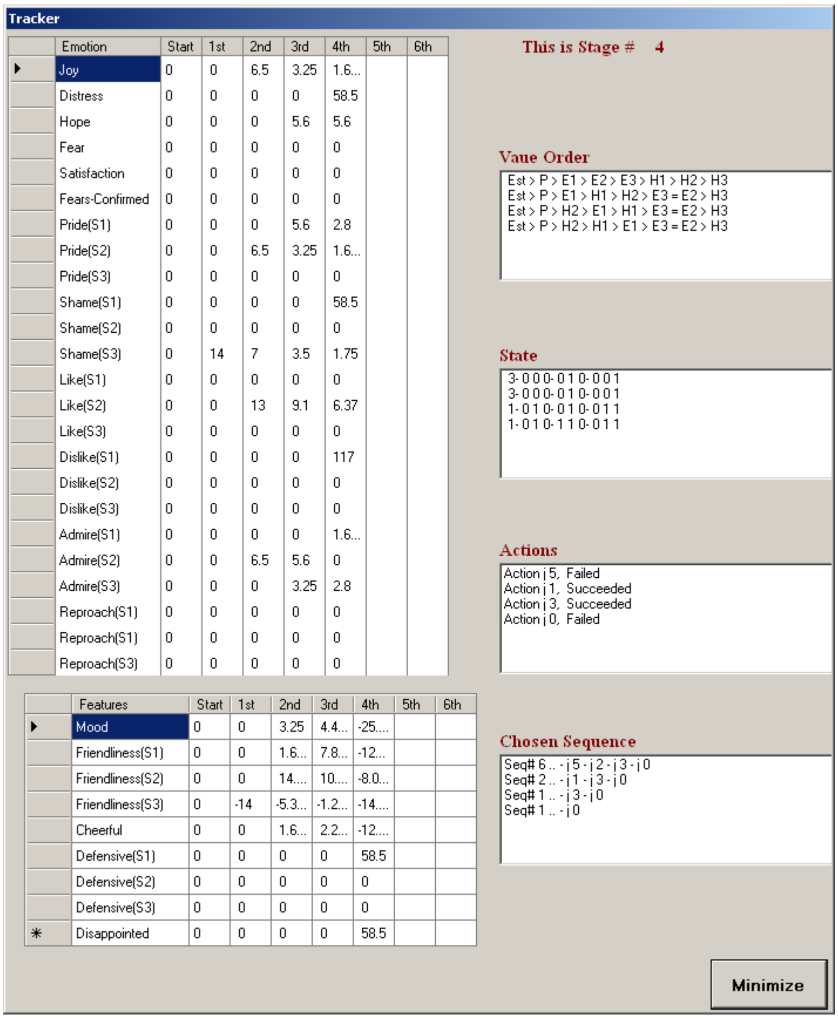
HoD will now perform j 5

The Agent wil: Fails

The Action Failed.. Click Replanning

Figure A.11: The Emotional model

Figure A.12: The tracker collapsed



	Emotion	Start	1st	2nd	3rd	4th	5th	6th
▶	Joy	0	0	6.5	3.25	1.6...		
	Distress	0	0	0	0	58.5		
	Hope	0	0	0	5.6	5.6		
	Fear	0	0	0	0	0		
	Satisfaction	0	0	0	0	0		
	Fears-Confirmed	0	0	0	0	0		
	Pride(S1)	0	0	0	5.6	2.8		
	Pride(S2)	0	0	6.5	3.25	1.6...		
	Pride(S3)	0	0	0	0	0		
	Shame(S1)	0	0	0	0	58.5		
	Shame(S2)	0	0	0	0	0		
	Shame(S3)	0	14	7	3.5	1.75		
	Like(S1)	0	0	0	0	0		
	Like(S2)	0	0	13	9.1	6.37		
	Like(S3)	0	0	0	0	0		
	Dislike(S1)	0	0	0	0	117		
	Dislike(S2)	0	0	0	0	0		
	Dislike(S3)	0	0	0	0	0		
	Admire(S1)	0	0	0	0	1.6...		
	Admire(S2)	0	0	6.5	5.6	0		
	Admire(S3)	0	0	0	3.25	2.8		
	Reproach(S1)	0	0	0	0	0		
	Reproach(S1)	0	0	0	0	0		
	Reproach(S3)	0	0	0	0	0		

	Features	Start	1st	2nd	3rd	4th	5th	6th
▶	Mood	0	0	3.25	4.4...	-25....		
	Friendliness(S1)	0	0	1.6...	7.8...	-12...		
	Friendliness(S2)	0	0	14....	10....	-8.0...		
	Friendliness(S3)	0	-14	-5.3...	-1.2...	-14....		
	Cheerful	0	0	1.6...	2.2...	-12....		
	Defensive(S1)	0	0	0	0	58.5		
	Defensive(S2)	0	0	0	0	0		
	Defensive(S3)	0	0	0	0	0		
*	Disappointed	0	0	0	0	58.5		

This is Stage # 4

Vaue Order

```
Est > P > E1 > E2 > E3 > H1 > H2 > H3
Est > P > E1 > H1 > H2 > E3 = E2 > H3
Est > P > H2 > E1 > H1 > E3 = E2 > H3
Est > P > H2 > H1 > E1 > E3 = E2 > H3
```

State

```
3-0 0 0-0 1 0-0 0 1
3-0 0 0-0 1 0-0 0 1
1-0 1 0-0 1 0-0 1 1
1-0 1 0-1 1 0-0 1 1
```

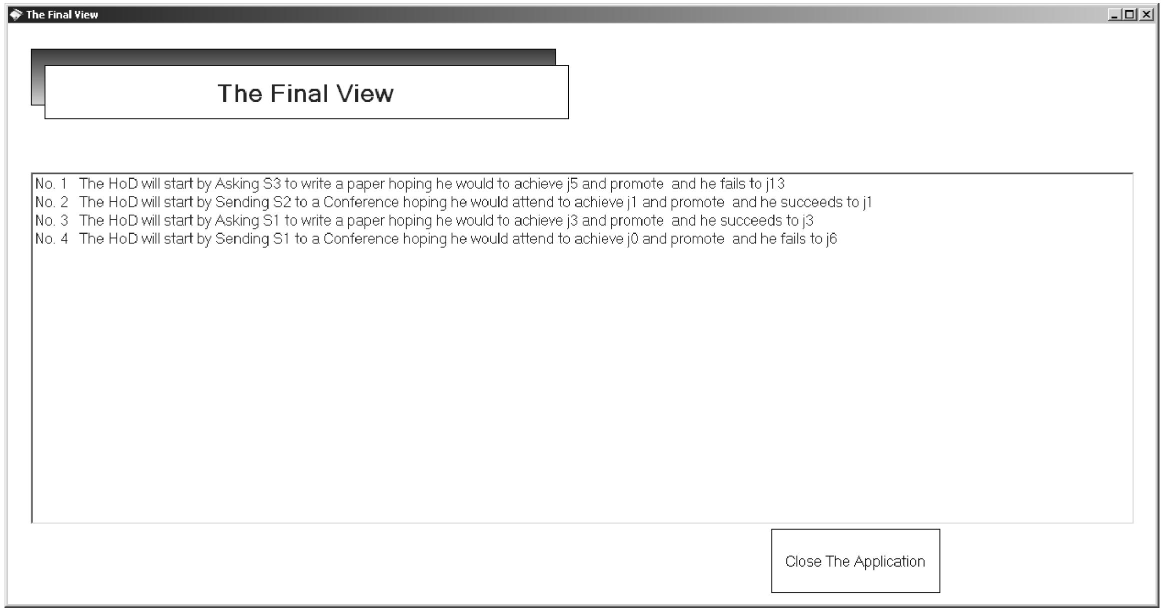
Actions

```
Action ¡ 5,  Failed
Action ¡ 1,  Succeeded
Action ¡ 3,  Succeeded
Action ¡ 0,  Failed
```

Chosen Sequence

```
Seq# 6 .. - ¡ 5 - ¡ 2 - ¡ 3 - ¡ 0
Seq# 2 .. - ¡ 1 - ¡ 3 - ¡ 0
Seq# 1 .. - ¡ 3 - ¡ 0
Seq# 1 .. - ¡ 0
```

Figure A.13: The tracker expanded

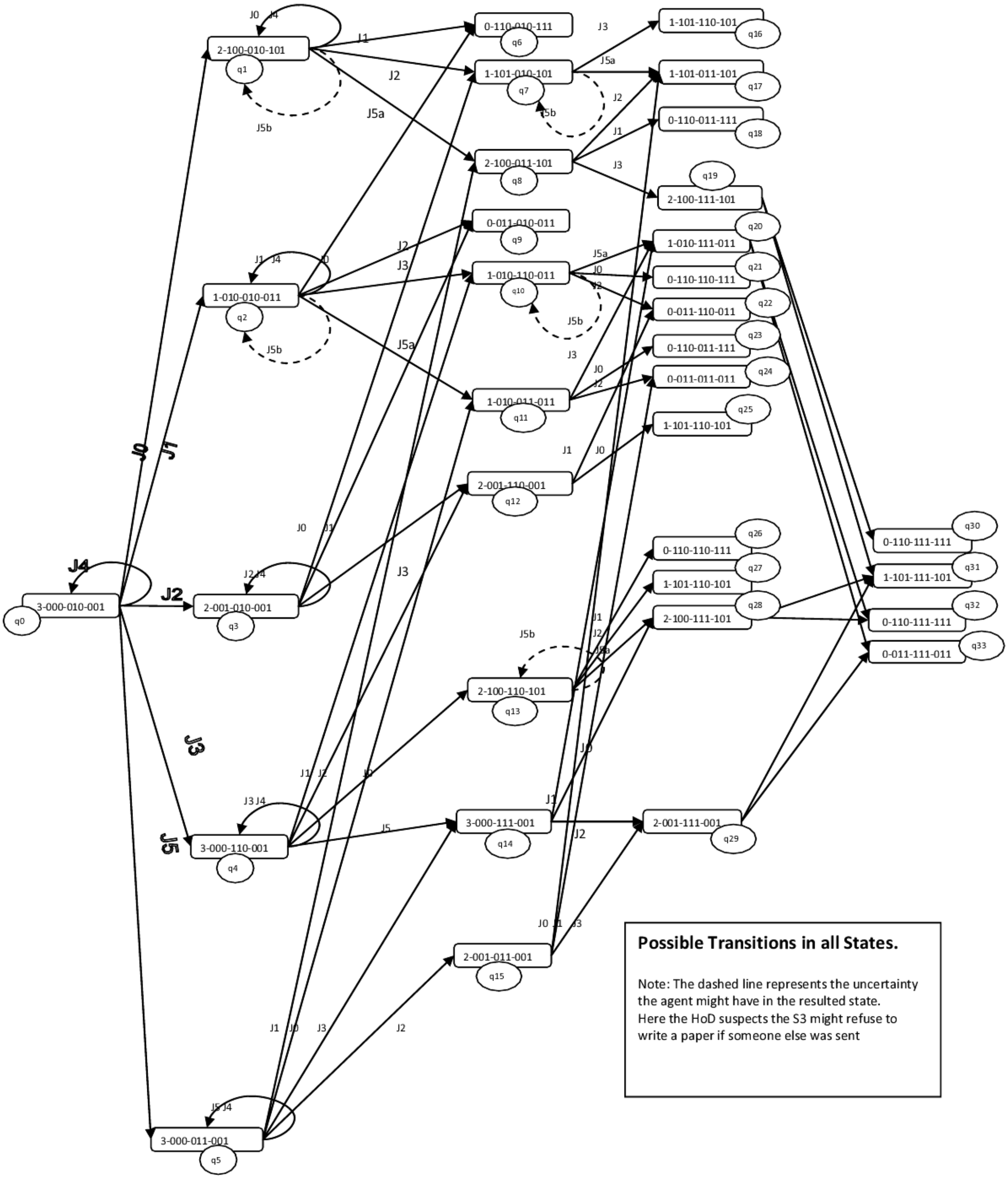


```
No. 1   The HoD will start by Asking S3 to write a paper hoping he would to achieve j5 and promote  and he fails to j13
No. 2   The HoD will start by Sending S2 to a Conference hoping he would attend to achieve j1 and promote  and he succeeds to j1
No. 3   The HoD will start by Asking S1 to write a paper hoping he would to achieve j3 and promote  and he succeeds to j3
No. 4   The HoD will start by Sending S1 to a Conference hoping he would attend to achieve j0 and promote  and he fails to j6
```

Figure A.14: The Final Results

Appendix B

A complete view of the AATS



Possible Transitions in all States.

Note: The dashed line represents the uncertainty the agent might have in the resulted state. Here the HoD suspects the S3 might refuse to write a paper if someone else was sent

Figure B.1: A Complete view of the AATS of the example in Chapter 4

Appendix C

Emotion Types

This appendix will describe each emotion type of the OCC emotions model [93] in more details. Ortony, Clore and Collins (OCC) [93] addresses the structure of emotions and how they are related and identifies twenty-two emotions organised into a hierarchy. The OCC model also provides a specification of the conditions which give rise to each emotion in terms of the notions of objects, events and actions. The OCC model includes notions of intensity of emotions, and identifies a number of variables that influence the intensity of the emotions experienced.

Tokens represents other emotions and naming conventions that are covered inside this emotion type. Variables are the different elements that influence the generation and intensity of these emotions, which are all aligned to decision-making aspects and in particular the model this thesis presents.

The OCC hierarchy consists of three main branches that represent how an agent perceives the world, which are (Consequences of events, Actions of agents and Aspects of objects). Under consequences of events, emotions then are divided into either focusing on self such as (Hope, Fear, Joy and Distress, ...etc) or focusing on others (Gloating, Pity, Resentment,...etc). The second branch is Actions of agents, which are emotions related to approving and disapproving of an agent's actions such as Pride and Shame or other actions such as Admiration and Reproach. The third branch (Aspects of objects) relates to attraction towards objects in a generic form and includes the emotions of love and hate.

joy and distress Pleased/Displeased about a desirable/undesirable event.

> Tokens: Contented, cheerful, delighted, escatic/depressed, distressed, displeased.

Variables :

1. The degree to which the event is desirable.

These emotions relate to an important event happening, which is either desirable and would cause “Joy” or undesirable to cause “Distress”. This is mapped to the agent’s “Values” in VAF. Whenever a goal that is of importance to the agent succeeds or fails, one of these emotions will occur with an intensity equal to the level of importance of that goal.

hope and fear Pleased/Displeased about the prospect of a desirable event.

Tokens: anticipation, excitement/anxious, dread, fright.

Variables:

1. The level of desirability of this event
2. The level of likelihood of this event occurring

These are the probabilistic emotions where they are instantiated with the probabilities of success of a “Goal” increasing or decreasing. Whenever the probability of a goal increases the emotion “Hope” will be generated with an intensity that is relevant to the change in probability and the desirability of the event. Whenever the probability of a goal decreases the emotion “Fear” will be generated with an intensity that is relevant to the change in probability and the desirability of the event.

satisfaction and fears-confirmed Pleased about the confirmation of a prospect of a desirable event/Displeased about the confirmation of the prospect of an undesirable event.

Tokens: hopes realise/fears realised

Variables :

1. The intensity of Hope/Fear emotions
2. The level of desirability of this event

These relate to an event happening similar to (Joy/Distress) with the difference that they have a precondition of the existence of (Hope or fear) for them to occur. Whenever an agent achieves a goal that he was initially hopeful to achieve the emotion "Satisfaction" will occur. On the other hand whenever an agent fails at achieving a goal that he was initially fearful of achieving the emotion "Fears-Confirmed" will occur.

relief and disappointment Pleased about the disconfirmation of a prospect of an undesirable event/Displeased about the disconfirmation of the prospect of a desirable event.

Variables:

1. The intensity of Hope/Fear emotions
2. The level of desirability of this event.

These are similar to (Satisfaction and Fears confirmed) with the difference that they relate to undesirable events. Whenever a goal is achieved that the agent was initially fearful about the emotion "Relief" will occur. Whenever a goal fails that the agent was initially hopeful of achieving the emotion "Disappointment" will occur.

Similar to (Satisfaction and Fears-Confirmed) the intensity is the degree of the Hope and Fear emotions and the desirability of this goal.

happy-for and pity Pleased/Displeased about an event that is presumed to be desirable/Undesirable for someone else.

Tokens: Delighted-For, pleased-for,...etc/compassion, pity, sad-for, sorry-for, sympathy.

Variables:

1. The level of desirability of this event
2. The level of liking toward the other agent

These emotions relate to an event that took place regardless of the actions that caused them. This is mapped to "Goals" in the decision-making methodology. These emotions also have a precondition that another emotion "Like" must also exist toward the other agent.

Whenever a goal succeeds that is relevant to another agent where Like toward this agent is not 0, the emotion of Happy-for will be generated and the intensity of this emotion will depend on how liked the other agent is (the intensity of Like) and how desirable this goal is (the Importance of the goal). Whenever a goal fails that is relevant to another agent where Like toward this agent is not 0, the emotion of Pity will be generated and the intensity of this emotion will depend on how liked the other agent is (the intensity of Like) and how desirable this goal is (the Importance of the goal).

gloating and resentment Pleased about an undesirable event for someone else / Displeased about a desirable event for someone else.

Tokens: Schadenfreude/envy, jealousy.

Variables:

1. The level of desirability of this event
2. The level of disliking toward the other agent

These emotions are the opposites of (Happy-For/Pity) where an event has occurred to someone else, but on the other hand, this someone is not liked. These emotions have a precondition of another emotion "Dislike" of being there.

If a disliked agent failed in achieving a certain goal, the emotion "Goating" will occur with an intensity that is equal to how disliked the agent is and the importance of this event. If a disliked agent succeeded in achieving a certain goal, the emotion "Resentment" will occur with an intensity that is equal to how disliked the agent is and the importance of this event.

pride and shame Approving / Disapproving of one's own action

Tokens: pride/embarrassment, guilt

Variables :

1. The degree of judged praiseworthiness
2. The expectedness of this action succeeding

These emotions do not relate to events, but to actions. It has a link with the expectedness of these actions succeeding or failing. These emotions related to the “Values” in VAF where a desirable action is the one that promotes a value, an undesirable action is the one that demotes one.

When an action that has low expectation succeeds and promotes a value, the emotion of pride will be generated with an intensity that has the components of expectation and importance of this event. When an action that has low expectation succeeds and demotes a value the emotion of shame will be generated with an intensity that has the components of expectation and importance of this event.

admiration and reproach Approving/Disapproving of someone else’s action.

Tokens: appreciation/contempt, despise

Variables:

1. The degree of judged praiseworthiness.
2. The expectedness of this action succeeding

These emotions are the same as (Pride and Shame) but with relation to a different agent’s actions. When an action that has low expectation succeeds and promotes a value the emotion of “Admiration” will be generated with an intensity that has the components of expectation and importance of this event. When an action that has low expectation succeeds and demotes a value the emotion of “Reproach” will be generated with an intensity that has the components of expectation and importance of this event.

like and dislike Liking/Disliking an appealing object.

Tokens: adore, love/hate, disgust

In this study, an assumption is made that the appealingness of an agent increases and decreases along with the values being promoted and demoted by it.

So, If another agent performs an action that promoted some value, the value of “Like” will increase. When the action demotes a value “Dislike” will increase. The intensity is dependent on the importance of the event and the expectedness.

gratification and remorse Approving/Disapproving of one's own action and being please/displeased about the related event

Tokens: self-satisfaction/self-anger

Variables:

1. The degree of judged praiseworthiness
2. The expectedness of this action succeeding
3. The importance of the event

These are the same as (Pride and Shame) with the difference that that the events involved are also desirable or undesirable not only action, this relates in VAF to both Values and Goals. We can see from the variables above that it is the same with the addition of the degree of importance of the event.

When another agent performs an action that resulted in a desirable event, first "Pride" will be generated. If the goal had importance to the emotional agent itself it will also raise "Joy" and the intensity will be the addition of Joy and Pride. When another agent performs an action that resulted in an undesirable event first "Shame" will be generated. If the goal had importance to the emotional agent itself it will also raise "Distress" and the intensity will be the addition of Distress and Shame.

gratitude and displeasure Approving/Disapproving of someone else's action being pleased/displeased about the related event

Tokens: appreciation, thankful / anger, fury, rage

Variables:

1. The degree of judged praiseworthiness
2. The expectedness of this action succeeding
3. The importance of the event

These are the same as (Admiration and Reproach) with the difference that the events involved are also desirable or undesirable not only action, this relates in VAF to both Values and Goals. We can see from the variables above that it is the same with the addition of the degree of importance of the event.

When another agent performs an action that resulted in a desirable event, first “Admiration” will be generated. If the goal had importance to the emotional agent itself, it will also raise “Joy” and the intensity will be the addition of Joy and Admiration. When another agent performs an action that resulted in an undesirable event, first “Reproach” will be generated. If the goal had importance to the emotional agent itself, it will also raise “Distress” and the intensity will be the addition of Distress and Reproach.

Bibliography

[1] Rajeev Alur, Thomas A. Henzinger, and Orna Kupferman. Alternating-time temporal logic. *Journal of the Association for Computing Machinery (ACM)*, 49(5):672–713, 2002.

[2] Leila Amgoud. Using preferences to select acceptable arguments. In *European Conference on Artificial Intelligence*, pages 43–44, 1998.

[3] Leila Amgoud and Claudette Cayrol. On the acceptability of arguments in preference-based argumentation. In *UAI*, pages 1–7, 1998.

[4] Leila Amgoud and Claudette Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *J. Autom. Reasoning*, 29(2):125–169, 2002.

[5] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.

[6] Leila Amgoud, Claudette Cayrol, and Marie-Christine Lagasquie-Schiex. On the bipolarity in argumentation frameworks. In James P. Delgrande and Torsten Schaub, editors, *Non-monotonic Reasoning Workshop (NMR)*, pages 1–9, 2004.

[7] Leila Amgoud and Nabil Hameurlain. A formal model for designing dialogue strategies. In *International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 414–416. Association for Computing Machinery (ACM), 2006.

[8] Leila Amgoud and Henri Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 173(3-4):413–436, 2009.

[9] Leila Amgoud and Srdjan Vesic. Repairing preference-based argumentation frameworks. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 665–670, 2009.

[10] Stig K. Andersen. Judea pearl, probabilistic reasoning in intelligent systems: Networks of plausible inference. *Artificial Intelligence*, 48(1):117–124, 1991.

[11] Aristotle. *The Nicomachean Ethics.* Oxford University Press, 1998.

[12] Katie Atkinson. *What should we do? Computational representation of persuasive argument in practical reasoning.* PhD thesis, University of Liverpool, 2005.

[13] Katie Atkinson and Trevor Bench-Capon. Addressing moral problems through practical reasoning. In *nternational Workshop on Deontic Logic in Computer Science (DEON)*, pages 8–23, 2006.

[14] Katie Atkinson and Trevor Bench-Capon. Action-based alternating transition systems for arguments about action. In *Association for the Advancement of Artificial Intelligence (AAAI)*, pages 24–29, 2007.

[15] Katie Atkinson and Trevor Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171(10-15):855–874, 2007.

[16] Katie Atkinson and Trevor Bench-Capon. Addressing moral problems through practical reasoning. *J. Applied Logic*, 6(2):135–151, 2008.

[17] Katie Atkinson and Trevor Bench-Capon. Action state semantics for practical reasoning. In *Association for the Advancement of Artificial Intelligence (AAAI) Fall Symposium. Technical Report SS-09-06*, pages 8–13, 2009.

[18] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Arguing about cases as practical reasoning. In *International Conference on Artificial Intelligence and Law (ICAIL)*, pages 35–44, 2005.

[19] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. A dialogue game protocol for multi-agent argument over proposals for action. *Autonomous Agents and Multi-Agent Systems*, 11(2):153–171, 2005.

[20] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Generating intentions through argumentation. In *International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1261–1262, 2005.

[21] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Multi agent argumentation for edemocracy. In *European Workshop on Multi-Agent Systems (EUMAS)*, pages 35–46, 2005.

[22] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Computational representation of practical argument. *Synthese*, 152(2):157–206, 2006.

[23] Katie Atkinson, Trevor Bench-Capon, and Sanjay Modgil. Argumentation for decision support. In *International Conference on Database and Expert Systems Applications (DEXA)*, pages 822–831, 2006.

[24] Tom Bailey. On trust and philosophy. Technical report, Department of Philosophy, University of Warwick, 2002.

[25] Joseph Bates. The role of emotion in believable agents. *Communcation of the Association for Computing Machinery (ACM)*, 37(7):122–125, 1994.

[26] Joseph Bates, A. Bryan Loyall, and Scott Reilly. An architecture for action, emotion, and social behavior. In *MAAMAW*, pages 55–68, 1992.

[27] Trevor Bench-Capon. The ideal audience and artificial intelligence and law. In *International Conference on Database and Expert Systems Applications (DEXA) Workshop*, pages 763–767, 2001.

[28] Trevor Bench-Capon. Value based argumentation frameworks. In *Non-monotonic Reasoning Workshop (NMR)*, pages 443–454, 2002.

[29] Trevor Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[30] Trevor Bench-Capon, Sylvie Doutre, and Paul E. Dunne. Audiences in argumentation frameworks. *Artificial Intelligence*, 171(1):42–71, 2007.

[31] Trevor Bench-Capon and Paul E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007.

[32] Trevor Bench-Capon and Henry Prakken. Justifying actions by accruing arguments. In *Computational Model of Arguments*, pages 247–258, 2006.

[33] Philippe Besnard and Anthony Hunter. *Elements of Argumentation.* MIT, 2008.

[34] Ken Binmore. *Fun and Games: A Text on Game Theory.* D. C. Heath and Company, 1992.

[35] Andrei Bondarenko, Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[36] Andrew Botterell. Trust. *Stanford Encyclopedia of Philosophy*, 2006.

[37] Michael Bratman. *Faces of Intention.* Cambridge Studies in Philosophy. Cambridge University Press, 1999.

[38] Michael Bratman. *Intentions, Plans and Practical Reason.* Harvard University Press, 2004.

[39] Michael Bratman. *Structures of agency.* Oxford University press, 2007.

[40] David Chapman. Planning for conjuctive goals. *Artificial Intelligence*, 32:333–377, 1987.

[41] Carlos Chesnevar, Jarred McGinnis, Sanjay Modgil, Iyad Rahwan, Chris Reed, Guillermo Simari, Matthew South, Gerard Vreeswijk, and Steven Willmott. Towards an argument interchange format. *The Knowledge Engineering Review*, 0:1–25, 2007.

[42] Alison Chorley, Trevor Bench-Capon, and Peter McBurney. Automating argumentation for deliberation in cases of conflict of interest. In *Computational Model of Arguments06*, pages 279–290, 2006.

[43] Philip Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 3:42, 1990.

[44] James Coleman. *Foundations of Social theory.* Harvard University press, 1990.

[45] Antonio Damasio. *Descartes' Error.* G P Putnams Sons, 1994.

[46] Mehdi Dastani and John-Jules Meyer. Programming agents with emotions. In *European Conference on Artificial Intelligence*, pages 215–219, 2006.

[47] Sylvie Doutre, Trevor Bench-Capon, and Paul E. Dunne. Determining preferences through argumentation. In *International Conference on Artificial Intelligence (AI*IA)*, pages 98–109, 2005.

[48] Sylvie Doutre, Trevor Bench-Capon, and Paul E. Dunne. Explaining preferences with argument positions. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1560–1561, 2005.

[49] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning and logic programming. In *Thirteenth International Joint Conference on Artificial Intelligence*, 1993.

[50] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[51] Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. Dialectic proof procedures for assumption-based, admissible argumentation. *Artificial Intelligence*, 170(2):114–159, 2006.

[52] Paul E. Dunne. Computational properties of argument systems satisfying graph-theoretic constraints. *Artificial Intelligence*, 171(10-15):701–729, 2007.

[53] Paul E. Dunne. Uncontested semantics for value-based argumentation, ulcs07013. Technical report, Department of Computer Science, The University of Liverpool, 2007.

[54] Paul E. Dunne and Trevor Bench-Capon. Complexity in value-based argument systems. In *The European Conference on Logics in Artificial Intelligence (JELIA)*, pages 360–371, 2004.

[55] Paul E. Dunne and Trevor Bench-Capon. Identifying audience preferences in legal and social domains. In *International Conference on Database and Expert Systems Applications (DEXA)*, pages 518–527, 2004.

[56] Paul Faulkner. An epistemic model of trust. In *5th Workshop on Deception, Fraud and Trust In Agent Societies*, 2002.

[57] Richard Fikes and Nils Nilsson. Strips a new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:198–208, 1971.

[58] Nico Frijda. *The Emotions.* Cambridge University Press, Cambridge, 1986.

[59] Nico Frijda. Emotions in robots. In H.L. Roitlab and J.A. Meyer, editors, *Comparative approaches to cognitive sciences*, pages 501–516, 1995.

[60] Nico Frijda. *The Laws of Emotion.* Lawrence erlbaum associates, 2007.

[61] Nico Frijda, Anthony Manstead, and Sacha Bem. *Emotions and beliefs.* Cambridge University press, 2006.

[62] Nico Frijda and Jaap Swagerman. Can computers feel? the theory and design of an emotional system. In *Cognition and emotion.* Lawrence Erlbaum Associates Limited, 1987.

[63] Dorian Gaertner and Francesca Toni. Computing arguments and attacks in assumption-based argumentation. *IEEE Intelligent Systems*, 22(6):24–33, 2007.

[64] Dorian Gaertner and Francesca Toni. Preferences and assumption-based argumentation for conflict-free normative agents. In *Argumentation in Multi-Agent Systems (ArgMAS)*, pages 94–113, 2007.

[65] Michael P. Georgeff and Amy L. Lansky. Reactive reasoning and planning. In *Association for the Advancement of Artificial Intelligence (AAAI)*, pages 677–682, 1987.

[66] Thomas F. Gordon, Henry Prakken, and Douglas Walton. The carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10-15):875–896, 2007.

[67] Charles Hamblin. *Imperatives.* Basil Blackwell, 1987.

[68] Richard Hare. *Freedom and Reason.* Oxford University Press, 1963.

[69] Richard Holton. Deciding to trust, coming to believe. *Australasian Journal of Philosophy*, 72:63–76, 1994.

[70] Ronald Howard and James Matheson. Influence diagrams. *Readings on the Principles and Application of Decision Analysis*, pages 719–762, 1984.

[71] Anthony Hunter. Making argumentation more believable. In *Association for the Advancement of Artificial Intelligence (AAAI)*, pages 269–274, 2004.

[72] Anthony Hunter. Towards higher impact argumentation. In *Association for the Advancement of Artificial Intelligence (AAAI)*, pages 275–280, 2004.

[73] Nicholas Jennings and Michael Wooldridge. *Agent Technology Foundation application and markets.* Springer, 1997.

[74] Hong Jiang and Jose M. Vidal. From rational to emotional agents. In *Association for the Advancement of Artificial Intelligence(AAAI) workshop on cognitive modeling and agent-based social simulations*, 2006.

[75] Dionysis Kalofonos, Nishan C. Karunatillake, Nicholas Jennings, Timothy J. Norman, Chris Reed, and Simon Wells. Building agents that plan and argue in a social context. In *Computational Model of Arguments06*, pages 15–26, 2006.

[76] Antony Kenny. *Practical Reasoning and Rational Appetite.* 1975.

[77] Sarit Kraus. Negotiation and cooperation in multi-agent environments. *Artificial Intelligence*, 94:79–97, 1997.

[78] Arthur Kuflik. Computers in control: Rational transfer of authority or irresponsible abdication of autonomy? *Ethics and Information Technology*, 1:173–184, 1999.

[79] Victor Lesser and Daniel Corkill. Functionally accurate distributed problem-solving systems. In *DARPA Workshop on Distributed Sensor Nets*, pages 21–26, "CMU, Pittsburgh", 1978.

[80] John-Jules Meyer. An introduction to agent technology. In *8th European Agent systems summer school (EASSS)*, 2006.

[81] John-Jules Meyer. Towards a quantitative model of emotions for intelligent agents. In *KI07 Workshop on Emotion and Computing - Current Research and Future Impact. Osnabruck, Germany*, 2007.

[82] Sanjay Modgil. Value based argumentation in hierarchical argumentation frameworks. In *Computational Model of Arguments06*, pages 297–308, 2006.

[83] Sanjay Modgil and Trevor Bench-Capon. Integrating object and meta-level value based argumentation. In *Computational Model of Arguments08*, pages 240–251, 2008.

[84] David Moffat and Nico Frijda. Where there's a will there's an agent. In *European Conference on Artificial Intelligence (ECAI) Workshop on Agent Theories, Architectures, and Languages*, pages 245–260, 1994.

[85] David Moffat, Nico Frijda, and Hans Phaf. *Analysis of a computer model of emotions.* IOS Press, 1993.

[86] Fahd Saud Nawwab, Trevor Bench-Capon, and Paul E. Dunne. A methodology for action-selection using value-based argumentation. In Philippe Besnard, Sylvie Doutre, and Anthony Hunter, editors, *Computational Models of Argument: Proceedings of Computational Model of Arguments 2008*, volume 172 of *Frontiers in Artificial Intelligence and Applications*, pages 264–275. IOS Press, 2008.

[87] Fahd Saud Nawwab, Trevor Bench-Capon, and Paul E. Dunne. Emotions in rational decision making. In Peter McBurney, Iyad Rahwan, Simon

Parsons, and Nicolas Maudet, editors, *Argumentation in Multi-Agent Systems*, volume 6057 of *Lecture Notes in Computer Science*, pages 273–291. Springer, 2010.

[88] Allen Newell, J. Shaw, and Herbert Simon. Report on a general problem-solving program. In *Internation conference on information processing*, pages 256–264, 1960.

[89] Søren Holbech Nielsen and Simon Parsons. Computing preferred extensions for argumentation systems with sets of attacking arguments. In *Computational Model of Arguments*, pages 97–108, 2006.

[90] Timothy Norman and Chris Reed. Group delegation and responsibility. In *International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 491–498, 2002.

[91] Timothy Norman and Chris Reed. A model of delegation for multi-agent systems. In *Foundations and Applications of Multi-Agent Systems*, pages 185–204, 2002.

[92] Timothy Norman and Chris Reed. A logic of delegation. *Artificial Intelligence*, 174(1):51–71, 2010.

[93] Andrew Ortony, Gerald Clore, and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

[94] Lin Padgham and Guy Taylor. A system for modelling agents having emotion and personality. In Lawrence Cavedon, Anand S. Rao, and Wayne Wobcke, editors, *The Pacific Rim International Conferences on Artificial Intelligence (PRICAI) Workshop on Intelligent Agent Systems*, volume 1209 of *Lecture Notes in Computer Science*, pages 59–71. Springer, 1996.

[95] Simon Parsons, Carles Sierra, and Nicholas R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.

[96] Simon Parsons and Michael Wooldridge. Game theory and decision theory in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 5(3):243–254, 2002.

[97] Chaim Perelman and L. Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press, Notre Dame, 1969.

[98] Plato. *The Republic.* Oxford World Classics, 2007.

[99] Iyad Rahwan and Leila Amgoud. An argumentation based approach for practical reasoning. In *International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 347–354, 2006.

[100] Iyad Rahwan and Bita Banihashemi. Arguments in owl: A progress report. In *Computational Model of Arguments08*, pages 297–310, 2008.

[101] Sarvapalid D. Ramchurn, Dong Huynh, and Nicholas R. Jennings. Trust in multi-agent systems. *The Knowledge Engineering Review*, 19:1–25, 2004.

[102] Anand S. Rao and Michael P. Georgeff. Modeling rational agents within a BDI-Architecture. In *International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 473–484, 1991.

[103] Joseph Raz, editor. *Practical Reasoning.* Oxford University press, 1978.

[104] Chris Reed and Douglas Walton. Towards a formal and implemented model of argumentation schemes in agent communication. In *Argumentation in Multi-Agent Systems (ArgMAS)*, pages 19–30, 2004.

[105] Scott Reilly. Building emotional characters for interactive drama. In *Association for the Advancement of Artificial Intelligence (AAAI)*, page 1487, 1994.

[106] Scott Reilly. The art of creating emotional and robust interactive characters. Technical report, International Joint Conference on Artificial Intelligence (IJCAI) workshop on AI art and entertainment, 1995.

[107] Scott Reilly. *Believable social and emotional agents.* PhD thesis, Carnegie Mellon University (CMU), 1996.

[108] Scott Reilly. A methodology for building believable social agents. In *Agents*, pages 114–121, 1997.

[109] Scott Reilly and Joseph Bates. Building emotional agents. Technical report, Computer Science Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1992. Available at http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/oz/web/papers/CMU-CS-92-143.ps.

[110] Scott Reilly and Joseph Bates. Natural negotiation for believable agents. Technical report, Department of Computer Science, Carnegie Mellon University, 1995.

[111] Jeffrey Rosenchein. *Cooperation among Intelligent Agents.* PhD thesis, Stanford University, 1985.

[112] Jeffrey Rosenschein and Gilad Zlotkin. *Rules of encounter: designing conventions for automated negotiation among computers.* MIT Press, Cambridge, MA, USA, 1994.

[113] Stuart Russel and Peter Norvig. *Artificial Intelligence a modern approach.* Prentice hall, 1995.

[114] Ed Sacerdoti. *A structure for plans and behavior.* American Elsevier, 1975.

[115] Tuomas W. Sandholm. Distributed rational decision making. pages 201–258, 1999.

[116] John R. Searle. *Rationality in Action.* The MIT Press, Cambridge, Massachusetts, 2003.

[117] Herbert Simon. *The Sciences of the Artificial.* MIT press, 1996.

[118] Bas R. Steunebrink, Mehdi Dastani, and John-Jules Meyer. Emotions as heuristics for rational agents. Technical Report UU-CS-2007-006, Department of Information and Computing Sciences, Utrecht University, 2007.

[119] Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer. A logic of emotions for intelligent agents. In *Association for the Advancement of Artificial Intelligence (AAAI)*, pages 142–147, 2007.

[120] Gerald Sussman. *A Computational model of skill acquisition.* American Elsevier, 1975.

[121] Pancho Tolchinsky, Sanjay Modgil, Ulises Cortés, and Miquel Sànchez-Marrè. CBR and argument schemes for collaborative decision making. In *Computational Model of Arguments*, pages 71–82, 2006.

[122] Francesca Toni. Assumption-based argumentation for closed and consistent defeasible reasoning. In *The Japanese Society for Artificial Intelligence (JSAI)*, pages 390–402, 2007.

[123] Douglas Walton. *Argumentation Schemes for Presumptive Reasoning.* Lawrence Erlbaum Associates, Mahwah, New Jersey, 1996.

[124] Douglas Walton. *The New Dialectic: Conversational Contexts of Argument.* University of Toronto Press, Toronto, 1998.

[125] Gerhard Weiss. *Multiagent Systems a Modern Approach to distributed artificial intelligence.* MIT Press, 1999.

[126] Steven Willmott, Gerard Vreeswijk, Carlos Chesnevar, Matthew South, Jarred McGinnis, Sanjay Modgil, Iyad Rahwan, Chris Reed, and Guillermo Simari. Towards an argument interchange format for multi-agent systems. *Argumentation in Multi-Agent Systems (ArgMAS)*, 1, 2006.

[127] Michael Wooldridge. A BDI logic of planning agents. Submitted, 1995.

[128] Michael Wooldridge. Intelligent agents: The key concepts. In *Multi-Agent-Systems and Applications*, pages 3–43, 2001.

[129] Michael Wooldridge. Reasoning about rational agents. *J. Artificial Societies and Social Simulation*, 5(1), 2002.

[130] Michael Wooldridge. *An introduction to multiagent systems.* John wiley and sons, 2005.

[131] Michael Wooldridge and Paul E. Dunne. On the computational complexity of coalitional resource games. *Artificial Intelligence*, 170(10):835–871, 2006.

[132] Michael Wooldridge and Wiebe van der Hoek. On obligations and normative ability: Towards a logical analysis of the social contract. *J. Applied Logic*, 3(3-4):396–420, 2005.