957
2006
002

The examination of an information-based approach to trust

Master's thesis

Maaike Harbers
Student number: S1217763

May 2006

External advisor: Prof. Dr. C. Sierra (IIIA)
Internal advisor: Dr. L.C. Verbrugge (AI RuG)

Artificial Intelligence
Rijksuniversiteit Groningen

Foreword

About a year ago, I had two desires about the final project for finishing my study Artificial Intelligence: I wanted to do something with multi-agent systems and I wanted to do this abroad. The first wish made me ask Rineke Verbrugge as internal advisor. She brought me into contact with Carles Sierra, which provided the opportunity to perform the project in Barcelona and so fulfil my second desire. After that, the subject of the project was easily picked. My research would be about trust and reputation in multi-agent systems and in September 2005 I was ready to go.

Spanish people have the name to be good cooks, to have a lot of parties and to enjoy life. This reputation gave me a lot of confidence that I would have a good time in Spain. On the other hand however, Spaniards are also known for not being too fast and efficient when something has to be arranged. Despite some little doubts about this mañana mañana culture, I really looked forward to go to Spain and to start with the project. Full of trust I took a plain to Barcelona…

In the meantime I can say that my trust was not groundless. I did an interesting project, enjoyed working at the research institute, met a lot of nice people and Barcelona was great. As will become clear in this thesis, after having new experiences or information an old opinion can be updated. My experiences can only confirm a positive view of Spanish people and working and living in Spain.

I want to thank the people that contributed to this thesis. First of all, my gratitude goes to Carles Sierra and Rineke Verbrugge. Carles thanks for your hospitality and enthusiasm, Rineke thanks for your conscientiousness and involvement, and both thanks for the useful comments and the nice cooperation. I would like to thank Jordi Sabater for his help (and patience!) with the ART test-bed. Further, I owe thanks to the research institute of artificial intelligence (IIIA) in Barcelona for providing me the facilities to perform the project. Finally, I want to thank my parents, my boyfriend Joost and many other friends for listening to me and giving support.

Contents

1 Introduction

This chapter will introduce the subject of this project. In the second section the research question will be mentioned and explained. Finally, an overview about the way this thesis is built up will be given.

1.1 Motivation

Negotiation is a process in which a group of negotiation partners tries to reach a mutually acceptable agreement on some matter by communication. It constantly takes place: people negotiate about big deals of millions of dollars, but also about smaller matters like what to eat for dinner. Besides humans, software agents and robots also negotiate. Negotiation plays an important role in multi-agent systems, in which it might even be the most fundamental and powerful form of interaction between different agents. Agents in a multi-agent system are autonomous, so they have no direct control over other agents and must negotiate in order to control their interdependencies.

In negotiations, one tries to obtain a profitable outcome. But what is a profitable outcome: pay little money for many goods of high quality? Although it seems to be a good deal, this might not always be the most profitable outcome. If negotiation partners will meet again in the future, it could be more rational to focus on the relationship with your negotiation partners, to make them trust you and to build up a good reputation.

If we take the future into account, another question arises: how will the opponent behave in the future? In the context of negotiations, agents have to make decisions about the acceptability of a deal. One of the determining factors in these considerations is the agent's opinion on the probability that the bargains made in a deal will be really accomplished after accepting the deal. Will the other agent deliver products of good quality? Will they be delivered on time, too late or maybe even not at all? Beforehand, an agent cannot know for sure whether the negotiation partner will fulfil his promises or not, so the agent has to deal with uncertain information. The modelling of trust and could help to make good predictions about the future.

This thesis will discuss the computational modelling of trust and reputation, an investigation topic receiving a lot of attention in the field of distributed artificial intelligence lately. The thesis will especially focus on a new way to deal with these topics, based on the information-based model for trust introduced by Sierra and Debenham (2005). In this thesis, their information-based approach will be discussed and tested. Further, their model of trust will be extended with algorithms for dealing with reputation information and social information.

1.2 Research question

The main question of this project will be the following:

> *Is the information-based approach a good way to deal with trust and reputation in multi-agent systems?*

In order answer this question the project is divided into two main parts. The first part will be a theoretical discussion of Sierra and Debenham's information-based model for trust, in which extra attention will be paid to the modelling of reputation and the role of social information. The second part will be more practical, the model will be tested by implementing an information-based agent and performing experiments with it. Concretely, the graduate project will consist of the following two tasks:

- Investigate how Sierra and Debenham's information-based model for trust could be extended with a more sophisticated way to deal with the influence of reputation and social information.
- Implement a negotiation agent making use of Sierra and Debenham's model of trust and test it with the Agent Reputation and Trust (ART) test-bed.

By the execution of these tasks, the model is examined in a theoretical and in a practical way. The results of the two parts together, should help in giving a founded answer to the research question of the project.

1.3 Structure of the thesis

The thesis starts with a theoretical discussion of Sierra and Debenham's information-based model of trust. First a general overview of the research in computational trust and reputation models is given (chapter 2), then the information-based model itself will be introduced (chapter 3). In the following two chapters, possible ways to extend the model with more sophisticated ways to deal with reputation (chapter 4) and social information (chapter 5) will be proposed.

The description of the practical part starts with the introduction of the ART test-bed (chapter 6), the test-bed that will be used for the experiments. Then the translation of the information-based model to a test-bed agent will be discussed (chapter 7). The next chapter (chapter 8) will describe the experiments with the test-bed, followed by a discussion (chapter 9), the conclusions and some suggestions for further research (chapter 10).

2 Computational models of trust and reputation

In this chapter an introduction will be given to computational models of trust and reputation. Several possible design choices will be discussed. Extra attention will be paid to the meaning of trust and reputation and the relation between these two concepts. Finally, three examples of existing models will be given.

2.1 What is a model of trust and reputation?

In computer science and especially in the area of distributed artificial intelligence, many models of trust and reputation have been developed over the last years. This relatively young field of research is still rapidly growing and gaining popularity. What exactly is a computational model of trust and reputation? And why are these models getting so much attention lately?

In multi-agent systems information is distributed among different parts of the system, and the different entities of the system, agents, are having interactions with each other. From the point of view of a single agent, it has to interact with other agents in a constantly changing environment. The agent has to make all kinds of decisions, for example the agents with which it will interact and the way to treat them. The agent does not know how other agents will behave in the future, so it has to make these choices based on uncertain information. The aim of trust and reputation models in these kinds of systems is to support the decision making in these kind of uncertain situations. A computational model of trust or reputation derives trust or reputation values from the agent's past interactions with its environment and possible extra information. These trust or reputation values influence the agent's decision making process, in order to facilitate the dealing with uncertain information. For example, if two agents offer the product the agent needs for the same price, this agent could choose to commit itself to the one with the highest reputation in delivering products of good quality.

Applications of computational trust and reputation systems are mainly found in electronic markets. In comparison to face-to-face negotiation, trading partners in electronic markets often have less information about each other's reliability or the product quality during the transaction. A trust or reputation system gives different parties the opportunity to rate each other, can derive a trust or reputation score from the aggregated ratings and provide this score to possible future trading partners. The trust or reputation score can assist agents in selecting negotiation partners, but it also promotes good behaviour (Jøsang et al. 2005). This is how a trust or reputation system could increase the efficiency and quality of a market as a whole. Several research reports have found that seller reputation has significant influences on on-line auction prices, especially for high-valued items (Mui et al. 2002). Besides electronic markets, then notions of trust and reputation play important roles in distributed systems in general.

A trust or reputation model has to be based on a theory or conceptual model of reference. Many present models of trust and reputation make use of game-theoretical concepts (Sabater and Sierra 2005). The trust and reputation values in these models are the result of utility functions and numerical aggregation of past interactions. Some other approaches use a cognitive model of reference, in which trust and reputation are made up of underlying beliefs. The trust and reputation values in these models are a function of the degree of these beliefs.

Trust and reputation of an individual can either be seen as a global property or as a subjective property[1]. In the first case, the trust or reputation of an individual is calculated from the opinions of the individuals that interacted with it. The value is publicly available and the trust or reputation of an individual is a property shared by all the other agents in the community. Trust or reputation is a subjective property when each agent assigns its own trust or reputation value to each member of the community, based on its own experiences.

Trust and reputation values can be based on different kinds of information sources. Sabater and Sierra (2005) distinguished four different sources: direct experiences, witness information, sociological information and prejudice. Information from *direct experiences* is the most relevant and reliable information for a trust or reputation model. Experiences based on direct interactions are used by almost all trust and reputation models. A less common form of direct experience is experience based on the observed interaction of other members of the community. *Witness information* is information assessed from other members of the community. The information can be based on their direct experiences or on information they gathered from other sources. Witness information is difficult for models to deal with, because information providing agents might hide information, change information or even tell complete lies. *Sociological information* is information provided by the society and might exist of social relations between agents or the role that agents play in the society. The power of a particular individual for example, might influence its reputation or the trust we have in that individual. Currently, only a few models take this kind of information into account. The last information source is *prejudice*, not very common in present trust and reputation models either. Prejudice is the mechanism of assigning properties to an individual, based on signs that identify the individual as member of a given group.

Besides decisions about the use of different sources of information, more choices about the presentation of information in the model have to be made. Is the information exchanged boolean information or a more sophisticated type of information? Does the model allow agents hiding information or providing false information or not? Are trust and reputation values accompanied by a reliability measure, indicating the probability of the information being true? Section 2.3 will provide some examples of computational trust and reputation models to make the ideas more concrete, but first the difference between trust and reputation will be discussed.

2.2 The relation between trust and reputation

The words trust and reputation are widely used by many people in many situations, but it is difficult to define the exact meanings of these concepts. With a look in a dictionary one will find out that both terms have more than just one meaning. Also in the context of distributed artificial intelligence, several different meanings of reputation (Mui et al., 2002) and trust (McKnight et al., 1996) have been discerned. The complexity of the terms

[1] Sabater and Sierra call this different ‘visibility types’ (Sabater and Sierra 2005)

makes it difficult to describe the relation between trust and reputation, but we at least know that trust and reputation are two different things. In some cases one can trust someone with a bad reputation, for example in a very close relation. Sometimes it is better to distrust someone with a good reputation, for example because a person once cheated on you.

In the context of computational models, the meanings of trust and reputation are determined by the way they are derived from a set of values. Therefore, this section will not provide *the* definitions of trust and reputation, but it will remark some of the important elements. To start with reputation, according to Jøsang et al. (2005), reputation is what is generally said or believed about a persons' or things' character or standing. The word 'generally' is important here, reputation is usually not based on the opinion of one individual. The examination of some important models of trust and/or reputation (Sabater and Sierra 2005) indeed shows that all models using the word reputation at least make use of witness information, information provided by other agents. So reputation values are mostly determined by the opinions of a whole set of agents.

Jøsang et al. (2005) define trust as the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible. In contrast to reputation, trust is something personal, the amount of trust one has in a given agent is a specific property of each individual. Sabater and Sierra's overview (2005) shows that models called models of trust rely on several information sources, but all of them at least use direct experiences to determine levels of trust. So trust values are in general based on the information gathered by one individual.

Conte and Paolucci (2002) give an extensive analysis of the relation between trust and reputation and some other related concepts. In their cognitive approach they make a distinction between *image* and *reputation*, where image is the direct evaluation of others and reputation is indirectly acquired. Image is based on an agent's direct experiences with other agents and reputation is based on information received from other agents. This information tells about other agents' direct experiences and reputation is the output of image spreading. According to Conte and Paolucci, image and reputation both contribute to trust.

Most existing models of trust and reputation do not differentiate between trust and reputation and only use one of the two concepts, and if they differentiate, the relation between trust and reputation is often not explicit (Sabater and Sierra 2005; Mui et al. 2002). The model proposed by Yu and Singh (2001) does distinguish between trust and reputation. Direct information is used to determine the trust in the target agent and witness information to determine the reputation of the target agent. However, the two information sources are not combined, the model only appeals to witness information when direct information is not available. The ReGreT system (Sabater 2002) is one of the few models that does combine trust and reputation. The reputation information (purely based on witness information) in this model is used to improve the calculation of trust values, which are also determined by other types of information. Mui et al. (2002) also proposed a model of trust and reputation in which both concepts are related with each other. According to them, increase in an agent α's reputation in its embedded social network **A** should also increase the trust from the other agents for α, decrease should lead to the reverse effect.

The few approaches that distinguish between trust and reputation and combine the two concepts (Conte and Poalucci 2002; Sabater 2002; Mui et al. 2002), seem to agree on a relation between trust and reputation in which reputation is (one of) the factor(s) that determine(s) trust. The strength of the influence of reputation depends on the specific context. This point of view about the relation between trust and reputation will also be taken in this thesis. In chapter 4, Sierra and Debenham's trust model will be evaluated according to this criterion.

2.3 Three examples of trust and reputation models

EBay is one of the world's largest *online market places* with a community of over 50 million registered users (Jøsang et al. 2005). It allows sellers to list items for sale, and buyers to bid for those items. EBay uses a reputation mechanism that is based on the ratings users give after the completion of a transaction. The user can choose between the three values positive (1), negative (-1) or neutral (0). The reputation value is calculated as the sum of the ratings over the past six months, the past month and the past seven days. Reputation thus is considered as a global property. Studies of eBay's reputation system report that buyers rate sellers 51.7% of the time and that the observed ratings are very positive, about 99% is positive (Jøsang et al. 2005). Although the system is quite primitive and can be misleading, the reputation system seems to have a strong positive impact on eBay as a marketplace.

A second example is Castelfranchi and Falcone's (1998) *cognitive model of trust*. According to them, the decision of agent α to delegate a task to agent *β* is based on a specific set of beliefs and goals, and this mental state is what we call trust. To build a mental state of trust the agent needs the following basic beliefs: competence belief (agent *β* can do the task), dependence belief (it is necessary or better when that *β* performs the task), disposition belief (*β* will actually do the task), willingness belief (*β* decided and intends to do the right actions) and persistence belief (*β* is stable in its intentions of doing these actions). The first two beliefs compound 'core trust' and together with the third belief also 'reliance'. If agent α has all these beliefs, it trusts the agent *β* on performing the task, and it could decide to delegate the task to that agent.

The last model of trust and reputation discussed here is proposed by Sabater (2002) and the system is called *ReGreT*. This system takes three different sources of information into account: direct experiences, information from third party agents and social structures. The direct trust module in the system deals with direct experiences and how these experiences can contribute to the trust on other agents. The reputation module of the system is divided in three types of reputation: witness reputation (calculated from information from other agents), neighbourhood reputation (calculated from information about social relations between partners) and system reputation (calculated from roles and general properties). A third module of credibility measures the reliability of witnesses and the information they provide. All these modules can work together to calculate trust. Because of the modular design it is also possible to use only some of the parts.

The three examples above are all computational models of trust and reputation, and the big differences among them give an indication of the broadness of the research area. The usefulness of trust and reputation seems obvious and literature around it is rapidly growing. Several articles providing an overview of the field conclude however that the research activity is not very coherent and needs to be more unified (Sabater and Sierra 2005; Jøsang et al. 2005; Mui et al. 2002; Fullam et al. 2004). In order to achieve that, test-beds and frameworks to evaluate and compare the models are needed.

3 An information-based model for trust

This chapter introduces the information-based model for trust that will be examined in this thesis. The language, the methods and the trust model of this approach will be discussed. In the last section a comparison between the information-based approach and a game-theoretical approach will be given.

3.1 Information Theory

Computational models of trust and reputation are always based on a certain theory or conceptual model. As mentioned in the preceding chapter, most present models of trust and reputation make use of game theoretical concepts. The model of trust that will be introduced in this chapter and that will be central in the thesis has another frame of reference. The model proposed by Sierra and Debenham (2005) is the first trust and reputation model based on information theory. Before the discussion of Sierra and Debenham's model specifically, this section provides a short introduction to the most important concepts of information theory.

Flipping a coin, throwing a dice and picking a blind card from a pile are all actions of which the outcome is uncertain beforehand. If the probability of one possible outcome is known, information theory provides a way to derive the *information content*[2] of that particular event. The information content h(x) of an outcome x is defined to be:

$$\mathrm{h(x)} = \log_2 \frac{1}{P(x)}$$

According to this definition, infrequent events give more information (have bigger information contents) than frequent events. If the probabilities of all possible events are known, another information theoretical concept can be calculated: the entropy of all possible outcomes. *Entropy* H is a measure of the uncertainty in a probability distribution for a discrete random variable X. The *entropy* of X, H(X), is the average information content of all possible events:

$$\mathrm{H(X)} \equiv \sum_i p(x_i) \log_2 \frac{1}{p(x_i)}, \quad \text{where } \mathrm{p(x_i) = P(X = x_i)}$$

In the following example, the probability distribution (p_i) of each letter (a_i) being randomly selected in an English document is provided. The last column gives the corresponding information contents $h(p_i)$.

[2] Information content is also called Shannon information content (MacKay 2003)

i	a_i	p_i	$h(p_i)$
1	a	.0575	4.1
2	b	.0128	6.3
3	c	.0263	5.2
4	d	.0285	5.1
5	e	.0913	3.5
6	f	.0173	5.9
7	g	.0133	6.2
8	h	.0313	5.0
9	i	.0599	4.1
10	j	.0006	10.7
11	k	.0084	6.9
12	l	.0335	4.9
13	m	.0235	5.4
14	n	.0596	4.1
15	o	.0689	3.9
16	p	.0192	5.7
17	q	.0008	10.3
18	r	.0508	4.3
19	s	.0567	4.1
20	t	.0706	3.8
21	u	.0334	4.9
22	v	.0069	7.2
23	w	.0119	6.4
24	x	.0073	7.1
25	y	.0164	5.9
26	z	.0007	10.4
27	-	.1928	2.4

(MacKay 2003, p32)

Not so often used letters like 'x' and 'q' have a low probability of being selected and thus a high information content. An often used letter like 'e' in contrast, gives less information according to information theory. Averaging all the information contents in the example gives the following entropy.

$$H(x) = \sum_i p_i \log_2 \frac{1}{p_i} = 4.1$$

In a probability distribution with many low probabilities the average information content will be higher, and this explains why another name for the entropy of X is the uncertainty of X.

In the example, all probabilities of all possible outcomes of randomly selecting a letter are known. There are however many situations in which these data are not available. Without any information about probabilities of possible outcomes the best option is to take the uniform probability distribution, in which the probabilities of all possible outcomes are equal ($P(x_i) = 1 / n$). Another possibility is that only a part of the information about possible outcomes is available. The exact probability distribution is unknown, but information about some constraints on this distribution is available. In these cases, the *maximum entropy principle* offers a rule for choosing a distribution that satisfies all constraints posed to the distribution. According to this rule one should select the distribution p that maximizes the entropy. This constructs the "maximally non-committal" probability distribution (Sierra and Debenham, 2005).

3.2 A negotiation language

In Sierra and Debenham's model (Sierra and Debenham 2005), agent α can negotiate with agent β and together they aim to strike a deal δ. In the expression $\delta = (a,b)$, a represents agent α's commitments and b represents β's commitments in deal δ. A is the set of all possible commitments by α and B the set of all possible commitments by β. All agents have two languages, language C for communication and language L for internal representation. The language for communication consists of five illocutionary acts, which are actions that can succeed or fail. The illocution particle set ι = {Offer, Accept, Reject, Withdraw, Inform} has the following syntax and informal meaning.

- Offer (α,β,δ) Agent α offers agent β a deal $\delta = (a,b)$ with action commitments *a* for α and *b* for β.
- Accept (α,β,δ) Agent α accepts agents β's previously offered deal δ.
- Reject (α,β,δ,[*info*]) Agent α rejects agents β's previously offered deal δ. Optionally, information explaining the reason for the rejection can be given.
- Withdraw (α,β,[*info*]) Agent α breaks down negotiation with agent β. Extra *info* justifying the withdrawal may be given.
- Inform (α,β,*info*) Agent α informs agents β about *info*.

Sierra and Debenham use *info* for referring to: (1) the process used by an agent to solve a problem, or (2) an agent's data including preferences. For this, they propose the following content language (*info* Є *L*) in Backus-Naur form:

info	::= *unit* [**and** *info*]
unit	::= *K*|*B*|*soft*|*qual*|*cond*
K	::= **K**(WFF)
B	::= **B**(WFF)
soft	::= **soft**(f,{V^+})
qual	::= V=D [>V=D]
cond	::= **If** *DNF* **Then** *qual*
WFF	::= *any wff over subsets of variables* {V}
DNF	::= *conjunction* [**or** DNF]
conjunction	::= *qual* [**and** conjunction]
V	::= $\mathbf{v_1|...|v_n}$
D	::= **a|a'|b|...**
f	::= *any function from the domain of subsets of V to a set A. For instance a fuzzy set membership function if A = [0,1]*

K and *B* refer to the agent's knowledge and beliefs. A WFF is a well-formed formula and DNF refers to the Disjunctive Normal Form. *Soft* and *qual* are used to express quantitative and qualitative preferences, respectively. A soft constraint associates each instantiation of its variables with a value from a partially ordered set. For example: "The probability I will choose a red book is 30% and the probability I will choose a blue book is 20%". A qualitative constraint expresses a preference relation between variable assignments. For example: "I prefer red books to blue books". The other expressions in the list make it possible to express sophisticated preferences. Some concrete examples of expressions are:

- "I prefer slippers to boots when it is summer"

  Inform (α, β, *if Season=summer then Shoe=slipper* > *Shoe=boot*)

- "I prefer more shoes to less shoes"
  Inform (*α, β, soft(tanh,{Shoes})*)
- "I prefer black shoes to green shoes"
  Inform (*α, β, if thing=shoe then colour=black* > *colour=green*)
- "I reject your offer since I cannot pay more than 200"
  Reject (*α, β, Money=200, hard(Money* < *200, {Money})*)

This section should give a basic idea of the language that is used in Sierra and Debenham's model of trust. The language is especially rich in expressing preferences. However, this thesis will not focus on the effect of information about preferences, so a deep understanding of the language will not be necessary to understand the thesis. For further details of the language is referred to Sierra and Debenham's article (2005).

3.3 Information-based negotiation

With an agent's internal language *L*, many different worlds can be constructed. A possible world represents for example a specific deal for a specific price with a specific agent. To be able to make grounded decisions in a negotiation under conditions of uncertainty, the information-theoretic method denotes a probability distribution over all these worlds. If an agent would not have any beliefs or knowledge, all worlds would have the same probability to be the actual world. Often however, agents do have knowledge and beliefs which put constraints on the probability distribution. The agent's knowledge restricts 'all worlds' to all *possible worlds*, the agent knows that some worlds are not possible. A possible world *v*, element of the set of all possible worlds *V*, is consistent with the agent's knowledge. Worlds inconsistent with the agents knowledge are believed to be false and do not have to be considered any further. The notation of the set of all possible worlds consistent with an agent's knowledge is $V|K = \{v_i\}$. An agent's set of beliefs *B* determine its opinion on the probability of possible worlds, according to its beliefs some worlds are more probable to be the actual world than others. A *random world*, $W|K = \{p_i\}$, is a probability distribution over all possible worlds, where p_i expresses the degree of beliefs an agent attaches to each possible world to be the actual world.

From the probability distribution over all possible worlds, the probability of a certain sentence or expression in language *L* can be derived. For example the probability *P (executed δ | accepted δ*) of whether a deal, once accepted, is going to be executed or not can be calculated. This *derived sentence probability* is always a probability with respect to a random world, a particular probability distribution over all possible worlds. A sentence *σ*'s probability is calculated by taking the sum of the probabilities of the possible worlds in which the sentence is true. For all sentences that can be constructed in language *L* counts:

$$\mathrm{P}_{\{W|K\}}(\sigma) \equiv \sum_{n} \{p_n : \sigma \text{ is true in } v_n\}$$

An agent with a set of beliefs has attached *given sentence probabilities* to all statements φ in its set of beliefs B. A random world is consistent with the agent's beliefs if for all statements element of the set of beliefs the attached probabilities to the sentences are the same as the derived sentence probability. Expressed in a formula, for all beliefs φ element of B:

$$B(\varphi) = P_{\{W|K\}}(\varphi)$$

So the beliefs of the agent impose linear constraints on the probability distribution. To find the best probability distribution consistent with the knowledge and beliefs of the agent, *Maximum entropy inference* states that the entropy of the probability distribution has to be maximized. The found probability distribution should have maximum entropy and be still consistent with the knowledge and beliefs. This distribution is used for further processing when a decision has to be made.

When the agent obtains new beliefs, the probability distribution has to be updated. This happens according to the principle of *minimum relative entropy*, which searches a probability distribution satisfying the new constraints and that has the least relative entropy with respect to the prior one. The relative entropy between probability distribution p and q is calculated as follows.

$$D_{RL}(p||q) = \sum_{i-1}^{n} p_i \log_2 \frac{p_i}{q_i}$$

The principle of maximum entropy is equivalent to the principle of minimum relative entropy with a uniform prior distribution.

While an agent is interacting with other agents, it obtains new information. Sierra and Debenham (2005) mention the following types of information from which the probability distribution can be updated.

- Updating from decay and experience. This type of updating takes place when the agent derived information from the direct experiences it had with other agents. When such an update takes place, the evaporation of beliefs as time goes by is taken into account. Negotiating people or agents forget about the behaviour of a past negotiation partner.
- Updating from preferences. This updating is based on past utterances of a negotiation partner. If agent β prefers a deal with property Q_1 to a deal with property Q_2, he will be more likely to accept deals with property Q_1 than deals with property Q_2.
- Updating from social information. Social relationships between agents, social roles and positions held by agents influence the probability of accepting a deal. Two ways to model the updating from social information are the modelling of power and the modelling of reputation.

3.4 The trust model

Once the probability distribution is constructed and up to date, it can be used to derive trust values which can be used in the decision process. From an actual probability distribution, the trust of agent α on deal δ with agent β at the current time, or the trust on agent β in general at the current time can be calculated. Sierra and Debenham (2005) propose two ways to calculate trust values. The first way to model trust is trust as conditional entropy. In this case the trust value, a value between 0 and 1, represents the dispersion of the expected observations: the closer to 1 the value of trust, the less dispersion of the expected observations. This formulation of trust is useful when any variation from the agreed contract is undesirable. The trust that an agent α has in agent β with respect to the fulfilment of a contract (a,b) is calculated.

$$\mathrm{T}(\alpha,\beta,b) = 1 + \frac{1}{B^*} \cdot \sum_{b' \in B(b)^+} P^t(b'|b) \log P^t(b'|b),$$

where $\mathrm{B(b)}^+$ is the set of contract executions that agent α prefers to b. $\mathrm{B}^* = 1$ if $|\mathrm{B(b)}^+| = 1$ and $\log |\mathrm{B(b)}^+|$ otherwise. The trust of α in β in general is the average of α 's trust in β in all possible situations.

$$\mathrm{T}(\alpha,\beta) = 1 + \frac{\sum_{b \in B} \left\lfloor P^t(b) \cdot \sum_{b' \in B(b)^+} P^t(b'|b) \log P^t(b'|b) \right\rfloor}{B^* \cdot \sum_{b \in B} P^t(b)}$$

The other way of modelling trust is trust as relative entropy. This models the idea that the more the actual executions of a contract go in the direction of the agent's preferences, the higher the level of trust. Therefore the relative entropy between the probability distribution of acceptance and the distribution of the observation of contract execution is taken.

$$\mathrm{T}(\alpha,\beta,b) = 1 - \sum_{b' \in B(b)^+} P^t(b') \log \frac{P^t(b')}{P^t(b'|b)}$$

Similarly to the previous trust calculations, the trust of α in β in general is the average of all possible worlds.

$$\mathrm{T}(\alpha,\beta) = 1 - \sum_{b \in B} P^t(b) \sum_{b' \in B(b)^+} P^t(b') \log \frac{P^t(b')}{P^t(b'|b)}$$

After making observations, updating the probability distribution and calculating the trust, P(Accept (α,β,δ)) can be derived from the trust and an agent can decide about the acceptance of a deal.

3.5 Information theory compared with game theory

Instead of using information theory, trust and reputation could also be modelled with game theory. An important concept in game theory is utility, the amount of satisfaction an agent derives from an object or an event. In game theoretical models, the goal is often to maximize utility. In the context of negotiations, an agent should accept a proposal if the utility *u* of the deal is higher than a particular margin value *m*. The utility can be calculated by taking the profits of a deal minus its costs. So the basic idea of game theoretical negotiation is that if $u > m$ in a given situation, the agent accepts the deal.

However, when the utility of accepting a deal is unknown or uncertain, this method will not work. Game theory solves this problem by using a random variable *S*, assigning probabilities to all possible outcomes after accepting the deal. The higher *S*'s standard deviation, the higher the uncertainty in the process will be. Now the agent can calculate $P(S > m)$, the probability that the utility of the outcome will be higher than the margin value. Taking its willingness to take risks into account, the agent is able to calculate $P(accept\ \delta)$, the probability that the agent accepts a deal.

In contrast to game theoretical approaches, Debenham and Sierra's information-based approach does not make use of the concept of utility and information-based agents are not 'utility aware'. The probability of acceptance, $P(accept\ \delta)$, is not an indication of how good deal δ is in the information-based method. In contrast, $P(accept\ \delta)$ is a combination of properties of the deal *and* the of integrity of the information against which δ has been evaluated. So $P(accept\ (\alpha,\beta,\delta_1)) > P(accept\ (\alpha,\beta,\delta_2))$ does not mean that δ_1 is a better deal than δ_2, it means that agent α is more certain that δ_1 is acceptable than δ_2 is acceptable.

Game theory and information theory both have some restrictions in the kinds of information they process. In order to calculate a utility, the game theoretical agent has to know the exact certainty of an event. This might be a problem, in the realistic world people are not always sure about uncertainties. In the information-based approach this is not required, without certainty about the uncertainty, probability distributions can also be calculated. However, the information-based approach has to deal with other problems. When an agent's language is restricted it is no problem to calculate probabilities for all possible worlds, but when the amount of possible worlds grows this can be a problem. Moreover, the information-based approach cannot deal with infinite domains and it can only deal with continuous values by representing the domain as a finite set of intervals. As long as the probabilities of different possible worlds are known, game theory does not have this problem.

Game theory is successfully applied in many different models. The concept of utility is intuitively very appealing and easily to understand. The game-theoretical approach does however suppose that agents are totally rational, which is not always the fact. And when few information is available its methods become less appealing. For example if uncertainties are high and an agent is willing to take great risk, the calculated utilities do not really make sense. In situations of few information, the information-based approach might be a better option as a guide for making decisions. Information-based approaches do not calculate utilities, but look directly to the information with which the decision is made. Even with only very few beliefs, a probability distribution can be calculated.

4 Reputation in the information-based model

Suggestions to extend the part about reputation in the information-based trust model will be given in this chapter. After an analysis of the role of reputation in the model, two possible approaches to deal with reputation will be worked out. The last section discusses the results of this work.

4.1 Trust and reputation in the model

Sierra and Debenham's information-based model of trust does not yet provide a fully developed way to deal with reputation, it only offers some ideas. In section 5.3 of their article, Sierra and Debenham (2005) propose to update trust from reputation. The probability distribution from which trust values are calculated is updated from reputation information and the result is a new probability distribution and thus new trust values. This relation between trust and reputation is found in some other models of trust and relation as concluded in chapter 2, for example in the ReGreT model (Sabater 2002). In section 7 of Sierra and Debenham's article reputation is mentioned again, this time in the calculation of the probability that a deal will be accepted. Sierra and Debenham do not give a definitive calculation of the probability of acceptance, but they "can imagine the probability of acceptance of a deal as a composed measure" (Sierra and Debenham 2005). Here they propose to add the weighed values of trust and reputation, to together determine the value of P(Accept(α,β,δ)).

In a combination of both proposals of the role of reputation (section 5.3 and section 7, Sierra and Debenham 2005), the relation between different concepts could be represented with the following figure.

$$\begin{tikzcd}
\textbf{Trust}(\alpha,\beta,\delta) \arrow[dr, "\text{reputation update}"'] & & \textbf{Reputation}(\alpha,\beta,\delta) \arrow[ddl] \\
& \textbf{Trust\_new}(\alpha,\beta,\delta) \arrow[d] & \\
& \textbf{P(Accept}(\alpha,\beta,\delta)) &
\end{tikzcd}$$

Studying this figure, the following question comes up. Why is reputation information used to determine P(Accept(α,β,δ)), if the information is already processed in the calculation of Trust_new? It seems redundant to use the same reputation information twice in the calculation of P(Accept(α,β,δ)). Sierra and Debenham do not discuss this issue in their article and they do not provide a clear way to deal with it.

Because of the seeming redundancy, both ways to handle reputation in the model of trust are examined separately in this chapter. Below the two options that will be discussed are represented in a figure.

$$\begin{tikzcd}
\mathbf{Trust}(\alpha,\beta,\delta) \arrow[d, "\text{reputation update}"] & \mathbf{Trust}(\alpha,\beta,\delta) \arrow[d] & \mathbf{Reputation}(\alpha,\beta,\delta) \arrow[d] \\
\mathbf{Trust\_new}(\alpha,\beta,\delta) \arrow[d] & {} \arrow[d] & {} \arrow[d] \\
\mathbf{P}(\mathbf{Accept}(\alpha,\beta,\delta)) & \mathbf{P}(\mathbf{Accept}(\alpha,\beta,\delta)) & {}
\end{tikzcd}$$

4.2 Updating trust from reputation

The first proposal to deal with reputation information is to consider it as one of the factors determining the level of trust. Reputation information here, is information an agent receives from other agents with their opinions of other agents. So besides an agent's own experiences with for example agent β, witness information could also influence the agent's opinion about agent β's behaviour. A lot of positive stories about agent β, might increase its trust in agent β.

By the illocution Inform (γ, α, *info*), agent α receives information from agent γ. In the case that the information content is an opinion of γ about another agent, the received information is reputation information. Sierra and Debenham (2005) represent this type of information Θ with Reputation (Φ, β), where β represents the agent the information is about and Φ the institution or domain the information applies to. An extension to the expression could be a variable r, to express the reliability of the provided information. This results in a more sophisticated type of information Θ, Reputation (Φ,β,r). After receiving Reputation (Φ, β) or Reputation (Φ,β,r), agent α will update $p(b'|b)$, which represents the prior probability that the contract execution will be preferred by α to β's commitments b. The new $p(b'|b)$, given the reputation information, can be calculated by the following formula (Sierra and Debenham 2005):

$$p(b'|b,Reputation(\Phi,\beta,r)) = p(b'|b) + g_3\,(b'|b,Reputation(\Phi,\beta,r))\;(1- p(b'|b))$$

In the formula, $g_3\,(b'|b,Reputation(\Phi,\beta,r))$ represents the strength of agent α's belief that the probability that the execution of contract b at time $t + 1$ will be preferred to b should change, given that Reputation(Φ,β,r) was received at time t.

Sierra and Debenham do not specify in their article how to calculate $g_3(b'|b,Reputation(\Phi,\beta,r))$, the strength of belief. Some factors that could influence the strength of belief are the following.

- *The content of the reputation information.*
  Very positive or very negative information will have more effect than slightly positive of slightly negative information. The content of reputation information is a value between -1 and 1. The bigger the absolute value, the more effect the information will have.
- *Possibly provided reliability information.*

The informing agent might provide a value between 0 and 1, assigning the reliability of the reputation information given. Reliability is an estimate of the extent to which information is correct. The higher the reliability, the more it will affect the trust value.

- *Persuasive power of the source agent.*
  This is a value between 0 and 1 stored in the agent's belief set. Initially this value will be 1, but when an agent has had negative experiences with the informing agent this value will decrease. Negative experiences could be the hiding of information or the providence of false information. The higher the persuasive power of the source agent, the more effect its provided information will have.
- *Similarity with other information.*
  Information that agrees with knowledge or beliefs about the agent's performance on similar domains, will have more effect than information that does not. If agent β is a good singer, he will probably also have feeling for rhythm. This similarity could also be represented by a value between 0 and 1.

The effect of the first aspect, the content of the reputation information Reputation(Φ,β,r) on a new probability distribution seems clear. As mentioned before, very positive or very negative information will have more effect than slightly positive of slightly negative information. If the reputation information is neutral, a value of 0, it will not have any effect at all. However, the value of this information could decreases for several reasons, reasons mentioned in the second, third and fourth factor. If the information is not 100% reliable, the value of the reliability r is not 1, the information looses influence on the probability distribution. If the agent for some reason lost persuasive power, for example because he provided bad information in the past, the effect of the reputation information will also decrease. The last factor, in the case that stored information on a highly similar domain is totally different with the provided information, can also cause a decrease of influence. The desired effects of increase and decrease of influence on the new probability distribution are reached by calculating $g_3(b|Reputation(\Phi,\beta,r))$ by multiplying the four factors with each other. The result will be a value between -1 and 1.

A remark has to be made on the second factor, the reliability information. This information could be false and could in an unjustified way decrease or increase the influence of the reputation information on the probability distribution. A way to solve this problem is to ignore the reliability information and not use it. However, throwing away of information is usually not the way to make better decisions and there are arguments that the use of the information will not lead to worse results. An agent could provide false reliability information r in two possible ways: the provided value is too high or too low. When the value is too low, correct reputation information could unjustly be ignored. This situation however is highly improbable, because it does not bring any advance to the other agent. The other possibility is that bad information with a high reliability value is provided, which could be of advantage of other agents and thus is quite probable. During the first interactions, the calculation of $g_3(b|Reputation(\Phi,\beta,r))$ would deliver the same answers as when the reliability information would be ignored. But when an agent continues providing high reliability measures to bad opinions, the receiving agent will start to 'learn' about the information providing agent. Because it provides bad information, the persuasive power of the information providing agent will decrease. Then a high reliability value does not matter any more, because the low value of persuasive power will already decrease the influence of the information.

So *g₃(b|Reputation(Φ,β,r))* can be calculated by multiplying the factors of content, reliability, persuasive power and similarity. Then, as proposed by Sierra and Debenham (2005), agent *α* revises its estimate of *p(b '|b)* by using the principle of minimum relative entropy.

$$(P_C^t(bj \mid b))_{j-1}^{n} = \arg\min \sum p_i \log \frac{p_i}{P^t(bi \mid b)}$$

This revision is subject to the constraint:

$$\sum_{x \in B(b)^-} P_C^t(x \mid b) = p(b' \mid b, \mathit{Reputation}(\Phi,\beta,r)),$$

where B(*b*)$^+$ is the set of contract executions that agent *α* prefers to *b*.

Here again a remark has to be made, a more fundamental and a more difficult one to solve than the problem with reliability information. Imagine the case that agent *α* receives very positive reputation information from agent *β* about agent γ. The provided reliability information is maximal, the persuasive power of agent *β* is maximal and the provided information is highly similar with *α*'s other beliefs about γ. In this case the reputation information has a maximal effect on the probability distribution. But how much effect will it have, as much as a direct experience? Although the circumstances of receiving the reputation information might be fine, it remains second hand information. First hand information, obtained by direct experiences, should be of more influence than reputation information although the circumstances are perfect.

More general, the problem is that each change of the probability distribution means a loss of information. With every update of the probabilities, old values stored are replaced by new ones. So it is very important to consider carefully whether a particular update really improves the predictable power of the probability distribution instead of throwing away valuable information that was in the model. In the case of an update from reputation information, this update should not replace all information obtained by direct experiences. In contrast, with some slight changes it should perfect the probabilities obtained.

Some ways to achieve such a proportional contribution of reputation information on the calculation of trust are the following.

- Reputation information is only used for updating, if the strength of agent *α*'s belief that the probabilities should change (*g₃ (b '|b,Reputation(Φ,β,r))*) is above a certain threshold.
- When updating from reputation information, multiply the strength of agent *α*'s belief that the probabilities should change (*g₃ (b '|b,Reputation(Φ,β,r))*) with a factor between 0 and 1. This factor indicates the importance of updating from reputation information in comparison to updating from direct experiences (with an importance of 1).

- Reputation information is only used for updating, if agent α received a certain amount of opinions from different agents about the same agent. In this case a way to aggregate different 'reputation informations' is needed.
- Reputation information is stored and when the probability distribution is updated from reputation, it is updated from all reputation information received in a specific period of time. In this case a way to aggregate different entities of received information is needed.

One can apply one of the four points mentioned here, or use some combination of them. A proposal to aggregate reputation information will be given below. Firstly, because the aggregation of different opinions about reputation before updating, prevents too much change of probabilities described to all possible worlds. A second reason is that aggregation of opinions of different agents about reputation conceptually makes a lot of sense. In section 2.2, the notice that reputation is usually not based on the opinion of one individual was stressed. So by aggregating reputation information received from different agents, a reputation value for a particular agent can be derived. Reputation updates will not longer be updating from 'reputation information', but updating from 'a reputation'.

Before different pieces of information can be aggregated, they have to obey to certain conditions. The first condition is that the different pieces have to contain information about the same agent. If one agent provided more than one opinion about another agent, the most recent opinion should be taken. Furthermore, only opinions with a reliability value higher than a given x, from a providing agent with a certain minimum trustworthiness value of y and a similarity value of at least z should be taken into account. The values of x, y and z are variable and can be set according to the users wishes. Finally, the amount of contributing agents given the whole population has to be chosen. One has to decide which percentage of agents has to provide opinions about an agent, to speak of the reputation of that particular agent. When all these parameters are set and an agent has enough valuable information according to the parameters, the different opinions can be aggregated. The reputation of an agent can now be calculated by taking the average of the opinions about that agent. The standard deviation of all the opinions indicates the probability of the reputation being good. The smaller the standard deviation, the more different agents agree with each other, the higher the probability that the reputation value is useful. The more it should influence the probability distribution in an update.

Sierra and Debenham (2005) want to model the idea that beliefs evaporate as time goes by. In their proposal, the natural decay of belief is offset by new observations. One could choose to also update from decay when an agent receives other types of information, for example reputation information. If the agent takes the evaporation of beliefs into account at any time it receives new information, it will always use the most up to date probability distributions for deriving its trust values.

4.3 Combining trust and reputation

Instead of reputation being one of the aspects updating trust, the two concepts could also be seen as two factors determining the probability that a deal will be accepted. In this case, reputation information still refers to information an agent receives from other agents. The meaning of trust slightly changes, in this case trust is only determined by an agent's direct experiences or observations. This picture highly resembles Conte an Paolucci's (2002) approach to trust and reputation. They would use the word 'image' for what is called trust here. According to them, reputation and image (information based on direct experiences) together determine the level of trust. In the naming in this section, reputation and trust together determine the probability of acceptance.

This second method keeps witness information and information from direct experiences separate till the probability P(Accept(α,β,δ)) is calculated. This is achieved by calculating two separate probability distributions. One of them determines the level of trust and is only updated from direct experiences. A second one deals with reputation information, *R*, the set of all information received in the form Reputation(Φ,β,r). Whereas the constraints in the first probability distribution are given by the agent's beliefs *B* derived from direct experiences, the constraints in the second probability distribution are only given by reputation information *R*. The probability distribution of trust is updated from direct experiences as described by Sierra and Debenham (2005) and the probability distribution of reputation is updated by one of the ways described in the previous section.

After the calculations of trust and reputation[3] from the probability distributions, the two probabilities are combined to determine the probability of accepting a deal. Sierra and Debenham (2005) propose the following formula to combine the two values.

$$P^t\,(\text{Accept}\,(\alpha,\beta,\delta)) = \kappa_1\ T\,(\alpha,\beta,\delta) + \kappa_2\ R(\alpha,\beta,\delta),$$

where $\kappa_1 + \kappa_2 = 1$, and they are constants or the result of a function depending on the environment. κ_1 and κ_2 represent the importance an agent gives to both aspects. In the case of trust or reputation the values of κ_1 and κ_2 could depend on the amount of experience of an agent. The more experience agent α has with agent β on deals like deal δ, the more its decision of accepting the deal depends on trust, its own opinion. When agent α has no experience at all in this field, its decision is purely based on reputation information, opinions of others.

The formula could easily be extended with other dimensions influencing the probability P^t(Accept (α,β,δ)). An extension should deliver the following form.

$$P^t\,(\text{Accept}\,(\alpha,\beta,\delta)) = \kappa_1\ T\,(\alpha,\beta,\delta) + \kappa_2\ R(\alpha,\beta,\delta) + \ldots + \kappa_n\,X(\alpha,\beta,\delta),$$

in which the condition $\kappa_1 + \kappa_2 + \ldots + \kappa_n = 1$ has to be satisfied. An extra dimension determining the probability of acceptance could for example be social information, about the power or social relationships between agents. Any new dimension should satisfy the condition that its information is not already being processed in another dimension.

4.4 Conclusions

Sierra and Debenham (2005) define trust as a measure of how uncertain the outcome of a contract is. So according to them, trust should incorporate the overall opinion of an agent about another agent or about a certain deal. This overall opinion should be based on all the information an agent has. In this sense, the approach in section 4.2 is preferable to the

[3] The reputation being calculated in the same way as trust is calculated from a probability distribution

approach in section 4.3. In section 4.2 trust values indeed are updated from all the information sources available. In section 4.3 trust is only based on the direct experiences of an agent with other agents. However, if Conte and Paolucci's (2002) terms would be used for the approach in section 4.3, trust would also contain all available information sources.

The approach of updating trust from reputation has some other advantages in comparison to combining trust and reputation. The first method only uses one probability distribution, which is simpler to handle than two or more probability distributions as in the second approach. The reputation update of this single probability distribution runs similarly to updating from other information sources, which is also a plus according to simplicity reasons. Finally, if all information is already processed in the probability distribution determining trust, than the trust value on a specific deal is automatically its probability of acceptance in Sierra and Debenham's model. Because all other dimensions are already integrated in the value of trust, there are no other factors left to determine P^t (Accept (α,β,δ)). So the method saves a calculation step.

A difficulty in the first approach however, is to control the contributions of direct experiences and witness information on P^t (Accept (α,β,δ)). By keeping trust and reputation separate till the end of the calculation, it is easier to see the effects of both aspects on the final decision. The second method provides a way to isolate the influence of reputation and to better investigate its role in the final decision. Influences of other aspects, like for example social information, could also be investigated this way. In most situations however, the profits of separating trust and reputation will not outweigh the conceptual arguments of the first approach.

5 Social information in the model

This chapter proposes to incorporate social information in the information-based trust model. After the discussion of social aspects that might play a role, a possible way to deal with social information is explained.

5.1 Social information

The updating of trust from an agent's own experiences and from preferences are worked out well in Sierra and Debenham's information-based model. They way they treat updating from reputation has been discussed in the previous chapter. Besides these factors, another information source could influences the level of trust. Lately, in the research field of computational models of trust and reputation, the role of social information is stressed (Sabater and Sierra 2005, Ashri et al. 2005, Mui et al. 2002) and becomes more and more important. Although Sierra and Debenham reckon reputation information under updating from social information, social information is more than just that. In this chapter, social information not directly based on an agent's own experiences or on information based on the experiences of other agents will be discussed.

Social information for example could tell something about the relationship between two negotiation partners. A negotiation about the division of rooms in an office between two employees with the same status would change if one of the two becomes the other's boss. An agent would prefer negotiating with an agent who needs products he sells, to negotiating with an agent without this dependency on his products. Ashri et al. (2005) denote two important aspects in the rise of social relationships: *interactions* and *organisational structures*. In their article, they provide tools for identifying and characterizing relationships between agents. They identify the following relationships or interaction types which are relevant with regards to trust.

• Trade	Agent α is able to buy a product from agent β within the same market.
• Dependency	Agent α is selling goods in a market that agent β can view, and at the same time β has the goal to buy the goods α is selling in that market.
• Competition	Agent α and β are selling the same goods in the same market, or α and β have the same goal, they want to buy the same products.
• Collaboration	Agent α is selling goods to agent β and at the same time, β is selling different goods to α.
• Tripartite relationships	Relationship between two agents if at least one more agent is added to the analysis.

According to Ashri et al. (2005), an agent should distrust its counterpart whenever the latter has an opportunity to defect. In a situation where agent α is dependent on agent β for example, β may have an opportunity to exploit α because α has no other choice than β

as an interaction partner. Agent α's trust in β should be lowest possible. According to Ashri et al. (2005), the different *types of relation patterns*, together with the *context* in which the relationship is developing, determine the *intensity* of the relationship between two agents. The context of a relation is determined by issues such as the abundance of a product, the number of sellers of the product and the amount being bought. The instantiation of an intensity calculation function will depend on the type of application (Ashri et al., 2005).

Other kinds of social information could inform the agent about the position of an agent in an organisation or institution, for example information about the power relationships between different agents. A system could also provide information about the agent's reputations. This information is different from the reputation information discussed in chapter 4, in the sense that the social reputation information is not directly based on other agents' experiences (witness information), but on objective reputation measures used by the system.

5.2 Social constraints in an agent's knowledge base

Sierra and Debenham (2005) give some initial ideas about how to deal with social information. Besides the influence of reputation information, they mention the influence of power on trust. According to them, the power of a negotiation partner influences the probability its opponent will accept a deal. In the model they accomplish this effect by adding the following constraint to an agent's knowledge K:

$$\text{Power}(\beta) > \text{Power}(\gamma) \rightarrow \text{P}(\text{Accept}(\alpha,\beta,\delta)) > \text{P}(\text{Accept}(\alpha,\gamma,\delta))$$

Here the assumption is made that power can be modelled as a function from agents to real values (Sierra and Debenham 2005). So this assumption presupposes a linearly ordered set of agents. There may be situations however, in which the order of agents according to power can be tree-like or in which an agent only has a lot of power over one specific agent and not over others. Therefore a partial order seems more appropriate to express differences in power between agents than a linear one. This can be achieved by representing power by a value between -1 and 1 attached to the predicate Power(α,β), indicating the strength of the power of one agent has over another. The expressions Power(α,β) = 1 and Power(β,α) = -1 are equal and mean that agent α has absolute power over agent β. A small absolute power value, means that none of the two agents has much power over the other agent. According to this representation, the power constraint in an agent's knowledge base would become the following.

1 $$\text{Power}(\beta,\alpha) > \text{Power}(\gamma,\alpha) \rightarrow \text{P}(\text{Accept}(\alpha,\beta,\delta)) > \text{P}(\text{Accept}(\alpha,\gamma,\delta))$$

Following this method, constraints modelling Ashri et al.'s interaction types discussed in the previous section could easily be added to the knowledge base of an agent. To add these constraints to an agent's knowledge base K, the operators Dep(α,β), Comp(α,β) and Coll(α,β) have to be introduced. The first one, Dep(α,β), is a dependency relation in which agent β is selling goods in a market and agent α has the goal to buy that goods from β. Comp(α,β) defines a competition relation between agent α and agent β. This means that either α and β are selling the same goods in the same market, or that α and β want to buy the same products. Comp(α,β) is a symmetric relation, the order of α and β makes no difference in the meaning of the expression. The last predicate Coll(α,β) represents a collaboration relation between α and β and it actually consists of two dependency relations Dep(α,β) and Dep(β,α). Later will be shown that Coll(α,β) is not a symmetric relation, but for now the order of α and β does not matter.

Ashri et al. (2005) state that an agent should distrust its counterpart whenever the latter has an opportunity to defect. Below the four possible relations agent α could have with β are put in the order of the trust α should have in β. In the first situation α should strongly distrust in β, in the relation types following, α can have more and more trust in β.

- Dep(α,β) (α should strongly distrust β)
- Comp(α,β) (α should distrust β)
- Coll(α,β) (α should trust β)
- Dep(β,α) (α should strongly trust β)

According to this principle of distrusting a counterpart which has the opportunity to defect, the following constraints could be added to the knowledge of an agent.

2 $\text{Comp}(\alpha,\beta) \wedge \text{Coll}(\alpha,\gamma) \rightarrow P(\text{Accept}(\alpha,\gamma,\delta)) > P(\text{Accept}(\alpha,\beta,\delta))$
3 $\text{Dep}(\beta,\alpha) \wedge \text{Comp}(\alpha,\gamma) \rightarrow P(\text{Accept}(\alpha,\beta,\delta)) > P(\text{Accept}(\alpha,\gamma,\delta))$
4 $\text{Dep}(\beta,\alpha) \wedge \text{Coll}(\alpha,\gamma) \rightarrow P(\text{Accept}(\alpha,\beta,\delta)) > P(\text{Accept}(\alpha,\gamma,\delta))$

Constraint number 2 states that an agent should prefer deals with agents with which it has collaboration relations to agents with which it has competition relations. In the case of a collaboration relation, both agents gain by not defecting during their interactions since they both depend on each other to achieve their goals. When two agents are in competition it is in their interest to undermine each other in all possible ways. So it is more probable that an agent will defect when it has a competition relation than when it has a collaboration relation. That is why an agent should prefer deals with collaborating agents to deals with competing agents.

Constraint number 3 and 4 compare a dependency relation with competitive and collaboration relations. When agent α itself is dependent on another agent for a particular service, it can not have more relations about that same service. So no constraints of preference have to be added for this situation. When another agent is dependent on agent α, the change that this agent will defect is very low because it has no other options that agent α. In competitive or collaborative relations the consequences of defection will be less severe for the other agents, or their defections will also harm agent α. Therefore deals with dependent agents are always preferable to deals with competing or collaborating agents.

In addition to the type of relationship pattern, Ashri et al. (2005) also take the context in which a relationship is developing into account. This can be represented by attaching a value between 0 and 1 to each instant of a relationship. Two examples of expressions are Comp(α,β) = 0.9, which means that agent α and β are strongly competing, and Dep(α,β) = 0.5, in which α is average dependent on β. Now it also becomes clear why Coll(α,β) is not a symmetric relation. A collaboration relation exists of two dependency relations, Dep(α,β) and Dep(β,α), which both have their own context values and these can be different. With this new information, the following constraints can be introduced.

5 $\quad \mathrm{Dep}(\beta,\alpha) > \mathrm{Dep}(\gamma,\alpha) \rightarrow P(\mathrm{Accept}(\alpha,\beta,\delta)) > P(\mathrm{Accept}(\alpha,\gamma,\delta))$
6 $\quad \mathrm{Comp}(\alpha,\beta) > \mathrm{Comp}(\alpha,\gamma) \rightarrow P(\mathrm{Accept}(\alpha,\gamma,\delta)) > P(\mathrm{Accept}(\alpha,\beta,\delta))$

From this the following constraint for collaboration relations Coll((Dep(α,β),Dep(β,α)) and Coll((Dep(α,γ),Dep(γ,α)) can be derived.

7 $\quad (\mathrm{Dep}(\alpha,\beta) - \mathrm{Dep}(\beta,\alpha)) > (\mathrm{Dep}(\alpha,\gamma) - \mathrm{Dep}(\gamma,\alpha)) \rightarrow P(\mathrm{Accept}(\alpha,\gamma,\delta)) > P(\mathrm{Accept}(\alpha,\beta,\delta))$

Constraint number 5 states that deals with agents strongly dependent on agent α are preferred to deals with agents not so strongly dependent on α. That is because the consequences of defection for strongly dependent agents are even worse than for normal or little dependent agents. So the stronger β depends on α, the smaller the probability that will β defect. A contrary relation is expressed in constraint 6: the more intense two agents are competing, the bigger is the probability that they will defect. Defecting will happen more often in more intense competition relations because it is becomes more important to undermine other agents.

The last information type discussed in the previous section was social information about reputations, this will be represented with SocReputation(β). The following constraint could be added to the knowledge of an agent.

8 $\quad \mathrm{SocReputation}(\beta) > \mathrm{SocReputation}(\gamma) \rightarrow P(\mathrm{Accept}(\alpha,\beta,\delta)) > P(\mathrm{Accept}(\alpha,\gamma,\delta))$

SocReputation differs from Power in the sense that the latter applies to two particular agents, whereas the former is a general property. The reputation of an agent provided by social information is independent of reputation values of other agents. Constraint 8 advises to prefer deals from agents with higher social reputations to deals with agents with lower social reputations.

5.3 The presentation of new social information

The previous section described some general constraints that could possibly be added to an agent's knowledge base. These constraints are however added to the knowledge base beforehand, not during a negotiation or a session of negotiations. During a negotiation or a session of negotiations new social information can become available from which the agent then has to be updated.

Before the introduction of how new social information will be presented, some choices have to be made. The first question concerns the source of social information: who actually provides social information? Here is assumed that social information is provided by some public institution and that there is only one such an institution. The information it provides is publicly available for all agents. A second question is: can social information be wrong? In the case of reputation information, other agents could cheat on each other and they could provide false information. In most cases, the reason for cheating behaviour is that this would be in the advantage of the cheating agent. An agent can for example present itself more positively, or save money and time by not executing all the things it promised. However, a public institution normally does not have goals such as making a lot of money or selling most products as possible. So the advantages for agents do not apply to a public institution. Another reason for providing false information could be the lack of right knowledge. One possibility is that knowledge is (partly) absent, another possibility is that knowledge is false. To justify the existence of a social institution, it is expected to have a different role than agents and to have access to different information than negotiating agents. Here is assumed that a social institution does not have access to false information and thus never has false knowledge. This makes the answer to the question whether social information can be wrong negative. The two reasons for providing false information, own advantage and wrong information, are argued to not hold for a social institution. The possibility of knowledge being (partly) absent is allowed in this system. In order to deal with this kind of knowledge, a reliability factor is introduced. This factor indicates the amount of information on which a social institution based its statements.

By the illocution Inform (γ, α, *info*), now social information can be introduced in the form SocialInfo(Φ, Ψ, r). SocialInfo(Φ, Ψ, r) returns a value which is the intensity factor of the social information, for example SocialInfo(Φ, Ψ, r) = 0.7. Here Φ indicates the particular domain or institution, Ψ informs about the type of social information and the agent(s) about which the information is given and r informs about the reliability of the information. Five possible Ψ 's (possible information types) are distinghished. For each of them, the meaning and form of SocialInfo(Φ, Ψ, r) is given.

- Ψ = Power(α,β) SocialInfo is a value between -1 and 1 representing the strength of agent α's power over β, in domain or institution Φ, with a reliability r.
- Ψ = SocReputation(β) SocialInfo is a value between -1 and 1 representing the social reputation of agent β, in domain or institution Φ, with a reliability r.
- Ψ = Dep(α,β) SocialInfo is a value between 0 and 1 representing the intensity of agent α's dependency relation with β, in domain or institution Φ, with a reliability r.
- Ψ = Comp(α,β) SocialInfo is a value between 0 and 1 representing the intensity of agent α's competition relation with β, in domain or institution Φ, with a reliability r.
- Ψ = Coll(α,β) SocialInfo is a value between 0 and 1 representing the intensity of agent α's dependency relation with β, in domain or institution Φ, with a reliability r.

5.4 Updating from social information

Updating of the probability distribution from social information will run similar to updating from reputation information received from other agents. The following formula will be used.

$$p(b'|b,\ SocialInfo(\Phi,\Psi,r)) = p(b'|b) + g_4\ (b'|b,\ SocialInfo(\Phi,\Psi,r))\ (1- p(b'|b))$$

In the formula, $g_4\ (b'|b,\ SocialInfo(\Phi,\Psi,r))$ represents the strength of agent α's belief that the probability that the execution of contract b at time $t + 1$ will be preferred to b should change given that the information was received at time t. The calculation of $g_4(b'|b,SocialInfo(\Phi,\Psi,r))$ depends on the specific information type Ψ of the received message. For each possible information type the calculation of $g_4(b'|b,SocialInfo(\Phi,\Psi,r))$ is presented below.

- Ψ = Power(α,β) If the updating agent is α: the value of SocialInfo multiplied by reliability r multiplied by -1.
  If the updating agent is β: the value of SocialInfo multiplied by reliability r.
- Ψ = SocReputation(β) The value of SocialInfo multiplied by reliability r.
- Ψ = Dep(α,β) If the updating agent is α: the value of SocialInfo multiplied by reliability r multiplied by -1.
  If the updating agent is β: the value of SocialInfo multiplied by reliability r.
- Ψ = Comp(α,β) If the updating agent is α: the value of SocialInfo multiplied by reliability r multiplied by -1 multiplied by intensity factor f.
  If the updating agent is β: the value of SocialInfo multiplied by reliability r multiplied by intensity factor f.
- Ψ = Coll(α,β) If the updating agent is α or β: the value of SocialInfo multiplied by reliability r multiplied by intensity factor f.

The second bullet, when the social information informs about the social reputation of another agent, is the only update that can always be applied. The other four information types only lead to an update if the information also includes the updating agent itself. In all of the calculations, the value of SocialInfo is multiplied by the possibly added reliability. If the reliability of the information is not hundred percent the influence of the social information on the probability distribution is somewhat weakened. When the social information is one of the types Power(α,β), Dep(α,β) or Comp(α,β), sometimes the value of SocialInfo(Φ,Ψ,r) is multiplied by one. For example if Power(α,β) = -0.8, agent β has quite a lot of power over α. Now if α is the updating agent, β's power over α should increase α's trust in β. Therefore the information should be multiplied by -1. The last aspect to be explained is the intensity factor in the information types Comp(α,β) and Coll(α,β). This factor is added to distinguish between these updates and updates from dependency relations. In a dependency relation an agent should strongly trust or distrust the other agent; in competition or collaboration relations the intensity of the trust or distrust is somewhat weaker. The intensity factor *f* is a value between 0 and 1, indicating the influence an update from social information about a competition or collaboration relation should have in comparison to an update from social information about a dependency relation.

In each of the five distinguished cases the calculation of $g_4(b'|b, SocialInfo(\Phi,\Psi,r))$ results in a value between -1 and 1. With this value, the probability distribution can be updated according to the principle of minimum relative entropy. This revision is subject to the constraint:

$$\sum P_C^t(x \mid b) = p(b' \mid b,\ SocialInfo(\Phi,\Psi,r))$$

After the update, the social information is applied to the probability distribution.

In the previous chapter, two ways of processing reputation information were discussed: updating trust from reputation information and combining trust and reputation. Both approached were explored, but preference was given to the updating of trust from reputation information. In this chapter about social information, only the updating of trust from social information is discussed. However, social information could also be processed according to the formula below.

$$P^t\,(\text{Accept}\,(\alpha,\beta,\delta)) = \kappa_1\, T\,(\alpha,\beta,\delta) + \kappa_2\, R(\alpha,\beta,\delta) + \kappa_3\, S(\alpha,\beta,\delta),$$

where $S(\alpha,\beta,\delta)$ is representing the influence of social information. Because of the disadvantages of the approach mentioned in section 4.4, this possibility will not be worked out here. A last remark is that when reputation information and social information are used, at least the same method for the processing of both information types should be chosen.

6 The ART test-bed

The ART test-bed is used for testing the information-based trust model. In this chapter, first the choice of this test-bed is argued, then an overview of the ART test-bed and its rules will be given and in the last section will be described how the test-bed will be used in this project.

6.1 The choice of a test-bed

A model of trust and reputation represents trust and reputation and the behaviour involved with these concepts. In a good model, the behaviour it defines and the behaviour it intends to represent should be the same. Therefore, to evaluate a model it has to be applied to a practical situation to test whether the application really results in the desired behaviour. test-beds provide these practical situations: they provide specific domains to which theoretical models can be applied. Tests with a concrete test-bed supply practical experimentation results and if different models are applied to the same domain, they can be compared with each other. So test-beds provide researchers with tools to compare and validate their models and to make more objective and standardized judgements.

Research in computational models of trust and reputation has grown fast in the recent years and a unified research still has to be set. Different models are now being tested with many experimental domains and metrics, and most researchers agree that there are no objective standards. Many of the test-beds are only used in one project, so these test-beds miss wide acceptance. The only sets of experiments used by several authors to compare reputation and trust models under the same conditions are the Prisoner's dilemma and the SPORAS experiments. The problem with these two experimental domains is that they are not rich enough and do not test all facets of trust and reputation. (K. Fullam et al., 2004).

For these reasons, a new attempt has been made to provide transparent and recognizable standards for trust and reputation. The *Agent Reputation and Trust (ART) test-bed* has recently been developed to serve as a test domain for models of trust and reputation (K. Fullam et al., 2004). The ART test-bed tries to satisfy the following characteristics and requirements. The first one is *modularity*, the parameters of the test-bed should be easily adjustable. The design is *multi-purpose*, the test-bed serves as an experimentation environment for a single approach and it can be used for the competition among different approaches. The *accessibility* of the test-bed is high, that is, it is usable for a wide range of approaches and a various numbers of participating agents. The chosen domain is hoped to be *exciting and relevant*, to improve likelihood the domain will be accepted. The test-bed involves *objective metrics*, objective success measures tied directly to the domain problem. The test-bed aims to focus on *one problem*: relevant trust and reputation problems should be addressed, while other research areas are excluded. The developers hope that their test-bed will be known and accepted by a big group of researchers.

The choice of a relative young and new test-bed for the evaluation of the information-based trust model has some disadvantages. It is not yet sure whether the

ART test-bed really will become a widely accepted and used test-bed. Secondly, only a first version of the test-bed will be available and this version might still have some errors. And finally, only very few people will have worked with the test-bed, so there will not be much material of others to compare the results with. On the other hand, the acceptance of the test-bed has to start somewhere and using the test-bed contributes to a wider use of the test-bed. Moreover, problems encountered in experimenting with the test-bed could be used to improve next versions of the test-bed. So in the prospects of a widely accepted standard for trust and reputation models with the broad requirements the ART test-bed states to satisfy, the ART test-bed is chosen for the practical examination of the information-based model of trust.

6.2 Overview of the ART test-bed

The domain of the ART test-bed is art appraisal. Participating agents have to valuate paintings for clients. Each painting in the test-bed has a fixed value, unknown to the participating agents. Agents receive more clients and more profit for producing more accurate appraisals. In the competition mode, each participating researcher controls a single agent which plays against every other agent in the system. After a random amount of game rounds, the winning agent is selected as the appraiser with the highest bank account balance.

All agents have varying levels of expertise in different artistic eras (e.g. classical, impressionist, post-modern), which are only known to the agents themselves and which will not change during a game. The clients request appraisals for paintings from different eras. If an appraiser thinks that it does not have the expertise to make an accurate appraisal by itself, it may gather opinions from other agents to produce better appraisals. Other appraisers provide an estimation of the accuracy of their opinions, determined by the cost they choose to invest in generating an opinion. Appraisers produce their final appraisals by using their own opinion and the opinions received from other appraisers. However, opinion providing agents may lie about the estimated accuracy and they may give false opinions. To help appraisers to know from which other appraisers to request for useful opinions, they may also purchase reputation information from each other. The winning agent, the one with the highest bank account balance, will be the agent who (1) is able to estimate the value of its paintings most accurately and (2) purchases information most prudently.

The ART test-bed is implemented in Java and consists of four components:

- A simulation engine, controlling the simulation environment.
- An agent skeleton, to assist players in implementing strategy codes.
- A database, which stores data calculations and experiment analysis.
- A user interface, through which the games are set up and viewed.

In one period of the game, the following steps are being processed by the simulation engine. First the simulation engine provides *client allocations* to the agents. Appraisers receive lager shares of clients if they have produced more accurate appraisals in the past.

Then the simulation controls *reputation transactions* between different agents, which consist of the following actions. First all agents send their reputation requests, for which they have to decide how many requests to send, to which agents and about whom. Then the receiving agents can accept or decline the incoming requests. In the case a request is accepted, the requested agent is paid and normally reputation information is exchanged between agents. Agents can however choose to cheat and not send the promised and paid reputation information.

The *opinion transactions* that follow are similar to the reputation transactions, although not exactly the same. Agents send their opinion requests to other agents, and then agents reply by sending a decline or a certainty assessment. Certainty assessments inform the other agent about the time (= money) they want to invest in creating the opinion. Agents respond to certainty assessments with a decline or payment to the requested agents. As in reputation transactions, accepted and paid requests normally lead to the exchange of opinions between agents. Here again agents have the possibility to cheat. Before receiving requested opinions, the agents have to send *reputation weights* to the simulation engine, representing the contribution of each agent to the final appraisal. A high weight value means that the opinions of that agent will have much influence on the final appraisal. This order enforces agents to rely on their trust values based on previous experiences in generating the weights, instead of determining their trust after having seen the received opinions. With the reputation weights and the received opinions, the simulation engine then calculates the *final appraisals*. Finally, agents can ask the simulation engine for the true values of the paintings, so they can update their trust model of opinion-providing agents.

During all these transactions the following data of each appraiser at each time-step are stored in the database:

- Client share (the amount of clients of an agent)
- Bank account balance
- Reputation weights the agent associated with other appraisers
- Generated opinions
- The calculated final appraisals

For the environment the database stores for each time-step:

- True painting values
- All the messages exchanged between appraisers.

After executing experiments with the test-bed, the stored data can be analysed and conclusions can be drawn from them.

6.3 Rules of the competition game

In the competition mode (Fullam et al., 2005), several game sessions are played and the winner is selected by averaging the results over all game sessions. The duration each session, in the test-bed represented by *timesteps-per-session,* is randomly determined by the simulation and unknown to the agents. In each game session dummy agents with unknown strategies may be included in the competition.

Initially, clients are evenly distributed among appraisers. When a session proceeds, appraisers whose final appraisals were most accurate are rewarded with a larger share of the client base. The parameter *average-clients-per-agent* is fixed, so the number of all the clients in one game session stays the same. To calculate each appraiser's client share, first an appraiser's relative appraisal error ε_a is calculated. Then each appraiser is assigned a preliminary client share according to its average relative appraisal error. Finally, each appraiser's actual client share is calculated, taking the appraiser's client share from the previous timestep into account. The strength of the influence of the previous client share can be varied by adjusting a *previous-client-share-influence* value.

After the distribution of client shares, appraisers can sell and buy reputation information and opinions from each other. The cost of each reputation transaction is *reputation-cost* c_r and the cost of each opinion transaction is *opinion-cost* c_p; these are both non-negotiable accounts. In general, c_r is lower than c_p to promote the exchange of reputation information. A client pays a fixed *client-fee* f for an appraisal, in general c_p will be smaller than f. If an appraiser accepts a reputation request, it is free to report its own belief or any other opinion about the subject agent. If an appraiser accepts an opinion request, it has to decide about how much time it wants to invest in creating an opinion. The more time it spends in studying a painting, the more accurate the opinion. The appraiser has to pay a variable cost c_g, dependent on the time taken to examine a painting. Then the simulation creates an opinion according to the following error distribution:

$$s = (s^* + \frac{\alpha}{c_g})\, t$$

The expertise of an agent in a certain artistic era is represented by s^*, t is the true value of the painting to be appraised and *sensing-cost-accuracy* α is a parameter which affects the relationship between opinion-generating cost and resulting accuracy. If an appraiser receives an opinion request, it has to provide a certainty assessment to the requesting agent. This certainty depends on c_g, the time an appraiser takes to study a painting; the appraiser is however free to report any certainty.

An agent's final appraisal is calculated by the simulation, to ensure that appraisers do not employ strategies for selecting opinions to use after receiving all purchased opinions. The final appraisal p^* is calculated as a weighted average of received opinions:

$$p^* = \frac{\sum_i (w_i * p_i)}{\sum_i (w_i)}$$

In the formula, w_i is the appraiser's weight for provider i and p_i is the received opinion from provider i. The true painting value t and the calculated final appraisal p^* are revealed by the simulation to the agent. The agent can use this information to revise its trust models of other participants.

6.4 Use of the ART test-bed in this project

Sierra and Debenham's information-based model of trust provides a tool for calculating probabilities, for example the probability that the promises made in a certain deal will be met and the probability that the deal will be accepted. With a not further specified negotiation strategy, actions that depend on the probabilities can be taken. As Sierra and Debenham call it themselves, "the probability distributions provide the fuel for the negotiation strategy" (Sierra and Debenham, 2005).

Agents in the ART test-bed are assessed by the ability to estimate the value of their paintings accurately and by the prudent purchase of information. The estimation of the painting values mainly depends on the ability to estimate the probabilities of the outcomes of certain actions. Will agent α provide the requested opinion or will it cheat by not sending any opinion? Will agent β be able to provide opinions of a good quality in era e or does it not have a high expertise in that era? Is it worth to invest in the opinion agent

γ requested or is there a big chance it will not buy the opinion? A prudent purchase of information is accomplished by making the right decisions about the actions to take, the strategy of an agent. What is the minimum level of trust for buying opinions from other agents? Will I cheat on other agents in order to gain more money? Will I sell opinions to agents that cheat on me?

The information-based trust model provides a way to estimate the probabilities of uncertain outcomes. So in the ART test-bed, its performance can be evaluated by paying attention to the accuracy of the appraisals. The better probabilities of possible future actions can be calculated, the easier it will be to deliver correct appraisals. The trust model does not tell how to deal with the second ability requested in the test-bed, the prudent purchase of information. The model does not provide a strategy of how to act in different situations. The goal in this research is to evaluate the information-based trust model, so most attention will be paid to the accuracy of the agents' appraisals and less attention is paid to their bank account balances.

The test-bed can be used in two modes, the competition mode and the experimentation mode. In the competition mode agents with the highest bank account balance perform best and win the game, in the experimentation mode one can choose the object of evaluation. In testing information-based trust model, the first interest does not go to the agent with the highest bank account balance so the experimentation mode of the test-bed is used. However, in the experiments most parameters are set as in the competition mode. For example, the same opinion and reputation costs are used and the final appraisals are calculated in the same way as explained in the previous section.

7 The test-bed agents

Several agents for the ART test-bed have been build. In the first section of this chapter will be explained how to build a test-bed agent. The second section will describe how the methods of the information-based trust model has been applied to an agent that can participate in the ART test-bed. Section three describes the behaviour of the information-based test-bed agent. The last section discusses some of the other agents that have been built; all of them are variations on the information-based agent.

7.1 Building a test-bed agent

The ART test-bed provides an abstract base class *Agent* for constructing new agents that can participate in the test-bed. To create a new agent, users have to implement a class *Participant* that extends the *Agent* class. A participant consists of ten methods in which a researcher can code the agent's strategy. In a game, the simulator calls the same method for all agents before moving to the next method. The ten methods are called by the simulator in the following order.

```
initializeAgent()

prepareReputationRequests()
prepareReputationAcceptsAndDeclines()
prepareReputationReplies()

prepareOpinionRequests()
prepareOpinionCreationOrders()
prepareOpinionCertainties()
prepareOpinionRequestConfirmations()
prepareOpinionProviderWeights()
prepareOpinionReplies()
```

The first method *initializeAgent()* is different from the other nine methods. It is not a strategic method and it will only be called once, at the start of a new game. This method gives the agent the opportunity to initialize data structures. All other methods are strategic methods and called each round of a game.

An agent's behaviour during the reputation transactions are determined by three methods. In the first one, the method *prepareReputationRequests()*, an agent has to decide which agents to request for reputation information about which other agents in what eras. When the request messages are sent, the simulator redistributes them and sends each message to the right agent. In *prepareReputationAcceptsAndDeclines()*, an agent replies to each request with an acceptance or a decline message. From the moment an agent accepts a request, the requesting agent has to pay the fixed reputation cost to the reputation providing agent. In *prepareReputationReplies()*, agents generate replies for the requests they accepted. The replies are sent to the simulator and then the simulator processes the exchange of all the generated replies.

The opinion transactions in one game round are covered by six methods. The first of these methods, *prepareOpinionRequests()*, runs similar to the preparation of reputation requests. An agent has to decide which agents to ask for opinions about paintings in what eras. The opinion requests are sent and redistributed by the simulator. After receiving opinion requests, in *prepareOpinionCreationOrders()* agents have to decide for which agents they will order opinions from the simulator and how much they are prepared to pay for these opinions. The more money is paid for an opinion, the better it will approximate the true value of the appraised painting. At this point an agent is not sure whether the requesting agent will really buy the opinion, so it should consider carefully in which requests it will invest. After sending the opinion orders to the simulator, agents have the opportunity to send opinion certainties to the requesting agents in *prepareOpinionCertainties()*. These certainties give information about the probability the created opinions will be true or close to the truth. The requested agent can choose to lie and send a higher certainty than it expects to provide. After receiving certainties from the requested agents, an agent has to decide which deals to accept and which to decline in *prepareOpinionRequestConfirmations()*. When an agent accepts a deal it has to pay for that opinion, even when the provider later does not provide the promised opinion. When the deals are fixed, agents have to send weights to the simulator for all agents in all eras in *prepareOpinionProviderWeights()*. The agent's final appraisals are based on the weighted sum of the opinions of the requested agents. An agent might also weight its own opinions based on its own expertise. The simulator calculates with zeros for not received weights. The last method of a game round, *prepareOpinionReplies()*, is for generating and sending opinion replies to requesting agents. This is done after sending the weights, so that agents cannot eliminate bad or not received opinions afterwards and they have to rely on their trust in other agents when they determine the weight values.

7.2 Application of the information-based model to an agent

A key decision in applying the information-based model to an agent in a specific situation is the choice of what probability distributions to use. This decision is made on the basis of what the agent is supposed to be able to do. Then, when new information arrives, the probability distributions have to be updated. Functions are needed to translate actions and events of the agent and its environment into constraints on a probability distribution. After this translation step, the probability distributions can be updated from sets of linear constraints.

The probability distributions of the information-based test-bed agent describe probabilities to the quality of the opinions other agents could provide. The agent keeps up a probability distribution for each era of expertise for each agent, so the amount of probability distributions in the model is the number of agents multiplied by the number of eras. The different possible worlds in a probability distribution represent the possible grades of the opinions an agent might provide in a specific era. An opinion of high grade means that the appraised value of a painting is close to the real value of the painting. A low grade means that the agent provides very bad opinions in the corresponding era or that the agent does not provide opinions at all. The quality of an opinion actually is a continuous variable, but to fit the model it is made discrete. All possible opinions are grouped into ten different levels of quality. The act of promising but not sending an opinion is classified in the lowest quality level.

The probability distributions are updated during the course of a session each time the agent receives new information. These updates are from the following three types of information.

- Updating from direct experiences
- Updating from reputation information
- Updating from the evaporation of beliefs

The first two types of updating take place when the agent receives the true values of paintings (updating from direct experiences) and when it receives witness information (updating from reputation information). Both types of messages are translated into constraints that can be put on the probability distributions, and then the updated probability distributions are calculated with these new constraints. The third type of updating, updating from the evaporation of beliefs, is performed each time before a probability distribution is updated either from direct experiences or from reputation information. There is no updating from social information, because social information is not provided in games of the ART test-bed.

Direct experiences and reputation information are translated into the same type of constraints. Such a constraint is for example: "agent α will provide opinions with a quality of at least seven in era e with a certainty of 0.6". This constraint is put to the probability distribution of agent α and era e. After updating from this constraint, the probabilities of the worlds 7, 8, 9 and 10 should together be 0.6. Constraints on a probability distribution are always of the type 'opinions of *at least* quality x'. In the information-based agent it is not possible to put constraints like 'opinions with a quality between x_1 and x_2' or 'opinions worse than quality x' on a probability distribution. This could be an improvement in more advanced versions of the agent.

This way of expressing might cause a positive bias in the probability values and in the trust values derived from these probability values. But because all the probability distributions have the same bias, the bias will disappear when comparing different probability distributions with each other. Another solution could be to multiply trust values by a correcting factor. One more point of using constraints of the type 'at least quality x' is that constraints could sometimes be counterintuitive. A constraint with a quality of eight with a certainty of 0.9 expresses more information than a quality of two with the same certainty. A quality of at least two can still be a quality of three or a quality of ten. So to express strong negative expectations, although high certainties might be expected, low certainties should be used. For example a quality of three with a certainty of 0.1 means that the probabilities of world 1 and 2 together should be 0.9. In the implementation this problem has been overcome by always using constraint certainties of 0.5. A constraint with a quality of two or one with a quality of eight with a certainty of 0.5 do express the same amount of information.

The value of a constraint (the quality grade) derived from a direct experience is obtained by comparing the real value of a painting to an agent's estimated value of a painting. The relative error of an opinion is calculated by taking absolute difference between the real and appraised value divided by the real value. The value of one minus this error, multiplied by ten represents the quality of the opinion and a new constraint can be added to the set of beliefs.

$$constraintValue = 10 * (1 - \frac{|appraisedValue - trueValue|}{trueValue})$$

In most cases this formula turned out to deliver values between one and ten. Ten is the highest possible constraint and can directly be added to the set of beliefs. If a value lower than one is found, a constraint with the value of one is added to the set of beliefs.

The value of a constraint is derived from reputation information by taking the average of the reputation values in all messages received at a specific time from trusted agents about a specific agent and era multiplied by ten.

$$constraintValue = 10 * \sum_{r \in reps} \frac{r}{n},$$

where r is a message with reputation information, *reps* the set of messages with useful reputation information and n the size of *reps*.

With a set of constraints and the principle of maximum entropy, an actual probability distribution can be calculated. Ideally, the principle of maximum entropy finds a solution that satisfies all the separate constraints, but with the code[4] used this was often not the case. Even with a quite small number of constraint often no solution was found, only with just one constraint it always found a satisfying probability distribution. So the choice has been made to derive one general constraint from all the stored constraints for calculating the probability distribution. Besides the practical advantage of always finding a solution, two more advantages will be explained in the next two paragraphs.

First, it becomes easier to influence the relation between updating from direct experiences and from reputation information. When the general constraint is calculated, constraints obtained from reputation information are weighed with a factor. This factor determines their importance in relation to constraints obtained from direct information. In the information-based agent the relation of importance between constraints from direct experiences and from reputation information is 1 in proportion to 0.3. This ratio is based on the assumption that constraints from direct experiences are more important than constraints from reputation information, because the former constraints are derived from first-hand experiences and therefore more trustworthy. The values 1 and 0.3 have been chosen after some tests with several ratios.

A second advantage of calculating one general constraint is that this gives a nice opportunity to update from the evaporation of beliefs. Constraints derived from direct experiences or reputation information are always stored with the timestep in which they were obtained. This time information is used when the general constraint composed of all stored constraints is calculated. The stored timesteps of the constraints are subtracted from the current timestep and this way the 'age' of each constraint can be determined. In composing the general summarizing constraint, all constraints are weighed according to their age. By giving younger constraints more influence on the probability distribution than older constraints, the evaporation of beliefs is modelled. The general constraint is a

[4] Code written by John Debenham

weighed average of all the constraints stored so far, calculated according to the following algorithm.

$$general\ constraintValue = \frac{1}{n} \times \sum_{c \in C} \frac{1}{(\mathrm{c}(\mathrm{t}_{\mathrm{obtained}}) - t_{current}) + 1} \times \mathrm{c(value)},$$

where constraint *c* is an element of the set of stored constraints *C* and *n* the total amount of constraints. Each contraint *c* consists of the time it was obtained $\mathrm{c}(\mathrm{t}_{\mathrm{obtained}})$ and a quality grade c(value) calculated with one of the formulas *constraintValue* on the previous page. The constraints are weighed with a factor of one divided by their age plus one. The one is added to their age to avoid fractions with a zero in the denominator, in the case of new constraints. The general constraint can be applied to John Debenham's maximum entropy code and a new and updated probability distribution will always be found.

Finally, when all information available has been processed and the probability distributions are up to date, trust values can be derived from the probability distributions. There are two types of trust, the trust of a particular agent in a specific era and the trust of a particular agent in general. The trust value of an agent in a specific era is calculated from the probability distribution of the corresponding agent and era. To derive trust, the notion of a most ideal probability distribution is used. The most ideal probability distribution is one in which the probability of getting opinions with the highest quality is very high and the probability of getting opinions with qualities lower than the highest quality is very low. In the implementation, in the 'most ideal' probability distribution the quality categories one through nine all have a probability of 1/18 and quality category ten has a probability of 1/2. Now trust can be calculated by taking one minus the relative entropy between the most ideal and the actual probability distribution.

$$trust(agent, era) = 1 - \sum_{i=1}^{n} P_{actual}(i) \times \log \frac{P_{actual}(i)}{P_{ideal}(i)},$$

where *n* is the number of probabilities. In the implementation the number of probabilities is always 10, one probability for each quality category. The trust of an agent in general is calculated by taking the average of the trust values of that agent in all the eras. Because all eras of expertise have the same importance in the test-bed, the trust values of the different eras equally contribute to the general trust value of an agent.

$$trust(agent) = \sum_{e \in eras} \frac{1}{n} \times trust(agent, e),$$

where *e* is an era, *eras* the set of all eras and *n* the size of *eras*.

7.3 Behaviour of the information-based agent

At each moment of the game, the agent can consult its model to determine the trust value of an agent in general or the trust value of an agent in a specific era. These trust values guide the behaviour of the agent. The information-based agent uses 0.5 as the critical value for trust, it only trusts agents with a trust value of 0.5 or more. This value is chosen after some experimentation with different trust values and values around 0.5 turned out to deliver the best results.

At the beginning of a new session the agents starts with trusting all agents, the probability distributions are initialized such that all derived trust values (for each agent in each era) are 1.0. The information-based model is updated with new constraints two times in a game round. In the method *prepareReputationRequests()*, the first method after the opinion transactions in the previous round, the model is updated from direct experiences. This can either be a comparison of a true value with an appraised value or an update from a promised but not received opinion. In *prepareOpinionRequests()*, the first method after the reputation transactions, the model is updated from reputation information. Updating from reputation information only takes place if the trust in the reputation information providing agent is higher than 0.5.

The general behaviour of the information-based agent is honest and it is cooperating towards the agents it trusts. Therefore it makes use of the trust values several times in a game round. In the methods *prepareReputationRequests()* and *prepare-OpinionRequests()* the agent buys relevant opinions and reputation messages from all agents it trusts (agents with a trust value of 0.5 or higher). Only if the provided certainty of an opinion is smaller than 0.3 it does not accept the opinion, in all other cases it will buy the requested reputation or opinion. In the experiments of this research project this will not lead to enormous costs, because the maximum amount of participating agents will be four. However, in experiments with bigger amounts of participants, the purchase of opinions should be restricted. The agent only accepts (*prepareReputationAcceptsAnd-Declines()*) and invests in (*prepareOpinionCreationOrders()*) requests from agents it trusts, and if the agent accepts a request it provides the best possible requested information. If the agent does not trust a requested agent and will not provide requested information, it informs the other agent by sending a decline message (*prepare-ReputationAcceptsAndDeclines()*) or by sending a low opinion certainty of 0.3 (*prepareOpinionCertainties()*). In the case of a reputation request from a trusted agent, the agent provides the trust value its model attaches to the subject agent. If the agent trusts an agent requesting for opinions, it always highly invests in ordering opinions from the simulator for that agent. The corresponding certainty value the agents sends, is the trust it has in itself in that era. Finally, the agent uses the model for generating weights in *prepareOpinionProviderWeights()*, it weights each agent (including itself) according to the trust in that agent in that era.

7.4 Variations on the information-based agent

The information-based agent described in the previous sections updates from direct experiences, from reputation information and from the evaporation of beliefs. To test the influences of the use of different types of information, three variations on this agent have been made. Chapter 8 will describe the experiments that will be done with these agents. The first agent is an agent only updating from direct experiences; a second agent updates from direct experiences and from the evaporation of beliefs; a third agent updates from reputation information and the evaporation of beliefs. The suffixes in the names of the agents indicate the information types they use for updating: *de* means updating from direct experiences, *rep* means updating from reputation information and *time* is updating from the evaporation of beliefs. This naming results in the following four information-based agents.

- Info-de
- Info-de-time
- Info-rep-time
- Info-de-rep-time

The agent Info-de does not take the evaporation of beliefs into account and all information from the past equally contributes to the probability distributions. Info-de and Info-de-time do not update from reputation information and consequently they do not request for reputations. The choice is made to keep these two agents from any participation in reputation transactions at all, so they neither promise nor provide reputation information to other agents. Info-rep-time does sell and purchase opinions, but it does not update from the information it could derive from comparing these opinions with the real values. Its model is only updated from reputation information.

Info-de-time has been picked out for a comparison of the information-based model with other ways to guide behaviour and obtain final appraisals. Info-de-time is chosen instead of the more complete Info-de-rep-time because of simplicity reasons. Building agents updating from two information types is easier than building agents updating from three information types. More important, the behaviour and processes of agents that update from two information types is easier to explain than that of agents updating from three information types. Two variations on Info-de-time have been made, namely two agents which are not based on information theory, but do resemble Info-de-time in some aspects. One is a very simple agent that does not have a model at all and the other agent has a theoretical model with a more game-theoretical background. A second set of experiments will be performed to compare the three agents below with each other.

- Info-de-time
- Basic
- Game

In general the behaviour of these three agents is equal, but the lack of a theoretical model in the agent Basic has some consequences for its behaviour. The agent is not updated from reputation information, direct experiences or the evaporation of beliefs, so it does not have any beliefs based on its history. The Basic agent cannot base its behaviour towards other agents on a model, so a general attitude has to be selected beforehand. In order to prevent the agent from misuse by other agents, it has been chosen to never trust other agents. Without a model and the capability to learn, the only way to protect oneself against cheating agents is to never trust any agent.

The resulting behaviour is that the Basic agent never makes reputation or opinion requests itself and that it never accepts requests from other agents. As the Info-de-time agent the Basic agent is honest, so it also informs other agents by sending a decline message in the case it received a reputation request and a low certainty in the case of an opinion request. The Basic agent only orders opinions from the simulator for itself and for its final appraisals the agent just relies on its own expertise.

As the agent Info-de-time, the agent Game updates its model from direct experiences and from the evaporation of beliefs. From its model it can also derive the general trust value and the era specific trust value of each agent. The way of updating and deriving trust values in the Game is different from the information-based agents and will be explained in the next paragraph. As Info-de-time, the agent Game uses the value 0.5 as a threshold for trust. Its decisions based on the derived trust values are the same as the decisions the agent Info-de-time would make with the same values; it acts exactly in the same way to agents it does and it does not trust. The Game agent also uses its model equally as the information-based agents to provide information to other agents and to determine the weights for final appraisals.

However, the model of the Game agent is not based on information theory and updating from direct experiences works different than in the model of the information-based agent. Both agents store each past experience, in Info agents these are stored as constraints (section 7.3) and in the Game agent they are called interactions. The key information of such a belief is one minus the relative error of an opinion[5].

$$belief = 1 - \frac{appraisedValue - trueValue}{trueValue}$$

Information-based agents weight all the constraints according to their age, calculate probability distributions from them and then they can derive trust values. The Game agent also weights its beliefs according to their age and then it is able to directly derive trust values. Trust in the agent Game is the average of the weighed beliefs.

$$trust = \sum_{i=1}^{n} \frac{b_i}{n}$$

In the formula, b_i is a belief and n is the total amount of all beliefs. Beliefs in the agent Game are weighed according to their age in the same way as constraints are weighted in Info-de-time (see section 7.3). So updating from the evaporation of beliefs is the same for both agents, and updating from direct experiences runs differently.

The Game agent is not designed with game theory as its starting point, its model is derived from the information-based model. However, the model can be described in game-theoretical terms very well and that is why the agent is called Game. In game-theoretical terms, all values stored as beliefs are called utilities. One minus the relative error represents the utility or amount of satisfaction the agent gets from a received opinion: the smaller the error the more satisfaction for the agent. The utility (or error) a requested opinion will deliver is unknown beforehand, therefore the Game agent

[5] In the information-based model, the value of a belief/constraint is multiplied by ten.

calculates a probable utility the requested opinion could provide. This probable utility u is the weighted average of all beliefs about agent α in a certain era: the trust in α in that era. If this u is higher than a certain margin m, the agent is willing to cooperate. In the Game agent a margin value of 0.5 is used; this value is the cut-off for trust.

8 The experiments

In this chapter eight hypotheses will be introduced and explained, the methods of the experiments to test the hypotheses will be described and the results of the experiments will be presented.

8.1 Hypotheses

The information-based model is claimed to deal especially well with cases of uncertain information (Sierra and Debenham 2005). Despite the availability of only little information, the model should be able to derive sensible conclusions. That is why the most important measurement of success in the test-bed experiments will be the accuracy of the information the agent provides. If Sierra and Debenham are right, an agent based on information theory should be able to make good appraisals even though information is only partly available.

The evaluation of the information-based model of the agent will consist of two parts. First the contribution of different types of information in the model will be investigated. The three types of information that will be examined are the different types of updating discussed in section 7.3: updating from (1) information based on direct experiences, (2) information based on reputation and (3) information based on the evaporation of beliefs as time goes by. In the second part of the experiments, the information-based agent Info-de-time will be compared with two agents with other theoretical models, the agents Game and Basic. From now on, the name *Info* will refer to the agent Info-de-time. Besides the accuracy of these three agents' appraisals, they will be compared on some other points. Their overall functioning will be described by measures on the following four aspects.

- Accuracy of provided information
- Performance in the ART test-bed
- Adaptation to new situations
- Efficiency

The expectations about the outcomes of the experiments are formulated in eight hypotheses, four about the use of different information kinds and four about the comparison of agents with different theoretical models. In the rest of this section these hypotheses will be presented, each of them followed by a short discussion.

1 The use of information from direct experiences will increase the average appraisal accuracy of an information-based test-bed agent.

Most agents participating in the ART test-bed will behave according to certain patterns, so their future behaviour can be predicted by using information from the past. The most direct way an agent knows about another agent's past behaviour is by examining its own past experiences with that agent. Hypothesis 1 supposes that an agent with an information-based model updating from direct experiences is able to predict other agent's behaviour and to use this information for anticipating its own behaviour towards that agent in order to obtain a higher average appraisal accuracy.

2 *The use of information from the evaporation of beliefs as time goes by will increase the average appraisal accuracy of an information-based test-bed agent.*

Sierra and Debenham stress the importance of modelling the evaporation of beliefs as time goes by. They argue that without an ongoing relationship we somehow 'forget' how good the opponent was (Sierra and Debenham 2005). Moreover, agents might change their behaviour during a game. To quickly adapt to this new behaviour, it is important to pay more attention to recent experiences than to past experiences.

3 *The use of reputation information will increase the average appraisal accuracy of an information-based test-bed agent.*

Every agent in a game has its own experiences with other agents. Information derived from these experiences is often used for decisions about how to act in the future. For a particular agent, knowledge about the experiences of other agents could also be useful for making its own decisions. Estimates are usually more accurate when they are based on more (reliable) information. In the test-bed, the interchange of reputation values is a way to learn about other agents' experiences. Not all agents keep up reputation values and not all provided reputation information is useful. However, if an agent finds out during a game which agents provide valuable reputation information, it could use this information to make more profitable decisions and improve its performance.

4 *Optimum average appraisal accuracy will be reached by using all available types of information: information from direct experiences, information from the evaporation of beliefs as time goes by and reputation information.*

It is expected that the three discussed information types all contribute to an increase of the average appraisal accuracy. Accordingly it is expected that a combination of the three types will yield the best test-bed results concerning average appraisal accuracy.

The next four hypotheses are expectations about experiments in which the agents *Info*, *Game* and *Basic* discussed in section 7.4 will be compared with each other.

5 *On average, appraisals of the agent Info will be the most accurate and appraisals of the agent Basic will be the least accurate.*

The information-based agent Info and the game-theoretical agent Game both learn from their previous experiences, whereas the agent Basic without a theory does not. That is why the agent without a theory is expected to provide appraisals with the least accuracy. The information-based agent is expected to be the most accurate agent, because its trust model takes the certainty of information into account. The Game agent's trust values of other agents are just determined by the quality of their previous appraisals. Trust values of the information-based agent represent the uncertainty of a specific expectation. So the agent Info is not only guided by quality, but also by the certainty of that quality.

6 *On average, the test-bed performance of the agent Info will be the highest and the test-bed performance of the agent Basic will be the lowest.*

Purchase behaviour is the same for the agents Info and Game, so they will spend almost the same amount of money on buying opinions. Their final bank account balances, which indicate the test-bed performance, are therefore only determined by their client shares. The size of a client share as described in chapter 6 is determined by the accuracy of an agent's appraisals, so as in the previous hypothesis it is expected that the information-based agent will perform better than the game-theoretical agent. The agent Basic will probably spend less money on buying opinions than the other two agents, but it is expected that this saving will not compensate for its worse appraisals.

7 *The agents Info and Game will show adapting behaviour when other agents change strategy, the Basic agent will not.*

The agent Basic does not update a model, so in equal situations it will show the same behaviour during the whole course of a test-bed game. Although another agent might change its strategy, the agent Basic will not adapt its behaviour and it will act in the same way before and after the change. The agents Info and Game update their model from new information during the whole course of a game. Moreover, they pay more attention to recent information than to older information. This yields changes in their own behaviour when their environments changes.

8 *The agent Info will have the highest computational costs and the agent Basic will have the lowest.*

The agent Basic does not use a theory and therefore does not have to update any theoretical model, so its computational costs will be the lowest. The agent Info has the most complex model; updating this model and determining trust values from it cost more calculation steps in this agent than in the agent Game. It is possible to prove this hypothesis theoretically, but an experiment indicating the computational costs an agent uses will also be performed.

8.2 Methods

In order to verify the hypotheses, besides the agents discussed in section 7.4, four extra agents have been implemented. The agents to be tested will participate in test-bed games together with one or more of these test-agents and then their behaviours will be measured and compared with each other.

The first test-agent is called *Cheat*. This agent never makes reputation or opinion requests itself, but when the agent receives requests it always promises to provide the requested reputation information or opinions. The agent even promises to provide opinions with a certainty of 100 percent. As its name suggests, the agent cheats on the other agents and it never sends any promised information. Its final appraisals are just based on its own expertise.

The agent *Naive* bases its behaviour on the idea that all agents it will encounter are trustworthy and Naive keeps on trusting other agents during the whole course of a game. This agent always requests every other agent for reputation information and opinions, it accepts all requests from other agents and it highly invests in creating the requested opinions. Its final appraisals are based on its own expertise and on the (promised but not received) opinions of all other agents.

A third agent is developed to investigate other agents' ability to adapt to new situations. This agent *Changing* shows the same behaviour as Naive during the first ten rounds of a game. Then it suddenly changes it strategy and from the eleventh game round till the end of the game it behaves exactly the same as the agent Cheat.

Updating from reputation information is only of use if there are agents in the game that provide reputation information. So to test the updating from reputations, a reputation information providing agent *Providing* has been implemented. This agent actually is almost the same as the agent Info-de-time. The only difference is that the Providing agent always accepts reputation requests and provides the wished reputation information, whereas the agent Info-de-time only provides reputation to agents it trusts. All the agents that will be used in the experiments have been discussed now. As a reminder below follows a short overview with the ten agents: Information-based agents (section 7.2, 7.3 and 7.4), non information-based agents (section 7.4) and agents used for testing (section 8.2).

Information-based agents:
Info-de
Info-de-time
Info-rep-time
Info-de-rep-time

Non information-based agents:
Basic
Game

Agents used for testing:
Cheat
Naive
Changing
Providing

For the evaluation of the different information-based agents, the following combinations will be run in the test-bed. Each X represents a different test-condition.

	Cheat + Naive	**Changing**	**Cheat + Naive + Providing**
Info-de	X	X	
Info-de-time	X	X	X
Info-rep-time			X
Info-de-rep-time	X		X

Table 8.1

To compare the agents Info, Game and Basic, they will participate together in test-bed games in the following five combinations. In this table, each X represents a participating agent.

	Info	Game	Basic	Cheat + Naive
Condition 1	X	X	X	
Condition 2	X	X	X	X
Condition 3	X	X		X
Condition 4	X		X	X
Condition 5		X	X	X

Table 8.2

Finally, the three agents Info, Game and Basic will be evaluated on how they perform apart from each other in games with test-agents. In this table an X again represents a different test-condition. In the last condition the three agents have to play against (or with) themselves.

	Cheat + Naive	Changing	Itself
Info	X	X	X
Game	X	X	X
Basic	X	X	X

Table 8.3

The ART test-bed allows researchers to set a lot of parameters, of which the relevant ones have been discussed in section 6.3. In this research the following values will be used and kept constant over all the sessions.

Timesteps-per-Session	=	20
Number-of-Painting-Eras	=	3
Average-Clients-Per-Agent	=	20
Client-Fee	=	100.0
Opinion-Cost	=	10.0
Reputation-Cost	=	1.0
Sensing-Cost-Accuracy	=	0.5
Previous-Client-Share-Influence	=	0.5

Besides that, for all the agents used in the experiment it holds that if an agent invests in creating an opinion, it pays a price of 10.

The most important evaluation aspect of the experiments will be the *accuracy of information* the agents provide. This measure will be used in both parts of the experiments, the investigation of different versions of the information-based agent and the comparison of the information-based agent with other agents. Accuracy will be measured by an agent's average appraisal error and the number of clients it has. After each game round when an agent sent his weighting factors to the simulator and the simulator calculated the final appraisals, average appraisal errors and client shares are determined by the simulator. The average appraisal error is the average of the relative appraisal errors of all final appraisals of a particular agent in one game round. The error will be low if an agent estimated well which agents to request for opinions and the weights to attach to them. A disadvantage of the use of the average appraisal error is that these values are only shown in graphs which are hard to exactly read, and not stored as numbers in the database. This problem has to do with the newness of the ART test-bed and will probably be remedied in proximate versions. Fortunately, an agent's clients share can also be used as an indicator of accuracy. A client share is determined by an agent's average appraisal error and its client share in the last game round, and these values are stored in the database each game round. In the results, the client share of the last game round will be displayed.

The next three evaluation points will only be used in the second part of the experiments, to describe the comparison between the information-based agent and agents based on other approaches. In the competition mode of the ART test-bed, agents are judged on their general *performance in the ART test-bed* and this is indicated by an agent's bank account balance. Bank account balance is partly determined by the accuracy of an agent's appraisals; the other part is determined by the purchase of information. In the case of equal quality of appraisals, the agent spending least money on buying opinions will have the highest test-bed performance. The information-based model does not provide strategies to manage opinion purchase, it only helps to provide useful information in order to make good strategic decisions. So to test the information-based model, test-bed performance plays a subordinate role to accuracy in the experiments.

The third evaluation point, *adaptation to new situations*, refers to an agent's ability to adapt to a changing situation. Agents have to participate in games together with the agent Changing, which starts cheating after the tenth game round. The weight values the agents ascribe to the agent Changing and their average appraisal errors will be examined. The weight values indicate whether an agent notices the change. Moreover, from information about the average appraisal errors it can be concluded whether the agent is able to adapt to the new situation.

The last aspect, *efficiency*, will be derived from the computational costs an agent needs for decision making. The more computational costs its processes use, the less efficient the agent. Computational costs will be derived with help of the java method *System.currentTimeMillis()*, which returns the current time in milliseconds. The method is called at the beginning and at the end of each of the ten agent methods. In most cases the same time will be returned, but sometimes a time difference of 15 or 16 milliseconds between the beginning and the end of a method will be found. In a game with the agents Info, Game and Basic these 'time-jumps' will be counted for each agent. The agent with the most time-jumps in its methods is supposed to have the highest computational costs and will be the least efficient agent.

8.3 Results

The results of the experiments will be presented in two forms. The graphics of some representative sessions will show how the agents' average appraisal errors and bank account balances develop during a game. Secondly, tables with the averages[6] of all the sessions per condition will give information about the agents' situations in the final game round. In the tables, *Client* refers to the final client share of an agent and *Bank* means its final bank account balance.

[6] The original results are presented in the Appendix

In the first experiment, each of the agents Info-de, Info-de-time and Info-de-rep-time participated in a test-bed game together with the agents Cheat and Naive. The graphics in figure 1 and figure 2 show an example of a session with the agents Info-de-time (blue), Cheat (green) and Naive (red).

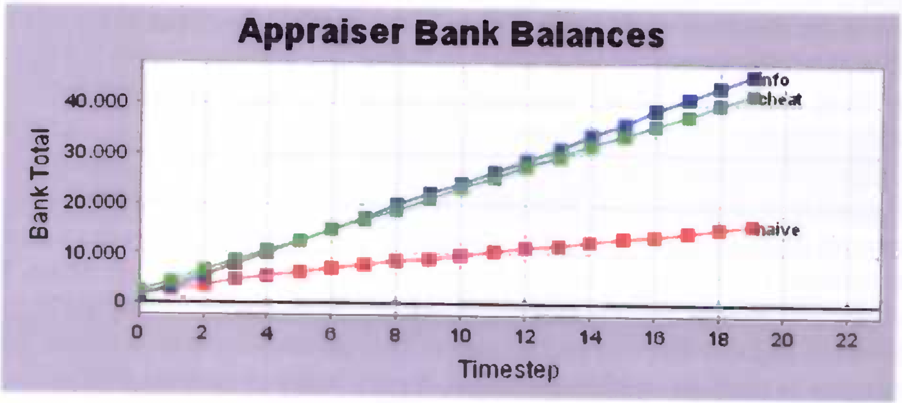

Figure 1

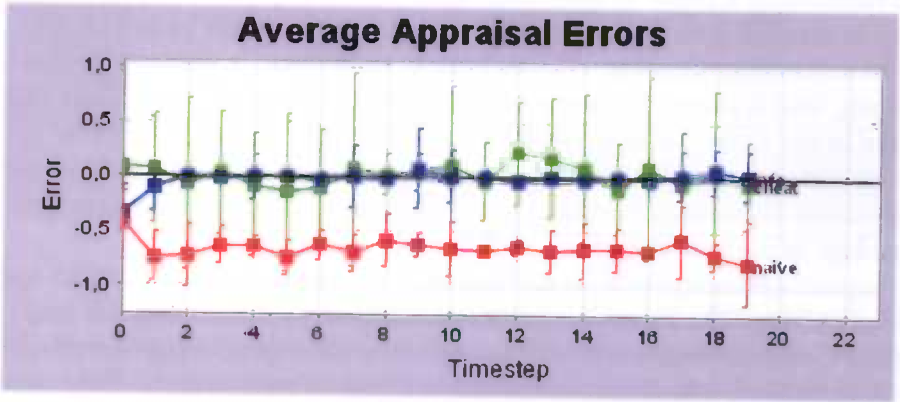

Figure 2

Figure 1 shows the agents' bank account balances during the whole game. Info-de-time ends the game with the most and Naive with the least money. The right figure shows the average appraisal errors of the agents in each round. The appraisals of Naive are obvious less accurate than the appraisals of the other two agents. This can be explained by Naive's behaviour to keep on trusting the cheating agent during the whole game. The agent Info-de-time provides its least accurate appraisals the first game round; there it still has to 'learn' that it cannot trust the agent Cheat. After that its appraisals are the most accurate, the errors are close to the zero line and show the least deviation.

The information-based agent does not always perform as successfully as in the previous example. Sometimes the agent seems to be 'confused' and to ascribe wrong trust values to the other agents. Figure 3 and 4 show an example of such a session with the agents Info-de-time (blue), Cheat (green) and Naive (red).

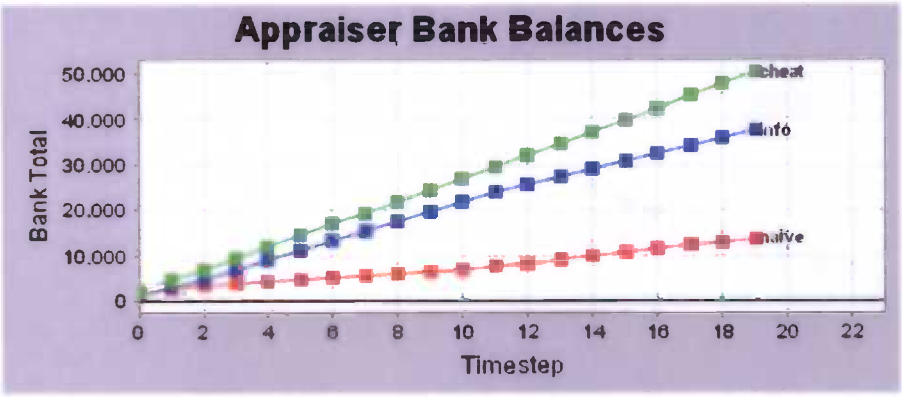

Figure 3

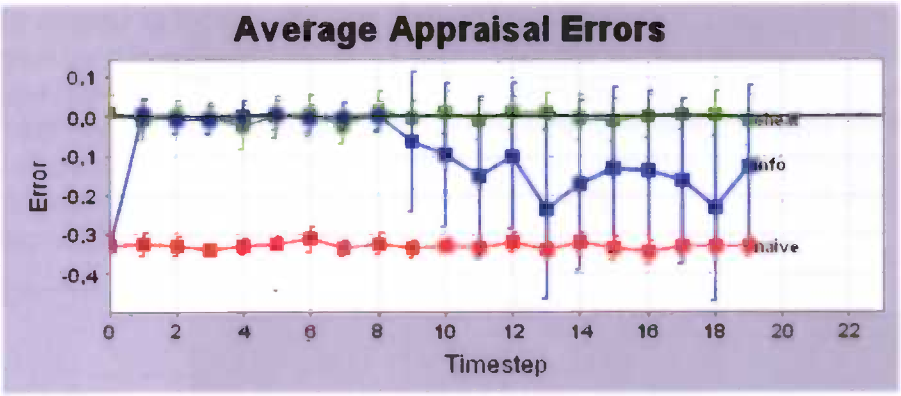

Figure 4

In the first part of this session everything seems to go well, but after round xx Info-de-time starts making errors. These errors are not caused because Info-de-time ascribes too high trust values to Cheat, but because the values it ascribes to itself and Naive are too low. The average results of all sessions in the three conditions are presented in the table below.

	Cheat		**Naive**		**Agent**	
	Client	Bank	Client	Bank	Client	Bank
Info-de	23.6	45070	11.2	16298	**25.5**	40192
Info-de-time	22.4	42828	11.6	16480	**26**	41772
Info-de-rep-time	21.2	41974	10.4	15804	**28**	43070

Table 8.4

In another condition, especially designed to test the effect of updating from the evaporation of beliefs as time goes by, Info-de and Info-de-time both participated in a game with the agent Changing. The graphics in figure 5 and 6 show a representative example of the development of the bank account balances and the average appraisal errors of the agents Info-de (red) and Changing (blue).

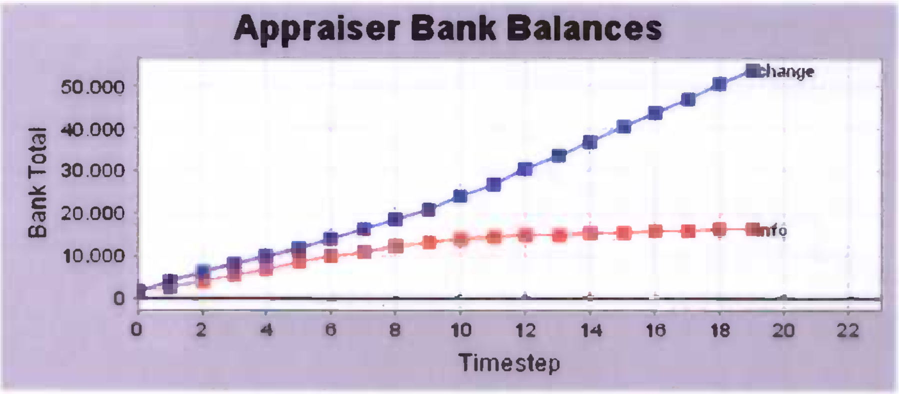

Figure 5

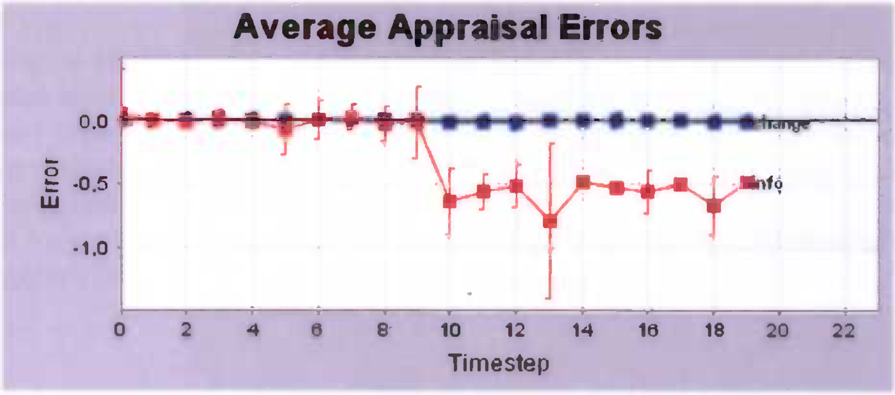

Figure 6

From the tenth round of the game the agent Changing starts to cheat. This is clearly visible in figure 6, after the tenth[7] round the accuracy of Info-de's appraisals decreases a lot. Although the agent learns from new information, information from the past also strongly contributes to the trust value it ascribes to Cheating. Figure 7 and 8 show an example of the bank account balances and appraisal errors of Info-de-time (red) and Changing (blue).

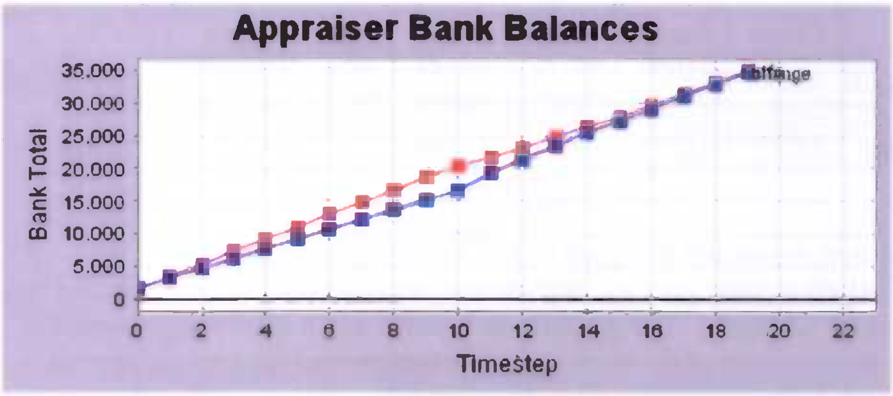

Figure 7

[7] Note that the first game round is round 0, so the tenth game round is round 9 in the figures.

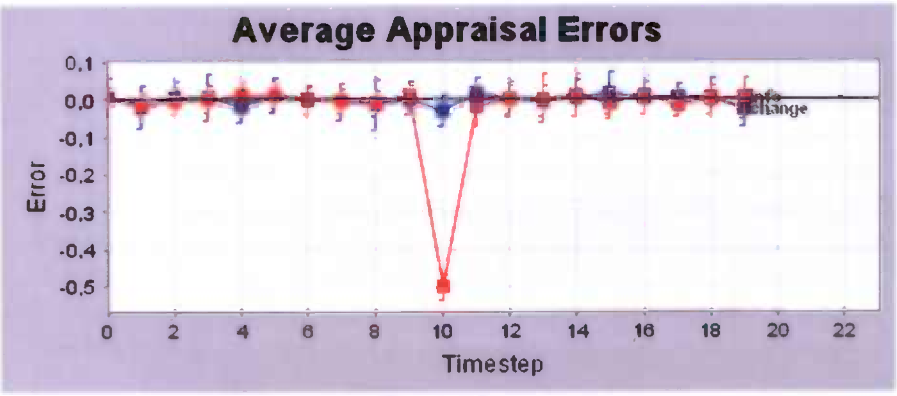

Figure 8

In contrast to Info-de, the agent Info-de-time does take the evaporation of beliefs as time goes by into account. As time goes by, information gathered in the past becomes less and less important. The difference is clear, after a first big decrease in appraisal accuracy when the agent Changing starts cheating, Info-de-time learns from Changing's new behaviour and adjusts its trust values. Its past beliefs about a trustworthy agent Changing do not overrule the new information it gathers. The averages of all the sessions with the agent Changing are presented in table 8.5.

	Changing		**Agent**	
	Client	Bank	Client	Bank
Info-de	34.7	49320	5.3	21140
Info-de-time	24	41557	**16**	28470

Table 8.5

In the last condition of this first part of the experiments, the agent Providing (providing reputation information) was introduced to focus on the updating from reputation information. Different versions of the information-based agent were tested in sessions with the agents Cheat, Naive and Providing. Figure 9 and 10 show the graphics of a session with the agents Info-rep-time (green), Providing (red), Cheat (yellow and black) and Naive (blue).

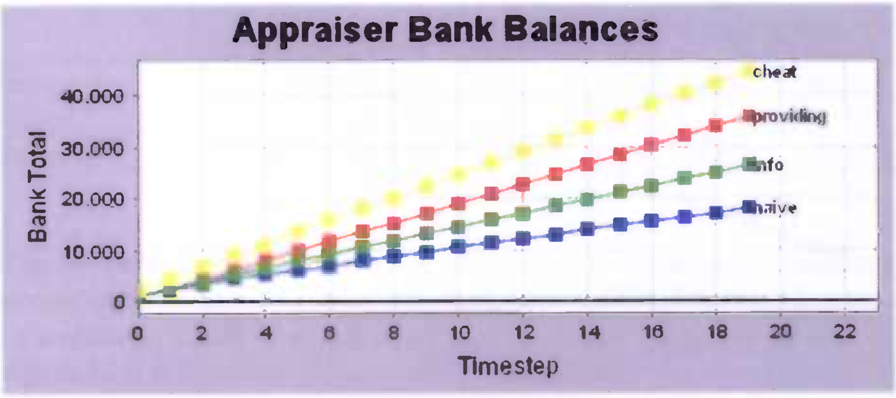

Figure 9

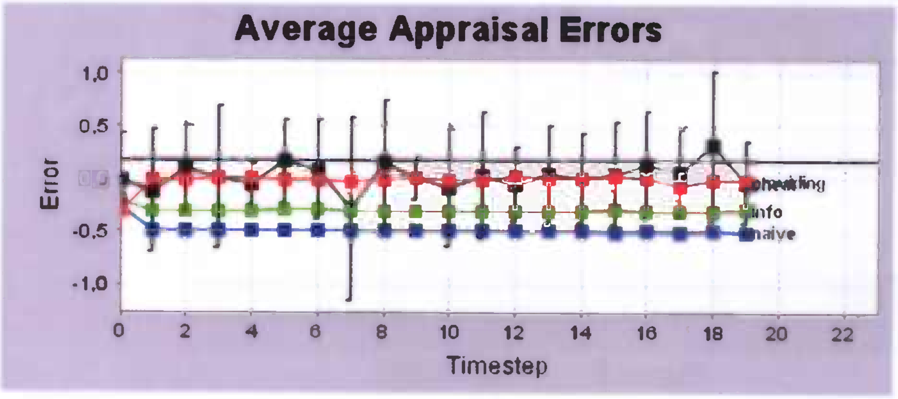

Figure 10

Figure 10 shows that the agent Providing provides the most accurate appraisals (the black line is not the zero line). It also shows that although the agent Info-rep-time does not use any information from its own experiences, its appraisals are more accurate than Naive's appraisals. The explanation must be that the agent Providing passes useful reputation information to Info-rep-time and that Info-rep-time is able to use this information. Table 8.6 shows the average final client shares and bank account balances of the set of experiments performed with the agent Providing.

	Cheat		**Naive**		**Agent**		**Providing**	
	Client	Bank	Client	Bank	Client	Bank	Client	Bank
Info-de-time	24.7	45537	10.3	12940	**25**	33867	20.7	30950
Info-rep-time	18.7	42960	15.7	18900	**19.3**	24810	25.7	35195
Info-de-rep-time	20.7	40153	11.7	14244	**23.7**	33767	23	32477

Table 8.6

The goal of the second set of experiments is to make a comparison between the agents Info, Game and Basic, where Info refers to Info-de-time. In a first experiment Info (green), Game (red) and Basic (blue) participated together in one game. Figure 9 and 10 show the results.

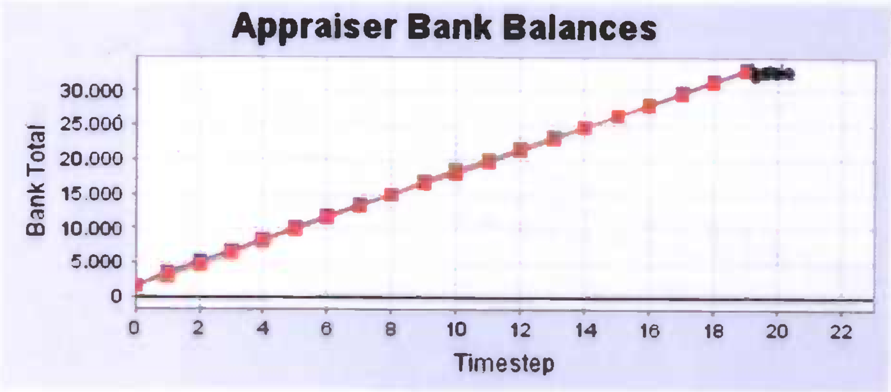

Figure 11

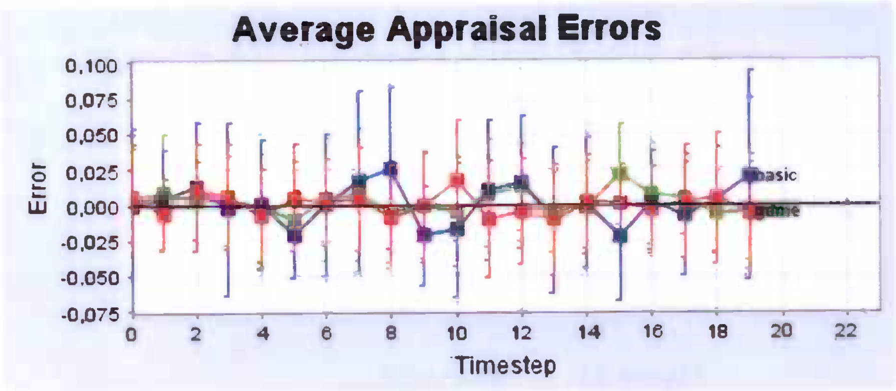

Figure 12

In the figures it is difficult to distinguish the different agents from each other. In this specific session Basic ended with 18 clients and Game and Info ended both with 21 clients. The graphics of the other sessions with four and five agents are even harder to read, so these are the results in numbers.

Sessions with:	**Cheat**		**Naïve**		**Info**		**Game**		**Basic**	
	Client	Bank	Client	Bank	Client	Bank	Client	Bank	Client	Bank
3 Agents					22.4	35792	19.4	32688	18	29716
4 Agents	20	40583	12	14960	25.7	37740	22	33797		
4 Agents	22.3	42723	8.7	13713	25	39520			23.3	41100
4 Agents	22.7	43270	7.7	12710			23.7	38120	25.7	44220
5 Agents	23.6	44400	7.8	11016	22.8	36380	22.2	34661	22.8	40248
Average					**24.1**	**37797**	**21.7**	**34718**	**22.4**	**38299**

Table 8.7

Besides these experiments with the three agents together, they have also been tested apart from each other. Table 8.8 shows how the agents performed with Cheat and Naive in a game.

	Cheat		**Naive**		**Agent**	
	Client	Bank	Client	Bank	Client	Bank
Info	22.4	42828	11.6	16480	**26**	41772
Game	25.4	47360	7.2	11674	**27**	43550
Basic	26.6	48094	4.4	9884	**28.2**	49230

Table 8.8

The next experiment tested the agents' reactions to the agent Changing. In this test condition the course of the session is important. The following figures show an example of a session of each of the three agents. In figure 13 Info is blue and Changing is red, in figure 14 it is the other way around.

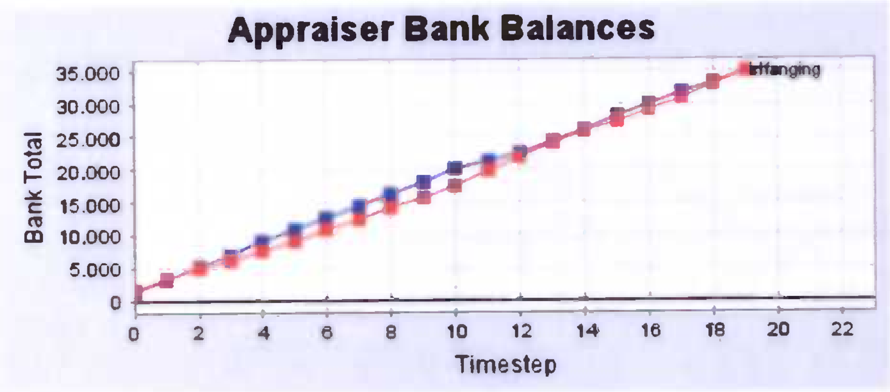

Figure 13

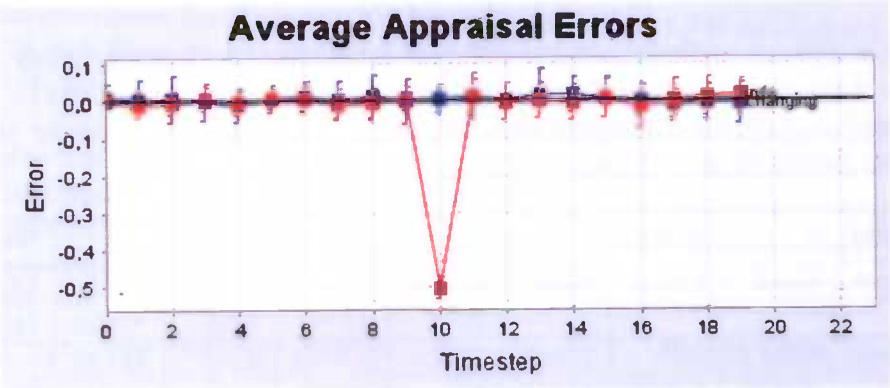

Figure 14

After round nine, the agent Changing starts cheating. Figure 14 shows that the accuracy of Info's appraisals decreases enormously in round ten. After that, Info adapts to the new situation and from round eleven it provides accurate appraisals again. The agent Info learned not to trust Changing anymore and stopped requesting opinions from Changing. Info's appraisals after round ten are purely based on its own expertise.

In the following two figures, the same confusion concerning the colours takes place with the agents Game (blue/red) and Changing (red/blue).

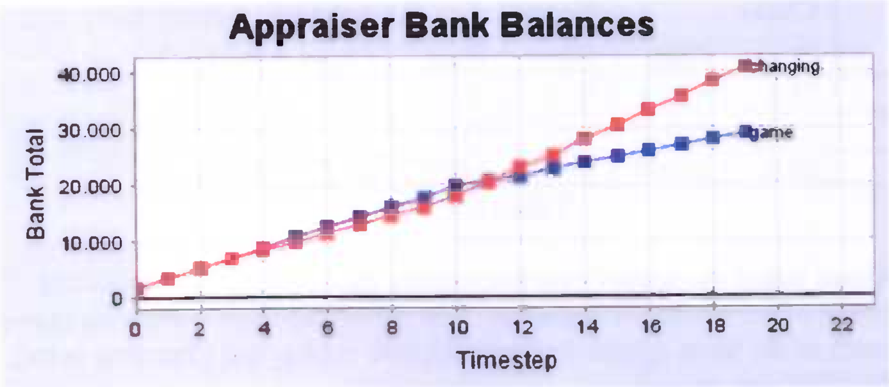

Figure 15

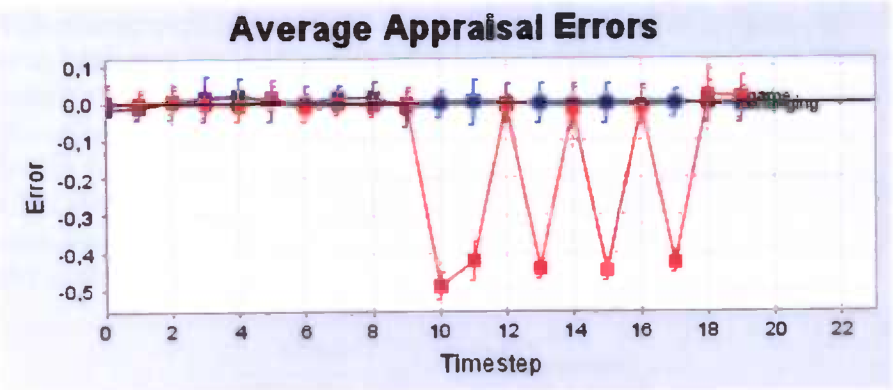

Figure 16

After round ten, the agent Game sometimes provides accurate appraisals and sometimes it provides inaccurate appraisals. In the first case it ignored the agent Changing and based the appraisals just on its own expertise. In the rounds with inaccurate appraisals it unjustly did trust the agent Changing. After a game round in which Changing got the opportunity to cheat on the agent Game, Game learns to not trust Changing. The effects of this learning are not strong enough however, because one game round later Game already forgot Changing's misbehaviour. It is expected that Game will learn to totally distrust Changing if the game would continue. In round 18 and 19 Game already managed to distrust Changing to rounds in a row.

Figure 17 and 18 show the results of the agent Basic (blue/red) and Changing (red/blue) in a game.

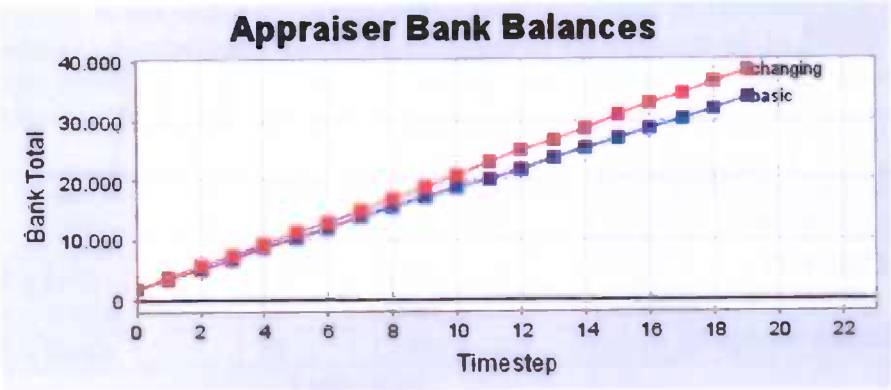

Figure 17

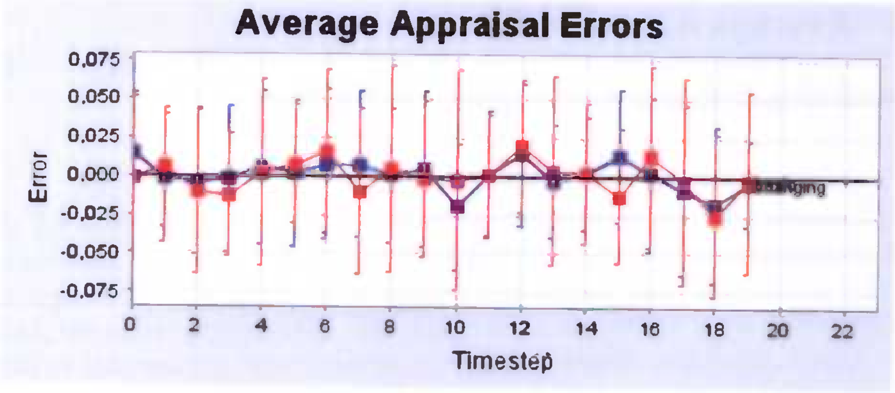

Figure 18

In contrast to the sessions with the agents Info (figure 14) and Game (figure 16), in figure 18 with Basic no change or effect on the accuracy of the appraisals is visible after Changing changed its behaviour.

The following figures provide some extra information about the three sessions displayed above. The lines in these graphics represent the reputation weights the participating agent subscribes to Changing in each of the three artistic eras. Figure 19 shows the reputation weights Info attached to Changing, figure 20 shows the reputation weights Game used and figure 21 the ones of Basic.

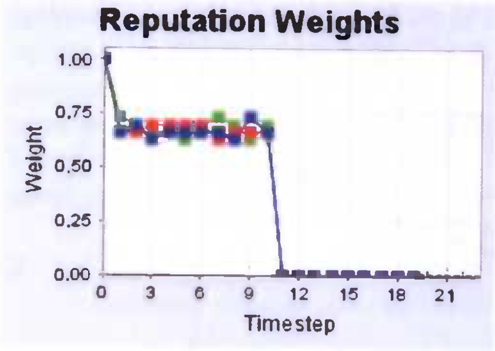

Figure 19

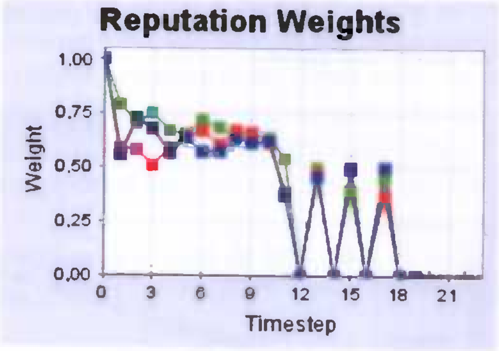

Figure 20

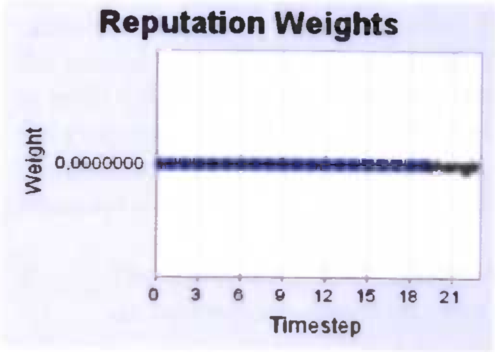

Figure 21

The examination of these figures corresponds with the findings of the figures 13 to 18 with bank account balances and appraisal errors. Info and Game changed the reputation weights they described to Changing after the tenth game round. Info (figure 19) immediately puts all weights to zero. Game (figure 20) also decreases the values but not totally to zero in all the game rounds. Figure 21 shows that Basic did not trust the agent Changing from the start and thus never used its opinions. This explains why Changing's change of behaviour did not affect the accuracy of Basic's appraisals. Table 8.9 displays the averages of all sessions.

	Changing		**Agent**	
	Client	Bank	Client	Bank
Info	24	41557	**16**	28470
Game	25.3	49857	**14.7**	20883
Basic	17.7	28860	**22.3**	43140

Table 8.9

The low results of Info and Game, an average client share of 16 and 14.7 respectively, can be explained by errors in their learning process such as demonstrated in figure 3 and 4. These failures happened more often in this condition than in other conditions, which might be explained by the small amount of agents participating in the game. If an agent unjustly attaches low trust values to another agent in the game, the effects are bigger when there are no other agents left to trust and to use opinions from. The agent Game had these kind of problems more often than Info, which suggests that Info's model is more robust than the model of Game.

In the final experiment Info, Game and Basic participated in a game with (or against) themselves. The table below shows the averages of the clients and the money the two agents were able to collect together. The column *Difference* shows the difference between the final client shares and bank account balances of the two agents.

	Average		**Difference**	
	Client	Bank	Client	Bank
2 x Info	**20**	33290	16	24620
2 x Game	**20**	35010	2.7	14473
2 x Basic	**20**	37000	10.7	14700

Table 8.10

The last table in this section presents the results about the computation time each agent needed for its decisions. The numbers the table 8.11 represent the 'time jumps' in the methods of each agent (for a description see section 8.2).

	Jumps
Info	52
Game	44
Basic	4

Table 8.11

9 Discussion

In the different sections of this chapter, different parts of this research project will be discussed. Starting with a discussion of the results of the experiments, followed by the design of the experiments, the information-based agent and testing with the ART test-bed, the information-based model of trust itself will be subject of discussion. Finally the use of information theory for the modelling of trust will be discussed.

9.1 Results of the experiments

In this section the results of the experiments will be discussed on the basis of the hypotheses made in section 8.1. Each hypothesis will be repeated and then a discussion about the corresponding results will be given.

1 The use of information from direct experiences will increase the average appraisal accuracy of an information-based test-bed agent.

In table 8.4 it can be seen that the information-based agents that all update from direct experiences provide more accurate appraisals than the agents Cheat and Naive that do not update from direct experiences. The third experiment is even more convincing: two information-based agents, one with and one without updating from direct experiences, were tested in the same condition. Table 8.6 shows that the agent that updated from direct experiences had a bigger final client share and therefore must have produced more accurate appraisals. So the first hypothesis is supported by the experimental results.

2 The use of information from the evaporation of beliefs as time goes by will increase the average appraisal accuracy of an information-based test-bed agent.

In the first and the second experiment two information-based agents updating from direct experiences were tested in the same condition. In the first experiment the two agents each participated together with the agents Cheat and Naive in a game (results in table 8.4); in the second experiment they each participated with the agent Changing in a game (results in table 8.5). Of these information-based agents, Info-de-time did use information from the evaporation of beliefs, Info-de did not. The results of both experiments support hypothesis 2, especially in the second condition the difference between the two agents becomes obvious.

3 The use of reputation information will increase the average appraisal accuracy of an information-based test-bed agent.

In the condition in which the agents Cheat, Naive, Providing and Info-rep-time participated together in one game (table 8.6), the agent Providing provides reputation information to Info-rep-time. Providing itself performs very well, so the reputation information it provides is supposed to be useful. Info-rep-time does not update from any of its own experiences, so its performance only depends on updating from reputation information. Info-rep-time ended with bigger client shares than Cheat and Naive, so it seems to well use Providing's reputation information. This observation supports hypothesis 3: the use of reputation information increases the average appraisal accuracy of an information-based test-bed agent. Of course this conclusion only holds when there is at least one agent in the game that is able and willing to provide useful reputation information.

4 Optimum average appraisal accuracy will be reached by using all available types of information: information from direct experiences, information from the evaporation of beliefs as time goes by and reputation information.

As seen in the discussions of the previous hypotheses, the use of information from direct experiences, information from the evaporation of beliefs as time goes by and reputation information all seem to improve the accuracy of an agent's appraisals. The question is whether they also work well in combination with each other, especially direct experiences and reputation information. Updating from the evaporation of beliefs as time goes by can by used in combination with the other two types of updating without hindering them. But updating from information from direct experiences and from reputation information go at the expense of each other. When more reputation information is used less information from direct experiences can be used and vice versa.

In table 8.6 the results of a comparison between an information-based agent using all types of available information, an information-based agent using all types of information except reputation information and an information-based agent using all types of information except information from direct experiences were compared with each other. The client shares they obtained were in the order from most to least: Info-de-time, Info-de-rep-time, Info-rep-time. These data do not support the hypothesis: according to this experiment the use of all available types of information does not yield the most accurate appraisals. However, the results in table 8.5 contradict the results in table 8.6. Here, the agent updating from reputation information (Info-de-rep-time) provides more accurate appraisals than the agent that does not (Info-de-time). However, in the second condition the agent Naive is the only agent providing reputation information, and Naive always provides the same reputations. Naive assumes that each agent is trustworthy, so it always provides reputations with the value 1. Because all the reputations Naive provides are equal, they do not give a lot of information and are not very useful. So the good performance of the agent using reputation information in this condition cannot be due to its updating from reputation information.

Although updating from reputation information leads to better results than no updating at all, it does not seem to work well in a combination with updating from direct experiences. Probably, information from direct experiences is much more valuable than reputation information. The use of reputation information does not counterbalance the loss of information from direct experiences. So the results of the experiments do not support hypothesis 4. Not the agent Info_de_rep_time as hypothesised, but Info_de_time appeared to provide the most accurate appraisals on average.

5 *On average, appraisals of the agent Info will be the most accurate and appraisals of the agent Basic will be the least accurate.*

This hypothesis has been tested in several experiments of which the results are shown in table 8.7, 8.8, 8.9, and 8.10. The different test conditions gave very different results. In the condition in which Info, Game and Basic participated together in a game, the order of accuracy of their appraisals was as expected (table 8.7). In de condition in which they participated with the agents Cheat and Naive one by one (table 8.8), the order of the quality of their appraisals was the opposite of the expectation. A closer look at the results teaches us that the variable outcomes are especially due to the varying performance of the agent Basic. In all but one of the conditions the agent Info provides more accurate appraisals than the agent Game, so this part of the hypothesis is supported by the results of the experiments. Now the question is why the agent Basic performs so variably and why it provides more accurate appraisals than Info and Game in some conditions, against the expectations.

Actually, the last question should be turned around and this is part of the explanation. The agent Basic always bases its appraisals on its own expertise, so it should always provide appraisals with more or less the same average accuracy. In contrast, Info and Game sometimes base their appraisals just on their own expertise (for example if they only participate with cheating agents), and sometimes they also use other agents' expertise. Appraisals based on the expertise of more than one agent should on average be more accurate than appraisals based on the expertise of just one agent. Sometimes the results did not seem to support this assumption; the same test-conditions then yielded very different results. This could be explained by the distribution of expertise levels in the first version of the ART test-bed. The average level of expertise over all eras was not always the same for all agents, so sometimes one agent could have high levels of expertise on all eras. In future versions of the test-bed this will be controlled, so that it will always be more profitable to make use of other agents' expertise.

Another remark that should be made in this discussion is that the total amount of clients in a game is independent of the agents' performances: the total amount is always the number of participants multiplied by twenty clients. So if all agents in a game perform equally badly or if they all perform equally well, every agent gets twenty agents. This is demonstrated in the condition in which Info, Game and Basic participated with themselves in a game, here the final client shares of the two agents together does not say anything about which of the agents Info, Game or Basic provides the most accurate appraisals (table 8.10). These observations show that the final client share might not be a good way to measure the accuracy of the agents' appraisals. At least it should not be used to compare performances between different test conditions: other agents in a game have too much influence on the number of clients.

So whether the results support or do not support the hypothesis depends a lot on the situations that will be considered. But what are actually representative and important test situations? This question is difficult to answer, but with most possible answers the results of the experiments would not be overwhelmingly supporting the hypothesis. The agent Basic provided more accurate appraisals than expected.

A last remark is that the more accurate appraisals of Info than Game does not prove that the information-based is a better approach than the game-theoretical one for the modelling of trust.

6 *On average, the test-bed performance of the agent Info will be the highest and the test-bed performance of the agent Basic will be the lowest.*

The argumentation for this hypothesis was that the order of test-bed performance would be the same as that of accuracy of appraisals. The result of the test condition with Cheat, Naive, Info and Basic forms an exception to this supposition; here Basic made more money than Info even though Info had more clients (table 8.7). In most cases however, the agents' bank account balances correspond with their client shares. So because the same order as in hypothesis 5 was supposed and this hypothesis was not supported by the results, hypothesis 6 neither is. A possible explanation for these unexpected results is the same as the explanation given with hypothesis 5.

7 *The agents Info and Game will show adapting behaviour when other agents change strategy, the Basic agent will not.*

This has been investigated by letting the three agents participate in a game with the agent Changing. Although Info and Game gather less clients than the agent Basic in this test condition (table 8.9), an analysis of the pictures in figure 13 till 21 demonstrates that Info and Game do adapt to the new situation and Basic does not. Figure 19 and 20 also show that Info and Game react to the cheating behaviour of Changing by decreasing the weight values they ascribe to it. This adaptation probably prevents worse results for both agents. From this can be concluded that hypothesis 7 is supported by the experiments.

8 *The agent Info will have the highest computational costs and the agent Basic will have the lowest.*

This follows from a theoretical analysis of the code of the agents. The agent Info calls the most and the most complex methods and the agent Basic calls the least and the least complex methods. The results in table 8.11 also support this hypothesis. Especially the difference between the agent Basic and the agents Info and Game is obvious. This can be explained by the fact that the agent Basic does not have a theoretical model whereas Info and Game do.

To summarize, hypotheses 1, 2, 3, 7 and 8 are supported by the experiments and the hypotheses 4, 5 and 6 are not. The wrong expectations in hypothesis 6 can be explained by the disappointing outcome of hypothesis 5. So the two main unexpected outcomes of the experiments are that updating from reputation information does not seem to be valuable in combination with updating from direct experiences and that the information-based and the game-theoretical agent do not always perform more accurate appraisals than the agent without a model.

9.2 The design of the experiments

The ART test-bed allows its users to vary a lot of parameters. In the experiments most of the parameters were kept constant over all the games and the only thing that changed were the agents that participated in the games. The used values of Timesteps-per-Session, Number-of-Painting-Eras, Average-Clients-Per-Agent, Client-Fee, Opinion-Cost, Reputation-Cost, Sensing-Cost-Accuracy and Previous-Client-Share-Influence were not based on extensive experimentation, but taken from the initial values in the file *Game Parameters*. More experimentation would give more information, but because the values were provided by the ART test-bed team they are supposed to be appropriate for experimentation.

Besides fixed values for the parameters in the test-bed, some aspects of the agents were also kept constant over all the experiments. For example, all agents paid a price of ten when they ordered opinions from the simulator. Secondly, all the information-based agents and the agent Game used the value of 0.5 as a threshold for trusting other agents. A final example is the ratio between the influence of information from direct experiences and reputation information in the agent Info-de-rep-time, the only ratio that was used was 1 (experiences): 0.3 (reputation). For all of these examples, more experiments to investigate the effects of these values would deliver new and maybe valuable information, but as in the case of the test-bed game parameters the choices are expected to be realistic.

The behaviour and performance of an agent depends a lot on the other participants in a test-bed game. For example, an agent with a very sophisticated model for dealing with reputation information only profits from its model when other agents in the game are prepared to provide reputation information. A cooperative agent might function very well with other cooperative participants, but perform very bad if a non-cooperative agent participates in the game. So to just let a test-bed agent play against itself or very similar agents does not give a complete picture of the agent. For a more complete evaluation, it should also be tested against agents with very different kinds of behaviour. In the experiments four test-agents were used, the agents Naive, Cheat, Changing and Providing, which show quite simple and obvious behaviour. The use of more different and more complex test-agents would provide more information.

A related point is that the number of participants in a game was always two, three, four or five. Exploring conditions with larger numbers of participants would create new situations and might yield extra information. Here again applies that in more complex situations, more aspects of the tested agent will become visible. However, for the purpose of this research the reactions to the different test conditions with the four test-agents already gave a lot of valuable information about the tested agents.

The tested agents were judged on four points: their test-bed performance, the accuracy of the information they provided, their ability to adapt to new situations and their efficiency. An agent's bank account balance obviously is a good indicator of test-bed performance; the test-bed defines an agent's performance by its bank account balance. The third measure, the development of the agent's average appraisal errors in a game with the agent Changing together with the weight values it describes to Changing surely shows something about an agent's ability to adapt. However, it only measures how an agent adapts in one specific situation, the reaction to an agent that changes is strategy from very cooperative to highly non-cooperative. The test gives a lot of information, but for a more complete evaluation it would be interesting to also examine an agent's behaviour in reaction to different and more subtle changes. The final evaluation point, efficiency, seems to be validly evaluated by measuring an agent's computational costs.

In comparison to the previous three aspects, the accuracy of the provided information was more difficult to measure. The trust values attached to other agents would be the most direct measurement on this point, but it was not possible to verify the accuracy of these values. So instead, the average appraisal error was planned to be used. This value involves information about which agents have been consulted and which weights have been attached to them. The better an agent makes these choices, the more accurate its processed information will be. A drawback of the use of the average appraisal error was that the values were not displayed in the test-bed's database. Therefore the final client share, which is shown in the database and derived from the average appraisal errors, was also used. However, as the discussion in the previous section showed, client share is not as appropriate to measure the accuracy of the provided information as the average appraisal accuracy. Another drawback is that final client share only gives information about how a session ends and not about the course of a session.

Finally, the experimental set-ups were repeated five and sometimes three times, which is not very much. To make grounded statements, more repetitions per condition would be preferable. However, in the first version of the test-bed it was not possible to alter the variable *Number-of-sessions* and to gain the results of several runs with the same conditions in one go. Each session had to be set by hand and this was rather time consuming. That is why unfortunately only a small number of sessions per condition has been run. In future versions of the test-bed it will be possible to vary the number of sessions and this will save a lot of work.

9.3 The information-based agent

The information-based model for trust as presented in chapter 3 is a model of trust in general, so it does not describe how to apply it to an ART test-bed agent. To implement an information-based test-bed agent, many choices of how to exactly apply the model have to be made. Which parts of the model should be used, how to apply these parts, how to translate that in the agent's code and what to do with test-bed requirements for which the model does not provide a theory? It was not always practical or possible to exactly map the theory to code of the information-based agent and in some aspects the model and the agent differ from each other. But to still say something about the information-based model, the information-based agent derived from the model should not differ too much from the original model. Below some differences and similarities between the model and the agent will be discussed and explained. This will show that the core of the information-based model has been transferred to the model of the information-based agent.

Sierra and Debenham (2005) provide a language for negotiation which gives the possibility to make offers, accept and reject offers, break down negotiations and inform other agents. Besides that, it also gives the possibility to express quantitative and qualitative preferences. This language enables agents to express complex sentences and have rich negotiation dialogues[8]. In the ART test-bed however, much of the possibilities of the language are not used. Prices are fixed in the game and the agents can just make offers, accept offers and decline offers, so there are almost no negotiation dialogues in the test-bed.

Another part of the trust model that was not needed by test-bed agents was Sierra and Debenham's second type of updating, *updating from preferences*. Because agents do not express preferences in test-bed games, it was not necessary to write methods for the updating from these preferences. In the ART test-bed all agents have an equal relationship with each other, so applying a part about power relations[9] between agents was not needed either. A last example of a theoretical idea in the model that was not used for the implementation of the agent was the modelling of trust as conditional entropy[10]. In the information-based agent trust was modelled as relative entropy. This choice was made because the quality of an opinion can be more or less high, so it makes sense to speak of the relative quality of an opinion. In some situations a certain outcome is just good or bad, then trust should be modelled as conditional entropy.

In contrast to the previous examples, Sierra and Debenham's trust model also lacks theory for some topics needed in the ART test-bed. Reputation plays a very important role in the test-bed, but the information-based model only shortly mentions the notion of reputation and gives some initial ideas of how to deal with it. To deal with reputation in the test-bed, a simplified version of the model discussed in chapter four of this thesis was implemented in the information-based agent. A simplification was for example that the information-based agent does or does not trust other agents, whereas the theory allows a continuum of trust.

The information-based trust model does not provide a negotiation strategy, it is just a system to keep up values of trust. However, in practical situations such as in the ART test-bed an information-based agent cannot do without a strategy. The goal of the research was to investigate the information-based model and not much attention was paid to the strategy. But still, to investigate the usefulness of the trust values the agent derived, it had to act according to some rules using these trust values. In the experiments for example, the information-based agent acted cooperatively against agents it trusted and non-cooperatively against agents it did not trust. But the agent's behaviour influences other agents' behaviour and this could influence the trust value it ascribes to these agents, so it is very difficult to draw a strict border between the performance of the model and that of the strategy.

An important step in applying the model to a test-bed agent is the choice of the probability distributions. The model does not prescribe how to make this choice in a practical situation. For test-bed agents it is very practical to predict what opinions other agents will provide. The agents cannot negotiate about the price or the time of delivery of the opinions. The only variable aspect of opinions is their closeness to the real value of the corresponding painting, their quality. Therefore, the probability distributions in the information-based agent were about the quality of opinions: each possible world represented a quality level. For each agent in each era, a probability distribution was kept up. An interesting extension to this would be to keep up an extra probability distribution

[8] See section 3.2 of this thesis for some examples

[9] Section 5.3 in Sierra and Debenham 2005

[10] Section 6.1 in Sierra and Debenham 2005

for each agent about their reliability, with information about whether an agent keeps his promises. As it is now, this kind of information is employed in the other probability distributions.

When the probability distributions had been chosen, direct experiences and reputation information had to be translated to constraints that could be applied to the probability distribution. As discussed in section 7.2, constraints in the information-based agents are always of the type agent α in era e provides opinions of *at least* quality x and all have a belief value of 0.5. The possibility of applying more flexible constraint types to the probability distributions might improve the performance of the agent. For example the possibility to vary belief values or to express the expectation that agent α will provide opinions with a quality of *maximally* x. Although these sophistications might yield better results, the general idea would not change.

The last two examples of this section will treat examples about parts of the model which were changed or adapted in the information-based agent. A difference between the model and the agent for example, is the way updating from the evaporation of beliefs as time goes by takes place. The model uses an equation inspired by pheromone like models (Sierra and Debenham, 2005), whereas the agents use a simpler formula (section 7.2 of this thesis). The algorithms are not the same, but both cover the idea that recent experiences are more important than older experiences.

According to the model, probabilities should be updated by applying the principle of minimum relative entropy. An updated probability distribution has to satisfy its new constraints, but further it resembles the old distribution as much as possible. The agent, on the other hand, calculates new probability distributions regardless of the old ones. Instead of the principle of minimum entropy it uses maximum entropy and with all current constraints it calculates a new probability distribution. For this project, code of maximum entropy was available and code of minimum entropy was not. Although the calculation of the two methods are different, the results are not expected to differ a lot. Instead of the model, the agent calculates new probability distributions regardless of the old ones, but the new distributions will contain much of the same information, because all previous constraints are taken into account.

9.4 Testing with the ART test-bed

A general problem of all test-beds is the question of validity: does the system test what it is supposed to test? Especially when complicated concepts are involved, it is very difficult to prove that a test-bed just examines the performance of a model on that particular concept. The aim of the ART test-bed is to compare and evaluate trust- and reputation-modelling algorithms (Fullam et al., 2005). But what do the developers exactly understand by trust and reputation? When is a trust-modelling algorithm a good algorithm: when it is a good predictor of the future? Or when it provokes favourable behaviour of other agents? In the test-bed, a winning agent is able to estimate the value of its paintings most accurately and purchases information most prudently (Fullam et al., 2005). Do agents that perform well on these tasks automatically have a good trust or reputation model? These questions are extremely difficult to answer and the complicatedness of the test-bed makes it even more difficult.

Critics of the ART test-bed team on the Prisoner's dilemma, another test-bed for trust, is that it is not rich enough to test all facets of trust and reputation. However, the ART test-bed itself is quite complicated and allows so many variables that it is sometimes difficult to explain why something happened and for what cause. With the huge amount of publications about the relatively simple Prisoner's dilemma, one could question whether the ART-Test-bed might be too complicated. Models are usually simplifications of the real world, to make it easier to understand the real world. When a model or test-bed better approximates the real world, it becomes so complicated that the advantage of simplification is lost. So although it is valuable to test a lot of different facets of trust and reputation, test-bed designers should take into account that sometimes simpler models and test-beds lead to more understanding than very complicated ones. One could for example design a test-bed only based direct experiences or just taking reputation information into account.

Although the ART test-bed is very rich and tests a lot of facets, still not all facets of Sierra and Debenham's trust model could be tested. As seen in the previous section, there are no different negotiation topics and agents cannot express their preferences, which are both strong points of Sierra and Debenham's trust model. Social information treated as in chapter five, discussing agents with different social roles and publicly available reputation information, does not play a role in the test-bed. So even though the ART test-bed is already complicated, it still does not test all aspects of trust and reputation. This demonstrates how difficult is to make a good test-bed of trust and reputation.

As already mentioned several times in this thesis, at the time of working on this project the development of the ART test-bed was not yet totally finished. For the experiments a first beta release of the test-bed was used, and some functions of this version still needed more perfection or were even missing. An example of a missing function is the introduction of dummy agents in a game, which still was not possible in the beta release. Nor was it possible to vary the number of sessions to run; each session had to be set manually. A last example is that agents should have access to their own levels of expertise on different artistic eras, but in the beta version this was not the case. Another already mentioned drawback due to the newness of the ART test-bed was the lack of material for comparison. All the agents in the test-bed have been programmed for this research, the information-based agent could not be compared with agents from other researchers. Despite these problems, experiments have been performed and some interesting results have been obtained.

9.5 The information-based model of trust

Sierra and Debenham's information-based model of trust provides some initial ideas about how to deal with reputation and other types of social information. Social aspects are becoming more and more stressed in the field of multi-agent systems lately, so a contemporary model of trust should give an account of it. Chapter four and five of this thesis discuss a more extensive account of how to deal with social information within the information-based approach. This attempt showed that the model is easily extendible, adding new parts to the model gave no problems and it should be possible to extend the model even more. The flexibility of the model was also proved from the extension Sierra and Debenham made to their model themselves. In a recent article (Sierra and Debenham 2006), they integrated the notion of honour in the information-based approach.

Sierra and Debenham define trust as a measure of how uncertain the outcome of a contract is, which is a very clear definition. Focussing on reputation and social information on the other hand, exposed some conceptual problems in their model. The double use of reputation and power discussed in section 4.1 of the thesis seems redundant, but Sierra and Debenham do not give an explanation for it. Because the right use of reputation and social information is therefore not obvious, the exact meanings of the concepts also stay unclear. A better explanation of the meanings of and the relations between some concepts in Sierra and Debenham's approach is desirable. In chapter 4 and 5 of this thesis it has been tried to provide such a clearer account of reputation and (other types of) social information.

9.6 The use of information theory

In the previous sections of this chapter, different aspects of the information-based model and the experiments about it have been discussed. This last section of chapter 9 will focus on the model as a whole and compare the use of information theory to other approaches.

A very usual approach to model trust is game theory, already discussed in section 3.5 and 7.4 of this thesis. Sierra and Debenham (2005) say that with uncertain information and decaying integrity, the 'utility calculation' as in the game-theoretical approach is a futile exercise. This seems to be a quite strong statement, because game theory proved to yield good results in many different applications. In reality, Sierra and Debenham are somewhat vague, because 'uncertain information and decaying integrity' is a very broad notion. In situations of very uncertain information and a very low integrity they are probably right, but it is not clear where to draw the border. The information-based approach is at least not the most appropriate approach for highly rational situations. In that case one could better use game theory (Debenham, personal communication). To find out in what situations exactly the information-based model outperforms game-theoretical ones, more experimentation should be done.

A second alternative to information theory, shortly seen in the second example of section 2.3 of this thesis, is the cognitive approach. In comparison with the information-based approach and game-theoretical approaches, this approach puts more emphasis on the analysis of social and cognitive aspects of reputation. Conte and Paolucci for example, investigate reputation in order to find a solution for the problem of social order (Conte & Paolucci 2002). They think reputation plays a central role in the existence of altruism, reciprocity, cooperation and norm obedience. Many other approaches do not concentrate on these aspects and Conte and Paolucci (2002) critique present applications of reputation in multi-agent systems for using an intuitive and still vaguely defined notion of reputation. This critique could also apply to the information-based approach: reputation is used although its role and its relation with trust are not very well defined in Sierra and Debenham's article (Sierra and Debenham 2005).

However, maybe it is better to see cognitive approaches to reputation as complementary instead of contradicting other approaches. The information-based approach does not reason in terms of beliefs, goals and intentions, which the cognitive approach does. In contrast to cognitive approaches, the information-based approach does provide means for designing computational applications using trust or reputation, by showing a way to calculate numerical values of trust. Sierra and Debenham's information-based model of trust is a model for guiding what Conte and Paolucci (2002) call epistemic decisions, decisions about whether to accept a certain belief. From this perspective, the two approaches model different aspects of reputation and trust, and might be combined with each other. After a cognitive analysis of reputation and related concepts, information theory could take care of the numerical calculations within the cognitive framework.

10 Conclusions and further research

In this final chapter the conclusions of the project and some suggestions for further research will be given.

10.1 Conclusions

The goal of this project was to examine Sierra and Debenham's information-based model for trust. This has been done in a theoretical way and in a practical way. First, a general overview of models of trust and reputation was given, then the information-based model has been discussed and finally it was extended on some parts. This resulted in some suggestions to improve the model and in two new modules in the model: one about updating from reputation information and one about updating from social information. For the practical examination of the model, an agent based on the information-based trust model has been implemented. Several experiments in the ART test-bed have been performed with this information-based agent. In general, the agent performed well in comparison to other agents and the information-based approach seems to be appropriate for the modelling of trust.

From the theoretical part of the examination it can be concluded that the core of the model is very clear: updating probability distributions from a set of beliefs using the principle of minimum entropy. Not all parts of the model are very well elaborated, for example the influence of reputation and other social information is not (yet) fully worked out. However, the model is flexible and it allows to work out and add new parts (as done in chapter 4 and 5). Sierra and Debenham use a very clear definition of trust in their model, but the exact meanings of other concepts are not always so well defined. So in general, the model Sierra and Debenham propose seems to be a good and robust approach, but is not finished yet.

The practical part of the examination of the information-based model of trust had three main shortcomings. First, the ART test-bed used for the experiments was developed very recently and therefore it did not (yet) function completely optimally and there was no material for comparison available (section 9.4). Secondly, not all aspects of the theoretical model were exactly translated to the practical application (section 9.3). The last point is that bigger numbers of experiments testing more different aspects should be done (section 9.2). Despite these drawbacks, the results of the experiments were promising. The experiments showed that the information-based agent learned about its opponents during a game session and could distinguish between cooperating and non-cooperating agents. They also demonstrated that the three examined types of updating, updating from direct experiences, updating from reputation information and updating from the evaporation of beliefs as time goes by, all improved the agent. The best combination of different types of updating was found to be updating from direct experiences and updating from the evaporation of beliefs.

Finally, with the findings of the practical and theoretical investigations an answer to the main question of this project can be given. *Is the information-based approach a* *good way to deal with trust and reputation in multi-agent systems?* The amount and the quality of the experiments in this research are not sufficient to give a decisive answer to this question. Moreover, the research field of trust and reputation lacks unity, so it is difficult to say what is exactly 'a good way' to deal with trust and reputation in multi-agent systems. However, the experiments did yield some promising results and showed that the way of calculating trust values and updating the model from new information seemed to work. The meaning of trust is clearly defined and turned out to be useful in the experiments; the meanings of some other concepts are less clear. So, it can be said conclusively that the core of the model seems to be a good approach, but for a fully developed approach to trust and reputation more work should be done.

10.2 Further research

Most models of trust and reputation in multi-agent systems are based on Game theory and most of the experiments in this field have been performed with game-theoretical applications. Sierra and Debenham are the first ones who used Information theory for the modelling of trust and the experiments in this project are the very first experiments with their trust model. The information-based agent derived from the model does not contain all aspects of the theory and the experiments performed surely do not test all facets of the information-based model. In future research the investigations of Sierra and Debenham's information-based trust model could be improved and extended.

Many aspects of the trust model could be translated more literally to the implementation of the information-based agent. For example, for the evaporation of beliefs as time goes by a pheromone like model could be used and the possibility to vary belief certainties could be added. A lot of other suggestions have already been mentioned in section 9.3. Of all these suggestions the use of the principle of minimum relative entropy instead of using maximum entropy deserves some extra attention, because this method forms the core of the model. Finally, it would be very interesting to pay more attention to the strategy of the agent.

Future research could also be directed to extend the diversity and the amount of the experiments. In the future, improved versions of the ART test-bed can be used. This will make it easier to increase the amount of sessions per condition and gives the possibility to add dummy agents. When agents of other researchers become available, the information-based agent can be compared with them. And of course, the influence of varying the many parameters in the ART test-bed could be investigated. A final option is to choose another test-bed and make a whole new application.

Besides these suggestions about experiments for more practical research, there is also more theoretical work to do. The main conclusion of the theoretical discussion of the model was that the model lacks some conceptual background. Meanings of and relations between different concepts are not always obvious. As seen section 9.6, cognitive approaches pay a lot of attention to the analysis of the exact meanings of different concepts and social and cognitive aspects play a very important role in this. Therefore, research to the possibilities of a combination of both approaches integrating the strong points of both theories with each other, seems to be a very interesting topic for further research.

Appendix

To create table 8.4 the data of the following three tables were used:

	Cheat		**Naive**		**Info-de**	
	Bank	Client	Bank	Client	Bank	Client
Session 1	41400	20	14800	11	47270	29
Session 2	51180	28	17110	13	34370	19
Session 3	49420	27	9830	5	42040	28
Session 4	50640	28	17110	11	33480	21
Session 5	32710	15	22640	16	43800	29
Average	**45070**	**23.6**	**16298**	**11.2**	**40192**	**25.5**

	Cheat		**Naive**		**Info-de-time**	
	Bank	Client	Bank	Client	Bank	Client
Session 1	49380	27	9550	5	42360	28
Session 2	50350	28	13820	13	37570	19
Session 3	35500	18	22640	14	41160	28
Session 4	36980	18	20400	14	41880	28
Session 5	41930	21	15990	12	45890	27
Average	**42828**	**22.4**	**16480**	**11.6**	**41772**	**26**

	Cheat		**Naive**		**Info-de-rep-time**	
	Bank	Client	Bank	Client	Bank	Client
Session 1	50500	28	7244	3	45836	29
Session 2	34820	14	21708	18	42792	28
Session 3	50730	28	8866	3	43324	28
Session 4	38350	20	20612	13	40608	27
Session 5	35470	16	20588	15	42792	28
Average	**41974**	**21.2**	**15804**	**10.4**	**43070**	**28**

To create table 8.5 the data of the following two tables were used:

	Changing		**Info-de**	
	Bank	Client	Bank	Client
Session 1	53790	37	16560	3
Session 2	38240	30	32140	10
Session 3	55930	37	14720	3
Average	**49320**	**34.7**	**21140**	**5.3**

	Changing		**Info-de-time**	
	Bank	Client	Bank	Client
Session 1	52550	33	17950	7
Session 2	37410	20	32500	20
Session 3	34710	19	34960	21
Average	**41557**	**24**	**28470**	**16**

To create table 8.6 the data of the following three tables were used:

	Cheat		**Naive**		**Providing**		**Info-de-time**	
	Bank	Client	Bank	Client	Bank	Client	Bank	Client
Session 1	45940	25	14820	14	30700	16	34790	26
Session 2	44970	24	10080	7	33150	25	33260	24
Session 3	45700	24	13920	10	29000	21	33550	25
Average	**45537**	**24.7**	**12940**	**10.3**	**30950**	**20.7**	**33867**	**25**

	Cheat		**Naive**		**Providing**		**Info-rep-time**	
	Bank	Client	Bank	Client	Bank	Client	Bank	Client
Session 1	44610	19	18000	15	33956	25	26340	20
Session 2	50930	25	15060	11	36350	26	23610	17
Session 3	33340	12	23640	21	35280	26	24480	21
Average	**42960**	**18.7**	**18900**	**15.7**	**35195**	**25.7**	**24810**	**19.3**

	Cheat		**Naive**		**Providing**		**Info-de-rep-time**	
	Bank	Client	Bank	Client	Bank	Client	Bank	Client
Session 1	30780	14	18405	15	34358	25	33547	25
Session 2	45080	24	11922	10	33748	23	34488	21
Session 3	44600	24	12405	10	29324	21	33267	25
Average	**40153**	**20.7**	**14244**	**11.7**	**32477**	**23**	**33767**	**23.7**

To create table 8.7 the data of the following five tables were used:

	Info		**Game**		**Basic**	
	Bank	Client	Bank	Client	Bank	Client
Session 1	44410	27	30310	14	29250	19
Session 2	34210	21	33230	22	21590	17
Session 3	34020	21	33340	21	32040	17
Session 4	32960	22	32880	19	33210	19
Session 5	33360	21	33680	21	32490	18
Average	**35792**	**22.4**	**32688**	**19.4**	**29716**	**18**

	Cheat		**Naive**		**Info**		**Game**	
	Bank	Client	Bank	Client	Bank	Client	Bank	Client
Session 1	36680	16	14580	12	38030	25	35930	26
Session 2	46940	25	11880	9	35430	26	33420	20
Session 3	38130	19	18420	15	39760	26	32040	20
Average	**52810**	**20**	**14960**	**12**	**37740**	**25.7**	**33797**	**22**

	Cheat		**Naive**		**Info**		**Basic**	
	Bank	Client	Bank	Client	Bank	Client	Bank	Client
Session 1	46780	25	14180	9	42160	25	35190	20
Session 2	35460	17	16840	12	38920	26	44280	25
Session 3	45930	25	10120	5	37480	24	43830	25
Average	**42723**	**22.3**	**13713**	**8.7**	**39520**	**25**	**41100**	**23.3**

	Cheat		**Naive**		**Basic**		**Game**	
	Bank	Client	Bank	Client	Bank	Client	Bank	Client
Session 1	45920	25	8160	4	44100	25	41010	25
Session 2	46300	25	10820	5	44010	25	37000	25
Session 3	37590	18	19150	14	44550	26	36350	21
Average	**43270**	**22.7**	**12710**	**7.7**	**44220**	**25.7**	**38120**	**23.7**

	Cheat		**Naive**		**Info**		**Game**		**Basic**	
	Bank	Client	Bank	Client	bank	client	bank	client	Bank	Client
Session 1	45200	24	13440	10	35840	24	30290	18	42030	23
Session 2	45630	24	13860	11	38070	20	37700	24	34920	21
Session 3	43090	23	7800	5	36800	24	37040	24	40950	23
Session 4	44190	24	7380	4	38410	23	39614	24	42390	24
Session 5	43890	23	12600	9	32780	23	28660	21	40950	23
Average	**44400**	**23.6**	**11016**	**7.8**	**36380**	**22.8**	**34661**	**22.2**	**40248**	**22.8**

To create table 8.8 the data of the following three tables were used:

	Cheat		Naive		Info-de-time	
	Bank	Client	Bank	Client	Bank	Client
Session 1	49380	27	9550	5	42360	28
Session 2	50350	28	13820	13	37570	19
Session 3	35500	18	22640	14	41160	28
Session 4	36980	18	20400	14	41880	28
Session 5	41930	21	15990	12	45890	27
Average	**42828**	**22.4**	**16480**	**11.6**	**41772**	**26**

	Cheat		Naive		Game	
	Bank	Client	Bank	Client	Bank	Client
Session 1	48940	27	9620	5	42440	28
Session 2	50630	28	8220	4	44830	28
Session 3	51220	28	12350	9	41800	22
Session 4	35470	16	20588	15	42792	28
Session 5	50540	28	7590	3	45890	29
Average	**47360**	**25.4**	**11674**	**7.2**	**43550**	**27**

	Cheat		Naive		Basic	
	Bank	Client	Bank	Client	Bank	Client
Session 1	50250	28	7980	3	49050	28
Session 2	39530	21	17980	10	50130	29
Session 3	50220	28	7740	3	48960	28
Session 4	50310	28	8460	3	49050	28
Session 5	50160	28	7260	3	48960	28
Average	**48094**	**26.6**	**9884**	**4.4**	**49230**	**28.2**

To create table 8.9 the data of the following three tables were used:

	Changing		**Info**	
	Bank	Client	Bank	Client
Session 1	52550	33	17950	7
Session 2	37410	20	32500	20
Session 3	34710	19	34960	21
Average	**41557**	**24**	**28470**	**16**

	Changing		**Game**	
	Bank	Client	Bank	Client
Session 1	48000	22	22820	18
Session 2	59450	29	11970	11
Session 3	42120	25	27860	15
Average	49857	25.3	20883	14.7

	Changing		**Basic**	
	Bank	Client	Bank	Client
Session 1	35370	21	36630	19
Session 2	35550	20	36450	20
Session 3	15660	12	56340	28
Average	**28860**	**17.7**	**43140**	**22.3**

To create table 8.10 the data of the following three tables were used:

	Info 1		**Info 2**	
	Bank	Client	Bank	Client
Session 1	39360	21	28130	19
Session 2	46690	35	18300	5
Session 3	50750	28	16510	12
Average	**45600**	**28**	**20980**	**12**

	Game 1		**Game 2**	
	Bank	Client	Bank	Client
Session 1	46930	20	21800	20
Session 2	42390	20	27340	20
Session 3	37420	24	34180	16
Average	**42247**	**21.3**	**27773**	**18.7**

	Basic 1		**Basic 2**	
	Bank	Client	Bank	Client
Session 1	36090	21	35910	19
Session 2	36540	20	35460	20
Session 3	57420	35	14580	5
Average	**43350**	**25.3**	**28650**	**14.7**

Summary

Negotiation is a process in which a group of negotiation partners tries to reach a mutually acceptable agreement on some matter by communication. The negotiators can of course be humans, but software agents and robots also negotiate. Agents in a multi-agent system are autonomous, so they have no direct control over other agents and must negotiate in order to control their interdependencies. In negotiations, one tries to obtain a profitable outcome. But what is a profitable outcome: pay little money for many goods of high quality? Although it seems to be a good deal, this might not always be the most profitable outcome. If negotiation partners will meet again in the future, it could be more rational to focus on the relationship with your negotiation partners, to make them trust you and to build up a good reputation.

The computational modelling of trust and reputation has received a lot of attention in the field of multi-agent systems lately. Most of the current models of trust and reputation are based on game theory. This thesis focuses on a new approach to trust, which is based on information theory. In this graduation project Sierra and Debenham's information-based model for trust has been examined (Sierra and Debenham 2005). The main question of the project was whether the information-based approach is a good way to deal with trust and reputation in multi-agent systems.

In the information-based model, trust is defined as the measure of how uncertain the outcome of a contract is. All possible outcomes are modelled and a probability is described to each of them. If there is no information available all outcomes get the same probability to be the actual outcome, but if some information is available constraints can be put to the probability distribution. Sierra and Debenham distinguish three types of information from which probability distributions can be updated: updating from decay and experiences, updating from preferences and updating from social information. From an updated probability distribution, the trust that a specific agent will fulfil all aspects of the contract can be calculated. In their model, reputation is not very clearly defined and updating from reputation information is not elaborated very extensively. In this thesis, in contrast, two possible ways to deal with reputation information are discussed and worked out. Another lack of the information-based model is a worked out approach for dealing with other types social information (not reputation information). Therefore, another section in the thesis focuses on updating from social information.

Besides this theoretical discussion of the model, the project also consisted of a more practical part in which the ART test-bed was selected to test the information-based trust model. Participants in the ART test-bed have to estimate the value of their clients' paintings in several game rounds. Agents that provide the most accurate appraisals will have the most clients and have the opportunity to make the most money. In order to estimate the value of a painting, agents can study the painting themselves or they can request other agents for help. This could be profitable if other agents have more expertise about the specific artistic era of the painting than the agent itself. However, agents might (on purpose) provide bad opinions or not even provide a promised opinion at all. So the agents in the test-bed have to learn which agents to trust and which ones to distrust. To facilitate this process, agents can buy information about other agents' reputations from

each other. Here again it holds that agents do not always tell the truth or provide valuable information.

An ART test-bed participant based on the information-based model for trust has been implemented. Its model used updating from direct experiences, updating from the evaporation of beliefs as time goes by, and updating from reputation information. Some variations on the information-based agent were made to compare the effects of the different types of updating. Besides that, six other agents were programmed for the experiments. The results of the experiments showed that the information-based agent learned about its opponents during a game session and could distinguish between cooperating and non-cooperating agents. They also demonstrated that the three examined types of updating all improved the performance of the agent. The best combination of different types of updating was found to be updating from direct experiences and updating from the evaporation of beliefs.

From the theoretical and the practical examination of Sierra and Debenham's information-based model for it trust can be concluded that the approach is promising, but that more work should be done. The core of the model seems to work well, but more conceptual grounding is desirable.

Bibliography

Ashri R., Ramchurn S.D., Sabater J., Luck M., & Jennings N.R. 2005. Trust evaluation through relationship analysis. Dignum F., Dignum V., Koenig S., Kraus S., Singh M.P. & Woolridge M. (Eds.) *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, Utrecht, The Netherlands, ACM, 1005-1011.

Castelfranchi C. & Falcone R. 1998. Principles of trust for MAS: cognitive anatomy, social importance and quantification. Demazeau Y. (Ed.), *Proceedings of the Third International Conference on Multi-Agent Systems*, Paris, France, IEEE Computer Society Press, 72-79.

Castelfranchi C., Conte R., & Paolucci M. 1998. Normative reputation and the costs of compliance. *Journal of Artificial Societies and Social Simulation*, 1(3), <http://www.soc.surrey.ac.uk/JASSS/1/3/3.html>.

Conte R. & Paolucci M. 2002. *Reputation in Artificial Societies*. Dordrecht: Kluwer Academic Publishers.

Fullam K., Klos T., Muller G., Sabater J., Schlosser A, Topol Z, Barber K.S., Rosenschein J, Vercouter L, & Voss M 2004. The agent reputation and trust (ART) test-bed competition game rules. *Laboratory for Intelligent Processes and Systems Technical Report*. <http://www.lips.utexas.edu/art-testbed/pdf/SpecSummary.pdf>

Fullam K., Klos T., Muller G., Sabater J., Topol Z, Schlosser A, Barber K.S, Rosenschein J, Vercouter L, & Voss M 2005. A specification of the agent reputation and trust (ART) test-bed: experimentation and competition for trust in agent societies. Dignum F., Dignum V., Koenig S., Kraus S., Singh M.P. & Woolridge M. (Eds.) *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, Utrecht, The Netherlands, ACM, 512-518.

Fullam K., Klos T., Muller G., Sabater J., Topol Z., Barber K.S., Rosenschein J., & Vercouter L. 2005. A demonstration of the agent reputation and trust (ART) test-bed: experimentation and competition for trust in agent societies. Dignum F., Dignum V., Koenig S., Kraus S., Singh M.P. & Woolridge M. (Eds.) *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, Utrecht, The Netherlands, ACM, 151-152.

Fullam K., Klos T., Muller G., Sabater J., Topol Z., Barber K.S., Rosenschein J., & Vercouter L. 2005. The agent reputation and trust (ART) test-bed architecture. Dignum F., Dignum V., Koenig S., Kraus S., Singh M.P. & Woolridge M. (Eds.) *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, Utrecht, The Netherlands, ACM, 50-62.

Fullam K., Sabater J., & Barber K.S 2005. Toward a test-bed for trust and reputation models. Falcone R., Barber K.S., Sabater J. & Singh M. (Eds.) *Trusting Agents for Trusting Electronic Societies*. Springer, Melbourne Autralia and New York USA, LNCS 3577: 95-109.

Jennings N.R., Faratin P., Lomuscio A.R., Parsons S., Sierra C., & Wooldridge M. 2001. Automated negotiation: prospects, methods, and challenges. *Int. J. of Group Decision and Negotiation*, 10(2): 199-215.

Jøsang A., Ismail R., & Boyd C. (to appear in Decision Support Systems) A survey of trust and reputation systems for online service provision.

MacKay D.J.C. 2003. *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press.

McKnight, D. & Chervany, N. 2001. The meanings of trust. *Trust in Cyber-Societies*, Berlin, Springer, LNCS 2246: 27-54.

Mui L., Mohtashemi M., & Halberstadt A. 2002. A computational model of trust and reputation for E-Businesses. Sprague R.H. (Ed.) *Proceedings of the 35th Hawaii International Conference on System Science (HICSS)* Big Island Hawaii, USA, IEEE Computer Society Press, 188-196.

Mui L., Mohtashemi M., & Halberstadt A. 2002. Notions of reputation in multi-agent systems: a review. Castelfranchi C. & Johnson W. (Eds.) *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems*. Bologna, Italy, ACM, 280-287.

Sabater J. 2002. Trust and reputation for agent societies (Ph.D.thesis). Autonomous University of Barcelona.

Sabater J. & Sierra C. 2005. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24: 33-60.

Sierra C. & Debenham J. 2005. An information-based model for trust. Dignum F., Dignum V., Koenig S., Kraus S., Singh M.P. & Woolridge M. (Eds.) *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems* Utrecht, The Netherlands, ACM, 497-504.

Sierra C. & Debenham J. (to appear in Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, 2006) Trust and honour in information-based agency.