# THE EFFECT OF TRUST AND TEAM PERSONALITY ON MULTIAGENT TEAMWORK PERFORMANCE

by

Feyza Merve Hafızoğlu

B.S., Computer Engineering, Boğaziçi University, 2007

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Systems & Control Engineering
Boğaziçi University
2010

## ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor Pınar Yolum for all her support during my years as an undergraduate and graduate student. Besides leading my studies, she also taught me how a researcher should be; I am privileged to study with her. Second, I would like to thank the members of my thesis committee, Haluk Bingöl, and Hakan Yaşarcan, for giving me valuable criticism.

Several people play significant role on my graduate study. Especially, I would like to thank my teachers: Levent Akın, Suzan Üsküdarlı, Cem Ersoy, and Necati Aras for their useful feedback, suggestions, and guidance beyond academics at critical points. I would like to mention the memory of my primary school teacher, Mehmet Yıldız, thankfully. Without his initial encouragement I would not be where I am today.

With gratitude to all my friends: Reyhan Aydoğan, Özgür Kafalı, Akın Günay, İlker Yıldırım, Derya Dönmez, Murat Şensoy, and Bülent Taştan for their friendship and profitable discussions about research problems. There are many people from whom I have benefited beyond academics. I would like to thank Betül Ceran, who is my soulmate, roommate, friend, citizen, and many more, and Dorina Strori, who is the interpreter of my feelings.

Last, but not least, I thank Mum and Dad, for always being there when I needed to talk, for caring my problems whenever I need their presence, and for their endless love, support, and trust. My brother is always with me to make me smile.

This research has been supported by The Scientific and Technological Research Council of Turkey.

## ABSTRACT


## THE EFFECT OF TRUST AND TEAM PERSONALITY ON MULTIAGENT TEAMWORK PERFORMANCE

Composed services consist of interacting services. Generally each service in a composed service is brought out by a different service provider. The quality of the composed service depends not only on the individual capabilities of the providers but also on how well they work together. In the pursuit of establishing effective teams, researchers propose several cognitive factors such as personality, trust, and leadership to model teams. In this theses, we study two significant factors: trust model and team personality. Existing trust models are geared towards identifying single services rather than composed services. However, in many settings it is important to find a group of service providers that can be trusted for a composed service. To address this, we propose a trust model that captures how trustworthy a group of service providers is for a particular composed service. The approach is based on capturing relations between services. Our proposed approach is tested on a modified version of ART Testbed. We compare our proposed model with an existing approach in the literature and show that capturing relations between services pays off in finding useful groups of service providers. For the second factor, we investigate the relationship between the team personality and teamwork performance. A promising personality model is composed of what is called Big Five Personality Traits: Agreeableness, Conscientiousness, Emotional stability, Extraversion, and Openness. We experimentally study the effect of these traits on multiagent teamwork. To do so, we model these traits and implement them in agents that can participate in ART Testbed by including interdependency attributes of teamwork. In this setup, we specifically study which traits are more significant than others for better performing teams, whether more trusted teams actually achieve a higher success rate than others, and whether heterogeneous teams perform better than homogeneous teams.

# ÖZET

## GÜVEN VE TAKIM KİŞİLİĞİNİN ÇOK ETMENLİ ORTAMLARDA TAKIM ÇALIŞMASININ PERFORMANSINA ETKİSİ

Bileşik servisler birbiriyle etkileşim içinde olan servislerden meydana gelir. Genellikle bileşik servislerde her bir servis farklı bir servis sağlayıcısı tarafından sağlanır. Bileşik servisin kalitesi sadece servis sağlayıcılarının yeteneklerine değil birlikte ne kadar iyi çalıştıklarına da bağlıdır. Başarılı takımlar kurabilmek için araştırmacılar güven, kişilik, liderlik gibi birçok faktörü incelemişlerdir. Bu tezde güven ve kişilik faktörlerinin takım performansına olan etkisi incelenmiştir. Varolan güven modelleri bileşik servisler yerine tek servisleri gözönünde bulundurmuştur. Ancak, bir bileşik servis için güvenilecek bir takımın bulunması çoğu zaman önemlidir. Bu durumu ele almak amacıyla, bir bileşik servis için bir grup servis sağlayıcısına güveni gösteren bir model öneriyoruz. Yaklaşımımız farklı bileşik servisler arasındaki ilişkiye dayanmaktadır. Önerdiğimiz modeli litaratürde varolan modellerle kıyasladığımızda, önerilen modelin seçtiği servis sağlayıcısı grupların daha başarılı olduğu gözlenmiştir. İkinci faktör olarak, takım kişiliği ve takım çalışması performansı arasındaki ilişkiyi inceledik. Kişilik modellemesi için psikoloji literatüründe çok popüler olan Beş Büyük Faktör kuramını kullandık. Beş Büyük Faktör kişilik özelliklerini beş sınıfa ayırır: uyumluluk, sorumluluk, duygusal denge, dışadönüklük ve açıklık. Bu özelliklerin çok etmenli ortamlarda takım çalışmasına etkisini deneysel olarak çalıştık. Bunu gerçekleştirebilmek için, kişiliğe sahip etmenleri bağımlılık özelliğini gösteren takım çalışması ortamında geliştirdik. Bu ortamda, takım başarısı için hangi kişilik özelliklerinin önemli olduğunu, güveni yüksek takımların başarılarının da yüksek olup olmadığını, ve aynı yapılı olan takımların daha başarılı olup olmadığını inceledik. Bu çalışmayı gerçekleştirmek için tek servisleri içeren Agent Reputation and Trust (ART) Testbed, bileşik servisleri içerecek şekilde değiştirilerek güven ve kişilik modelleri geliştirilmiştir.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF SYMBOLS/ABBREVIATIONS

| | |
|---|---|
| $\alpha$ | Penalty weight |
| $\beta$ | Reward weight |
| | |
| $A_i$ | Agent $i$ that performs $T_j$ |
| A($A_i$) | Agreeableness level of $A_i$ |
| C($A_i$) | Conscientiousness level of $A_i$ |
| E($A_i$) | Emotional Stability level of $A_i$ |
| Ex($A_i$) | Extraversion level of $A_i$ |
| $P(A_i)$ | Performance of $A_i$ |
| $T_i$ | Task $i$ |
| $TW_i$ | Teamwork $i$ |
| | |
| ART | Agent Reputation and Trust |
| IMA | Individual Modeling Agent |
| TMA | Teamwork Modeling Agent |

# 1. INTRODUCTION

Multiagent systems are distributed computing systems that are composed of a number of interacting computational agents [3]. Agents are persistent computations that can perceive, reason, act, and communicate. They can be developed independently of each other and come together to work in a multiagent system. On the web, agents usually carry out services. When the agents take part in a multiagent system, they fulfill composed services.

Most real life needs are satisfied by composite services, rather than single services. Such composite services are realized by teams of service providers. Thus, teamwork is an essential part of multiagent systems. Teamwork needs to be considered in parts and different parts need to be performed by different agents with necessary competencies. Team formation, in other words selecting appropriate team members for a particular composed service, is important for teamwork performance. For the purpose of establishing successful teams, researchers study the dynamics of teamwork and try to characterize successful teams [8, 9, 10]. Based on the studies in the social sciences literature, the combination of several criterion among team members such as skills, cognitive ability, personality composition, leadership, or trust seem to be important. We study two factors of effective team formation: *trust* to a team for a particular composed service and *personality composition*.

In this study, the multiagent teamwork model is corresponding to composed services, and a teamwork consists of a number of tasks that corresponds to services of a composed service. Finally, a group of service providers are referred as a team of agents.

The first factor that we studied is *trust to a team*. In dynamic open systems, many agents interact with each other to achieve their goals. In such environments, a self-interested agent selects most trusted and suitable partners to interact with from a pool of agents, whose behaviors are not known. Ideally, an agent should interact with the agent, who fulfills the expectations of the requester agent. A trust model consists of opinions of an agent about other agents; it's formed by using its own experience with the related agent and the other agents' opinions about the related agent. Each agent builds its own trust model and uses this model to decide on whom to trust.

When an agent needs a team of agents rather than a single agent to fulfill its request, the agent should consider the trust to the team instead of the trust to each individual agent in this team. In this case, the agent needs a trust model to evaluate the trustworthiness of possible teams. Consider a team in the production line of a furniture workshop. The team contains four members: a designer, who designs the furniture, a carpenter, who produces the components of a furniture from wood according to the designer's drawing, an assembler, who combines the pieces of furniture in conformity with how the carpenter cut the pieces, and finally a painter, who paints the furniture. Individual tasks are accomplished sequentially in the order described above. The carpenter's task is strongly dependent on the furniture design, hence the carpenter and the designer frequently interact with each other. Assembling is dependent on the carpenter's work, and finally painting is strongly dependent on the furniture design. It is obvious that the manner in which one provider works will affect the service of another provider in the team. Even though, the service providers work well individually, they may not have the same performance when they work together. Hence, the owner of the furniture workshop have to establish a team of service providers that she can trust for the composed service.

Whereas a vast literature exists for modeling trustworthiness of individual service providers, there is not much work done in modeling teams of providers for a given teamwork. Developing a team model for trust requires the following questions to be answered. How can one agent model a team of other agents? How can previous experiences with the agents be used to predict their behavior in a team?

In addition to trust, this study investigates *the influence of team personality* on multiagent teamwork performance. In other words, we study how the personalities of the team members affect the success of the team. Considering the furniture workshop example: Is it important that the designer's personality is compatible with the personalities of the carpenter and the painter. Furthermore, what if the designer exhibits a depressed behavior, will the painter and the carpenter need to consume extra energy to deal with the designer's behavior? We use *The Big Five* [7] personality model that categorizes all personality traits into five fundamental factors: agreeableness, conscientiousness, emotional stability, extraversion, and openness. We examine which traits are predictive in teamwork performance. In addition to this, we characterize the successful teams whether they are homogeneous or heterogeneous according to certain traits.

We consider teams that have varying levels of task interdependence, high role differentiation, high task differentiation, and distributed expertise. Their tasks are interdependent on each other, and should be performed sequentially or simultaneously according to the interdependency relation. Each team member has high level of expertise on certain subjects, and these members frequently interact with each other.

In open distributed environments, the agents require teamwork to satisfy their needs. In order to carry out the tasks of teamwork, teams are established by using agents with personalities and varying job skills. Agents have distributed expertise that corresponds to the job specific skills. In this case, the service requester composes teams according to its team formation strategy, which is its trust model of others in this study. For the purpose of determining how an agent is suitable to carry out a task, an agent models trust to the other agents in the environment. Considering trust to a team rather than individuals, we develop agents that can effectively model trust to possible teams for particular teamwork based on several criterion. An agent can update its trust model by using its past experiences with the other agents and the reputation information that is obtained from the other agents. After accomplishment of teamwork, the performance of the team is measured with respect to a criteria and the agent updates opinions about the team members according to this performance.

We study the team formation process, where the agent decides which agents will participate in teamwork. In decision making process, there are several factors to consider: trust to other agents, the personalities of agents, the agents' experiences, etc. This setting is applied in a customized version of ART [1], which is modified to enable teamwork including interdependency relationship.

The rest of this theses is organized as follows: We give related work in Chapter 2. Chapter 3 develops a teamwork trust model that consists of new tools: team trust model and service graphs, for modeling trustworthiness of teams. Chapter 4 describes our experimental framework, ART, and explains how we enable teamwork in ART. In Chapter 5, we compare the teamwork trust model with traditional individual trust model. We introduce agents with personalities in Chapter 6. Then, we look at the influence of personalities on teamwork performance in Chapter 7. Finally, Chapter 8 concludes.

# 2. LITERATURE REVIEW

Most approaches to trust consider agents individually and predict their trustworthiness accordingly. However, in many real-life settings, an agent has to interact with a group of agents to receive a composed service. This study proposes a teamwork trust model to understand the behavior of such teams that carry out a teamwork.

Barber [18] presents a trust-based mechanism for team formation problem where agents selectively pursue partners of varying trustworthiness in a market-based environment, where a job consists of multiple subtasks and agents have different skills which correspond to subtasks. A certain percentage of the agents are randomly selected as contractors at each round and they decide to continue their current job or a new job, which turns to establish teams to work on their new job, by using a greedy heuristic. Candidate members of the team have different tendencies towards completing an assigned task. Results show that an agent may utilize better by selecting less trustworthy partners with comparison to more trustworthy partners. In contrast to our group trust model, this study proposes a trust model with the aspect of the participants of teamwork, the agents are modeled individually based on the tendency to complete a subtask and consider subtasks requiring different number of rounds to complete, and maximize the profit. The behavior of the agents in the team doesn't differ based on the team or teamwork, instead they have certain characteristics to continue or leave their current job based on maximizing their profit.

TRAVOS [16] is a probabilistic trust model that considers both trust and reputation in order to handle the possibility of inaccurate reputation information. Self-interested agents may betray the trust by not performing the requested action as required. In TRAVOS, trust is calculated using probability theory between agents considering the past interactions. Whenever there is little or no interaction with an agent, the agent uses the reputation information gathered from third parties. This study especially handles the possibility of inaccurate reputation information based on the interactions with the agent whom requests the reputation information. However, TRAVOS does not provide a modeling mechanism to evaluate teamwork.

Another solution [17] is developed by using Bayesian approach and deals with the sequential decision making problem of agents operating in computational economies. It allows agents to incorporate different trust priors and explore optimally with respect to their beliefs when choosing potential service or information providers. The trustworthiness of the agents in the environment is uncertain. A generic Bayesian Reinforcement Learning algorithm is applied to the exploration-exploitation problem where agents decide whether to keep interacting with the same "trusted" agents or keep experimenting by trying other agents with whom they haven't had much interaction so far. This algorithm considers the expected value of perfect information of an agent's actions to take optimal sequential decisions; it's applied to the ART Testbed scenario.

The proposed solution in Blizzard [15] is an action-based approach for modeling the environment; and it is also developed in ART Testbed. Blizzard differs from traditional agent-based trust models by modeling actions of the agent and their effect on the environment instead of models all agents individually. Q-learning method which originally deals with actions and states is used by removing state mapping since there is no state info in ART. Three versions of the Blizzard is developed and compared with Frost agent which is an agent-based trust model in the evaluation part, and it dramatically outperforms the agent-based approaches during evaluations.

Many studies that examine the relationship between personality and teamwork performance exist in the literature of social sciences such as psychology, and management. One of the most supported classification of personality characteristics is Big Five personality traits. Digman [7] prepared a comprehensive review of emergence of the Big Five personality traits. Many researchers [4, 5, 6] attempted to develop a personality model and found nearly the same big five personality traits independently. While, they seem to be agree on the dimensions of the model, their models differed in small details from each other.

Barrick and Mount [8] investigate the relationship between Big Five Personality traits and three job performance criteria (job proficiency, training proficiency, and personnel data) for five occupational groups (professionals, police, managers, sales, and skilled/semi-skilled) using meta-analytic methods. They look at the score correlations between personality dimensions and job performance criteria for different occupations in 117 studies. Results indicates that conscientiousness shows positive relations with performance for all occupations. Extraversion is found to be a valid predictor for occupations involving sociability, managers and sales (across all criterion). Openness and extraversion are valid predictors for training proficiency (across occupations). They also obtain other relations between personality dimensions and occupations however the magnitudes of the correlations are not remarkable($< 0.10$).

Neuman, Wagner, and Christiansen [9] study teamwork performance with two different aspects of personality compositions of teams: team personality elevation (TPE), a team's average level on a particular personality trait, and team personality dimension(TPD), the variance among team members for a particular personality trait. They examine which traits improve performance when the team is homogeneous, and which traits improve the performance when the team is diverse. Their research question is whether TPE and TPD uniquely predict the performance of work teams. The personalities of 328 retail assistants working in 82 teams were assessed via California Personality Inventory (CPI) and Personal Audit (PA) tests. They develop two performance measures: a rating of customer service and a rating of task completion. Results indicates that TPE was positively related to team performance for the traits of conscientiousness, agreeableness, and openness; TPD of extraversion and emotional stability was positively related to team performance.

Kichuk and Wiesner [10] study the relationship between big five personality factors and the performance of three-member product design teams. Their focus is to describe the optimal combination of personality. Successful teams differ from unsuccessful counterparts as having higher levels of general cognitive ability, higher extraversion, higher agreeableness, and lower neuroticism. Additionally, they deduce that the heterogeneity of conscientiousness is negatively and significantly related to the performance of the team.

Costa [11] presents a comprehensive study of the relationship between trust and team performance. Trust is represented as a multi-component variable with distinct but related dimensions. Relating trust with perceived task performance, team satisfaction, she tests a model by gathering the data of 112 teams of 395 people from three social care institutions in The Netherlands. The results advocates that trust within team is positively related with perceived task performance, which is assessed by objective measures, and team satisfaction, which assesses the extent to which team members are satisfied with their teamwork.

Kang [13] examines the relationship between agent activeness and cooperativeness with team decision efficiency by using a team model 'Team-Soar' consisting of four intelligent agents. Team-Soar models agent activeness, that is derived from extraversion, at two levels (active and passive), and agent cooperativeness, that is derived from agreeableness, at three levels (cooperative, neutral, and selfish). Results show that the effect of agent type depends on the amount of information during decision process, active agent improves the effects of agent cooperativeness on team decision efficiency, and finally active agents do not always perform better than passive agents.

Aghaee [19] studies fuzzy agents with dynamic personalities. The personality of agents are updateable and represented by using Big Five personality traits. These five dimensions are extended into thirty facets, where each agent has different levels of these facets, namely low, medium, and high. A set of if-then rules are geared toward personality descriptors, factors, characteristics, and modifications. The behavior of the fuzzy agents and the personality updates that correspond to the dynamic personality are performed according to the levels of agent's personality facets and the rules. Even though, they present a richer personality representation, they don't apply this model to any multiagent framework.

Sichman [20] proposes a dependency graph for the representation of dependency relations of multiagent systems. The agents that cannot accomplish every action due to their limited capabilities or resources search for partners to achieve their goals. The graph consists of actions and agents, where the edges are directed from the agent to an action if the agent is dependent on other agents for this action or from the action to the agent if the agent has a capability or resource to accomplish this action. As a result of these dependency relations, social structures namely groups and collectives emerge. Collectives and groups differ from each other by different levels of complexity and cohesiveness of the underlying dependence graph. When the complementary agents share the goal for which they are needed, full collective dependence occurs.

# 3. MODELING AGENTS WITH TEAM TRUST MODEL

When a group of providers are sought for a teamwork, considering the providers' individual behavior is not enough because in carrying out the teamwork providers will be participating in a team and their behavior in a team may be different. Teamwork trust problem emerges with the following issue: the behavior of an agent in teamwork environment may differ from its behavior in single task environment. In teamwork, the behavior of the agent depends on the teamwork, other agents in the team, and so on.

One can naively think that whenever a group of agents are required for a teamwork, we may select the most trusted agent for each task of the teamwork in the environment. But, there is no guarantee that an agent has the same performance when it is taking place in teamwork and when it is acting individually.

Being in a collaboration may have a positive or negative effect on the performance of agents. For example, an agent, who is a successful painter, works very well individually. However, it has a worse performance when it participates in teamwork as a painter. As a conclusion, the idea behind the teamwork is totally different from a single task and it is more complicated in the sense of both representation and reasoning. Hence, considering only the participant's individual trustworthiness is not going to be enough to understand the trustworthiness of a team.

Possible tendencies of agents, who participate teamwork, can be listed as the following:

- *Ideal Behavior:* The agent performs well both individually and in teamwork.
- *Group Antipathy:* The agent may dislike being in a team. Thus, whenever the agent participates in teamwork, its performance will worsen.
- *Group Motivation:* The agent performs well in teamwork even if it does not perform well individually, i.e. other agents may help this agent. Being a member of a team has a positive influence on the agent's behavior.

- *Colleague Effect:* The agent's behavior changes based on the other agents in the team. The agent may perform better with some agents but not with others.
- *Teamwork Effect:* The agent may have a bad performance due to the teamwork characteristics. For example, a painter may work well with another plumber but not so well with an electrician.
- *Familiarity Effect:* The agents in the team improve their performance as the number of times that they come together to carry out a teamwork increases.

## 3.1. Representation

In this section, we focus on agents that model trust to the agents in the environment. In the purpose of modeling, our agents use three different representations: service graphs, team trust model, and individual trust model to assess the expertise of the other agents and teams in the environment. The expertise of an agent with respect to a task in a trust model represents the expectation about how much this agent would be successful when it performs this task. Similarly, the expertise of a team with respect to a teamwork is equal to the expectation of the success of this team for this teamwork. The expertise is corresponding to the trustworthiness of an agent or a team. Service graphs define the relationship between different teamwork. Team trust model enables agents to classify their teamwork experiences with respect to the teamwork and the team that perform this teamwork. Finally, individual trust model keeps the experiences with respect to tasks and individuals. First, we explain the simplest model, individual trust model, and continue with team trust model and service graphs.

### 3.1.1. Individual Trust Model

The individual trust model includes the answers of the question: how can I trust $A_x$ for $T_y$? We use this basic model to support single tasks. If there are $n$ agents and $m$ tasks in the environment, there exist $n$ *times* $m$ individual trust model instances for each agent-task pair. A model instance contains the information of the past interactions with the agent for a particular task such as the expertise weight, number of interactions, that is the number of times an agent is requested to carry out this task and the task is accomplished by this agent, and so on. At the initialization of the agent, a default model, which has an expertise weight of value 0.5, is created for each agent-task pair in the environment. Individual trust model is used when an agent is assigned a single task rather than a teamwork. In this case, a provider agent carries out the single task and the corresponding individual trust model instance for this agent-task pair is updated. The expertise weight is increased or decreased based on the performance of the provider agent and the number of interactions is increased by one.

### 3.1.2. Team Trust Model

In team trust model, each agent classifies its experiences with respect to the requested teamwork and the agents that participate in carrying it out. A team trust model instance mainly consists of a task list and the corresponding agents, who are assigned these tasks: that is, a list of agent-task pairs.

A team trust model has the information about the agents that exist in the team and the tasks, which these agents are assigned. Each model has an expertise weight, which is updated after each experience of the team for the same teamwork and agent-task assignments, and the number of interactions, which is also an indicator of the accuracy of the expertise weight. As the number of interactions increases, the expertise weight would be more accurate as mentioned in several models in the literature. The expertise weight of a team trust model instance has a default value of 0.5 when it's created, and it's updated based on the teamwork performance.

### 3.1.3. Service Graphs

In order to characterize the tasks in a teamwork, a graph-based representation of services [2] is used. A service graph is a weighted, directed graph including nodes for teamwork and the edges for transitions, which represents the relationship, between two teamwork. Each teamwork is represented as a node in the graph. The weights on the edges show how likely the providers that are successful in a source node are likely to be successful in the target node. By using this relationship between different teamwork,

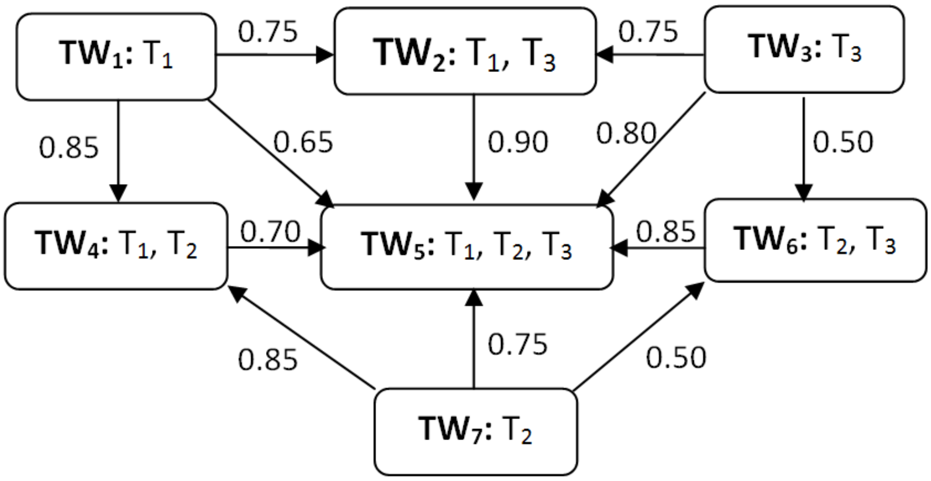


Figure 3.1. An example service graph for teamwork including $T_1$, $T_2$, and $T_3$

the agent composes a new group of agents as a team for a given teamwork by using past team experiences of other teamwork in the graph.

In the service graph, only the nodes or teamwork that have at least one common task are connected to each other. Otherwise, there is no relationship between two nodes. Each edge has a weight that has similar functionality with the expertise in the graph and a default weight of 0.5 is assigned when it is created. The weights capture the likelihood of a team in the source node to be successful in the target node. Each agent constructs its own service graph according to its teamwork experiences. Thus, the information in a service graph represents the opinions of the owner agent of this service graph.

When it's the first time that the agent is assigned a teamwork, the agent adds the directed edges to the service graph for this teamwork, which becomes the target node, and the other teamwork with combinations of the task set of this teamwork become source nodes. An example service graph is shown in Figure 3.1. $TW_5$, which consists of $T_1$, $T_2$, and $T_3$, is the target node, and $TW_1$, $TW_2$, $TW_3$, $TW_4$, $TW_6$, and $TW_7$ are the source nodes. In this service graph, the edges are experienced several times by the owner agent of the graph, and the current weights of the graph has different values according to their usefulness for establishing teams that will participate in $TW_5$.

For example, using the teams that participated in $TW_2$ possibly result with successful teams for $TW_5$ because the weight of the edge that is directed from $TW_2$ to $TW_5$ has a value of 0.90 that is a higher weight and an indicator of expectation of success. However, using teams that participated in $TW_4$ for the same purpose would not be a good idea because the weight of its edge has a value of 0.70.

As we asserted our motivation before, the agent may not behave the same way in two situations: when it is acting independently and when it is taking place in a team. The most important information for an agent, who is assigned a teamwork including more than one task, is the edge weights of the graph. Each edge of the graph has an information of the weight and the number of usages, which is number of times that an edge is used to establish a team for the teamwork in the target node of this edge by using the previous team experiences of the teamwork in the source node of this edge. These edges of the target teamwork are used to establish a team when this target teamwork is assigned to the agent. The agent selects a set of edges from the service graph of the assigned teamwork, where the union of tasks of these edges' source nodes is equal to the tasks of the target node. The agent selects the most useful edges with higher weights and composes teams, which are experienced before for the source teamwork of the selected edges, to form a team for the currently assigned teamwork. The weights of the selected edges are updated based on the performance of the actual team whenever it's experienced and the number of usages is increased by one for each selected edge.

*Example:* Let's say that an agent is assigned to accomplish a teamwork, which consists of $T_1$, $T_3$, $T_5$, and $T_7$, and decides to use the service graph of this teamwork to form a team. Related nodes (teamwork) are all subgroups of this teamwork, such as a teamwork, which consists of $T_1$ and $T_3$, or a teamwork, which consists of $T_5$ and $T_7$. The agent should follow the possible edges to establish a team for the assigned teamwork. Specifically, it selects a set of useful edges considering the weights of the edges. Now, the target node of these selected edges would be used to find the team instances. Let's say two edges, those come from this teamwork, which consists of $T_1$ and $T_7$, and the teamwork which consists of $T_3$ and $T_5$, are selected. These two teamwork are composed to obtain the currently assigned teamwork. Actually, the agent composes the teams of agents experienced before for these selected teamwork.

### 3.2. Updating Models

Whenever an agent uses its trust tools, the corresponding edge weights of its service graphs and expertise of team and individual trust model instances are updated according to the performance of corresponding teamwork. Assume that their performance is assessed as a real number between 0 and 1. The same update method given in Equation 3.1 is used for all three models.

$$expertise := \begin{cases} \alpha * performance + (1-\alpha) * expertise, & expertise > performance \\ \beta * performance + (1-\beta) * expertise, & expertise < performance \end{cases} \tag{3.1}$$

The reward and penalty weights can be monitored to obtain the most suitable strategy. In this study, we prefer to use high penalty weight and low reward weight in order to increase the accuracy of the models. For example, when an agent obtains a better result than its current expertise value, which is defined in the individual trust model, then we increase its expertise weight by a small amount. However, if it performs worse, we decrease its expertise by a higher amount to penalize this agent [14, 15].

### 3.3. Team Formation Strategies

The team formation strategy of the agent for a particular teamwork depends on the number of tasks in teamwork. If the number of tasks is one, the agent uses individual trust model to find the most suitable provider agent to carry out this task. Otherwise, the team trust model and service graphs are used to establish a team.

Remember that the individual trust model includes all agent-task pairs. When the agent is assigned a teamwork that contains only one task, it selects the most trusted agent for this task. The most trusted agent means that the agent with the highest expertise for the corresponding task in its individual trust model.

Finding trusted teams of agents for a teamwork is more complicated; there are four alternative strategies. The agent pursues these alternatives one by one in the order as shown below. Once it finds a suitable team of agents in any of the steps, it finalizes the team formation procedure. In order to use the first and the second strategies, the agent should have satisfactory experiences with the corresponding teamwork in the past.

i. *Using the exact experience which is higher than a certain threshold from the team trust model instances:* The first strategy is similar to the strategy used for teamwork with one task, the agent uses the expertise of the team trust model instances with the same teamwork. If there exist successful experiences, whose expertise weights are higher than a certain threshold, for exactly the same teamwork type, the team with the highest expertise can be used.

ii. *Using a set of edges of the service graph if edges with high weights exist in the graph:* The second strategy uses the service graph information, namely the weights of the edges. If there exists a set of edges, whose average weight is higher than a certain threshold value, the agent selects these edges, where the union of tasks of these edges' source nodes is equal to the tasks of the assigned teamwork. Then, the agent looks at its team trust models to find an appropriate team instance for each selected edges' source node (teamwork) by using the method that is used in the first strategy.

iii. *Using inexperienced edges whose team instances have an expertise value higher than a certain threshold:* Third strategy uses the organization of the service graph between teamwork, namely the task lists of edges' source nodes in the graph rather than the edge weights. The agent finds all possible edge sets for the current teamwork, and then finds the best team instances for each possible edge set by looking at its team trust model instances and compose these teams to obtain a final team. The edge set, which has the highest average expertise based on the expertise of the team instances, and the corresponding teams for this edge set are selected. If the highest average expertise of composed teams is higher than a certain threshold, these team instances of the selected edges' source nodes are used to carry out the teamwork.

iv. *Using the individual trust model:* If the agent cannot find an appropriate team after trying first three strategies, it uses the last strategy. In this case, it selects agents one by one for each task of the teamwork considering individual trust models without considering any threshold value, since no alternative strategy exists in this step. This strategy works well with the agents those have *ideal behavior* mentioned at the beginning of this chapter.

Note that if the agent uses the second or third strategies in the current timestep, it keeps the selected edges for the corresponding teamwork to update the edge weights of the service graph according to the teamwork performance. Furthermore, the agent also keeps the agent and task pairs to update the corresponding team trust model instances.

*Example:* Let's say an agent is assigned $TW_5$ that contains $T_1$, $T_2$, and $T_3$ in the $85^{th}$ timestep and its current service graph is given in Figure 3.1. The agent will establish a team by using its team formation strategies that are explained above. Assume that there are 4 team trust model instances, in other words teamwork experiences, for this teamwork type and they're given in Table 3.1.

Table 3.1. Team model instances for teamwork contains $T_1$, $T_2$, and $T_3$

| | Team | Expertise | Number of interactions |
|---|---|---|---|
| Instance-1 | $A_1$, $A_7$, $A_5$ | 0.60 | 2 |
| Instance-2 | $A_5$, $A_7$, $A_1$ | 0.72 | 3 |
| Instance-3 | $A_3$, $A_6$, $A_1$ | 0.65 | 1 |
| Instance-4 | $A_4$, $A_2$, $A_5$ | 0.78 | 4 |

- *Strategy1:* If there is a satisfactory team instance, the agent just uses the same team with the same agent-task assignments. To do so, the agent selects the team instance with the highest expertise, which is Instance-4, from its past experiences given in Table 3.1 and compares the success of this team with the threshold that has a value of 0.80 at the $85^{th}$ timestep. Unfortunately, Instance-4 is not sufficient in this step. Thus, the agent prefers to explore new teams by trying following methods rather then exploiting past teamwork experiences. In the next step, the agent will try to use the service graph of this teamwork.

- *Strategy2:* The agent considers the edge sets of the service graph for this teamwork. In Figure 3.2, the service graph of $TW_5$ and the corresponding teamwork and task experiences of the owner agent of the graph are provided to show an agent uses its service graph in team formation. Remember that the agent carries out a two step procedure to establish a team by using service graphs: (1) select the useful edges with the highest average weight, and (2) select the team experiences with highest expertise from the corresponding team trust models of selected edges' source nodes. In Figure 3.2, the team trust models and individual trust models are shown with thick arrows that are directed from the teamwork or task to team trust model or individual trust model. For example, there are four teamwork experiences of the agent for $TW_2$ and $TW_6$, and three teamwork experiences for $TW_4$. On the other hand, there are seven individual trust model instances for each task. Because there are seven agents in the environment and all possible individual trust models are created with the default expertise when the agent is initialized. However, the team trust model instances are created with default expertise when the agent experiences a teamwork for the first time.

  Now the agent that is searching for a new team for $TW_5$ evaluates the possible edge sets in the service graph and selects the edge set with the highest average weight. Four possible edge sets are given in Table 3.2. The tasks of the teamwork that is source of an edge are given in parenthesis. For example, *Edge Set-1* consists of two edges that come from two teamwork: $TW_4$ that consists of $T_1$ and $T_2$, and $TW_3$ that consists of only $T_3$. The average weight of an edge set is calculated for *Edge Set-1* by using the edge weights given in Figure 3.1 as the following:

  averageWeight = ( 0.70*2 + 0.80 ) / 3 = 0.73

  The average expertise of *Edge Set-2* has the highest value that is 0.85 among possible edge sets given in Table 3.2. Assume that the threshold for using service

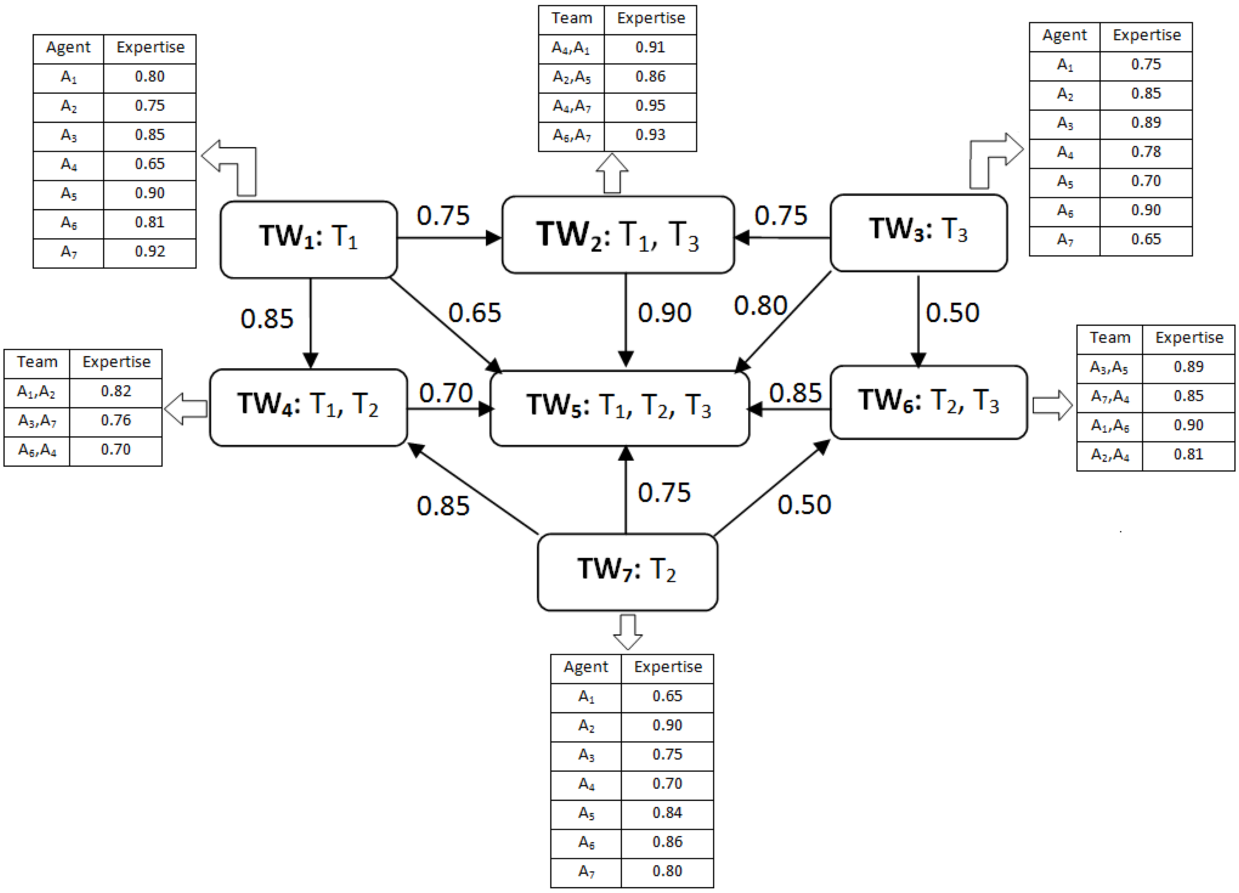


Figure 3.2. An example service graph for teamwork including $T_1$, $T_2$, and $T_3$ and past teamwork experiences

graphs has a value of 0.80 at the $85_{th}$ timestep. This means that the average weight of the best edge set is sufficient to form a team. The agent will establish the team by using its service graph for *Edge Set-2*. Now, the agent considers the team and individual trust model instances for two teamwork in the *Edge Set-2*, $TW_2$ and $TW_7$. The agent considers its team trust model instances for $TW_2$. The team with the highest expertise consists of $A_6$ and $A_7$, and their expertise is a value of 0.93. They carry out $T_1$ and $T_3$, respectively. The agent considers its individual trust model instances for $TW_7$, in other words $T_2$. The agent with the highest expertise, which is value of 0.90, is $A_2$. Finally, the agent composes a team consists of $A_6$, $A_2$, and $A_7$ for carrying out $T_1$, $T_2$, and $T_3$, respectively. In addition to this, a new team trust model instance for this teamwork is created with a default expertise of 0.5. The expertise value of this new team model

Table 3.2. Possible edge sets for $TW_5(T_1,\, T_2,\, T_3)$

| | Edges | Avg. Expertise of Edges | Number of usages |
|---|---|---|---|
| Edge Set-1 | $TW_4(T_1,\, T_2)$, $TW_3(T_3)$ | 0.73 | 1 |
| Edge Set-2 | $TW_2(T_1,\, T_3)$, $TW_7(T_2)$ | 0.85 | 2 |
| Edge Set-3 | $TW_1(T_1)$, $TW_6(T_2,\, T_3)$ | 0.78 | 3 |
| Edge Set-4 | $TW_1(T_1)$, $TW_7(T_2)$, $TW_3(T_3)$ | 0.73 | 1 |

instance and the weights of the edges in *Edge Set-2* are updated according to the overall performance of this team.

## 3.4. Non-cooperative Behavior

In this part of the study, we consider agents, whose behavior may change based on the particular teamwork that it's taking part of. This attitude corresponds to *teamwork effect* given at the beginning of Chapter 3. That is, even if an agent performs well independently, in certain types of teamwork, its performance may worsen. Each agent has a finite list of teamwork in which it is going to be non-cooperative. This is called the *noncooperativeness list* and each agent constructs its own list randomly. Thus, *noncooperativeness list* is different for each agent. The *noncooperativeness level* shows the extent of cooperation. If the noncooperativeness level of the agent is 0, then the agent cooperates with all assigned tasks. If it is 1, then the agent never cooperates in any of the possible collaborations.

# 4. EXPERIMENTAL FRAMEWORK

## 4.1. Technical Background: ART

ART is a well-known experimental framework that is originally designed and developed for comparing and evaluating different trust strategies that act in harmony within a common framework. We explain original architecture of the ART and how we modify ART to enable teamwork environment.

### 4.1.1. ART Game Overview

ART [1] consists of appraiser agents that appraise the value of paintings. Each painting belongs to a particular era from a set of eras and agents have varying expertise in these eras. An agent's expertise is its ability to generate an opinion about the value of a painting. Agents are not aware of the expertise levels of other agents in the environment but their own expertise levels. At each timestep, agents are assigned a number of paintings by the simulator to appraise. All appraiser agents are assigned the same number of paintings at the beginning of the game. Agents are paid a fee for each appraised painting. If an agent has low expertise about an era, which assigned painting belongs to, the agent asks opinions of other agents to come up with more accurate appraisals. Intuitively, the agent should query the agents who have higher expertise values of corresponding era to increase the profit. But, the expertise values of other agents are not directly known by the agent. So, each agent tries to model and learn the expertise of other agents for existing eras by using its past experiences with other agents and reputation information of an agent that is requested from other agents. As the accuracy of the appraisals of the assigned paintings increases, agents have more clients and consequently earn more money. The goal of agents is to maximize their *bank balance* by minimizing appraisal error in other words producing more accurate appraisals. In this setting, the agents that are assigned paintings to appraise their values are called appraiser agents and the agents, whose opinions are requested, are provider agents. Creating an opinion is a task, thus creating opinions for different eras can be thought as different task types. Finally, hte expertise of agents corresponds to skills of provider agents.

The ART game continues a number of timesteps that is defined by a parameter. At each timestep, ART simulator is responsible for assigning paintings to the appraiser agents for evaluations, receiving the answers of agents about the painting, calculating the appraisal error according to the true value of the painting, and informing the agents about appraisal error as the indicator of their performance. The agents, then, can act accordingly.

### 4.1.2. ART Parameters

At the beginning of the game, the simulator assigns the expertise levels of each agent for each era, where these expertise values are different from each other with respect to eras and agents. During the game, the expertise levels slightly change; the amount of change is determined by the parameter *Expertise Change Rate* that is given in Table 4.1. Appraiser agents may ask the certainty of provider agents, which is an information about the expertise of an agent about a particular era to decide from which agents to ask opinion. Provider agents are also paid fixed fees for each opinion and reputation that they provide. So, agents may also increase their profit by selling opinions and reputation information to other agents. Fees for selling opinion and reputation information are determined by the game parameters *Opinion Cost* and *Reputation Cost* in Table 4.1. During the game, the correctness of the replies are not guaranteed, and the strategies of other agents are not known due to the heterogeneity of agents. In fact, it is quite likely that agents provide incorrect information in order to decrease the requester's client base in such a competition environment.

### 4.1.3. ART Communication Protocols

There are three communication protocols between agents in ART: reputation protocol, certainty protocol, and opinion protocol. Communication protocols are strictly regulated and all agents follow the protocol synchronously. All three protocols are

Table 4.1. ART game parameters

| Parameter | Description |
| --- | --- |
| *Average Clients Per Agent* | The number of paintings that each agent is assigned at the first timestep. The number of assigned paintings of an agent decreases or increases based on average appraisal error of the agent in the rest of the game. However, total number of paintings at each timestep remains fixed. Default value is 20. |
| *Expertise Change Rate* | The amount of change in expertise levels. Default value is 0.01. |
| *Client Fee* | How much paid an agent for appraising the value of a painting. Default value is 100 |
| *Opinion Cost* | How much paid to the agent whose opinion is asked. Default value is 10. |
| *Reputation Cost* | How much paid for a reputation request about an agent. Default value is 1. |
| *Certainty Cost* | How much paid for asking expertise level of an agent. Default value is 2. |
| *Number of Timesteps* | How long the simulation continues. Default value is 100. |
| *Number of Painting Eras* | How many eras there are in the game. Default value is 10. |
| *Number of Agents* | Number of agents in the simulation. Default value is 6. |
| *Bank Balances* | The total bank account of the agent. |

performed in every timestep, respectively.

i. *Reputation Protocol:* The requester agent sends a reputation request message, where the agent about whom (and era with respect to which) it is requesting reputation information is given, to a possible reputation provider. After receiving the message, potential reputation provider sends an *accept* or *decline* message according to its will. Finally, the requester agent sends payment for the reputation information. The reputation cost is determined by the parameter *Reputation Cost* in Table 4.1. The accuracy of the reputation information depends on the provider agent's will. For example, honest agents may reply with 100% accuracy. However, cheating agents may send artifice information to beguile the requester agent.

ii. *Certainty Protocol:* The requester agent sends a certainty request message to a potential opinion provider agent in order to learn its expertise level for a particular era. Upon receiving a request message, the provider agent, whose certainty is requested, sends an expertise level back to the requester agent. The accuracy of the expertise level depends on the provider agent's nature whether it's honest or cheating.

iii. *Opinion Protocol:* The appraiser agent first determines the opinion provider agents to ask their opinion about a painting according to its decision strategies, which vary in a range from simpler strategies like random selection to more sophisticated strategies like Blizzard [15]. Then, it sends opinion request messages, where the painting is identified, to the selected provider agents. After receiving requests, opinion provider agents send their opinion creation orders to the simulator by identifying the amount of money for the opinion. The accuracy of the opinions that are created by the simulator, is proportional to the amount that is sent by the appraiser agent and the expertise of the provider agent on the era of the painting. The opinion provider agent is paid an opinion cost that is determined by the parameter *Opinion Cost* in Table 4.1. However, the amount that is sent to the simulator by the opinion provider agent, may be equal to or less than *Opinion Cost* according to its strategy. If the appraiser agent asks more than one agents' opinion for a painting, then it sends weights, that is actually an indicator of trust to an agent, for each provider agent to the simulator to obtain a weighted appraisal of the opinions of provider agents. Appraiser's final appraisal, which is calculated by the simulator, is the weighted average of these eligible opinions.

The whole protocol is given step by step respectively on the agents' side as the following:

- *prepareReputationRequests():* The agent sends reputation requests to potential reputation providers.
- *prepareReputationAcceptsandDeclines():* The agent accepts or declines reputation requests.
- *prepareReputationReplies():* The agent generates reply messages for the agents that request reputation information based on its strategies.
- *prepareCertaintyRequests():* The agent sends certainty requests to potential opinion providers by specifying the era of the painting.
- *prepareCertaintyReplies():* The agent replies the certainty messages according to its strategies.
- *prepareOpinionRequests():* The appraiser agent sends opinion request to opinion provider agents that are determined by using its strategies.
- *prepareOpinionCreationOrders():* The opinion provider agents order opinions with amounts that change in the range between 0 and *Opinion Cost* from the simulator via sending a message of type OpinionOrderMsg for all opinion requests.
- *prepareOpinionProviderWeights():* The appraiser agents send weights for the agents, whose opinion is requested, to the simulator within a message of type WeightMsg.
- *prepareOpinionReplies():* The agent sends messages of type OpinionReplyMsg by finding the appropriate opinions that are already sent to the simulator.

#### 4.1.4. ART Metrics

ART provides two basic performance metrics: appraisal error and bank balance. Appraisal error is the average appraisal error of an agent during the game and bank balance is the amount in the bank account of the agent at the end of the game. We are expecting successful appraiser agents to have higher bank balances and lower appraisal errors. Agents with different strategies compete with each other during the game; the winner of the game is determined according to either average error or bank balance at the final timestep.

## 4.2. ART with Teamwork

In original version of ART, the agents are expected to provide a single service, i.e., evaluating a single painting, rather than teamwork. However, to investigate teamwork dynamics, the environment needs to be modified so that an agent will be requested to offer a teamwork. To do so, we first modify ART Testbed to provide teamwork environment, and then develop our agents that use team trust model and service graphs.

### 4.2.1. Teamwork in ART

The fundamental task in ART domain is appraising the value of a painting. However, a teamwork consists of several tasks that act in combination and each task is fulfilled by a single agent. In order to achieve our goal, we extend the framework so that a teamwork is represented as a group of paintings, where each painting belongs to an era. *Appraising the value of a painting is corresponding to a task in ART.* Paintings that belongs to different eras are corresponding to different task types and they should be appraised by the agents having necessary skills. So, each teamwork consists of a number of tasks and each task of the teamwork has an effect on the overall effectiveness. The effect of each task is represented with normalized weights, whose sum is equal to 1.

In ART domain, each task type is characterized by a certain era, which the painting belongs to, and each teamwork is characterized by a group of task types. Hence, two teamwork that contain the same tasks should have similar weight distribution among their tasks. Creation process of teamwork begins with the generation of a number of tasks and their corresponding true values, where the era of each painting is randomly selected from a set of eras. Then, the weights of these tasks are generated randomly. Now, this teamwork becomes a unique type by its tasks and is specialized by its task weights. The weights are registered in order to reuse for this type of teamwork, which contains the same task types. After registration, whenever a teamwork that contains the same group of task types is created, the weights of tasks for this composition are taken from the registry and slightly perturbed (between 0 and 0.05). This perturbation is decided randomly for the weight of each task. Note that these weights are only known by the simulator.

In order to evaluate a teamwork, opinions related to all tasks of the teamwork need to be collected. That is, the appraiser agent asks the opinion of exactly one agent for each task of the assigned teamwork, and then opinion provider agents, whose opinion are requested, become a team and offer a teamwork. Each agent in the team appraises the value of the corresponding painting. The simulator calculates the overall appraisal error of the team and this error is sent to the appraiser agent without giving additional information about the appraisal errors of the individuals.

### 4.2.2. Changes in Data Structures

In the original version of ART, a painting object is featured by a particular era. In other words, a painting is distinguished from other paintings by the era attribute that it belongs to. We add a teamwork object that consists of a group of paintings as tasks and featured by the weights of the tasks. Consequently, a teamwork object has the following attributes: a list of tasks and a list of weights of these tasks.

### 4.2.3. Changes in Game Parameters

We change the usage and value of two parameters: the number of allowed certainty requests and the number of allowed opinion requests according to the nature of teamwork. The other parameters such as number of agents, number of timesteps etc will be given in experimental setups.

i. *Number of Allowed Certainty Requests:* The maximum number of certainty requests is set to three for each task, because we don't want to limit the requesters for asking the certainty levels of the other agents. Actually, the default value of the number of certainty requests in the original ART is 20 at each timestep, where the number of certainty requests per task is nearly one. With comparison to the original value, sending three certainty requests is much more flexible for requester agents. The exact number of certainty requests sent depends on the requester agent's will.

ii. *Number of Allowed Opinion Requests:* Only one opinion request is allowed per task. Note that one agent is responsible for carrying out each task in our teamwork model. We assume that none of the agents rejects to provide an opinion.

### 4.2.4. Changes in ART Simulator

The major changes are applied on the simulator, which is central to the game. In the original ART Testbed, the simulator only deals with single task. Now, it's adapted for creating and evaluating teamwork.

<u>4.2.4.1. Teamwork Creation.</u> The simulator performs a number of steps to obtain a teamwork. The whole teamwork creation process in *setMarketShares()* method of *Sim* object, is given in Figure 4.1.

First, the number of tasks are determined randomly between 1 and 4 unless it's given as a game parameter.

In the second step, the simulator selects the task types randomly from a set of types by guaranteeing that they are different from each other. Note that we enforce the tasks to be different from each other in a teamwork. True value of each painting of the corresponding task is generated by using the original true value generation formula of the simulator.

$TW_i.numberOfTasks$ =: RANDOM-BETWEEN$(1, 4)$<br>
$TW_i.taskList := \{\emptyset\}$<br>
**for all** $TW_i.taskList$ **do**<br>
    $task.era :=$ RANDOM-ERA$(eralList)$<br>
    $task.trueValues :=$ GENERATE-TRUE-VALUE()<br>
**end for**<br>
**if** $TW_i$ isn't seen before **then**<br>
    **for all** $TW_i.taskList$ **do**<br>
        $task.weight :=$ RANDOM-WEIGHT()<br>
    **end for**<br>
    normalize weights<br>
    register weights for teamwork type of $TW_i$<br>
**else**<br>
    get weights from registry for teamwork type of $TW_i$<br>
    **for all** $TW_i.taskList$ **do**<br>
        $task.weight :=$ PERTURB-BETWEEN$(0, 0.05)$<br>
    **end for**<br>
**end if**

Figure 4.1. Teamwork creation process

In the third step, the simulator decides whether this teamwork type has been seen before by looking at the registry. If this is the first time such a teamwork is created, weights of tasks are randomly generated and then normalized. If the newly generated teamwork type is seen before, the weights are obtained from the registry, where teamwork types are classified according to their task lists. In order to perturb the task weights, the simulator randomly generates real numbers between $-0.05$ and $0.05$ for each task and adds this real number to the original weight.

*Example:* Let's see how the simulator creates a teamwork step by step.

i. Determine the number of tasks randomly: 3
ii. Determine the types (eras) of the tasks (paintings): Era1, Era3, and Era7 for $T_1$, $T_2$, and $T_3$, respectively.
iii. Generate true values for paintings of $T_1$, $T_2$, and $T_3$: 1000, 2000, and 1500, respectively.
iv. This is the first time a teamwork that contains task types: Era1, Era3, and Era7. Then, determine the weights randomly and normalize them: 0.45, 0.25, and 0.30, respectively. Register these weights for the teamwork type that contains task types: Era1, Era3, and Era7.

After a while let's say another teamwork that is the same type as the previous teamwork, is created as the following:

i. Determine the number of tasks randomly: 3
ii. Determine the types of the tasks: Era1, Era3, and Era7 for $T_1$, $T_2$, and $T_3$, respectively.
iii. Generate true values for paintings of tasks: 2500, 1200, and 2000 respectively.
iv. This type of teamwork is created before by the simulator. So, the simulator accesses the weights from the registry: 0.45, 0.25, and 0.30, and perturbs these weights slightly: 0.40, 0.27, and 0.33 for $T_1$, $T_2$, and $T_3$, respectively.

4.2.4.2. Evaluating Appraisal Error. Originally, the simulator calculates the appraisal error of a task with Equation 4.1 in *distributeAppraisalData()* method of *Sim* object.

$$appraisalError = (|appraisedValue - trueValue|/trueValue) \tag{4.1}$$

In teamwork environment, there are tasks with weights that indicate the influence of the task on the teamwork. In this case, the overall appraisal error of the teamwork is calculated as the weighted sum of individual appraisal errors with Equation 4.2.

$$overallError = \sum_{i=1}^{n} weight(T_i) * (|appraisedValue(T_i) - trueValue(T_i)|/trueValue(T_i)) \tag{4.2}$$

*Example:* Let's continue with the previous example give in Section 4.2.4.1 to show how the appraisal error is calculated. The appraiser agent establishes a team with three members and opinion provider agents appraise the value of their own task as 1100, 1980, and 1560, for $T_1$, $T_2$, and $T_3$, respectively. The overall appraisal error is calculated using Equation 4.2 and has a value of 0.0545.

overallError = (0.45 * (|1100 - 1000|/1000)) + (0.25 * (|2000 - 1980|/2000)) + (0.30 * (|1560 - 1500|/1500)) = 0.0545

### 4.2.5. Changes in Communication Protocol

There is a generic *Message* object, from which the actual messages that are sent according to protocols, inherit. In the original ART, there are only one task and a single provider in the transaction. Thus, agents only send the task object within the necessary messages. In ART with teamwork, appraiser agents send teamwork information that contains tasks, team members, and the task, which the provider agent is asked about. These modifications are applied on *OpinionRequestMsg* and *OpinionOrderMsg* objects.

### 4.2.6. Changes in Agents

In agents side, we change the behavior of agents in two steps of the ART protocol: opinion requests and weights of opinions.

i. *Opinion Requests:* An appraiser agent, who is assigned a teamwork, forms a team based on its strategies in *prepareOpinionRequests()* of *Agent* object. The agent sends one opinion request for each task of teamwork to ensure that only one agent carries out a task.
ii. *Weights:* Previously in ART, the appraiser agents request opinions of more than one agent and send weights that shows the significance of an provider's opinion, for each opinion provider agent. However, only one agent is requested for its opinion for each task in teamwork environment. In this case, the appraiser agents simply send 1.0 as a weight for the opinion provider of each task not to disturb the protocol in *prepareOpinionProviderWeights()* of *Agent* object. Since, there will be only one provider agent for a task. In this way, we eliminate the effect of provider weights in ART.

## 4.3. Agent Strategies

We explain the representation of three models, individual trust model, team trust model, and service graph model, update procedure, non-cooperative behavior of agents, and opinion request strategies in Chapter 3. In this section, we present how the game evolves through certainty and opinion protocols in a single timestep. We exclude the reputation protocol because we don't use reputation information in this study.

- *prepareCertaintyRequests():* We set a high value for the maximum number of certainty messages. Though, the agent sends a certainty request for each task of each assigned teamwork. The agent especially prefers to request the certainty of agents, whose certainty value is unknown about the related task type (era).
- *prepareCertaintyReplies():* The agent replies the certainty messages according to its real expertise value for the related task type.

- *prepareOpinionRequests():* The agent uses individual trust model for single tasks. If the teamwork contains more than one task, the agent uses team trust models and service graph.
- *prepareOpinionCreationOrders():* Noncooperativeness property of the agent emerges in this step. When the provider agent's opinion is asked about a task of the teamwork, which exists in its noncooperativeness list, then the provider agent orders an opinion value of 1 from the simulator via sending a message of type OpinionOrderMsg. If the non-cooperative list doesn't contain this teamwork type, then the provider agent orders an opinion value of 10.
- *prepareOpinionProviderWeights():* Weights don't have any effect in team model setting, since the agent asks one opinion from only one agent for a task of a teamwork. This is the only opinion that effects the appraisal of the task. Formally, the appraiser agent sends 1 as the weight of opinion providers to the simulator via WeightMsg.
- *prepareOpinionReplies():* The agent sends messages of type OpinionReplyMsg by finding the appropriate opinions that are already sent to the simulator.

# 5. SIMULATIONS OF AGENTS WITH TEAMWORK TRUST MODEL

In Chapter 3, we have explained how an agent can decide on the trustworthiness of teams using teamwork trust modeling tools. Through experiments, we evaluate how well such an agent can indeed form a team. Intuitively, the better an agent models teams, the better the chosen team's performance will be for a teamwork. To evaluate this, we compare the performance of *TMA* (Teamwork Modeling Agent) that uses the teamwork trust modeling tools with the performance of *IMA* (Individual Modeling Agent) that uses traditional individual trust model with respect to bank balances and appraisal errors.

On one hand, TMA agents uses team trust model, service graphs, and individual trust model. On the other hand, IMA agents are modeled by excluding the capabilities of team trust model and service graphs of TMA agents. In other words, IMA agents use the same individual trust model, opinion request strategy for single tasks, adaptable parameters, and reward and penalty weights, which are used in update method, with TMA agents. The only difference is that when requesting an opinion, IMA agents use only the individual trust model to select agents one by one for each task of the teamwork.

## 5.1. Experimental Setup

Exploration vs. exploitation is an important issue for decision making mechanisms. In this study, we monitor the thresholds that determine whether the expertise of a team trust model or the weight of an edge set from the service graph is satisfactory or not, based on timesteps. We use increasing threshold values during the game. There are two places where the agent makes a decision by using these thresholds: selecting a successful team trust model instance and selecting an appropriate set of edges from the service graph. For example, an appraiser agent, who is searching for a team, uses a team trust model instance if its expertise value is higher than the value of the corresponding threshold.

The threshold values used in strategy (1), (2), and (3) that are introduces in Section 3.3 are adapted by the agents during the game based on the current timestep. Different threshold parameters are used for these strategies. The expertise weights of team trust models and the weights of the service graph are expected to increase for better teams during the game. Thus, an agent increases the threshold values by a small amount in certain periods.

For team trust models and service graph, the threshold values are adapted with the formula given in Equation 5.1. The value of *increaseRate* parameter in Equation 5.1 is 0.5 for team trust model threshold and 0.3 for the service graph threshold. The value of the threshold of service graphs is increasing more slowly than the value of the threshold of the team trust model. Because the team trust models are used more frequently, hence the expertise of successful teams increasing faster than the weights of the useful edges in the service graph. In this case, selecting better teams require increasing the values of the thresholds. During these steps, the value of the team trust model and the service graph threshold is increased up to 0.90. In the rest of the game, these threshold values are used to decide how successful a team is or how useful an edge set of the service graph is.

$$threshold = 0.50 + increaseRate * ((currentTimestep - 10)/10) \tag{5.1}$$

Another decision point, where the exploration and exploitation issue exhibits, is combining the certainty information with the expertise that is kept in individual trust model of the agent for a particular task. Our agents calculates an actual expertise as a weighted sum of the certainty of an agent and the expertise level of this agent for the corresponding task. The weights of the certainty and expertise values changes during the game and their sum is 1.0. Actually, these weights are adaptable parameters and their value depends on the current timestep. The weight of the expertise is increased by a small amount in certain timesteps, while the weight of certainty decreases by the same amount. This is because the importance of the agent's own experiences increases as the number of experiences increases, the agent with the highest weighted sum of certainty and expertise value is selected to request its opinion.

## 5.2. Effect of Trust Model

In order to understand the necessity of considering the trustworthiness of providers as a team rather than individuals, we compare TMA and IMA agents. In addition to TMA and IMA agents, we use honest agents, which randomly select the opinion provider agents to request opinion from. In other words, an honest agent forms a team by selecting team members randomly one by one, when it's assigned a teamwork. Our experimental setup contains 9 agents: 3 TMA agents, 3 IMA agents, and 3 honest agents. All agents exhibit noncooperativeness behavior and the same noncooperativeness level is used for these agents. All agent types, TMA, IMA, and the honest agents, always response honestly to all certainty, and opinion requests. The game continues 200 timesteps and is repeated 100 times for reporting average results.

Interdependency relationships are disabled in this part of our study. The number of tasks in a teamwork is randomly selected between 1 and 4 by the simulator at the creation time of the teamwork as explained in Section 4.2.4.1.

We both compare the bank balances and the average appraisal errors of TMA and IMA agents. The noncooperativeness level is fixed to 0.3 in this part of the experiments. The summary of the simulation parameters are given in Table 5.1.

- *Bank Balance:* Figure 5.1 depicts the average bank balances of TMA, IMA, and honest agents with the noncooperativeness level, which is a value of 0.3. The simulation parameters are given in Table 5.1. Accordingly, the average bank balances of TMA, IMA, and honest agents for each timestep are reported by averaging the values obtained from 100 runs. We are expecting successful agents

Table 5.1. Simulation parameters

| Parameter | Value |
|---|---|
| Timesteps | 200 |
| Number of Agents | 9 |
| Number of Repetitions | 100 |
| Noncooperativeness Level | 0.3 |
| Number of Tasks | 1,2,3,4 |
| Interdependency | N/A |

to model the other agents more properly and choose appropriate agents to ask opinion. In this case, these successful agents come up with lower appraisal errors, their client shares increase, and earn more money. Consequently, we are expecting successful agents to have higher bank balances and lower appraisal errors in the results. Note that TMA agents considerably outperform IMA agents by using teamwork trust modeling tools with respect to the bank balance that is considered as the indicator of success in ART competitions. The bank balances of IMA and honest agents are differing from each other in the first part of the game. However, IMA agents always have higher bank balances than the bank balances of honest agents in the second part of the game.

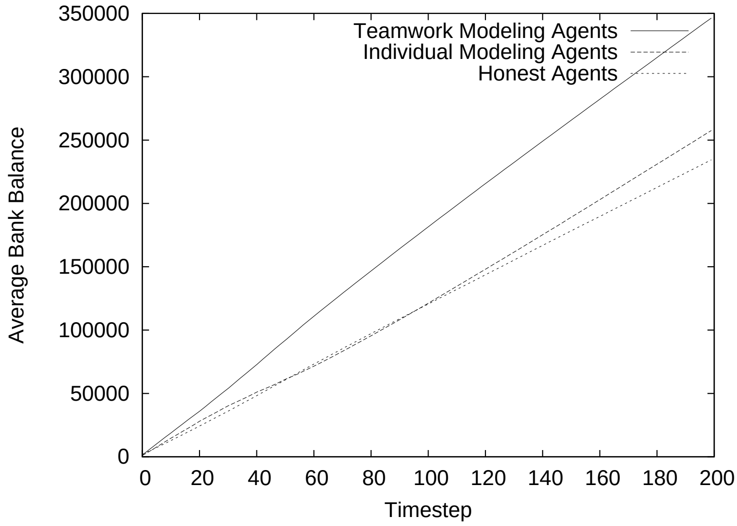


Figure 5.1. Bank balances of TMA, IMA, and honest agents

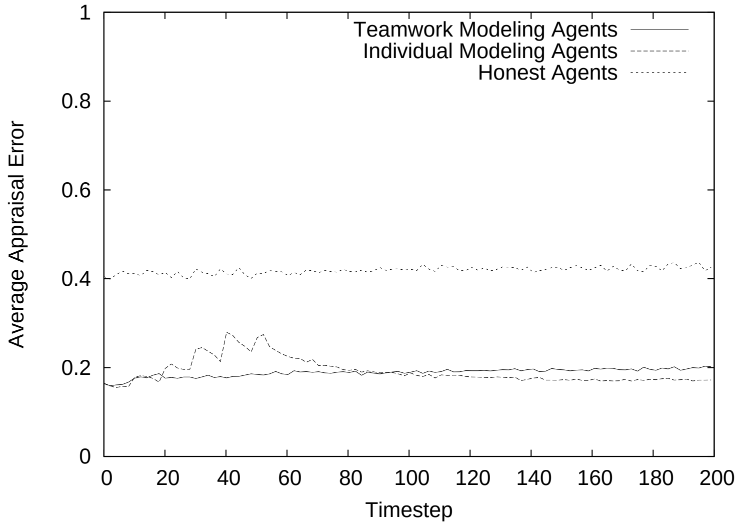


Figure 5.2. Average appraisal errors of TMA, IMA, and honest agents

- *Appraisal Error:* In Figure 5.2, the average appraisal errors of TMA and IMA agents are shown. TMA agents have nearly stable appraisal errors during the game. However, IMA agents start with high appraisal errors, then continue decreasing errors for a while, and finally the appraisal error rate nearly remains stable. In the second half of the game, the average appraisal error of TMA agents is slightly higher than the average appraisal error of IMA agents. On the other hand, honest agents always have higher appraisal errors with comparison to TMA and IMA agents.

## 5.3. Effect of Noncooperativeness Level

We also compare TMA, IMA, and honest agents based on their bank balances and the behavior of their final bank balances with respect to changing noncooperativeness level. In addition to final bank balances, we also look at the effect of the noncooperativeness level on the appraisal error. For this purpose, the average appraisal errors of TMA agents will be provided with varying noncooperativeness levels.

The noncooperativeness level is increased by 0.1 between 0 and 1. 100 simulations are performed for each noncooperativeness level of the set 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0. The simulation parameters given in Table 5.1 are used with different noncooperativeness levels.

### 5.3.1. Noncooperativeness Level vs. Final Bank Balances

In Figure 5.3, the final bank balances of TMA and IMA agents with respect to different noncooperativeness levels are shown. For low noncooperativeness levels, the difference between the final bank balances of TMA and IMA agents has higher values. This difference decreases as the noncooperativeness level increases, in other words the cooperativeness level of the agents decreases. Decreasing cooperativeness means that agents, from which opinions are requested, send opinions properly for very limited set of teamwork and they choose not to cooperate for most of the teamwork in the environment. Hence, TMA agents start to misclassify teams as noncooperativeness level increases. In contrast to TMA agents, IMA agents cannot handle even the smaller noncooperativeness levels. Because they just use individual trust models and they assume that the agents behave the same way for any teamwork in any environment.

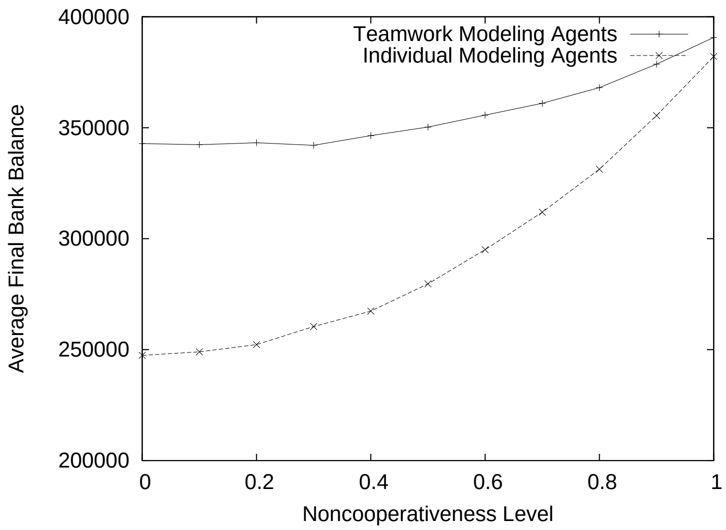


Figure 5.3. Noncooperativeness level vs. final bank balances

Another important result is that the higher noncooperativeness levels produce higher bank balances. Remember that the noncooperative behavior is paying the smallest amount to the simulator for opinion creation order in ART domain. Hence, as the noncooperativeness level increases, opinion costs (the amount paid by the opinion provider to generate the opinion) decrease, and provider agents save their money rather than sending high amounts of money that is close to 10. For example, the provider agents always pay 10 to the simulator for opinion creation order, when the noncooperativeness level is 0.0. Remember that the opinion fee is 10. So, they spend all money that they are paid for providing opinion, and earn no money. On the other hand, the provider agents always pay 0 for opinion creation order, when the noncooperativeness level is set to 1.0. In this case, they are paid 10 as the opinion fee and they spend an amount of 0 for opinion creation order. So, they earn an amount of 10. In this way, increasing noncooperativeness levels led to higher bank balances.

### 5.3.2. Noncooperativeness Level vs. Appraisal Errors

Average appraisal errors of TMA agents for different noncooperativeness levels are shown in Figure 5.4. Remember that being noncooperative means that if the current teamwork that the agent is invited to participate is in the noncooperativeness list of this agent, it sends the worst opinion creation order that is an amount of 1 to the simulator. Then, the simulator creates an opinion for this order accordingly. For the noncooperativeness level 0.1, the average appraisal error has values below 0.2. On the other hand, the average appraisal error has values above 0.3 for the noncooperativeness level 0.9. This is not a surprising result. Because as the noncooperativeness level increases, provider agents become more and more noncooperative, and this behavior leads to higher appraisal errors. In other words, as the provider agents in the environment send worse opinion creation orders due to the higher noncooperativeness levels, the simulator creates worse opinions. Consequently, the results show that the higher noncooperativeness levels cause higher appraisal errors as expected.

## 5.4. Effect of Population

In this section, we look at the influence of the larger populations on our performance metrics: bank balance and appraisal error. To do so, the number of agents for each agent type is increased to 10; there exist 10 TMA, 10 IMA, and 10 honest agents

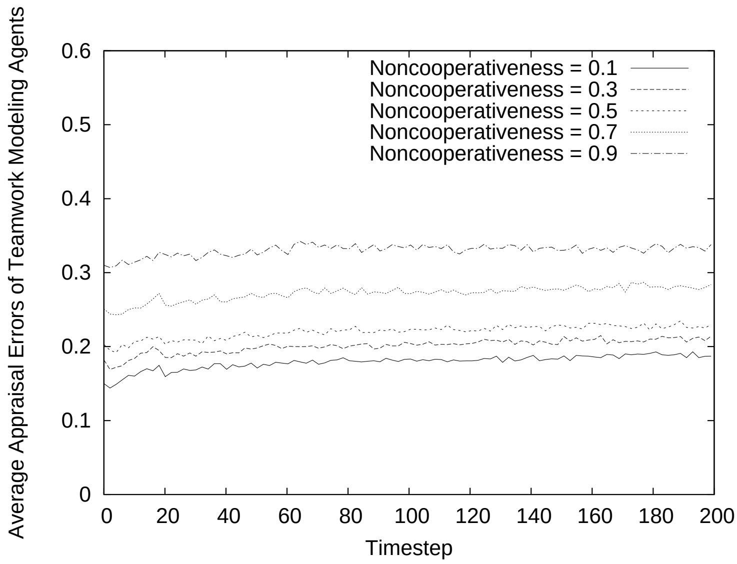


Figure 5.4. Noncooperativeness level vs. appraisal error of TMA

in the environment. The simulation parameters given in Table 5.1 are used, the only change is the number of agents.

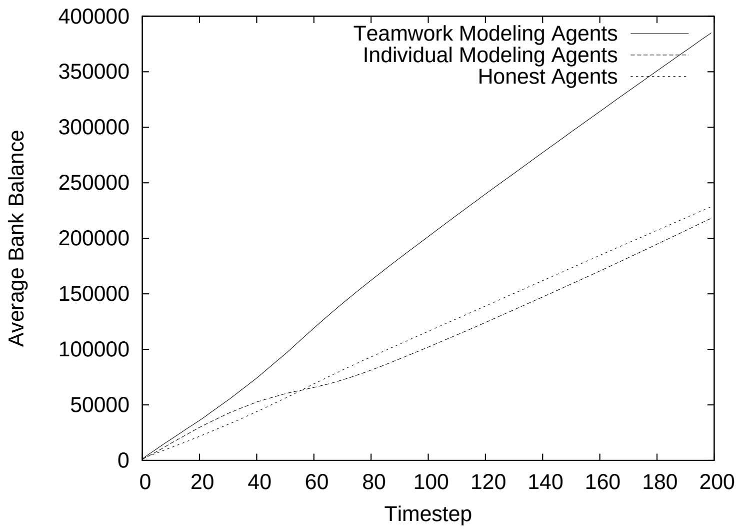


Figure 5.5. Bank balances of TMA, IMA, and honest agents (high population)

- *Bank Balance:* Figure 5.5 shows the average bank balances of 10 TMA, 10 IMA, and 10 honest agents. The difference between bank balances of TMA, IMA, and honest agents increases with respect to the difference obtained with smaller population given in Figure 5.1. However, the average bank balance of IMA agents is lower than the average bank balance of honest agents after $60^{th}$ timestep of the

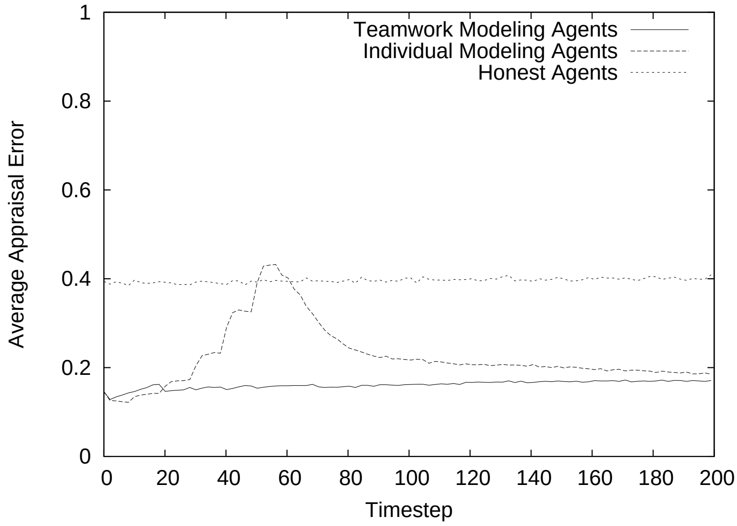


Figure 5.6. Average appraisal errors of TMA, IMA, and honest agents (high population)

game. These results show that considering trustworthiness of individuals is worse than choosing agents randomly.

- *Appraisal Error:* The influence of higher population can be seen more clearly in Figure 5.6 that depicts the average appraisal errors of TMA, IMA, and Honest agents. TMA agents always have lower appraisal errors except for the first 20 timesteps. On the other hand, the distribution of errors of TMA and IMA agents during the game are similar to the distribution for small population given in Figure 5.2. On the other hand, the honest agents again have higher appraisal errors.

These results show that our teamwork trust modeling tools are utilized better in the environments with higher populations. This is a great opportunity because we develop these models for open distributed environments.

# 6. MODELING AGENTS WITH PERSONALITY

In the first part of the study, we enable teamwork environment in ART domain and build agents that are able to model trust to teams successfully. Nevertheless, this teamwork model is lacking necessary features that exhibit in real life with respect to two factors: the interdependency relationship between tasks of a teamwork and the personality of the agents.

- *Interdependency:* Assuming that all tasks in a teamwork are totally independent from each other do not express the real life situations successfully. Because the tasks of a teamwork have varying degrees of interdependence to other tasks. In order to capture this factor, we enable the interdependency relationship for the teamwork model in this chapter.
- *Personality:* The noncooperativeness behavior is introduced to mimic the difference between single task and teamwork environments. However, this behavior is not sufficient to cover many real life situations occurred in teamwork environments. In the purpose of expressing the factors that affect the behavior of the agents in teamwork more comprehensively, we develop agents with personality and discard the noncooperativeness behavior, which is only able to represent the *teamwork effect*. We use a well-known personality model in psychology literature: Big Five Personality Traits.

## 6.1. Interdependency in Teamwork

A group of service providers offers a composed service by working cooperatively according to common goals. Interdependency, which represents the relationship, where a service is influenced to some degree by one or more services in a composed service, is an important concept for composed services. Interdependency relationships are determined at the job design phase; a service may be dependent on one or more services in a composed service. It is represented with a solid line and an arrowhead, and shows in which order the services will be performed. In this structure, a service is dependent

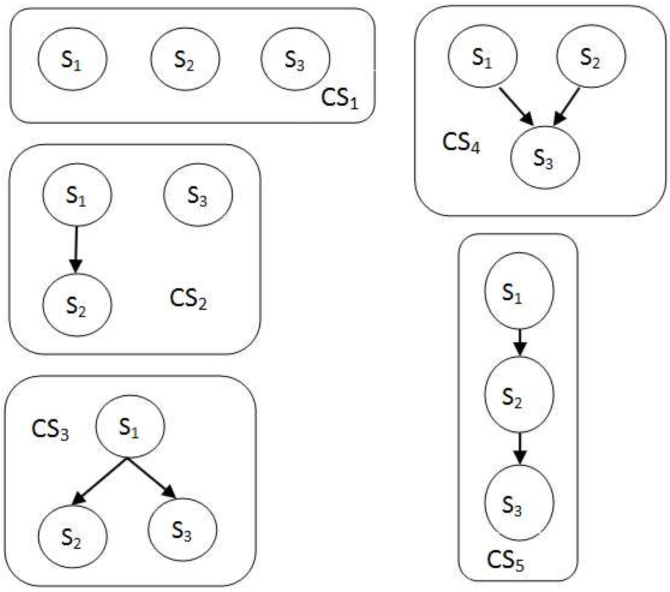


Figure 6.1. Service interdependency relationships

on previously performed service(s) which is connected to this service with a line.

In order to represent all core interdependency relationships, we need at least three services. Figure 6.1 depicts five different interdependency relationships between three services. These five composed services contains basic structures that are used in business process execution languages such as WS-BPEL[12].

In the first relation ($CS_1$), three independent services are performed concurrently, and this composition corresponds to the flow structure of WS-BPEL. The second relation ($CS_2$) comprises an interdependency relationship between the first and second service, that corresponds to the sequence structure of WS-BPEL. The third service is independent from the others. In the third relation ($CS_3$), both the second and third services are dependent on the first service, where this relation corresponds to the merge operation. In the fourth relation ($CS_4$), the third service depends on both the first and second service, which corresponds to the fork operation. Any other interdependency relationship between more than three services can be established by combining these basic structures. For example, the fifth composition ($CS_5$) is developed by using sequence structure, where three services are performed sequentially. The second service is dependent on the first service, and the third service is dependent on the second service.

Eventually, a composed service is specialized by its services. A teamwork exhibits one of the five interdependency relationships shown in Figure 6.1.

## 6.2. Technical Background: Big Five Personality Traits

The recent research indicates the significance of personality in addition to job specific skills of individuals for teamwork effectiveness [7, 8, 9, 10]. Personality-job performance relationship is important to predict the teamwork performance. Many researchers use a five-factor personality model called the Big Five [7] to represent the relationship between personality and job performance. Researchers agree on the robustness of Big Five as a meaningful model for classifying personality traits. The Big Five personality model consists of the following traits:

- *Agreeableness* (Likability) is tendency to be courteous, good-natured, cooperative, straightforward, altruist, complaint, modest, tender-minded, and trusting. Agreeableness measures the quality of interpersonal interaction. An agreeable person is interested in people, believes people are basically honest, decent, and trustworthy, expresses sympathy and empathy, demonstrates social humility and a lack of arrogance, and has a willingness to cooperate in conflict situations.
- *Conscientiousness* (Will to achieve) is tendency to be responsible, hardworking, competing, order, dutiful, achievement oriented, self-disciplined, and deliberative, which are important characteristics for accomplishing work tasks. A conscientious person is always prepared, strictly, obeys work ethics and moral obligations, works hard to achieve goals, begins and carries the tasks out to completion.
- *Emotional Stability* is tendency to be relaxed, secure, and calm. If a person is described as emotionally unstable that is called neuroticism, then she is easily disturbed, gets stressed out easily, worries a lot, and changes her mood a lot.
- *Extraversion* is tendency to be sociable, assertive, and active. This trait becomes

important for the occupation types such as managers, sales, which require communicating with people. When teams require frequent social interactions, sociability level of team members becomes deterministic for the teamwork performance.

- *Openness* (Intellect and Intelligence) is tendency to be imaginative, cultured, and original. It is the most difficult trait to identify and model. This dimension is not covered in this study.

## 6.3. Interdependency in ART

In ART with teamwork, each teamwork has an interdependency relationship. In teamwork creation process, if this is the first time that this teamwork type is created, one of the relations given in Figure 6.1 is randomly selected as an interdependency relationship between the tasks of a teamwork and the interdependency relationship is registered with the weights of teamwork to reuse for this type of teamwork, which contains the same task types. If the newly created teamwork type has been seen before, the interdependency relationship information is taken from the registry and is used for the newly created teamwork. The interdependency relationship knowledge is shared with the appraiser agent.

When the interdependency relationship is enabled in the simulations, the number of tasks in a teamwork should be fixed and has a value of three. Because, we only implement the elementary interdependency relations that are given in Figure 6.1. The interdependency relation is stored in *teamwork object* as a number between 1 and 5 that corresponds to the identification number for relationships in Figure 6.1. The treatment of the simulator according to these relations is embedded in the code.

### 6.3.1. Opinion Cost

An *Opinion Cost* has a value of 10 in the original ART Testbed. When we consider the interdependency relations, some of the tasks have more significant roles on the performance of the teamwork. For example, the first task in the fifth teamwork in Figure 6.1 affects both the second and third tasks. More attention should be paid

Table 6.1. Significance of tasks based on interdependency relationship

| | $T_1$ | $T_2$ | $T_3$ |
|---|---|---|---|
| Interdependency-1 | 0 | 0 | 0 |
| Interdependency-2 | 1 | 0 | 0 |
| Interdependency-3 | 2 | 0 | 0 |
| Interdependency-4 | 0.5 | 0.5 | 0 |
| Interdependency-5 | 2 | 1 | 0 |

for such critical tasks. Intuitively, the fee of a more important task should be higher than the fee of easier or relatively less important tasks such as the third task of the fifth teamwork in Figure 6.1. The indicator of the importance of a task is that how many tasks depend on this task. For example, the significance of the first, second, and third tasks are 2, 1, and 0 respectively for the fifth teamwork. Significance levels of tasks of teamwork with different interdependency relationships are given in Table 6.1.

Fees should increase proportional to the significance of the assigned task. A coherent fee for $T_i$ is calculated with the following equation:

$$Fee(T_i) = 10 + (10 * Significance(T_i)) \tag{6.1}$$

The appraiser agents may pay an opinion fee between 10, which is the minimum amount, and the coherent value that is calculated by Equation 6.1 according to its will. The exact amount of fee is determined based on the appraiser agent's strategies in *prepareOpinionRequests()* of *Agent* object. The most coherent fee should be paid for better appraisals, since higher fees result with better appraisals.

In this setting, appraiser agents send opinion fee and the teamwork information that contains tasks, team members, interdependency, and the task, which the provider agent is asked about, within the *OpinionRequestMsg* message.

### 6.3.2. Opinion Creation

Originally, the simulator creates an opinion, whose accuracy depends on the amount that is sent by the opinion provider and provider agent's expertise level on the particular era of the painting.

In ART with teamwork, the amount of money sent to the simulator for the opinion creation order differs between 0 and the opinion fee that is paid by the appraiser agent and may have a value between 0 and the coherent fee that is calculated according to the task's significance by using Equation 6.1. In order to create an opinion, the simulator first normalizes the payment of the order by using the coherent fee of the task with Equation 6.2. The normalized payment of the opinion creation order differs between 0 and 10 as in the original setting. This normalization is performed in order to reuse the existing opinion creation methods that is located in *fulfillOpinionCreationOrders()* method of *Sim* object. However, bank accounts of the appraiser and provider agents are updated according to the opinion fee that is paid to the provider and the payment that is sent to the simulator, respectively.

$$normalizedPayment = 10 * (payment/coherentFee) \tag{6.2}$$

## 6.4. Agents with Personality

We apply four of the Big Five Personality traits: agreeableness, conscientiousness, emotional stability, and extraversion, to agents in ART domain. Agents have different personalities from each other. Each personality trait is represented as a real number between 0 and 1 that correspond to the level of the corresponding trait. Rather than evaluating a trait with discrete levels such as low, medium, high, we prefer to use fuzzy personality representation. This way, we can express the real life situations more accurately. Many studies use fuzzy representation for big five personality traits [19].

The levels of personality traits of an agent are determined independently from each other. Agents are aware of their own personalities, but not the personalities of the other agents in the environment. Agents exhibit appropriate behaviors according to their personalities. For example, an agent, whose level of conscientiousness is high (close to 1), will do its best in teamwork, and another agent, whose level of agreeableness is low (close to 0), will have problems while working with other agents.

In this chapter, we frequently use performance concept for a team or an individual, it is corresponding to $(1 - appraisalError)$ in ART domain.

### 6.4.1. Agreeableness in ART

Team members, who are responsible for performing the interdependent tasks, will be in a close relation with each other. That is, two or more team members have to work cooperatively. Even selecting the most skilled people for the team doesn't work unless they work well together. We should answer the following question: how well they work cooperatively in such situations, where the answer corresponds to the level of agreeableness. Hence, the agreeableness levels of these team members, who are interacting with each other, have an influence on the team performance. If the interacting agents have high levels of agreeableness, then they work well together, and no problems would arise due to the interdependency relationship. Otherwise, lower agreeableness levels of interacting agents have a negative influence on the performance of these agents.

Consider the furniture workshop example, where the designer and the carpenter have to work together. The carpenter should learn every detail of the design from the designer to act properly. Sometimes, the carpenter may object some parts of the design if necessary. For example, a conflict may arise due to the physical impossibilities of the design. If the designer's willingness of cooperation in conflict situations is low, then a problem grows up between the carpenter and designer. In another case, the designer may not trust the carpenter, and believes that the carpenter just objects because of his self-interest rather than serious problems. These situations have a negative influence on the performance that is caused by the low agreeableness levels of interacting team members.

In this setting, the performance of interacting agents are penalized or decreased proportional to their average agreeableness levels. The role of the agreeableness is applied by multiplying the performance of the agent by the arithmetic average of the agreeableness levels of this agent and other agents that interact with this agent due to the interdependency relationship. Equation 6.3 shows how the simulator applies the influence of the agreeableness on the task performance of $A_i$, who interacts with $A_{i-1}$.

$$P(A_i) := P(A_i) * (A(A_i) + A(A_{i-1}))/2 \tag{6.3}$$

### 6.4.2. Conscientiousness in ART

Being conscientiousness is tendency to being achievement oriented and persevering to accomplish tasks. Conscientiousness is usually considered related to individual task performance. Hence, conscientiousness level of an agent shows how an agent makes effort to accomplish the assigned task. Consider the furniture workshop example, each member is responsible from its own task. If the carpenter has tendency to be undutiful, and careless, then he cannot exhibit a great performance, because he doesn't utilize his skills due to his low conscientiousness level.

Remember that the fundamental task in ART domain is appraising the value of a painting, which belongs to a certain era, by sending an opinion creation order to the simulator. If the conscientiousness level of the agent is high, then it would do its best. Otherwise, the agent cannot take advantage of its expertise even it has the highest expertise on the corresponding era. The best an opinion provider agent can do is sending the highest opinion creation order to the simulator. Considering the conscientiousness trait, the amount sent for the opinion creation order an agent is proportional to the provider agent's conscientiousness level. The amount sent to the simulator is calculated with Equation 6.4. For example, if the conscientiousness level of the agent, whose opinion is asked, has a value of 0.80, then the amount for opinion creation order becomes $8(= 0.80 * 10)$.

$$orderPayment := C(A_i) * fee \tag{6.4}$$

### 6.4.3. Emotional Stability in ART

Being emotionally stable is tendency to be secure and relaxed. In ART, we apply emotional stability in the task of appraising the value of a painting, which is sending an amount for opinion creation order to the simulator. Normally, the agent sends an amount proportional to its conscientiousness level for opinion creation order. Considering emotional stability, the agent exhibits normal behavior proportional to its emotional stability level. If the emotional stability level of an agent is low, then this agent frequently exhibits unpredictable behaviors in opinion creation process. In ART, unpredictable behavior corresponds to sending abnormal amounts for opinion creation order, numerically the laziest amount is 1 for an opinion. Before the agent sends an opinion creation order to the simulator, a real number is randomly generated between 0 and 1. If this number is smaller than the agent's emotional stability level, the agent responds normally. Otherwise, it sends the worst opinion creation order, which is an amount of 1. Emotionally stable agents rarely send such unexpected responses. In other words, these agents sends reasonable amounts according to their will. For example, honest agents mostly send high amounts and cheating agents send low amounts. The opinion creation order method, where the emotional stability has an effect, is shown in Equation 6.5.

$$order = \begin{cases} 10 * C(A_i), & rand < E(A_i) \\ 1, & rand \geq E(A_i) \end{cases} \tag{6.5}$$

### 6.4.4. Extraversion in ART

Extraversion is the personality trait related to being social and talkative. Considering the protocols in ART domain, the fundamental part of the protocol, where extraversion becomes significant, is the part, when the appraiser agent asks the certainty values of the agents in order to determine opinion provider agents. In ART domain, if the extraversion level of the appraiser agent is high, which is between 0.7 and 1, the appraiser agent sends three certainty requests. Else if the extraversion level is between 0.3 and 0.7, the agent asks two agents about their certainty levels. Otherwise, only one agent is requested for its certainty level.

So far, we explain how we apply Big Five personality traits in ART Testbed. For the purpose of higher teamwork performance, an agent models trust to other agents in the environment by using individual trust model that is explained in Section 3.1.1. Remember that an agent classifies its experiences with respect to the requested task and the agent carries it out. The models are updated based on the overall teamwork performance according to work roles of the team members. However, the agents don't consider the personality of the agents for modeling their trustworthiness.

## 6.5. Agent Strategies

- *prepareCertaintyRequests():* The extraversion level of the appraiser agent has an effect at this step. If the appraiser agent is highly social and talkative, it sends three certainty requests for each task of the teamwork. The agent prefers to ask especially the certainty of agents, whose certainty values are not known yet.
- *prepareCertaintyReplies():* The agent replies the certainty messages according to its real expertise value of the related task.
- *prepareOpinionRequests():* The appraiser agent uses individual trust model to find agents to request their opinions. For each task, only one agent is selected to carry it out. Opinion fees are not fixed in this part, because some tasks have great importance for teamwork. For example, the provider agent for the first task of the fifth composition in Figure 6.1 should be carefully selected and paid more

than the third provider. Because great effort is required to accomplish this task and it can be obtained by great motivation, which is opinion fee in this case. The opinion fee of a task increases proportional to the number of tasks that depend on this task. The actual fee of a task is fixed fee, 10, plus the additional amount, which is 10, multiplied by the number of tasks depend on this task. The agent is free about the amount that is paid as an opinion fee. However, better appraisals can only be provided by paying higher fees.

- *prepareOpinionCreationOrders():* The conscientiousness and emotional stability levels of the agent have an effect in this step. When the agent's opinion is asked about a painting, then the agent randomly generates a real number between 0 and 1. If this number is higher than its emotional stability level, the agent orders an opinion value of 1 from the simulator by sending a message of type *OpinionOrderMsg*. Otherwise, the amount paid is determined based on the conscientiousness level of the agent.
- *prepareOpinionProviderWeights():* Weights don't have any effect in this setting, since the agent asks one agent's opinion for each task. This is the only opinion that affects the appraisal accuracy. Formally, the agent sends 1 as a weight to the simulator within a message of type *WeightMsg*.
- *prepareOpinionReplies():* The opinion provider agent sends messages of type OpinionReplyMsg by finding the appropriate opinions that are already sent to the simulator.

## 6.6. Performance Evaluation

The performance of single task can be easily determined, because there is no influence in the environment. However, evaluating the teamwork performance needs considering several factors that arise from the environment or the team. In this section, we introduce two factors affecting the teamwork performance: interdependency and personality, and explain how we combine the individual task performances to determine the overall teamwork performance.

### 6.6.1. Interdependency Factor

The definition and the representation of interdependency relationship are explained in Section 6.1. When a task is dependent on one or more tasks, if the previous task is performed with a poor performance, than probably the dependent task cannot result with a great performance because it's affected by the previous task. Consider the furniture workshop example: if the furniture is not successfully designed at the beginning, the furniture would not be satisfactory when it's finished. Because the carpenter's task depends on the design and the assembler's task depends on the carpenter's task. Both these task are affected negatively by a worse design of furniture.

If the result of a task would affect the success of another task, intuitively this task has a great importance on teamwork performance. Hence, the provider agent for this task should be carefully selected. The fifth teamwork in Figure 6.1 is a dramatic example of interdependency, because the first task has an effect on the second task, and the second task has an effect on the third task. In this case, the success or failure of the first provider, which affects the remaining two tasks, has a significant role on teamwork performance.

In this study, the performance of single tasks and the teamwork performance are evaluated between 0 and 1. The best result is scored as 1, and the worst result is scored as 0. The score of the a task is computed by multiplying its individual performance by the performances of the services, which the current task is dependent on. If the task is affected by more than one task, the performance multiplier would be the arithmetic average of the performances of these affecting tasks.

*Example:* Let's say that an agent is assigned a teamwork that contains $T_1$, $T_2$, and $T_3$, and exhibits the fifth interdependency relationship in Figure 6.1, and $A_1$, $A_2$, $A_3$ carry out these tasks, respectively. Assume that, the individuals scores, which are calculated by considering only the agent's skill, emotional stability level, and conscientiousness level, are evaluated as 0.8, 0.7, and 0.5 for $T_1$, $T_2$, and $T_3$, respectively. These individual evaluations do not reflect the actual performances because $T_2$ and $T_3$ are

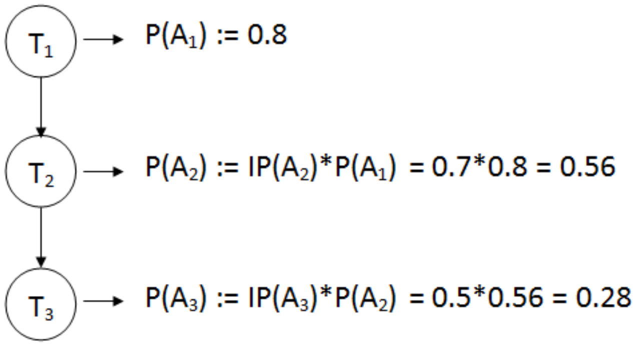


Figure 6.2. Interdependency factor in performance evaluation

dependent on the corresponding previous tasks. The success of $A_1$ has an effect on the success of $A_2$. Likewise, the success of $A_3$ is affected explicitly by the success of $A_2$, and implicitly by the success of $A_1$. The simulator applies the effect of interdependency to the individual performances (IP) as following: performance of $A_2$ is obtained by multiplying it by the performance of $A_1$: IP$(A_2)$ $= 0.7 * 0.8 = 0.56$; performance of $A_3$ is obtained by multiplying it by the performance of $A_2$: IP$(A_3)$=$0.5 * 0.56 = 0.28$; performance of $A_1$ remains the same because it isn't affected by any task. The process is shown in Figure 6.2.

### 6.6.2. Personality Factor

Agreeableness factor is applied on the performance evaluation of interdependent tasks as explained in Section 6.4.1. In ART domain, this corresponds to the part that the simulator calculates the actual performance of the agents, who send opinions creation orders to the simulator. The performance of the agent, whose task is dependent to other task(s), is multiplied by the average agreeableness level of the agent and the provider(s) of the affecting task(s).

*Example:* Continuing the previous example, we apply the effect of agreeableness factor on individual performances. Assume that the agreeableness levels of provider agents $A_1$, $A_2$, and $A_3$ are 0.6, 0.9, and 0.7, respectively. The performance of $A_2$ is affected by agreeableness levels of the first and the second providers. The arithmetic

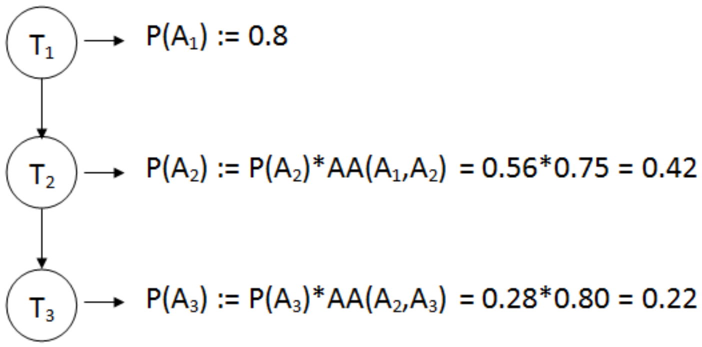


Figure 6.3. Personality factor in performance evaluation

average of agreeableness levels of these two providers is 0.75. The performance of $A_2$ is multiplied by this average agreeableness level to obtain the actual performance; IP($A_2$)=$0.56 * 0.75 = 0.42$. Similarly, the actual performance of $A_3$ is calculated by multiplying the performance of $A_3$ by the average agreeableness levels of the second and the third providers: IP($S_3$)=$0.28 * 0.80 = 0.22$. The process is shown in Figure 6.3

### 6.6.3. Level of Significance

So far, actual performances of individuals are calculated based on interdependency relationship and agreeableness levels of the interacting team members. There are several approaches for the computation of the overall performance of individual performances in the literature. For example, selecting the minimum/maximum performance, or averaging individual performances. In this study, each task of teamwork has a weight that is between 0 and 1 as explained in Section 4.2.1. Remember that the overall teamwork performance becomes a weighted sum of these individual performances.

*Example:* By continuing the previous example, the actual individual performances are obtained: 0.80, 0.42, and 0.22 for $A_1$, $A_2$, and $A_3$, respectively. Let's say that the weights of these tasks are 0.40, 0.35, and 0.25, respectively. The teamwork performance (TP) that is the weighted sum of individual task performances becomes 0.52.

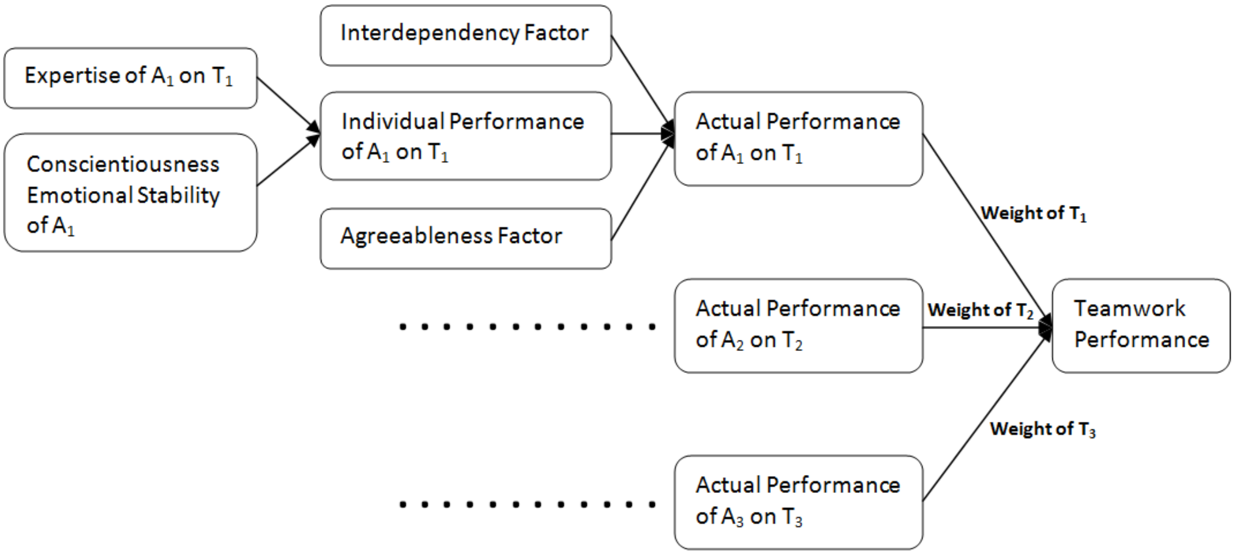


Figure 6.4. Performance evaluation of teamwork

$TP = (0.40 * 0.80) + (0.35 * 0.42) + (0.25 * 0.22) = 0.52.$

Figure 6.4 depicts the whole performance evaluation process of a teamwork that contains $T_1$, $T_2$, and $T_3$. These tasks are carried out by $A_1$, $A_2$, and $A_3$, respectively. At the beginning, $A_1$ sends an amount, which is determined based on the conscientiousness and emotional stability levels of $A_1$, to the simulator for opinion creation. Simulator creates an opinion according to the amount that is sent by $A_1$ and expertise of $A_1$ on particular era of $T_1$. The performance of this opinion is individual performance of $A_1$. Simulator applies the effect of interdependency and agreeableness factors on the individual performance of $A_1$ to obtain the actual performance. After $A_2$ and $A_3$ perform their tasks properly, their performances are computed in the same way. Finally, the weighted sum of these actual performances becomes the overall teamwork performance.

# 7. SIMULATIONS OF AGENTS WITH PERSONALITY

In Chapter 6, we explained how we model agents with personality by using Big Five Personality model in teamwork environment. Our research question was that which personality traits are significant for the success of the team. To do so, we look at the relationship between the team personality and performance. If such a relationship is found, this knowledge can be predictive for teamwork performance in team formation process. The personality of a team is defined as the personality model that corresponds to the arithmetic average of the corresponding traits of team members.

In addition to significance of traits, we look at the influence of the trust within the team on the performance of the teamwork in an open distributed environment, specifically in modified ART domain. During these simulations, we compare different interdependency relationships to capture different teamwork types in real life. Note that the teamwork in real life have varying degrees of interdependency between their tasks. Some of them have strongly interdependent tasks while some of them have loosely coupled tasks. In this way, we show that which personality traits are significant for which teamwork types.

## 7.1. Experimental Setup

Homogeneous agents with varying personalities are used in the simulations, and the number of agents in the simulation is 50 in order to increase the number of alternative groups. Each simulation continues 300 timesteps. We repeat the simulations 50 times and report the average data obtained from 50 simulations in the results. Table 7.1 summarizes the simulation parameters.

## 7.2. Significance of Traits

The simulator determines the personality of agents at the beginning of the simulation. The levels of the personality traits of an agent are generated independent from

Table 7.1. Simulation parameters

| Parameter | Value |
|---|---|
| Timesteps | 300 |
| Number of Agents | 50 |
| Number of Repetitions | 50 |
| Noncooperativeness Level | N/A |
| Number of Tasks | 3 |
| Interdependency | 1,2,3,4,5 |

each other. Having a high conscientiousness level doesn't suppose having a high agreeableness value. The levels of personality traits are distributed as low, medium, and high equally in the environment. Considering agreeableness trait, the simulator randomly selects one third of the agents, and randomly generates a real number between 0.90 and 1 as an agreeableness level for each selected agent. That is, these agents form the group of agents having higher agreeableness levels. Then, the simulator randomly selects one third of the agents, who are not assigned an agreeableness level yet, and randomly generates a real number between 0.60 and 0.70 as the agreeableness level for each of them. They become the group of agents with medium agreeableness levels. Finally, the simulator randomly generates lower agreeableness levels between 0.30 and 0.40 for each one of the rest of agents. We keep agreeableness levels high not to dramatically decrease the teamwork performances during the reflection of the effect of agreeableness of interacting teammates on their performance. For conscientiousness, emotional stability, and extraversion, high levels are generated between 0.90 and 1.0, medium levels are between 0.45 and 0.55, and finally low levels are between 0 and 0.10.

For the purpose of reporting, we keep a comprehensive teamwork information; agents, teamwork, interdependency type, agent-task assignments, average levels of agreeableness, conscientiousness, and emotional stability of the team members, extraversion level of the owner agent of this teamwork, average appraisal error, and number of occurrences that these agents come together to provide this teamwork with the same work role assignments.

In order to investigate the significance of each personality trait on teamwork performance, we examine teams having best performance (minimum overall appraisal error) and teams having worst performance (maximum overall appraisal error). If we can obtain the most significant traits for teamwork performance, this information can be used to select team members in future studies.

We keep top fifty teams for successful and unsuccessful teams. Note that we start to keep these top fifty teamwork after the game becomes mature (after $200^{th}$ timestep). After the simulation, average levels of personality traits of teamwork instances are calculated to obtain the average personality of the team, or team personality. Team extraversion level is equal to the extraversion level of the appraiser agent, who is assigned this teamwork by the simulator. During the simulation, there exists approximately $100,000$ teamwork instances for 50 agents and 10 task types (eras) in 300 timesteps.

### 7.2.1. Influence of Personality Traits

The average performance, trust within a team, and the average personality information of successful, unsuccessful, and most trusting teams are given in Table 7.2. The simulation parameters given in Table 7.1 are used.

Table 7.2. Average personality, performance, and trust of IMA teams

| | Successful | Unsuccessful | Most Trusting |
|---|---|---|---|
| Agreeableness | 0.67 | 0.67 | 0.69 |
| Conscientiousness | 0.60 | 0.36 | 0.57 |
| Emotional Stability | 0.57 | 0.56 | 0.55 |
| Extraversion | 0.50 | 0.50 | 0.50 |
| Std. of Consc. | 0.24 | 0.31 | 0.20 |
| Performance | 0.99 | 0.0 | 0.46 |
| Trust | 0.59 | 0.51 | 0.80 |

At first glance, the gap between the conscientiousness level of successful and unsuccessful teams becomes evident. Average conscientiousness level of successful teams (0.59) is remarkably higher than the conscientiousness level of unsuccessful teams (0.36). Conscientiousness has an influence on the individual performance in ART domain. That is, the agents having high conscientiousness levels accomplish their task successfully and do their best. Individual task accomplishment is the core of teamwork. So, high individual performances and implicitly high teamwork performances are obtained with high conscientiousness levels. Therefore, conscientiousness level of a team is strongly and positively related to teamwork performance. The average level of other personality traits, agreeableness, emotional stability, and extraversion are almost the same for successful, unsuccessful, and most trusting teams.

### 7.2.2. Different Occupational Groups

We examine the relation between personality traits and interdependency level. In simulations, we fix the interdependency relation of teamwork to one of the relations given in Figure 6.1. Actually, varying interdependency levels are corresponding to different occupational groups. On one hand, some occupations requires strongly connected team members, such as football teams, armies, surgery teams, etc. On the other hand, the members of teams such as party organizers and wedding planners are loosely coupled with each other.

Table 7.3. Average agreeableness of IMA teams for different interdependencies

| | Successful | Unsuccessful | Most Trusting |
|---|---|---|---|
| Interdependency-1 | 0.66 | 0.66 | 0.65 |
| Interdependency-2 | 0.78 | 0.64 | 0.70 |
| Interdependency-3 | 0.84 | 0.70 | 0.69 |
| Interdependency-4 | 0.76 | 0.69 | 0.70 |
| Interdependency-5 | 0.86 | 0.66 | 0.70 |

Table 7.3 shows the average agreeableness levels of teams with respect to different interdependency relations in Figure 6.1. The agreeableness level of three types of teams are almost the same for the first interdependency relationship, where tasks are

performed independently from each other, and interaction is not required between team members. However, the difference between the agreeableness levels of successful and unsuccessful teams obviously increases as the interdependency level of teamwork increases. When we look at the fifth interdependency relationship, where team members are strongly dependent on each other, the agreeableness level of best performing teams, which is 0.86, is higher than the agreeableness level of worst performing, which is 0.66, and most trusting teams, which is 0.70. Agreeableness trait is positively related to teamwork performance for teamwork having interdependent tasks, and the significance of agreeableness factor increases in proportion to the interdependency level.

Table 7.4. Average conscientiousness of IMA teams for different interdependencies

| | Successful | Unsuccessful | Most Trusting |
|---|---|---|---|
| Interdependency-1 | 0.62 | 0.36 | 0.62 |
| Interdependency-2 | 0.57 | 0.36 | 0.56 |
| Interdependency-3 | 0.60 | 0.36 | 0.57 |
| Interdependency-4 | 0.58 | 0.37 | 0.58 |
| Interdependency-5 | 0.61 | 0.35 | 0.55 |

Average conscientiousness levels of best performing, worst performing, and most trusting teams are shown in Table 7.4. The conscientiousness level of better performing teams are always remarkably higher than the conscientiousness level of better performing teams. The conscientiousness factor, which is also known to be the indicator of individual task performance, keeps the significance with respect to team performance for all interdependency levels. Either team members work independently as in the first interdependency relationship or they are strongly dependent on each other as in the fifth interdependency relationship, conscientiousness is always strongly and positively related to teamwork performance.

In Figure 7.1, the average appraisal error of all agents are depicted. The increase in the interdependency levels causes an increase in the average error. Remember that the agreeableness levels directly affect the performance of the interacting providers, and their performance decreases with respect to their level of agreeableness as given

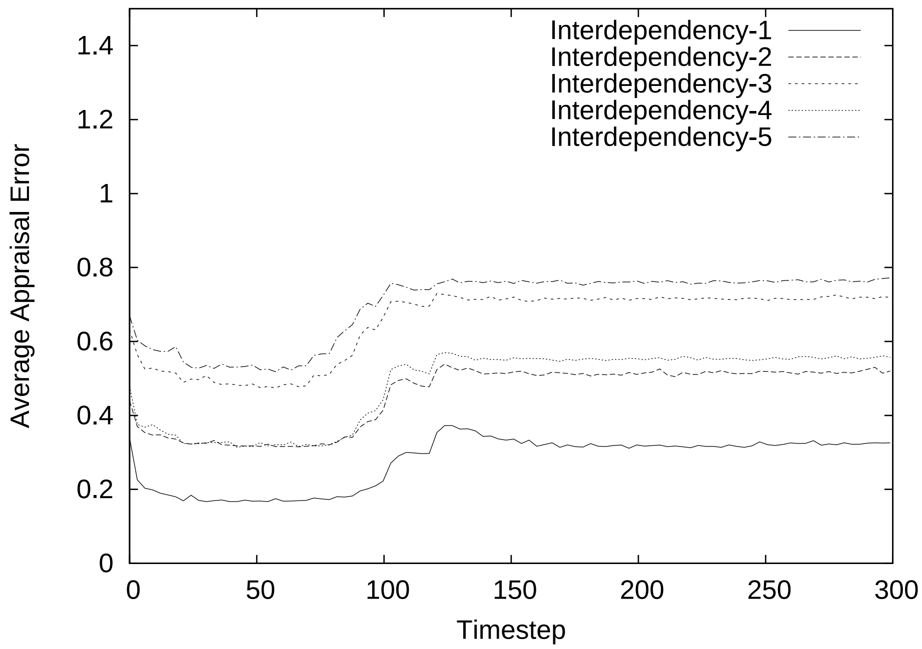


Figure 7.1. Average error of IMAs for different interdependencies

in Equation 6.3. Intuitively, if the team members frequently interact with each other, then this interaction becomes a factor that the agent is dependent on one or more agents in the team. Unless these interacting agents are perfectly compatible with each other, which means agreeableness levels are all 1.0 and is impossible in reality, problems with varying degrees always occur between interacting agents. Note that the highest amount of interaction required for the fifth interdependency. This results with the highest average appraisal error. Simulations with the first interdependency type result with the lowest error, since it doesn't require any interactions between team members. Consequently, the results are very reasonable.

### 7.3. Effect of Teamwork Trust

An additional attribute of a team instance is the average trust between team members. Trust among team members is calculated by averaging the sum of trust of each team member to other members on corresponding tasks. We are interested in teams having most trusted teammates (average trust is calculated between team members), thus keep top fifty team instances that have high average trust among team members. We compare the teams having higher average trust, with successful teams with respect to their performance to answer whether trust among team members is necessarily required for successful teams.

In Table 7.2, the average trust values of successful and unsuccessful teams shows that the average trust within team members of successful teams that is a value of 0.59 whereas for unsuccessful teams, it is 0.51 as given in Table 7.2. This indicates that successful teams consist of members trusting to each other more than the members of unsuccessful teams. (Even though the difference is not large) This is an interesting and surprising result because we didn't enforce such a behavior among agents. However, trust by itself is not sufficient to guarantee a high performance. Note that in Table 7.2 better performing teams achieve a performance of 0.99, whereas most trusting teams achieve a mere 0.46. Therefore, we come to the conclusion that establishing a team with members having highest trust to each other doesn't guarantee to have a great performance. Instead, we should consider the team personality in order to improve teamwork performance.

### 7.4. Homogeneity vs. Heterogeneity

In this study, teams that have low standard deviation among team members in terms of one of the personality traits are usually referred to as homogeneous with respect to this trait, whereas teams that are high in terms of standard deviation are described as heterogeneous. Thus, we try to investigate the relationship between the homogeneity and heterogeneity of the teams with respect to each personality trait. We look at all successful teams whether we can characterize them in terms of homogeneity or there is a relationship between the homogeneity/heterogeneity and the success of the team.

In Table 7.2, considering the standard deviation of conscientiousness level of successful, unsuccessful, and most trusting teams, there is some difference between successful and unsuccessful teams with comparison to the standard deviation of other traits. The standard deviation of conscientiousness level of successful teams is 0.24, while standard deviation of conscientiousness level of unsuccessful teams is 0.31 as given in Table 7.2. Besides conscientiousness, the homogeneity of conscientiousness is also found to be positively related to teamwork performance. On the other hand, the standard deviations of other traits have higher values that means the teams are heterogeneous with respect to these traits.

Table 7.5. Standard deviation of agreeableness of IMA teams for different interdependencies

| | Successful | Unsuccessful | Most Trusting |
|---|---|---|---|
| Interdependency-1 | 0.18 | 0.18 | 0.17 |
| Interdependency-2 | 0.16 | 0.18 | 0.15 |
| Interdependency-3 | 0.14 | 0.17 | 0.14 |
| Interdependency-4 | 0.15 | 0.17 | 0.15 |
| Interdependency-5 | 0.11 | 0.18 | 0.13 |

Table 7.5 shows the standard deviation of the agreeableness levels of teams with respect to the different interdependency relationships. Simulations are repeated fifty times for each interdependency relationship. The standard deviation of agreeableness levels of most successful teams has a value of 0.18 for the teamwork having the first interdependency relationship. Then, this value decreases to 0.16, 0.14, 0.15, and 0.11 for the teamwork having the second, third, fourth, and the fifth interdependency relationship, respectively. This shows that besides the significance of agreeableness, the homogeneity of the agreeableness trait also becomes significant as the interdependency level increases. Similar to the homogeneity of successful teams with respect to conscientiousness, the homogeneity of better performing teams in terms of agreeableness becomes significant when the agreeableness is significant for success. We only introduce the standard deviation of agreeableness trait because homogeneity/heterogeneity of teams in terms of other traits does not exhibit such a change with respect to different interdependency relations.

Conscientiousness and the homogeneity of conscientiousness are significant on success for all interdependency relationships. The role of agreeableness and the homogeneity of agreeableness has increasing significance for the success of teams in teamwork with increasing interdependency levels. These results show that when a trait is significant for success, the successful teams are also homogeneous with respect to this trait.

### 7.5. Effect of Team Maturity

Successful teams may be established by chance, and come together for the purpose of carrying out a teamwork for once. For example, even the emotional stability level of the team is low, they can come up with great result. However, the success of the same team for the second time has very low probability and nearly impossible. In this case, we can say that emotional stability is not significant for the success of the team. However, this result does not completely express the situation. For this reason, we examine mature teams whether any relations between team personality and performance are emerged with respect to the team maturity. A mature team is defined as a group of providers that come together more than once to carry out the same tasks. In the pursuit of observing mature teams, we consider only the teams that the number of times that they come together to perform a teamwork is more than one for reporting most successful, unsuccessful, and most trusting teams.

Considering only mature teams, Table 7.6 shows the personality, performance, and trust within team values of successful, unsuccessful, and most trusting mature teams. In this case, the average emotional stability level of successful teams that has a value of 0.64 is remarkably higher than the emotional stability level of unsuccessful teams that has a value of 0.48. This means that some teams that have low emotional stability levels may be successful by chance for once, but very low probability for the second or third successful teamwork. Because of this, the emotional stability levels of successful and unsuccessful teams that only come together for once are nearly the same with each other in Table 7.2. However, if a team successfully carries out a teamwork more than once, then their emotional stability levels become significant for their success as shown in Table 7.6. Actually, the emotional stability is a significant factor for the successful teams. The conscientiousness level for successful teams is again higher with comparison to the others. Average agreeableness level of successful teams is slightly higher that this level of unsuccessful teams.

Table 7.6. Average personality, performance, and trust of mature IMA teams

| | Successful | Unsuccessful | Most Trusting |
|---|---|---|---|
| Agreeableness | 0.72 | 0.66 | 0.71 |
| Conscientiousness | 0.63 | 0.46 | 0.58 |
| Emotional Stability | 0.64 | 0.48 | 0.58 |
| Extraversion | 0.51 | 0.51 | 0.52 |
| Performance | 0.90 | 0.12 | 0.60 |
| Trust | 0.60 | 0.59 | 0.71 |

## 7.6. Effect of Trust Modeling

In order to prove the accuracy of the results that is the relation between the personality and performance, we perform simulations by using the agents with a different trust model, which is teamwork trust model that is explained in Chapter 3. We try to show that our results are independent of the underlying trust model. The number of agents, the length of the simulation, the number of repetitions of the simulations, and other parameters are the same as in the first simulation.

In this section, the results of teamwork trust modeling agents are introduced to show whether the relation between the personality and teamwork performance changes if sophisticated agents are used. The performance, trust within team, and average personality of successful, unsuccessful, and most trusting teams are given in Table 7.2. We collect data from 50 simulations.

In Table 7.7, the average conscientiousness level of successful teams (0.60), is remarkably higher than the conscientiousness level of unsuccessful teams (0.34). The average level of other personality traits, agreeableness, emotional stability and extraversion are almost the same for successful, unsuccessful and most trusting teams. Furthermore, there is a remarkable difference between standard deviation of conscientiousness levels of successful and unsuccessful teams with comparison to the standard deviation of other traits. The standard deviation of conscientiousness level of successful teams is 0.24, while standard deviation of conscientiousness level of unsuccessful teams is 0.31 as shown in Table 7.7. Consequently, we compare the results of TMA (Teamwork Modeling Agents) having personality given in Table 7.7 with the results of IMA (Individual Modeling Agents) having personality given in Table 7.2; the conscientiousness and the homogeneity of conscientiousness are significant again for the success of teamwork.

Table 7.7. Average personality, performance, and trust of TMA teams

| | Successful | Unsuccessful | Most Trusting |
|---|---|---|---|
| Agreeableness | 0.66 | 0.66 | 0.69 |
| Conscientiousness | 0.60 | 0.34 | 0.59 |
| Emotional Stability | 0.58 | 0.55 | 0.57 |
| Extraversion | 0.50 | 0.47 | 0.54 |
| Std. of Consc. | 0.24 | 0.31 | 0.25 |
| Performance | 0.99 | 0.0 | 0.67 |
| Trust | 0.52 | 0.47 | 0.57 |

When we compare the average trust of successful and unsuccessful teams, we find that the average trust within team members of successful teams, that is 0.52, whereas for unsuccessful teams, it is 0.47 as given in Table 7.7. Even they seem to be very close to each other, the significance of this difference is understood when we compare the maximum possible trust, which has a value of 0.57. This indicates that trust among members of successful teams is higher than trust among the members of unsuccessful teams. Furthermore, better performing teams achieve a performance of 0.99, whereas most trusting teams achieve a mere 0.67. That means trust among team members is not sufficient for the success of teams. These results also match up the results obtained with IMA agents.

Table 7.8 shows the average agreeableness levels of teams with respect to different interdependency relations in Figure 6.1. These results are similar to the results in

Table 7.8. Average agreeableness of TMA teams for different interdependencies

| | Successful | Unsuccessful | Most Trusting |
|---|---|---|---|
| Interdependency-1 | 0.64 | 0.67 | 0.65 |
| Interdependency-2 | 0.76 | 0.66 | 0.69 |
| Interdependency-3 | 0.84 | 0.67 | 0.71 |
| Interdependency-4 | 0.76 | 0.65 | 0.69 |
| Interdependency-5 | 0.84 | 0.69 | 0.71 |

Table 7.3 with the following aspects. The average agreeableness level of successful teams increase as the interdependency level increases, while this level of unsuccessful and most trusting teams slightly increases. Actually, this increase is unnecessary when we compare with the increase of this level of success teams. Furthermore, the difference between successful teams and others become observable while the interdependency level increases towards the fifth interdependency. When we look at the teamwork having fifth interdependency relation, where team members are strongly dependent on each other, the agreeableness level of best performing teams, that is 0.84, is conspicuously higher than the agreeableness level of worst performing, which is 0.69, and most trusting teams, which is 0.71. The agreeableness trait is positively related to team performance for teamwork having interdependent task, and the significance of agreeableness factor increases in proportion to the interdependency level of teamwork.

Table 7.9. Average conscientiousness of TMA teams for different interdependencies

| | Successful | Unsuccessful | Most Trusting |
|---|---|---|---|
| Interdependency-1 | 0.60 | 0.32 | 0.60 |
| Interdependency-2 | 0.59 | 0.34 | 0.59 |
| Interdependency-3 | 0.60 | 0.35 | 0.54 |
| Interdependency-4 | 0.59 | 0.34 | 0.59 |
| Interdependency-5 | 0.60 | 0.37 | 0.61 |

The average conscientiousness levels of better performing, worse performing, and most trusting teams are shown in Table 7.9. Similar to results in Table 7.4, conscientiousness level of better performing teams are always remarkably higher than conscientiousness level of worse performing teams.

Table 7.10. Standard deviation of agreeableness of TMA teams for different interdependencies

| | Successful | Unsuccessful | Most Trusting |
|---|---|---|---|
| Interdependency-1 | 0.19 | 0.19 | 0.19 |
| Interdependency-2 | 0.17 | 0.19 | 0.18 |
| Interdependency-3 | 0.16 | 0.18 | 0.18 |
| Interdependency-4 | 0.15 | 0.18 | 0.18 |
| Interdependency-5 | 0.12 | 0.17 | 0.17 |

Table 7.10 shows the standard deviation of agreeableness levels of teams with respect to different interdependency relationships. The standard deviation of agreeableness level of successful teams within the teamwork with the first interdependency type, has a value of 0.19 and this value decreases to 0.17, 0.16, 0.15, and 0.12, for the second, third, fourth, and fifth interdependency relation respectively. The standard deviation of agreeableness levels decreases as the interdependency degree of the teamwork increases. In other words, the homogeneity of better performing teams with respect to agreeableness increases in proportion to the interdependency level. The behavior of homogeneity is similar to the results in Table 7.5 that are obtained by using agents with individual trust model.

In Figure 7.2, average appraisal error of TMA agents is depicted. The increase in interdependency levels causes increase in average error, though similar results with results in Figure 7.1 are obtained. However, the behavior of the appraisal error and its values differ between two trust models as expected. The average appraisal error of TMA agents slightly decreases at the beginning, and then stay stable during the game. Furthermore, the error values are smaller than the values of IMA agents. Because, TMA agents are more successful in modeling trust to agents in the environment, since their teamwork trust modeling tool is more sophisticated and powerful.

Obtaining parallel results with IMA and TMA agents indicates that our conclu-

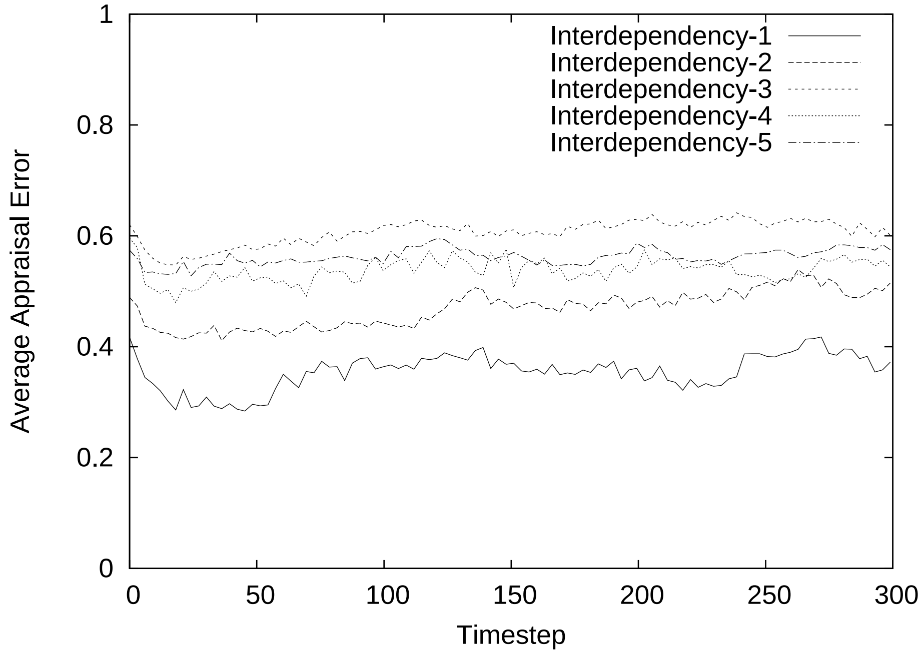


Figure 7.2. Average error of TMAs for different interdependencies

sions are model-independent. On one hand, we use agents with very simple trust model in the simulations. On the other hand, agents with very sophisticated trust model are used. The relation between the performance and personality of teams exhibits similar behavior for both trust models.

## 7.7. Significance Analysis

In previous sections of this chapter, we experimentally show the importance of traits on the performance of teamwork. In this section, we analyze the significance of the factors on the performance by using statistical methods, namely *ANOVA*. ANOVA, ANalysis Of VAriance, is a collection of statistical models, where the observed variance is partitioned into components according to different factors in the environment. In an ANOVA model, there are factors and responses that are dependent on the factors. Different ANOVA procedures exist according to the number of factors in the environment, such as one-way, two-way, three-way ANOVA.

In our model, the personality traits of the teamwork are factors and the performance of the teamwork is the response variable. Factors are independent from each other. We analyze the significance of three personality traits namely agreeableness, conscientiousness, and emotional stability. Because ANOVA models that include three or more factors are difficult to interpret; they are not very useful. That is why, we eliminate extraversion factor, since the experimental results in Section 7.2 show that these three factors are more significant than extraversion in ART teamwork environment.

Table 7.11. ANOVA design for teamwork environment in ART

| Data Type | Name | Levels |
|---|---|---|
| Factors | Agreeableness | Low, Medium, High |
| | Conscientiousness | Low, Medium, High |
| | Emotional Stability | Low, Medium, High |
| Response | Performance | |

ANOVA design of our teamwork model is given in Table 7.11. There are three factors and a response variable as shown in the table. Performance is corresponding to $(1 - appraisalError)$ in ART. The value of the factors are classified into three levels: low, medium, and high. The value ranges of the levels of factors are given in Table 7.12. We divide possible personality trait ranges into three parts for defining these three levels.

Table 7.12. Value ranges for the levels of factors in ANOVA model

| Factor | Low | Medium | High |
|---|---|---|---|
| Agreeableness | $0.30 - 0.60$ | $0.60 - 0.80$ | $0.80 - 1.0$ |
| Conscientiousness | $0.0 - 0.33$ | $0.33 - 0.66$ | $0.66 - 1.0$ |
| Emotional Stability | $0.0 - 0.33$ | $0.33 - 0.66$ | $0.66 - 1.0$ |

In our ANOVA design, there are 27 combinations for three factors, where each factor exhibit three levels. Note that we look at the average personality of the team and their performance. First, we classify all teamwork instances into 27 groups according to the distribution of their personality levels. Then, we randomly select 500 teamwork instances for each combination and obtain 13500 instances for three-way ANOVA with repetitions test. Teamwork instances are obtained from one simulation, where the experimental setup in Table 7.1 is used.

The analysis of variance for the team performance data is shown in Table 7.13. Main factors, two-way interactions of the factors, three-way interaction of factors, and their effects are given in the results. The information related to sum of squares, degree of freedom, F, and p-level of all factors and their combinations are provided. The 27 treatment combinations have 26 degrees of freedom. Each main effect has 2 degrees of freedom, each two-factor interaction has 4 degrees of freedom, and the three-factor interaction has 8 degrees of freedom.

Table 7.13. The analysis of variance for the team performance

| Source of Variation | Sum of Squares | Degree of Freedom | F | p-level |
|---|---|---|---|---|
| Main Effects | | | | |
| A, Agreeableness | 22.59 | 2 | 186.06 | $< 0.0001$ |
| B, Conscientiousness | 19.78 | 2 | 162.92 | $< 0.0001$ |
| C, Emotional Stability | 6.36 | 2 | 52.37 | $< 0.0001$ |
| Two-way Interactions | | | | |
| AB | 0.19 | 4 | 0.81 | 0.52 |
| AC | 2.26 | 4 | 9.32 | $< 0.0001$ |
| BC | 7.56 | 4 | 31.13 | $< 0.0001$ |
| Three-way Interactions | | | | |
| ABC | 0.53 | 8 | 1.09 | 0.37 |

The most important result is that agreeableness, conscientiousness, and emotional stability factors are statically significant for performance, because p-level values are very close to 0. Furthermore, two of the two-factor interactions, AC and BC, are also significant.

Remember that if F that is calculated using sum of squares is close to 1, the evidence favors the null hypothesis (the two population variances are equal). But if F is much larger than 1, then the evidence is against the null hypothesis. So, the results show that F values of all factors and interactions except AB and ABC are much larger than 1. That means variances of these factors are not equal.

The two-factor interactions between agreeableness, conscientiousness, and emotional stability are analyzed graphically in Figure 7.3, Figure 7.4, and Figure 7.5. The excluding factor is averaged in two-way interaction graphs. Increasing levels of both agreeableness, conscientiousness, and emotional stability result with higher performances of teams. So, the highest performance is reached when both two factors have values of high levels. That shows the significance of three traits for the teamwork performance.

Although higher levels of these traits result with higher performances in all three graphs, the maximum difference between the lowest and the highest performance is obtained in Figure 7.3. The lowest mean performance that is a value below 0.40 is seen when the levels of both conscientiousness and agreeableness are low. The highest mean performance that is a value above 0.60 is seen when these levels are high. This is because, agreeableness and conscientiousness traits have more important roles for the teamwork performance. The minimum difference between the lowest that is a value above 0.40 and the highest performance that is a value nearly 0.50 is obtained in Figure 7.5.

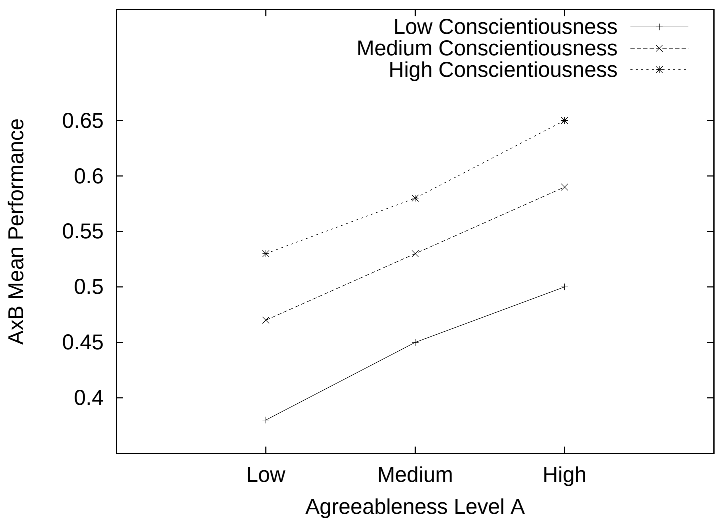


Figure 7.3. Agreeableness level vs. performance wrt different conscientiousness levels

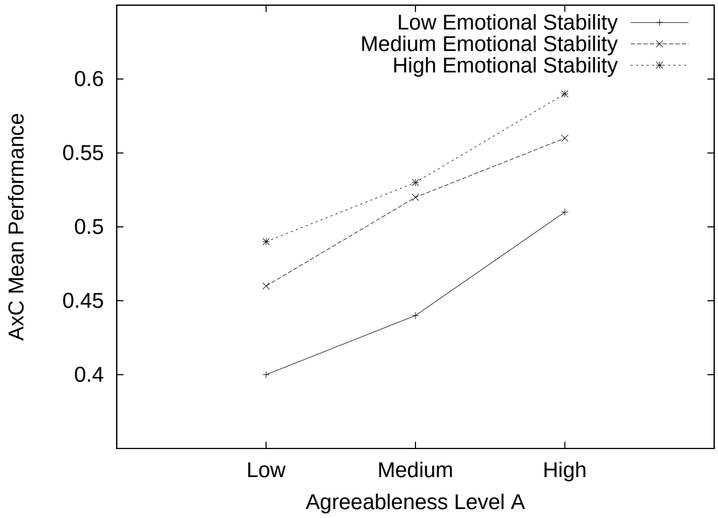


Figure 7.4. Agreeableness level vs. performance wrt different emotional stability levels

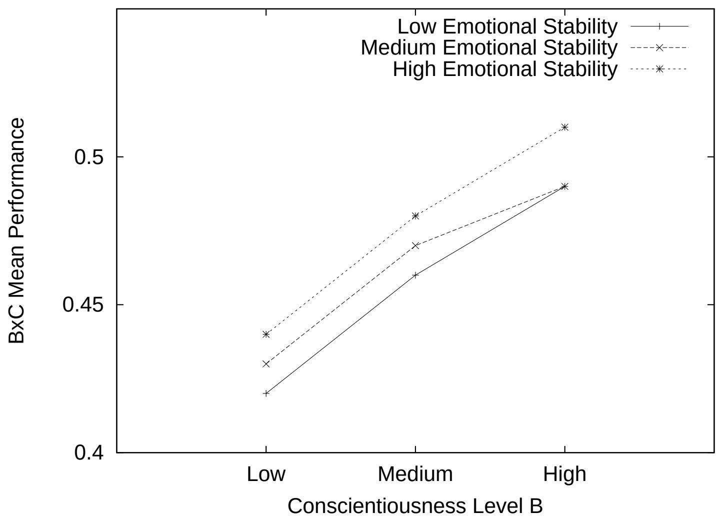


Figure 7.5. Conscientiousness level vs. performance wrt different emotional stability levels

# 8. CONCLUSIONS

In this theses, we mainly study on multiagent teamwork with the aspect of performance and effectiveness. Trust and personality composition are two of the most important factors that affect teamwork performance. First, we develop a teamwork trust model in order to model the trust to teams rather than individual providers in teamwork environment. Then, we study teams that consist of agents with personalities to examine the influence of personality on teamwork performance. We incorporate well-known teamwork structure into ART Testbed as an experimental framework.

Even though trust is a popular issue and many trust models exist in the literature, studies generally consider trust to individual providers. In this study, we show that trust to a team cannot be evaluated by using trust to individual providers. In other words, composing most trusted individual providers into a team may not always work. There is no guarantee that these successful individuals work in concord with the other team members. Right decision is considering trust to a team or how successful a team is when they work together. We develop teamwork trust model that consists of three tools: team trust model, service graphs, and individual trust model, and models the trust to a team for a particular teamwork. When we compare our model with individual trust model, teamwork trust model is remarkably more capable of modeling trust to teams.

The second aspect that is studied in this theses is personality composition of the team. Using ART, we study whether the well-known personality traits emerge as significant in teamwork. The results indicate that conscientiousness has the highest significance in the multiagent teamwork. Conscientiousness strongly influences the performance of an individual, where the teamwork performance is obtained by combining individual performances, higher individual performances result with higher teamwork performances. The homogeneity of teams with respect to conscientiousness is also distinctive with comparison to other traits. An interesting result, which is obtained without any enforcement to the environment, is that successful teams have higher trust between team members, than unsuccessful teams trust level. When we examine team performances with respect to different interdependency relations, the significance of agreeableness factor increases as the interdependency between tasks increases, while conscientiousness keeps its significance for all interdependency levels. These results are also supported by many studies in real life organizations.

As a future work, an agent that models the agents in the environment by considering both their expertise and personalities can be developed to establish effective teams. By this way, we obtain a more sophisticated agent model. This agent model can be compared with TMA agents that only consider the expertise of agents.

## REFERENCES

1. Fullam, K., T. B. Klos, G. Muller, J. Sabater, Z. Topol, K. S. Barber, J. S. Rosenschein, and L. Vercouter, "The Agent Reputation and Trust (ART) Testbed Architecture", *The Workshop on Trust in Agent Societies at the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 50–62, 2005.

2. Yolum, P. and M. P. Singh, "Service Graphs for Building Trust", *International Conference on Cooperative Information Systems (CoopIS)*, pp. 509–525, 2004.

3. Singh, M. P., *Multiagent Systems, A Theoretical Framework for Intentions, Know-How, and Communications*, Springer-Verlag, 1994.

4. Goldberg, L. R., "Language and Individual Differences: The Search for Universals In Personality Lexicons", *Review of Personality and Social Psychology*, Vol. 2, pp. 141–166, 1981.

5. Costa, P. T. and R. R. McCrae, *The NEO Personality Inventory Manual*, Psychological Assessment Resources, Odessa, 1985.

6. Cattell, R. B., *Personality and Motivation: Structure and Measurement*, New York: World Book, Oxford, 1957.

7. Digman, J. M., "Personality Structure: Emergence of The Five-Factor Model", *Annual Review of Psychology*, Vol. 41, pp. 417–440, 1990.

8. Barrick, M. R. and M. K. Mount, "The Big Five Personality Dimensions and Job Performance: A Meta-Analysis", *Personnel Psychology*, Vol. 44, No. 1, pp. 1–26, 1991.

9. Neuman, G. A., S. H. Wagner, and N. D. Christiansen, "The Relationship Between Work-Team Personality Composition and The Job Performance of Teams", *Group*

*& Organization Management*, Vol. 24, No. 1, pp. 28–45, 1999.

10. Kichuk, S. L. and W. H. Wiesner, "The Big Five Personality Factors and Team Performance: Implications For Selecting Successful Product Design Teams", *Journal of Engineering and Technology Management*, Vol. 14, pp. 195–221, 1997.

11. Costa, A. C., "Work Team Trust and Effectiveness", *Personnel Review*, Vol. 32, No. 5, pp. 605–622, 2003.

12. WS-BPEL: Web Services Business Process Execution Language Version 2.0, http://docs.oasis-open.org/wsbpel/2.0/Primer/wsbpel-v2.0-Primer.html, 2007.

13. Kang, M., "The Effects of Agent Activeness and Cooperativeness On Team Decision Efficiency: A Computational Simulation Study Using Team-Soar", *International Journal Human-Computer Studies*, Vol. 65, pp. 497–510, 2007.

14. Kafalı, Ö. and P. Yolum, "Trust strategies for ART Testbed", *Ninth International Workshop on Trust in Agent Societies at The Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 43–49, 2006.

15. Kafalı, Ö. and P. Yolum, "Action-Based Environment Modeling for Maintaining Trust", *Eleventh International Workshop on Trust in Agent Societies at The Seventh International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 23–32, 2008.

16. Teacy, W. T. L., J. Patel, N. R. Jennings, and M. Luck, "TRAVOS: Trust and Reputation In The Context of Inaccurate Information Sources", *Journal of Autonomous Agents and Multiagent Systems*, Vol. 12, No. 2, pp. 183–198, 2006.

17. Teacy, W. T. L., G. Chalkiadakis, A. Rogers, and N. R. Jennings. "Sequential Decision Making with Untrustworthy Service Providers", *Proceedings of The Seventh International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 755–762, 2008.

18. Jones, C. L. D., K. K. Fullam, and S. Barber, "Exploiting Untrustworthy Agents in Team Formation", *International Conference on Intelligent Agent Technology*, pp. 299–302, 2007.

19. Ghasem-Aghaee, N. and T. I. Oren, "Towards Fuzzy Agents with Dynamic Personality for Human Behavior Simulation", *Proceedings of the 2003 Summer Simulation Conference*, pp. 3–10, 2003.

20. Sichman, J. S. and R. Conte, "Multi-Agent Dependence by Dependence Graphs", *Proceedings of The First International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 483–490, 2002.