Universiteit Utrecht

Faculty of Science
Intelligent Systems Group / Division of Virtual Worlds

Master's thesis

Introducing Culture in the Multi-party Virtual Dialogue System

Andrey Obolenskiy, 5671108

1. Supervisor Dr. F. P. M. Dignum
Faculty of Science
Department of Information and Computing Sciences
Utrecht University

2. Supervisor Dr. Z. Yumak
Faculty of Science
Department of Information and Computing Sciences
Utrecht University

August, 2017

Abstract

Over the past twenty years, we have learned to visually create truly lifelike virtual characters. We can bring them to life with current modelling, animation and rendering technologies, and even populate vast virtual worlds with them. On the other hand, advances in the fields of artificial intelligence and computational anthropology can also amaze one's imagination. However, the unification of these areas, namely the empowerment of virtual characters with real human features, such as personality, social roles, interpersonal relationships and culture, still awaits for many new discoveries. In particular, these discoveries are needed for the field of human-machine interaction.

Thereby, in this thesis we propose an internal cultural model that serves as an extension to the existing multi-party virtual dialogue system. Our approach is based solely on Hofstede's concept of culture. To prove the effectiveness of our model, we conducted a cross-cultural user study, in which participants from Mexico, Russia and the Netherlands interacted with culturally strengthened virtual characters in two scenarios: doctor-patients and teacher-students conversations. Despite the fact that we got some positive results, we came to the conclusion that for a virtual incarnation of truly lifelike human behavior culture alone is not enough. To achieve excellence in such human-machine interaction systems, it is also necessary to take into account the personality, personal experience, and other social aspects.

Contents

1 Introduction

This Chapter is intended to indicate the main focus of our research and its structure is as follows. First, in section 1.1 we introduce the main subject domain of our proposed research, namely human-machine communication via virtual characters. Second, we describe our motivation and research objective in section 1.2. Finally, section 1.3 provides an outline of the remainder of this thesis.

1.1 Research Domain

Human-machine interaction has been a subject of research for more than five decades and become more elaborated each year. Various forms of implementation of human-machine interaction have found its applications in many domains, such as transportation, medicine and entertainment [1]. One of the most interesting and rapidly growing forms of interaction is human communication with virtual characters and human-like robots [2, 3, 4, 5]. To be more precise, our research is related to multi-modal, multi-party communication between humans and virtual characters.

Extensive formal and computational studies have been done on two-party human-machine communication. Relatively recently the focus of human-machine communication has shifted to multi-party settings [6]. Both two-party and multi-party interactions have a variety of issues, most of which are yet to be solved [7, 8]. Such issues, like conversational roles and turn-taking, are closely related to several factors of real human communication.

As stated above, many factors that affect human-human communication are also important in human-machine conversation. These factors may be external, such as facial expressions and gaze, and internal (social), such as social roles, culture and interpersonal relationships. Since it is almost impossible to model all the factors of human communication into a single multi-party dialogue system at once, we limit our focus on a single factor: modelling cultural influence on human-machine communication in a multi-party conversation setting.

1.2 Motivation, Research Question and Contribution

Current computational models of culture for virtual characters are focused on different aspects of cultural phenomena. Most of these models can be categorized into two major types: external and internal [9]. Models of the first type are intended to cause the observer to visually perceive culturally appropriate behavior of virtual characters [10, 11, 12, 13, 14, 15, 16]. However, with external cultural model only, a virtual character has no idea how to interpret its own behavior and the behavior of others.

Cultural models of the internal type are intended to represent culture in some sort of rules for virtual characters to make culturally appropriate decisions [17, 18, 19, 20, 21]. These models are more complex in development than external ones, as virtual characters should produce and interpret meaningful behavior influenced with cultural aspects. To model such behavior, the creators of these models base their work on different fundamental concepts of culture.

The notion of culture varies from one scientific work to another. Some researchers tried to establish a single definition of this phenomenon [22]. On the contrary, others try to define culture as conceptual models, in which cultural aspects are structured and systematized [23, 24, 25]. The most notable of this kind is Hofstede's concept of culture, a so called "Software of the Mind" [26, 27]. This approach focuses on different aspects of culture, such as manifestations and dimensions of cultures that influence our society.

Nevertheless, none of the internal models are perfect as in those only some specific facets of cultural influence are considered. Most of these models are not intended for multi-party scenarios. Moreover, cultural dependent phenomena like turn-taking, interruption and overlapping speech, which are very important for multi-party dialogues, are not considered in these models of culture [28]. They focus on more high level, abstract, behavior and give no indication how these have to be connected to the more concrete dialogue behaviors.

As it may be seen from the above, computational cultural models vary as well as the concepts of culture behind them. In our approach, we picked the most comprehensive concept of culture in our opinion, namely Hofstede's approach of cultural dimensions [26, 27]. More specifically, we consider four dimensions in our research: power distance, individualism, long-term orientation and indulgence, all of which are purposely simplified to represent two opposing extremes (stereotypes). Further, we indicate turn-taking and interruptions as important aspects of conversation with regard to cultural influence. We also described an existing multi-party virtual

dialogue system [29], in which our cultural model serves as an extension in a form of additional module.

Therefore, our research question is as follows:

How can we add culture in a form of four Hofstede's cultural dimensions to human-machine conversation in multi-party setting, such that the result will be a culturally appropriate behavior of virtual characters?

Hence, the primary contribution of our approach is a comprehensive but simple-to-use internal model of culture that is solely based on Hofstede's concept of cultural dimensions. Our secondary contribution is that our model is intended to serve as an extension to a recently developed multi-party dialogue system, and thus makes it one step closer to a universal system of realistic virtual characters.

1.3 Thesis Outline

The structure of the rest of the thesis is as follows. In Chapter 2, we depict the notion of culture and describe several cultural concepts. In Chapter 3, we describe related work such as the multi-party dialogue system in which our cultural model is integrated, followed by the description of existing external and internal cultural models. In Chapter 4, we introduce our approach, namely the model of culture for multi-party virtual dialogue system. In Chapter 5, we describe the implementation of our model in the form of a cultural module. In Chapter 6, we describe our evaluation process, experimental design and hypotheses, followed by the results analysis. Finally, in Chapter 7 we conclude this thesis with the assessment of our approach and how it can be extended in the future.

2 Theoretical Background

In this Chapter we provide a survey of related literature. First, in section 2.1 we depict the notion of culture and describe several most noticeable concepts of it. Then, in section 2.2 we provide a small insight on theory of mind. In section 2.3, we introduce several studies on how culture may influence turn-taking and interruptions in conversation. Finally, in section 2.4 we conclude this Chapter with a brief summary of the material presented here.

2.1 The Concepts of Culture

Numerous amount of definitions and concepts of culture exist. More than 60 years ago, Kroeber and Kluckhohn categorized the concepts and definitions of culture in several groups based on emphases of these definitions [22]. The most widely accepted and used paradigm states that culture represents the unwritten rules of society. However, as our interest is based on computational point-of-view, there is no need to excrete a single definition. It is more important to identify the best theoretical model of culture, on which a computational model should be based on.

2.1.1 Hofstede's Theory

Hofstede et al. [26, 27] provided the most influential and rigorous work to date, in which culture is characterized and categorized. Instead of trying to define the notion of culture in terms of other concepts, this work describes culture by means of cultural manifestations and dimensions that influence our societies [30].

The idea holds in the fact that every person carries within themselves patterns of thinking, feeling and potential acting, which are not innate and learned throughout the person's lifetime. These patterns are acquired in the person's social environment and life experiences: family, neighborhood, school, workplace, etc. Culture is always a collective phenomenon and mentioned above patterns are usually shared between those who belong to the same culture.

Nevertheless, as culture can only be learned and cannot be inherited, it should be distinguished from human nature and individual's personality. [Fig. 2.1] illustrates

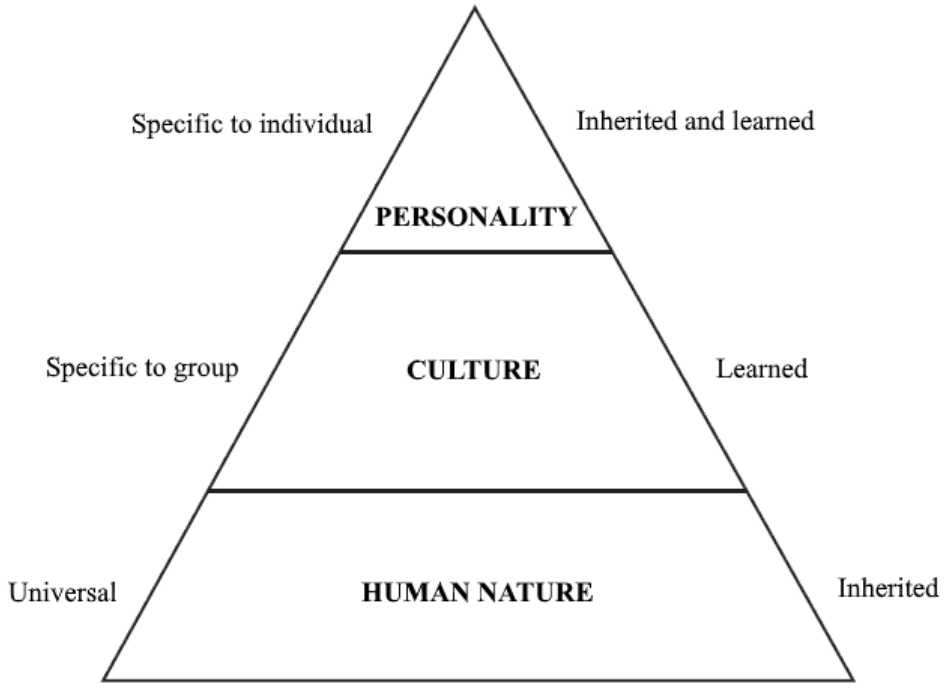

Fig. 2.1. Levels of humans' uniqueness [27].

the humans' levels of uniqueness. Human nature indicates universal patterns, those that are shared between all human beings and which we inherit within our genes. Our ability to feel fear, anger, joy, and sadness, all that belong to the human nature level. On the contrary, personality determines individual's set of patterns that are unique. These patterns are partly inherited and partly learned. However, the way in which we express all these patterns strongly depends on the second level - culture. Patterns of personality that are learned are usually affected by culture. Likewise, the way we express inherited patterns of human nature is also modified by culture.

Cultural Manifestations

Cultural differences may emerge in several ways. Hofstede et al. [26, 27] proposed four types of cultural manifestations that define the concept fairly accurately. These manifestations are symbols, heroes, rituals and values.

- ***Symbols*** are signs that contain a specific meaning, which can be recognized and understood only by those who share the same culture. Symbols are slang words, specific facial expressions and gestures, certain items of clothing and hairstyles. New symbols appear as fast as old ones disappear. In other

words, symbols are all that come with fashion and become replaced with new generations.

- ***Heroes*** are real persons or imaginary characters, who possess traits that are highly valued in a culture. Usually, heroes serve as models for behavior in society. In the past, heroes were actual heroes of wars, saints in religions, scientists, cosmonauts and others that inspired generations. Nowadays, in this age of television and the Internet, inspiration by heroes has shifted more to imaginary characters, from cartoons and movies.

- ***Rituals*** are collective activities that are socially essential for a certain culture. Rituals are social practices that indicate specific patterns of how to behave in certain situations. Rituals can be something routine, such as the ways of greetings and paying respect to elder ones, and can be something more specific, like religious and official ceremonies.

- ***Values*** represent the degree of importance of some things or actions over others, which is defined by the culture and acquired early in our lives. Schwartz defined values as desirable, trans-situational goals, varying in importance, which serve as guiding principles in our lives [33]. Furthermore, Schwartz et al. proposed a universal theory of human values that derives ten types of values by describing its central motivational goal [31, 32, 33, 34].

In conjunction, symbols, heroes and rituals form so called practices that are visible to an outside observer, but their cultural meanings are hidden and only comprehensible by the insiders. New practices are learned throughout our lifetime, and older ones can be replaced or disappear. On the other hand, values are sufficiently stable and are quite resistant to changes.

Cultural Dimensions

The most important aspect of Hofstede's concept are cultural dimensions. Dimensions are the cultural facets that can be measured relative to other cultures. Each dimension groups together several phenomena in a society that occur in combination and allows comparing different national cultures based on these phenomena. However, all the described below dimensions are just abstractions that capture behavioral trends across cultures, and therefore do not represent the cultural ground truth.

At first, based on the database gathered from multinational corporation IBM, which contained the results of questionnaires completed by the employees from 71 coun-

tries, Hofstede proposed four global dimensions: power distance, individualism and collectivism, masculinity and femininity, and uncertainty avoidance [35]. In his later work, Hofstede proposed two additional dimensions of culture: long-term orientation and indulgence [27].

- ***Power distance*** indicates inequality in societies; it highlights the importance of power statuses in society. In cultures with stronger power distance, individuals are treated depending on their power status. In cultures with weaker power distance, power status has no or little influence.

- ***Individualism and collectivism*** are the opposite poles that indicate the degree of individuality in society. Individualism refers to societies in which individuals expect to be treated independent of each other. Collectivism refers to societies in which individuals tend to be strongly integrated into groups (relatives, friends, workmates, etc.) and expect an in-group behavior (helping and protecting each other, working together, etc.)

- ***Masculinity and femininity*** refers to the importance and the degree of equality of gender roles in societies. A society is considered to be masculine when emotional gender roles are clearly distinct: men are supposed to be more focused on material success, while women are supposed to be more concerned with the quality of life. A society is considered to be feminine when emotional gender roles overlap: both men and women are supposed to be focused on material success and concerned with the quality of life.

- ***Uncertainty avoidance*** refers to the desire for explicit situations with predictable outcomes. Stronger desire leads to emergence of rules that reduce the likelihood of ambiguity. Cultures with weaker desire tend to avoid the presence of such rules and prefer situations with unspecified behavior.

- ***Long-term orientation*** refers to the amount of time taken into account when taking decisions. In cultures with long-term orientation each decision is accurately weighted, rewards can be sacrificed for better ones in the future. In cultures with short-term orientation, individuals expect an immediate success, avoiding failures and decisions relying on dogmatic rules.

- ***Indulgence and restraint*** indicate the degree of freedom in which individuals can act as they please or not. Indulgent cultures possess a higher degree of freedom that allow individuals to fulfil their desires related enjoying life and fun related activities. In restrained cultures, such desires are limited and regulated by strict social norms.

2.1.2 Kemper's Status-power Theory

Another cultural concept is the Status-Power theory by Kemper, the main purpose of which is to structure the ultimate motives of the overall socio-cultural behavior [23]. According to Kemper, individuals manipulate and navigate the order of interaction in accordance with two main relational dimensions: status and power. Both dimensions are meant to cover a wide range of interactional strategies and behaviors. Status denotes the degree of respect with which we voluntarily concern needs, wishes and interests of the others. Power, on the other hand, denotes the degree of coercion with which we want the others to behave in our favor.

2.1.3 D'Andrade's Schema Theory

Schema theory proposed by D'Andrade declares that culture can be represented as a collection of schemas, which are abstractions of some concept or culturally appropriate behavior [24]. There are at least two important classes of schemas: cognitive schemas and event schemas. Cognitive schemas represent compositions of knowledge about a certain concept that can be caused by specific symbols or images. Event schemas are behaviors for everyday situations. The notion of event schemas is similar to the notion of rituals in Hofstede's approach – culturally appropriate behavior.

Moreover, in an earlier work, D'Andrade proposed the constitutive rules system – a cognitive schema to account shared conceptual meaning [25]. That system is defined as a set of rules that are known and shared by the members of a particular culture and which define a culture-related concept. Adhering certain behaviors infers a cultural state that relates back to certain cultural concept. In addition to symbols, observing or performing any of the actions associated with the schema trigger a particular type schema.

2.1.4 Discussion

Hereinabove, we described several theoretical approaches that aim to conceptualize and structuralize the comprehensive notion of culture. Both Kemper's status-power and D'Andrade's schema theory concepts are fundamental and elaborated, and some internal cultural models, which we introduce in the next Chapter, are based on these theories [17, 18, 19]. However, in our opinion, these two concepts are not as well-defined and well-structured as the first one - Hofstede's "Software of the Mind".

We focus our research on Hofstede's concept of culture as it has established itself as a comprehensive and robust approach [27]. This approach of cultural manifestations and dimensions is well-structured, which is undoubtedly a huge benefit when one tries to create a computational model on the theoretical basis. As the aim of our research is to empower virtual characters with behavior appropriate to different cultures, cultural dimensions is the best choice, because manifestations of those can be measured in relevance to other cultures.

In the next two sections we discuss other worth-mentioning theoretical studies, some of which are not directly related to the notion of culture, but can be used in combination with it.

2.2 Theory of Mind

Theory of Mind, proposed by Nichols and Stich, states that individuals have an ability to ascribe mental states, such as beliefs, intentions and values, to themselves and to others [36]. In a social interaction individuals generate two types of mental representations. The first type of representations consists of the individual's own knowledge, beliefs, intentions and values. The second type of mental representations is a set of belief models of the other individuals' knowledge, beliefs, intentions and values. All these mental representations in combination are used in generating behavior. In the absence of personal information, many of our initial beliefs about another's norms, beliefs and intentions are based on stereotypes about the other person's culture.

As mentioned above, this theory has no particular interest to our research. However, its concept of mental representations is a very suggestive point of view of how real humans generate behavior when they do not have enough personal information about others. And consequently, understanding of such phenomenon can help in development of virtual characters' behavior, appropriate to similar situations. In addition, one of the cultural models introduced below combines this theory with D'Andrade's schema approach to define the idea of shared socio-cultural aspects [19].

2.3 Turn-taking and Interruption in Human-machine Conversation

Culture has great influence on most aspects of our social life. As our research is based on human conversation with virtual characters, it is important to distinguish what aspects of conversation can be affected by culture. Some of these aspects are turn-taking, interruption and overlapping mechanisms in conversation.

During real human conversation, it is essential that interlocutors take their turns appropriately. Motivation to take turn or remain silent in conversation depends on many factors, such as personal traits of interlocutors, desire to avoid overlaps, the conversational setting, cultural norms and expectations, and other interrelated factors. In our investigation, we are interested in the influence of cultural aspects on turn-taking.

The question of cultural influence on turn-taking has been addressed and studied in the past. More than twenty years ago, a research was conducted that compared Spanish and American turn-taking styles [28]. The results indicate that amount of overlap in Spanish conversation is much higher than in American conversation. Another empirical observation shows that turn-taking is mostly a universal phenomenon [37]. However, there are still measurable cultural differences in turn-taking.

Another interesting approach related to interruptions in was proposed by Cafaro et al. [38]. This work is based on the adopted version of Argyle's two-dimensional model of interpersonal attitudes [39] due to its solidity and simplicity. In this empirical study the authors investigate whether different interruption types and strategies in agent-agent interaction have influence on perceived agents' engagement, involvement, and interpersonal attitude. The authors found out that types of interruption have more influence on the perceived attitudes of agents than the strategies. Despite the fact that this study shows no correlations between culture and interruptions, it proposes an explicit structure of types and strategies of interruptions, so these correlations can be found in an additional user study.

2.4 Summary

In this Chapter we reviewed theoretical studies in a varying degree related to our research. Among the presented concepts of culture, we chose Hofstede's approach of cultural dimensions for its robustness and structure [26, 27]. Then we introduced theory of mind that is indirectly useful for our work [36]. Finally, we reviewed several studies related to turn-taking and interruptions in communication.

As for the latter, the studies mentioned above show that turn-taking and interruptions, as many other aspects of communication, are culture-dependent processes. Furthermore, an empirical study by Cafaro et al. [38] proved to be very useful for determining the dependencies between culture and interruptions in conversations.

3 Related Work

In this Chapter we provide a survey of related work. First, in section 3.1 we describe a universal multi-party dialogue system, which our work extends. Second, in section 3.2 we depict several previous methods of modelling culture in intelligent virtual characters, which are classified as external and internal models of culture. Finally, in section 3.3 we enclose this chapter with a short summary.

3.1 Multi-party Virtual Dialogue System

Multi-party virtual dialogue system is a framework for conversation between human and multiple virtual characters, and it aims for realistic multi-modal human-machine communication [29]. The key advantage of the system is its universality. It is possible to model various realistic conversation scenarios, real time or turn-based.

In its current state, the system involves two main components. First, a communication management system that supports multi-party, multi-modal scenarios. The purpose of this component is to facilitate the data transmission between actors. The second component is an agency system that controls agents' conversation behavior. [Fig. 3.1] illustrates the overall structure of the dialogue system.

A communication management system controls the flow of the virtual conversation and holds the structure of predefined scenario. It serves as a connection between agents, transmits the data, and relies on the agents' conversation graphs.

An agency system serves as an agent's 'brain' and is based on the idea that actors during conversation follow a routine process that guides their behavior through time. This routine is iterative and consists of three major steps: perception, deliberation, and action. In the first step, an agent reviews all recent (realized dialogue moves) and current activities (current dialogue moves) and labels other actors in the conversation as active or passive. In the second step, these recently-occurring changes are determined and apply their effects on the agent's internal state. Then, based on its new internal state and predefined expectations, an agent selects its target dialogue move. In the final step, based on predefined turn-taking rules, an agent makes a decision on whether to perform an action, proceed with the action selected in the

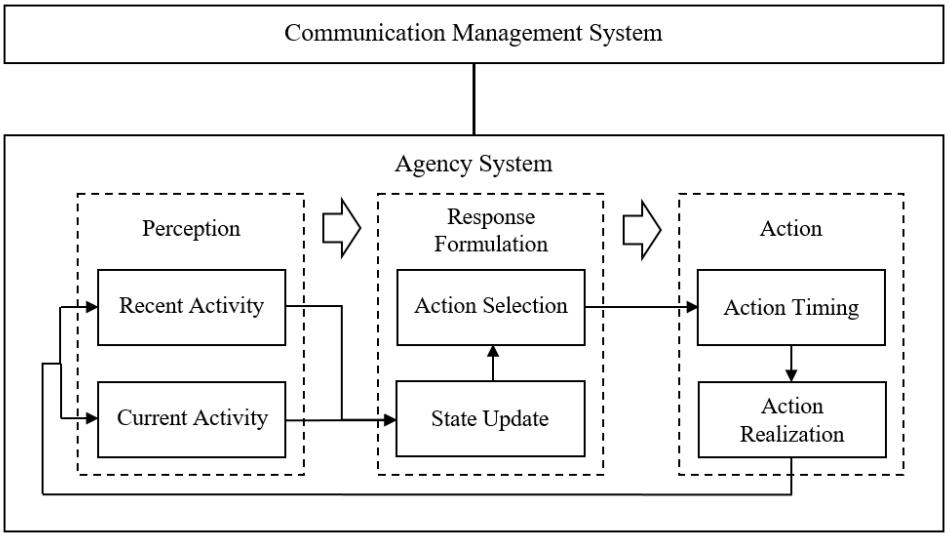

Fig. 3.1. The overall structure of the universal virtual dialogue system [29].

previous iteration, or to stop its current action. The other actors then perceive the resulting behavior, and the new iteration begins.

Moreover, the system holds a conversation graph that reflects the structure of the conversation. The structure of the conversation graph can be divided into several meaningful segments of conversation (for example: greetings, subject discussion, goodbyes), in which conversation states are represented in form of nodes. Each node holds the turn-taking rules and expectations, on which the transition to the next state depends. Conversation graph, expectations and turn-taking rules are the components of the dialogue system that are of the greatest interest to us.

3.2 Existing Cultural Models for Virtual Dialogue Systems

Due to the fact that human culture is very diverse phenomenon and there are so many differences between cultures, the ways of modelling culture in virtual characters are numerous and varied as well. The main question regarding cultural models is how such models should be implemented and therefore perceived by the observer and virtual characters. Thereby, there are two main types of cultural models: external and internal [9].

External cultural models are intended to cause an observer of the virtual character to see its behavior in accordance with that culture. However, with external cultural model only, a virtual character have no idea how to interpret its own and the behavior

of others. Internal cultural models imply the presence of a representation of culture in some form of rules for an agent to make decisions about its own and others behaviors. However, an observer might not recognize the behavior as appropriate for a specific culture. A perfect cultural model for virtual characters could include both external and internal models. Throughout this document, we call such approaches hybrid models and we aim to establish one.

Furthermore, recent research shows the importance of agent architectures to be modular, in which culture is an element that can be modified or replaced with another culture [40]. The research indicate that an agent could have the following components (but not necessarily all of them): internal representation (emotional model, personality), external representation (visual appearance, non-verbal communication, gestures), and independent cultural dimensions model. Internal elements drive the external representation of the agent, as it is difficult to recognize the differences between agents' cultures during interactions only from the computational aspect.

In addition, Hofstede and Pedersen introduced the notion of synthetic cultures [41]. The original concept of synthetic cultures belongs to Paul Pedersen and Ivey E. Allen [42]. Synthetic cultures are simplified forms of real national cultures that show only a single aspect of social behavior. Synthetic cultures do not describe the interdependency between dimensions of culture; they represent extreme manifestations (stereotypes) of the value orientations at the end of the dimensions of culture. Hofstede outlined ten profiles of synthetic culture based on five dimensions (a pair of extreme opposite cases for each dimension) and used them in the role-playing simulation game [43]. In the course of our research, we consider synthetic cultures as theoretical notions that can be used in our computational model.

3.2.1 External Cultural Models

Culture can be expressed externally in several ways. The easiest and most effective way to express culture in virtual characters is their actual virtual physical appearance and a set of gestures they can adopt to convey an appropriate cultural behavior. Such physical appearance attributes could be skin color, clothing and hair styles, and static facial expressions. Moreover, these visual exterior characteristics can be abstracted in a way to represent more stereotypical cultural manifestations. In addition, virtual environments can also have their impact on the perception of culture. By virtue of the current level of computing technologies, we are able to model and simulate highly detailed and plausible virtual characters and environments [10, 11, 12].

Other ways to express culture externally are proxemics, gaze and facial expressions. Some external cultural models have different culture-specific normal distances to

represent proxemics, as well as other factors such as group cohesion and noise level [13]. A culture related gaze model exist that relies on speech act, speaker type and function, and culture group to set probabilities for gaze targets [14]. In addition, a recent research shows that facial expressions of emotions are not culturally universal [15].

A model for simulating cultural differences in the conversational behavior of virtual characters is a significant approach with respect to turn-taking [16]. The model can express cultural differences in three conversational aspects: proxemics, gaze, and overlap in turn-taking, which is more important for our research. Each of the aspects has its own parameters that can be tweaked to better match a specific culture. Turn-taking specifically is based on Gaussian distribution with parameters that can be culturally defined.

Another important external approach is the intercultural training system Traveller. In Traveller, users interact with intelligent embodied virtual characters with cultural-specific behavior [44]. This system is designed in a form of visual narrative game and is based on an intercultural training theoretical framework. In Traveller, the behavior of the user determines the response of the characters during critical incidents. If the user acts appropriately, the interactions with the characters go smooth, probably with a few misunderstandings. Otherwise, conflicts may arise. The user perceives the characters behaving according to their culture. With properly adjusted cultural-specific parameters, this is a great external cultural approach.

Traveller is built on top of a well-established and flexible architecture for emotional agents – FatiMA Modular [45]. Originally, the purpose of this architecture was to use emotions and personality to influence the agent's behavior. The architecture was extended with several components in order to simulate particular cultural behavior of agents.

3.2.2 Internal Cultural Models

Internal models of culture for virtual agents are more comprehensive and complex in development. The core idea of internal models is to make agents so that they can produce and interpret behaviors, make inferences and decisions with respect to culture. In other words, internal models are cognitive agent models of culture. Below we examine and compare the most interesting of such models.

SID Model. Recently, the work on Traveller was expanded with an internal approach. The authors introduced a novel agent model, named the Social Importance Dynamics (SID) model, which is an extension to the fundamental BDI paradigm [17, 18]. The

aim of the SID model is to increase the social intelligence of regular BDI agents with respect to Hofstede's cultural dimensions, in particular – individualism. Additionally, the model is based on the notion of status from Kemper's theory, which the authors refer to as social importance (SI).

Based on the well-known BDI model, SID model uses its core concepts. Beliefs represent the knowledge of the agent about the environment, itself and other agents. In SID, each agent has a set of its own beliefs obtained through its sensors and a set of beliefs about the beliefs of other agents. Desires represent the agent's motivations, a set of abstract goals it wants to achieve. Each goal has a unique identifier, a list of preconditions, a list of success conditions, and the value that indicates the importance of that goal for an agent. Intentions represent a plan of actions the agent should perform to achieve a specific goal's success conditions.

When the aforementioned concepts are set, they are then processed in each agent's execution cycle through three major steps: beliefs revision, option generation, and action selection. During the first step, an agent updates its own beliefs and its beliefs about other agents' beliefs, including any new agents that appeared in the environment. In the second step, a list of options is generated that lists an agent's achievable abstract goals with regard to updated beliefs. Then, all generated options are filtered based on their expected utility (i.e. importance of success). During the third step, an agent composes a plan of achieving the success conditions of the goal with highest utility, and then decides the next action to perform. A plan must only contain actions, with which an agent has enough SI to perform. Otherwise, the action is dropped and a new plan is composed.

Each of the three steps are influenced by three SI components. Beliefs revision step is affected by SI Attribution Rules that determine an agent's relationship with others, i.e. the amount of SI that another agent gains or loses. Option generation is affected by SI Conferrals that defines an agent's motivation to perform an action, i.e. the amount of SI conferred by a particular action. Action selection is affected by SI Claims that determine the appropriateness of a plan based in a particular socio-cultural context, i.e. the amount of SI a particular action is claiming.

Finally, the culture is modelled based on the relation between SI and cultural influence (cultural dimensions). This is done through the following equation [18]:

$$V_{modified} = V_{initial} + |V_{initial}| * M * \frac{Score(D)}{100}$$

The value V corresponds to the associated SI value that is indicated in a particular SI attribution rule, SI claim, or SI conferral, M is a positive or negative multiplier, and $Score(D)$ denotes cultural dimension (individualism) score, assigned to an agent.

CAB Model. The second internal approach on the list is the Culturally-Affected Behavior (CAB) task-oriented model that aims to define a language for encoding ethnographic data in order to capture cultural knowledge and use it to affect human behavior models [19]. The model is based on D'Andrade's schema theory and theory of mind, and the authors define culture as a combination of the following aspects: shared appearance, shared external behavior, and shared internal knowledge and reasoning. However, their approach is focused on internal knowledge and reasoning aspects, which refer to socio-cultural norms and the mental representation of other people and cultures (biases and stereotypes). This approach is incorporated with a Bilateral Negotiation (BiLAT) training application that provides a game-like simulation, in which participants can practice their negotiation skills in a virtual environment on real world scenarios [46]. In this application, a participant interacts with an agent by using actions from four categories: say, ask, give, and do. To model negotiation behavior, the application uses PsychSim social simulation tool that employs a formal decision-theoretic approach using recursive models [47].

This approach explicitly models socio-cultural norms in a task model, which consists of domain states and socio-cultural network. The latter in turn comprises socio-cultural states and tasks. Domain states are focused on the external factors of a context (other agents and virtual environment), while socio-cultural states are focused on internal factors of an agent. In the aggregate, both domain and socio-cultural states represent a model of human behavior that reflects the state of the world. A huge advantage of this approach is its modularity: a specific cultural behavior may be applied by simply swapping socio-cultural network with another one, while domain human behavior remain the same.

In socio-cultural network task that may be performed by the agent are connected to states with association lines, which indicate the degree of effect, positive or negative, from a task to a state. Each socio-cultural state has a current utility and intrinsic utility values. Intrinsic utility values represent the shared importance that members of a particular culture on a socio-cultural norm weighted in relation to other norms. Values of intrinsic utility can be negative, which means that a state is undesirable by the members of a culture. A current utility value strongly depends on an intrinsic utility value, and if it becomes negative, then the meaning of a state becomes opposite.

The utility value can be changed by an effect, and if there are multiple effects on a single state, all of them are combined: the current utility value is summed with the

change in utility value caused by each effect. The change in the utility value after the task execution is calculated with the following equation [19]:

$$\Delta u = s_{eff} * d_{eff} * u_{int}$$

In the equation, s_{eff} denotes the sign of the effect (+/-), d_{eff} is the degree of effect, and u_{int} is the intrinsic utility value of the state.

The task, its effect on the state, and the state itself in combination model a rule of D'Andrade's constitutive rules approach. The overall task model evaluates the constitutive rules by assigning utility values and degrees of effects to the states, which allows this cultural model to reflect the fact some norms and effects are more important in a particular culture than others.

Extended ABMP Model. The next approach is an agent-based model for bargaining in the virtual trading context, in which the agents follow their strategies to maximize gain and minimize risks [20]. The agents also behave in culturally appropriate manner based on Hofstede's five dimensions: power distance, individualism, masculinity, uncertainty avoidance and long-term orientation. This model is an extension to the Agent-Based Market Place (ABMP) negotiation model [48]. The aim of this work is to simulate realistic human negotiation behavior, instead of just optimizing rational decision-making. The simulation environment was implemented in agent-based simulation framework – Cormas [49].

The original ABMP process is an exchange of bids, in which the agents must come to a trading compromise. Starting with a bid by one of the agents, the second agent evaluates the bid with a utility function, which combines the weight factors that indicate the agent's preferences. After evaluation, the second agent prepares its bid as a concession to the previous one. The utility weight factors are negotiation speed that an agent uses to determine the target utility of the next bid, concession factor that defines the minimum utility acceptable by an agent, acceptable utility gap that indicates the difference between an agent's utility and the last partner's bid utility and for which an agent can accept the partner's bid, and impatience that denotes the probability with which an agent can terminate the negotiation.

The culture is modelled by each of five cultural dimensions affecting the utility weight factors and ABMP parameters: they can be increased or decreased along with each dimension. For example, if a partner is from the same group, the concession factor increases and impatience decreases, while for an out-group partner these utility weight factors remain the same.

MARV Model. The last approach considered in this Chapter is Multi-Attribute Relational Value (MARV) model of decision-making for intelligent virtual agents to make a variety of decisions in a social setting [21]. The model takes into account a number of different metrics for evaluating a given situation. An advantage of MARV is that it can model an agent that cares about different aspects: self-interest, collective interest, fairness, and other. This approach uses Hofstede's model of culture for cultural modeling of decision-making by assigning weights for different valuations, based on the scores of five dimensions (power distance, individualism, masculinity, uncertainty avoidance and long-term orientation). To test the model, the authors applied it to the Ultimatum game in a virtual environment with human-like virtual characters.

MARV as a multi-attribute model consists of multiple metrics, several of which are more complex, functions of more primitive metrics. Each individual agent has a vector of weights that indicates the relative importance of metrics' valuations. Some of these metrics indicate personal and cultural aspects, such as self-interest, collective interest and uncertainty. For each agent's choice, the total value is the sum of weights (W_j) and values (V_j) of n valuation, as the following equation depicts [21]:

$$Value(Choice_i) = \sum_{j=1}^{n}(W_j * V_j(Choice_i))$$

In order to evaluate their approach, the authors provided two studies, each of which included two experiments: agents versus humans, and agents versus agents. In the first study, the authors manually configured the weights for different valuations, based on meanings of Hofstede's dimensions and their scores. They therefore generalized MARV model in such a way that the elements of the weight vector are divided into one component per dimension, and thus the model is considered as an overall matrix of n valuations of all five dimensions. The following equation indicates this generalized model [21]:

$$Value(Choice_i) = \sum_{j=1}^{n}((\prod_{d=IDV}^{UAI} W_{j,d}) * V_j(Choice_i))$$

In the second study, the authors applied an inverse reinforcement learning (IRL) technique that used a received data of an observed human behavior. Unlike reinforcement learning (RL) technique, for which the reward function should be defined, the agent with IRL learns the reward function from a set of data of interactions between the agent and the environment.

The results of the conducted experiments showed that the weights learned from IRL surpass the handcrafted weights of the first study. This shows the usefulness of IRL technique for decision-making mechanism in negotiation.

3.2.3 Discussion

Multi-party virtual dialogue system is a great multi-modal framework for virtual conversations. Its architecture is very comprehensive and promising. Due to the modular approach, it can be relatively easily extended with new specific modules if one wishes to expand the functionality with new aspects of communication.

As for the existing models of culture, internal ones are of the most interest to us. It appears to be a challenge to compare the aforementioned internal cultural models. We managed compare these models based on background cultural theory, AI technique that was used to model agents, application in which the model was implemented, the aspects of culture that were investigated with the model, and the possession of modelling multi-party scenarios and turn-taking. [Tab. 3.1] depicts an overview comparison of these internal methods.

CAB is a very robust model; its application is real-time and considers other modules useful for human-machine communication, such as speech recognition. However, it is not capable of multi-party conversation and it has a background theory different to which we consider in our research.

SID, extended ABMP and MARV models are all based on Hofstede's cultural dimensions approach, on which we are focused in our research. While MARV and ABMP consider five dimensions, SID model in its current form considers only individualism dimension. In addition, all of the considered models are modular, which is a huge advantage.

As it may be seen, none of the models possesses turn-taking, and moreover, applications of most of them are not even real-time. In addition, none of the models evaluate indulgence dimension (IVR). It seems that taking one of these solid models as the basis of our work, extending it with culturally sensitive turn-taking component and incorporating it into an existing dialogue system could be one of the possible solutions of our research. However, in our case we decided to develop our own simple but yet valuable computational model. Models discussed here serve mostly as an inspiration for our approach.

Model	**SID** (2015)	**CAB** (2008)	**Extended ABMP** (2012)	**MARV** (2014)
Cultural Theory	Kemper's theory (status) and Hofstede's dimensions (IDV)	D'Andrade's cognitive schemas theory and theory of mind	Hofstede's dimensions (PDI, IDV, MAS, UAI, LTO)	Hofstede's dimensions (PDI, IDV, MAS, UAI, LTO)
AI Technique	Extended BDI	Socio-cultural network	Generic Agent Model	Inverse RL
Application	Traveller	ELECT BiLAT	Cormas	Ultimatum game
Cultural Aspect	Cultural influence on collectivistic behavior in conversation	Socio-cultural values and attitudes of a culture	Cultural influence on decision-making in bargaining negotiation	Cultural influence on decision-making
Multi-party	▲	▽	▲	▲
Turn-taking	▽	▽	▽	▽

Tab. 3.1. Overview comparison of related internal cultural models.

3.3 Summary

To conclude, in this Chapter we reviewed related work. First, we described the framework, namely multi-party virtual dialogue system, which we expand with our cultural model. Second, we discussed several existing cultural models and grouped them as external and internal approaches. As for the latter, all of the discussed internal cultural models are robust but lack some of the core aspects of conversation. Therefore, in our cultural model these aspects should be considered.

4 The Model of Culture

In this Chapter we describe the methodology of our research. In section 3.1, we show how we came from a theoretical concept of culture proposed by Hofstede to our computational approach. In section 3.2, we propose the requirements for our model that helped in its design formulation. In section 3.3, we describe the design of our cultural model in the form of additional module for the dialogue system. Finally, in section 3.4 we discuss interruptions and why we eventually decided to avoid them in the current version of our model.

4.1 From Concept to Model

Since we established that our model is based on Hofstede's cultural concept, it is important to show how we applied it to our model. It is also important to note that, as our research question states, our work is based only Hofstede's cultural dimensions with no additions of other concepts or any existing computational approaches described above. In other words, for us it is interesting to see whether it is possible to develop a decent cultural model, which is based on a single but yet comprehensive cultural concept.

Previously we indicated that in case of our experiment we took only four of six cultural dimensions (power distance, collectivism, long-term orientation, indulgence) to construct culturally affected behavior. However, our model is not limited to those four dimensions, the parameters of the remaining two dimensions are already in the system and it is still possible to add them in a form of cultural norms (which are described later in this Chapter). The reason why we took only four of those lies in the evaluation design of our model. In the scenarios we made to verify our model the remaining two dimensions simply have no particular influence on similar situations in real life.

To draw a line, our work is based on four of six Hofstede's dimensions, but it is not limited to those only. In the following sections we propose the requirements to our approach and how cultural dimensions are actually applied to the model.

4.2 Requirements

Taking into account the capabilities of the multi-party virtual dialogue system and what the other existing cultural models are good and bad for, we determine a list of requirements for our cultural model. As the virtual dialogue system consists of several interdependent modules, our model should be developed in the same manner. Ultimately, it should be a separate cultural module that is flexible and affects some of the existing components of the dialogue system, such as conversation graph, action selection and action timing. In addition, it should be possible to easily disconnect cultural module, in case the architect of the virtual conversation desires it to be culturally independent.

Therefore, the cultural module should satisfy the following requirements:

- The cultural module must be modular, so that it uniformly fits into the dialogue system and consistently co-exist with the other modules of the system, but can be easily disconnected from it.

- The cultural module must be flexible, so that its parameters can be quickly and easily modified to receive a different cultural behavior.

- The cultural module must affect only the following components of the dialogue system: conversation graph, action selection and action timing.

The purpose of these requirements is to specify constraints for our approach that help us to clearly see the final result and whether it is holistic enough. In the next section we describe the design of our module and how we met the proposed requirements.

4.3 Cultural Module Design

In the previous two sections we proposed the influence of cultural concept on our model and three requirements that we must take into account in our module's development. In this section we sort of combine these propositions and define the structure of the module, followed by the description of its components.

In a nutshell, the module is pretty straightforward and consists of three main components. The first component contains logic of cultural virtual characters and a pair of structures that keep parameters, which define characters' culture. The next component is the most interesting one as it holds cultural norms, which are

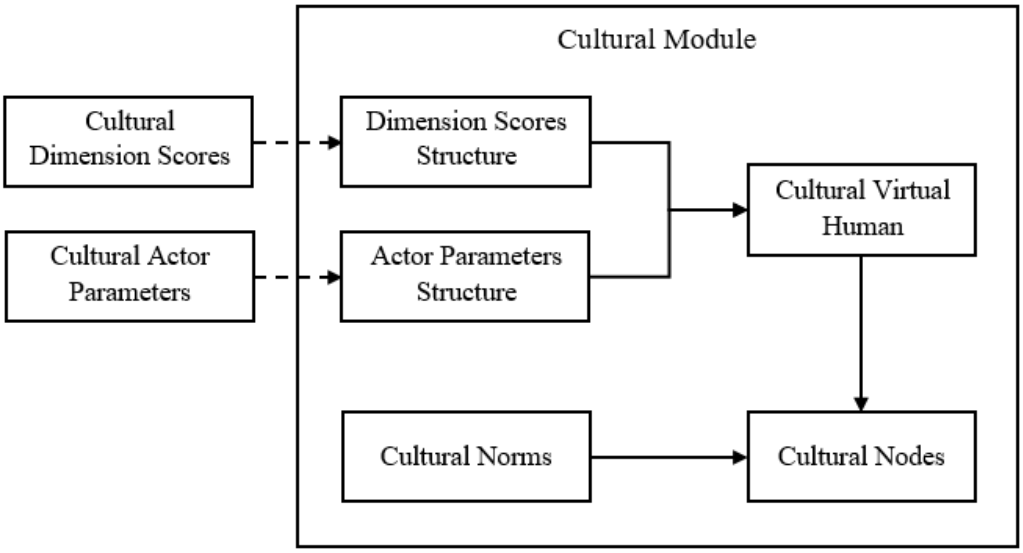

Fig. 4.1. Simplified structure of the cultural module.

used to affect dialogue moves based on cultural parameters through the following component. The last component is a new type of nodes for the dialogue system's conversation graph, namely the cultural node, which allows the execution of a dialogue move if related to it cultural norms are abided. [Fig. 4.1] illustrates the simplified cultural module structure.

The *first proposed requirement is satisfied* by definition, as our cultural module is implemented as a separate module but has the structure, similar to other modules of the system. In addition, some components of the dialogue system affect our module's components but not other way around, so it can be simply ignored by the architect if a non-cultural scenario is needed.

4.3.1 Cultural Virtual Human and Cultural Parameters

First, this component holds two sets of parameters involved in the module: six cultural dimensions parameters and five cultural actor parameters. These sets are stored in two separate structures. The first six parameters hold the scores of PDI, IDV, MAS, UAI, LTO and IVR dimensions (however, as it is mentioned above, only four dimension parameters are involved in our evaluation part). Dimension scores are the same as those introduced in Hofstede's work [27], but in our model these scores are automatically reassigned into two types (high and low) to represent two opposing extremes (stereotypes).

The second set consists of five parameters, three of which are related to particular dimensions. These parameters are as follows:

- **Actor's Name**. This parameter defines the name of the actor, which is used in virtual conversations.

- **Actor's Culture**. This parameter defines the nationality (culture) of the actor. Based on this parameter, the system takes dimension scores related to that culture, which form cultural view of the actor and its decision making process.

- **Power Status**. This parameter is related to PDI dimension and defines the status of the actor.

- **Group Identifier**. This parameter is related to IDV dimension and defines the group, which the actor is belongs to.

- **Gender Identifier**. This parameter is related to MAS dimension and defines gender of the actor. However, as previously mentioned, this parameter has no practical use in the evaluation of our cultural module.

The first parameter is just a name of the actor. The second one defines the actor's culture, so the appropriate dimension scores from the first set of parameters can be loaded. The third parameter is the actor's status; actors of high PDI cultures behave differently depending on their own and other actors' status. The fourth parameter is group id of the actor; actors of low IDV cultures behave differently depending on whether other actors are from the same group or not. The fifth and the last parameter is gender id of the actor; actors of high MAS cultures behave differently depending on their own and other actors' gender.

In combination, both sets of parameters define a specific culture for each actor. Depending on the specified parameters, the actors perform certain dialogue moves predefined in the conversation graph. Thus, if the actor's parameters are changed, its behavior also changes. In addition, it is also possible to make a virtual dialogue between actors that represent different cultures.

One of the advantages of our module (and this component in particular) is that both sets of parameters can be either loaded from predefined XML files or hard coded into the module directly (depending on scenario architect's preferences). The first option allows quick alteration of parameters in order to receive a different behavior of virtual characters, which proofs the flexibility of our module.

The purpose of cultural virtual human is simple. It loads parameters described in this section and assigns them to a specified actor. In its basis this component is

very similar to the one already implemented into the dialogue system [29]. Defined cultural actors and their parameters are used in the next two components.

4.3.2 Cultural Norms

Cultural norms are the rules, satisfaction of which allow the actor to perform a particular dialogue move. Each norm represents a set of conditions that verify actors' cultural parameters. These conditions can be very different: from a simple check of the source actor's parameter that is going to perform the dialogue move, to a complex multi parameters verification of source, target, or both actors. If all the conditions in the norm are satisfied, the next developed component, namely the cultural node, allows an actor to perform a dialogue move. Moreover, cultural norms can be used to simulate interruptions.

Abstractly, these cultural norms are simplified representations of real norms, those that are laid in our minds throughout life. Thereby, it is important to understand how Hofstede's dimensions influence these norms and how those are implemented into our module. In the previous section we slightly raised this matter, so it is not superfluous to dive a bit deeper into details. Therefore, we propose the following aspects of behavior that are influenced by cultural norms:

- PDI-dependent cultural norms influence **interpersonal attitude**: *high PDI* actors always pay respect to and treat others with higher status as superiors; *low PDI* actors treat others as equals regardless of their status.

- IDV-dependent cultural norms influence **group behavior**: *high IDV* actors behave individually and more often use "I", "me", "my" pronouns; *low IDV* actors behave as a group, in which an actor with the highest status is considered a leader that converse with out-group actors and more often uses "we", "us", "our" pronouns.

- LTO-dependent cultural norms influence **conversation duration**: *high LTO* actors converse shorter and do not participate in non subject discussions; *low LTO* actors converse longer and frequently participate in non-subject discussions.

- IVR-dependent cultural norms influence **interpersonal attitude**: *high IVR* actors behave more positive, friendly, personal and relaxed with others; *low IVR* actors behave more neutral, unfriendly and impersonal and less relaxed with others.

In its current state, the module holds 8 norms that are based on the four chosen dimensions (PDI, IDV, LTO, IVR). However, the list of those can be easily extended with new cultural norms based on any of the six dimensions if needed. This is another proof of module's flexibility, so the *the second proposed requirement can be considered satisfied*.

4.3.3 Cultural Node

Cultural node is an additional internal node to the dialogue system and it defines whether a specific dialogue move can be performed based on the assigned to it cultural norms. Specifically, as a condition it takes one or more predefined cultural norms and two actors (not necessarily the one who is going to perform this dialogue move and the one to whom the first actor addresses; both actors could be any of those who participate in the multi-party dialogue). As stated above, once all the conditions in the assigned norms are satisfied, the source actor of the attached dialogue move executes it.

As it may be seen, the *third proposed requirement is also satisfied*. Cultural virtual human and cultural norms components of the module affect action selection and timing components, while cultural node extends capabilities of conversation graph of the dialogue system.

4.4 Exclusion of Interruptions

Initially, interruptions were considered to be the part of our approach. In the second Chapter we introduced two related works that could help us to implement interruptions in our module. Moreover, in our first user study, which is described in Chapter 6, we found some correlations between culture and interruptions. And needless to say, it is probably clear for everyone that in different countries it is normal to interrupt someone with lower status during conversation, if status matters at all in that culture. However, we decided not to implement interruptions in the current version of our module for the following reasons.

The first reason lies in the fact that in the scenarios we took for evaluation of our approach the agents have the roles that do not interrupt the participant's role. They pay far more respect and interruption is simply impolite in most cases for most cultures.

The second reason is that participants simply will not interrupt virtual characters, even if it is normal for their roles to do so. It is because participants will be interested

in the context and will not skip utterances of any of the actors to have a complete image of the scenario. Moreover, even if the participants would interrupt virtual characters, they could simply miss cultural features that help them to recognize a specific culture. In other words, interruptions could do more harm than good for the evaluation of our approach in its current state.

Nevertheless, the virtual dialogue system still contains functionality for interruptions. It is still possible to add them based on particular cultural norms. But in the current state of our model it would not give much, while a considerable amount of additional work should be done. Cultural influence on interruptions and overlapping is a vast topic by itself, so it should be considered for the future work.

4.5 Summary

In this Chapter we described the way we came from the theoretical concept of culture to the formulation of our computational model. First, we described the integration of Hofstede's approach into the model. Second, we proposed the requirements for our model, so that it fits perfectly into the dialogue system for which it is made for. Then, we introduced the cultural module's design with descriptions of its major components. Finally, we discussed the reasons for the absence of interruptions in the current version of our cultural model. In the next Chapter we immerse into more technical details of the module's implementation.

5 Implementation

In this Chapter we depict the implementation of our cultural module. In section 4.1, we describe the architecture of the module, class by class, how it interacts with the rest of the dialogue system, and what software was used during its development. We then conclude this Chapter with a short summary.

5.1 Architecture

As it is mentioned previously, our cultural module consists of three main components. In this section we discuss technical aspects of these components and our module as a whole. The module was first designed as a set of UML diagrams in NClass and developed using Microsoft Visual Studio 2015 afterwards. Several classes and interfaces shown in this section are part of the dialogue system [29]. Appendix A holds an instruction manual for our cultural module.

5.1.1 ICulturalActor, CulturalVirtualHuman and Cultural Structures

As its name suggests, `CulturalVirtualHuman` is a class that forms the core of cultural virtual characters. This subsection provides a description of how it is formed. The first UML diagram illustrates this class and how it is related to other subcomponents of the dialogue system and cultural module in particular [Fig. 5.1].

The first important subcomponent implemented here for our module is `ICulturalActor` interface. `ICulturalActor` extends two other core interfaces of the dialogue system, namely `IManagedActor` and `IActor`. Other than that, it also adds two new properties on top of them: `ActorCulture` and `ActorCulturalParams`. These methods are invented to get dimensions scores and cultural actor parameters from two structures described below.

`Culture` and `CulturalActorParams` are the structures that hold cultural parameters described in Chapter 3. These parameters can be loaded through either predefined XML files or hard coded into the structures directly. The advantage of using XML file is that if the scenario is already defined and compiled as a set of DLL files, it is possible to change actors' parameters and thus change their behavior. In other words,

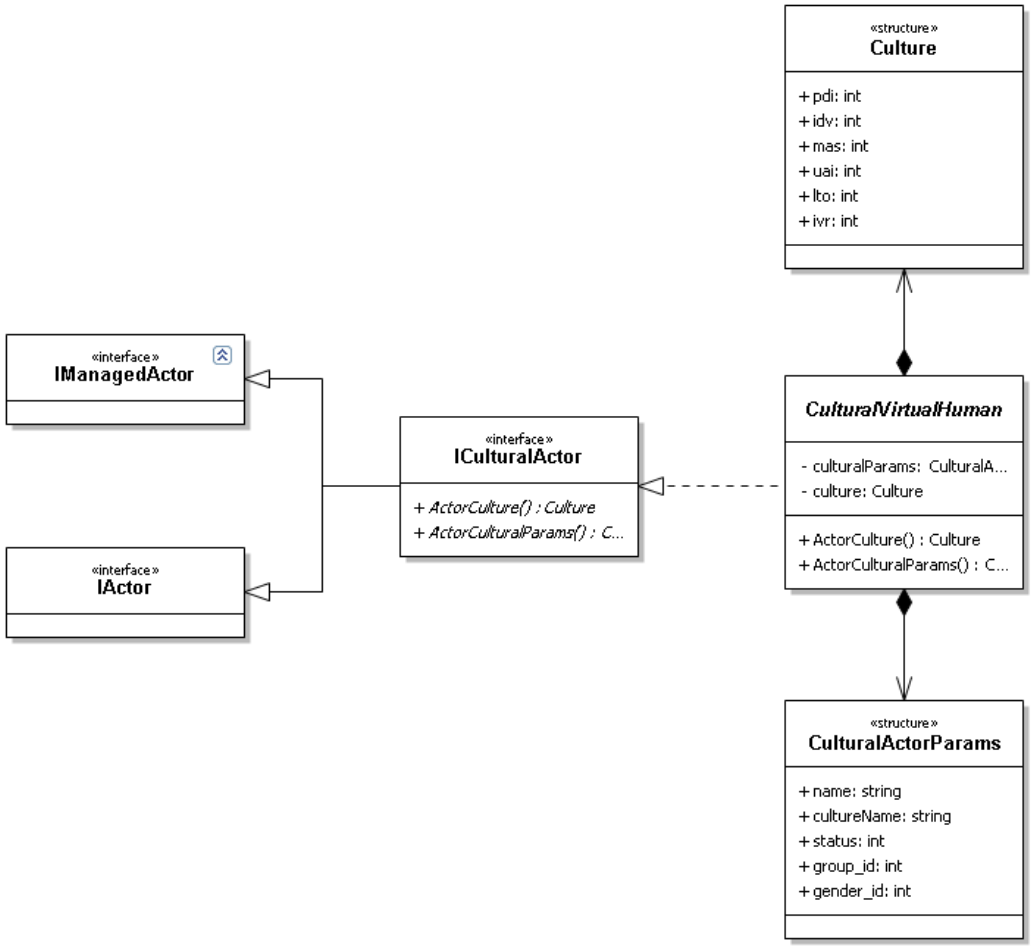

Fig. 5.1. Nclass simplified UML diagram of `CulturalVirtualHuman` and other subcomponents.

there is no need to re-compile the whole scenario if it is necessary to get the behavior of actors corresponding to another culture. Such an approach to get parameters is relatively easy to implement, whereas it provides the necessary flexibility.

`CulturalVirtualHuman` itself is an implementation of `ICulturalActor` interface. This class is realized in our module very similar to `VirtualHuman` class in the dialogue system, with only addition of cultural parameters implementation. `ActorCulture` and `ActorCulturalParams` public methods get parameters directly from structures through implemented in `CulturalVirtualHuman` constructor.

5.1.2 ICulturalNorms

`ICulturalNorms` is a completely new interface for all cultural norms. The implementations of `ICulturalNorms` should define `IsMovePossible` boolean method that

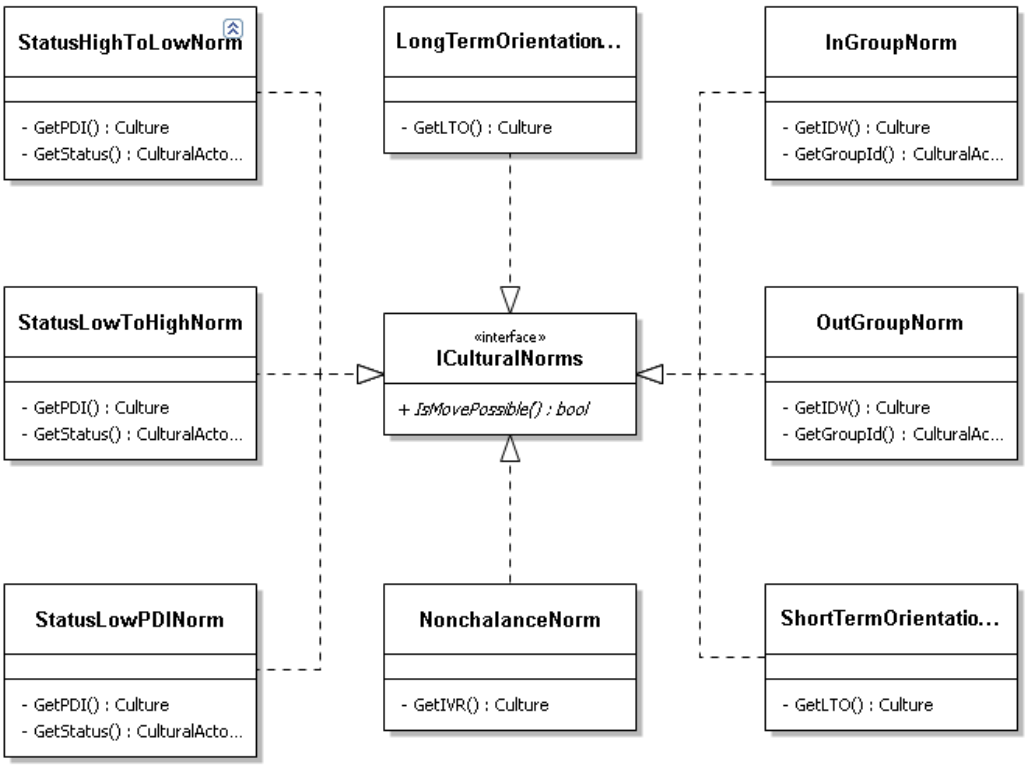

Fig. 5.2. Nclass simplified UML diagram of `ICulturalNorms` and eight implemented cultural norms.

permits actors to realize dialogue moves. The second UML diagram depicts this interface and eight implemented cultural norms [Fig. 5.2].

Each cultural norm is an implementation of `ICulturalNorms` interface, which gets parameters of both source and target actors based on cultural context. In addition, there are also specific dimension score limit parameters (`pdiLimit = 50, idvLimit = 50, ltoLimit = 50, ivrLimit = 50`) to reassign cultural actors' scores into one of two types (high or low). The specifications of each implemented cultural norm are as follows:

StatusHighToLowNorm. PDI-dependent norm. For cultures, in which status is essential for social interaction, it is important to invent a rule that permits actors with high status to act as superiors and perform corresponding dialogue moves. Therefore, `StatusHighToLowNorm` verifies whether the dialogue move can be performed based on the following conditions:

```
source.ActorCulture.pdi ⩾ pdiLimit;

source.ActorCulturalParams.status ⩾ target.ActorCulturalParams.status.
```

StatusLowToHighNorm. PDI-dependent norm. Pro rata to the previous norm, in case of status-dependent societies, there must be another rule that permits actors with low status to behave accordingly and realize respective dialogue moves. `StatusLowToHighNorm` verifies whether the dialogue move can be performed based on the following conditions:

$$\texttt{source.ActorCulture.pdi} \geqslant \texttt{pdiLimit};$$

$$\texttt{source.ActorCulturalParams.status} < \texttt{target.ActorCulturalParams.status}.$$

StatusLowPDINorm. PDI-dependent norm. Contrariwise to previous two norms, there must be a rule for cultures that do not welcome the power of status, which permits actors to behave equally with others, no matter what is their social position. Thereby, `StatusLowPDINorm` verifies whether the dialogue move can be performed based on the following conditions:

$$\texttt{source.ActorCulture.pdi} < \texttt{pdiLimit}.$$

InGroupNorm. IDV-dependent norm. In collectivistic societies, behavior and attitude with members within different groups differ from behavior and attitude in relation to those who do not belong to these circles. Therefore, there must be a rule that permits actors to perform specific dialogue moves in conversation with other in-group actors. `InGroupNorm` verifies whether the dialogue move can be performed based on the following conditions:

$$\texttt{source.ActorCulture.idv} < \texttt{idvLimit};$$

$$\texttt{source.ActorCulturalParams.group\_id} == \texttt{target.ActorCulturalParams.group\_id}.$$

OutGroupNorm. IDV-dependent norm. On the contrary to the previous norm, in case of collectivistic cultures as well, there must be a rule that permits actors to converse accordingly with out-group actors. Therefore, `OutGroupNorm` verifies whether the dialogue move can be performed based on the following conditions:

$$\texttt{source.ActorCulture.idv} < \texttt{idvLimit};$$

$$\texttt{source.ActorCulturalParams.group\_id} \; ! = \texttt{target.ActorCulturalParams.group\_id}.$$

LongTermOrientationNorm. LTO-dependent norm. Some societies treat time with great trepidation. And as we develop scenarios and evaluate our module based this cultural dimension, it is essential to invent a simple rule that permits actors represent-

ing such cultures to behave accordingly. Hence, `LongTermOrientationNorm` verifies whether the dialogue move can be performed based on the following conditions:

`source.ActorCulture.lto` $\geqslant$ `ltoLimit`.

ShortTermOrientationNorm. LTO-dependent norm. Other cultures, in contrast to those mentioned in the previous paragraph, treat time much more leniently. A simple rule that permits short-term oriented actors to behave in accordance to the norms of these cultures must exist in our module. `ShortTermOrientationNorm` verifies whether the dialogue move can be performed based on the following conditions:

`source.ActorCulture.lto` $<$ `ltoLimit`.

NonchalanceNorm. IVR-dependent norm. For people of some societies it is absolutely common to have more extroverted personalities and interact with others with positive and friendly attitude. People of other societies are more restraint: they are more pessimistic and cynical, restrict themselves with numerous moral constraints, and have less positive attitude with others. Thus, it is important to distinguish typical behavior for such cultures in a form of a simple rule. `NonchalanceNorm` Verifies whether the dialogue move can be performed based on the following conditions:

`source.ActorCulture.ivr` $\geqslant$ `ivrLimit`.

All the described above cultural norms were used in the scenarios for our evaluation process. Definitely, each of them is just a single or a combination of two simple, if-then conditions. However, is not it a simplified representation of the logic that is laid in our minds and which we use when to make decision on how to behave appropriately to our culture? It is similar, but in reality it is far more complex and we make large amount of such decisions in a second.

It is still possible to extend the list of those based on similar or different, more complex, conditions. In the current version of our model we just want to see whether such an approach can work properly at all. In case success, it can be expanded with numerous cultural norms for any culturally-dependent behavior ever possible.

5.1.3 Cultural (Node)

`Cultural` is a new type of internal nodes that evaluates a specific cultural norm and based on the results of this evaluation determines the consecutive flow of virtual conversation. Other than that, it is similar to `Conditional` node from the dialogue system. The last simplified UML diagram illustrates this new node [Fig. 5.3].

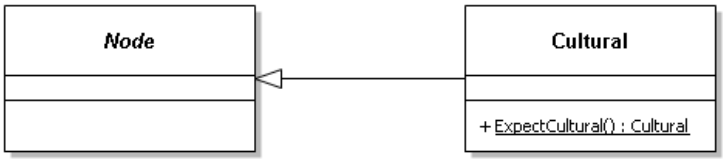

Fig. 5.3. Nclass simplified UML diagram of the `Cultural` node and its relation to `Node` class.

`Cultural` node inherits from `Node`, which is a base class for all types of nodes in the dialogue system. To use this node in conversation graph, our cultural module also holds a new type of expectations - `ExpectCultural`.

`ExpectCultural` is a new expectation (node) that gets a specific cultural norm, a body, and alternative body if needed. Cultural norm can be one of those introduced in the previous subsection, or any other implemented in the future. Body is another expectation, which is usually an `ExpectEvent` that realizes the dialogue move if all the conditions in specified cultural norm are satisfied (`true`). However, in case of more complicated dialogue moves that require two or more cultural norms (for example, an utterance based on power distance and individualism dimensions), it is possible to use two or more `ExpectCultural` nodes nested one inside the other (i.e. individualism cultural norm `ExpectCultural` node is a body of power distance cultural norm `ExpectCultural` node). Alternative body is used for another expectation that should be realized in case at least one of the conditions in specified cultural norm is unsatisfied (`false`).

5.2 Summary

In this Chapter we described a more technical side of our cultural module. First, we proposed the software we used to design and develop the module. Then, we went through all three components and their subcomponents. We also showed simplified structures of these components and their dependencies to other components of the dialogue system. In the next Chapter we the evaluation of our approach and its results.

6 Evaluation and Results

In this Chapter we discuss the evaluation part of our research. In section 6.1, we describe the design of our experiment, in particular its setup, the challenges that could arise, the number of participants, hypotheses and independent variables, and the questionnaire design. In section 6.2, we provide the results and analysis of our evaluation followed by the discussion. Finally, in section 6.3 we summarize this Chapter.

6.1 Experimental Design

This section contains the description of the experiments that we conduct in order to answer our research question. First, we describe our experiment setup and scenarios that are modelled with use of the virtual dialogue system and the cultural module in particular. Second, we mention several challenges that we could face during the evaluation process and how we try to avoid them. Next, we discuss the participants of our experiment. Then, we provide the description of independent variables and hypotheses, which help to evaluate the effectiveness of our approach. Finally, we describe the design and purpose of the questionnaire.

6.1.1 Setup and Scenarios

Our experiment setup is as follows. One human participant and two virtual characters engage in two sessions of conversations in two different scenarios, each of which with its own predefined context. For each session, all actors are culturally strengthened with parameters appropriate to a specific culture. However, participants do not know which culture is set in these scenarios. After each session, based on their experience, knowledge and perception, the participants indicate to which culture the behavior of the actors is more appropriate, how appropriate it is to their culture and their perception of several other, more specific aspects of conversation.

Our goal is to evaluate the effectiveness of the cultural module and see whether the participants of the experiment can recognize a specific culture in agents' behavior. Therefore, we choose three countries, cultural dimensions' scores of which are

sufficiently different. As mentioned previously, in our experiment these cultural dimensions are simplified to represent two opposing extremes (stereotypes).

For our user study we choose Mexico, the Netherlands and Russia. [Tab. 6.1] illustrates the scores of these countries in four chosen cultural dimensions (PDI, IDV, LTO, IVR). Mexico and Russia have high scores in PDI, which means that power status in these countries means a lot. The Netherlands, in contrary, scores quite low in PDI, which means that Dutch people mostly treat each other as equals regardless of their statuses. In IDV dimension Russia and Mexico are also close with low scores, so these countries are collectivistic. IDV scores of the Netherlands show that it is very individualistic. The LTO score of Mexico shows that this culture is short-term oriented, while the Netherlands and especially Russia are long-term oriented. Finally, the distributions of IVR scores is quite broad, but as the implementation of dimensions is simplified, we group Mexico and the Netherlands together as high IVR cultures, while Russia is low IVR culture.

Dimension	**Mexico**	**Netherlands**	**Russia**
PDI	▲ (81)	▽ (38)	▲ (93)
IDV	▽ (30)	▲ (80)	▽ (39)
LTO	▽ (24)	▲ (67)	▲ (81)
IVR	▲ (97)	▲ (68)	▽(20)

Tab. 6.1. Comparison of PDI, IDV, LTO and IVR scores for three selected countries [27].

The scenarios are typical doctor-patients and teacher-students conversations. In the first scenario, the participant takes role of the doctor (therapist), and the agents take roles of two patients (a couple). In the second scenario, the participant takes role of the teacher (supervisor), and the agents take roles of two students (group mates).

It is important to understand the features of conversations typical for a specific culture. In fact, dimension scores and key differences between societies proposed by Hofstede [27] are not enough to construct culturally appropriate scenarios. It is necessary to find key features of typical doctor-patients and teacher-students scenarios for three chosen cultures.

Characteristic	Mexico	Netherlands	Russia
Average Duration	**Long** (5-10 minutes)	**Medium** (3-5 minutes)	**Medium** (3-5 minutes)
Doctor's Attitude	**Positive**, friendly and personal	**Positive**, friendly and personal	**Neutral** and impersonal
Treatment	Patients treat doctors as **superiors**, but they actively supply the doctor with various information	Patients treat doctors as **equals**, but conversation is mostly controlled by the doctor	Patients treat doctors as **superiors** and conversation is mostly controlled by the doctor
Non-subject Discussion Commonness	**Very common** and can appear before, after or mid-subject discussion	**Not common** or extremely rare; strongly depends on personalities	**Not common** or extremely rare; strongly depends on personalities
Non-subject Discussion Average Duration	**Medium** (2-4 minutes)	**Short** (1-2 minutes)	**Short** (1-2 minutes)
Non-subject Discussion Typical Topics	Medicine related topics, non-medicine related topics, personal stories discussion	Medicine related topics, non-medicine related topics	Medicine related topics
Gratitude	Gratitude is a norm, but only immaterially	Gratitude is a norm, but only immaterially	Gratitude is a norm, also materially
Doctor's Interruptions	Any strategy/type of interruption	Any strategy/type of interruption	Any strategy/type of interruption
Patient's Interruptions	To ask for clarification	To ask for clarification, to express personal opinion (agree/disagree), to politely complete what is being said	To ask for clarification

Tab. 6.2. The characteristics of the doctor-patients scenario, typical to each of the three cultures.

Characteristic	Mexico	Netherlands	Russia
Average Duration	**Medium** (3-5 minutes)	**Medium** (3-5 minutes)	**Medium** (3-5 minutes)
Teacher's Attitude	**Neutral** and impersonal	**Positive**, friendly and personal	**Neutral** and impersonal
Treatment	Students treat teachers as **superiors** and conversation is controlled by the teacher	Students treat teachers as **equals**, but conversation is mostly controlled by the teacher	Students treat teachers as **superiors** and conversation is controlled by the teacher
Non-subject Discussion Commonness	**Not common** or extremely rare; strongly depends on personalities	**Common** but only once, before or after subject discussion	**Common** but only once, before or after subject discussion
Non-subject Discussion Average Duration	**Medium** (2-4 minutes)	**Long** (4+ minutes)	**Medium** (2-4 minutes)
Non-subject Discussion Typical Topics	Studies related topics, non-studies related topics	Studies related topics, non-studies related topics, personal stories discussion	Studies related topics, non-studies related topics, personal stories discussion
Teacher's Interruptions	Any strategy/type of interruption	Any strategy/type of interruption	Any strategy/type of interruption
Student's Interruptions	To ask for clarification	To ask for clarification, to express personal opinion (agree/disagree), to politely complete what is being said, to reflect awareness and summarize what is said	Any kind of interruptions are unacceptable

Tab. 6.3. The characteristics of the teacher-students scenario, typical to each of the three cultures.

In order to find these features of conversation, we conducted a pre-experiment user study, in which we ask the participants to describe what makes their culture remarkable in typical doctor-patients and teacher-students scenarios. The results of this user study helped us to determine the key characteristics of the scenarios for each of the three cultures. [Tab. 6.2, Tab. 6.3] reflect these results.

Initially, participants were supposed to take more associative roles, i.e. patients and students, so that they are better immersed in the scenarios. However, as it is already stated above, we decided to make it other way around. It is much easier to perceive and judge the behavior for the participants in such a setting, as they look at the virtual characters as if they look at the mirror. In other words, can better recognize culture in the behavior of virtual characters as if they were in their place.

It is very important to note that each of the scenarios is based on a specific set of cultural norms, which are related to particular dimensions. With the help of Hofstede's work and the pre-experiment user study conducted, we found out that in real-life scenarios influence and priority of dimensions are different (i.e. have different weights for various situations). For example, it became clear to us that power distance is much more important than indulgence in a typical teacher-students conversation. At the same time, in a typical doctor-patients scenario it is other way around. Thereby, our virtual scenarios are based on the following cultural norms (dimensions are listed in order of priority):

- Doctor-patients scenario: **IVR**, **LTO** and **PDI** related norms.
- Teacher-students scenario: **PDI** and **IDV** related norms.

While with the first scenario everything is pretty clear, the second one is interesting to evaluate in case of Mexican and Russian cultures. One could already notice that PDI and IDV dimension scores for both these cultures are nearly the same, which in our model makes actors with parameters set for Russian and Mexican cultures to behave the same way (i.e. produce the same dialogue moves). We assume that Mexican and Russian participants perceive their own culture in both cases.

The simulation of the scenarios described above is provided by an application that is built in Unity 5 Engine with dialogue system and cultural module implemented in form of external libraries. The scenes for both scenarios are made in such a way, so that interiors and virtual characters look appropriately for all participants. It is very important for our evaluation that participants do not pay attention to the visuals, but only make their judgements based on actors' dialogue moves.

Fig. 6.1. A doctor-patients scenario scene built in Unity 5: upper left speech bubble outputs the utterances of the first patient, upper right outputs for the second patient, and the bottom one outputs for the doctor.

Doctor-patients scenario scene is shown in [Fig. 6.1]. Each of the speech bubbles output dialogue lines of a particular actor; font colors also indicate who makes a dialogue move. Teacher-students scenario scene is made in a similar fashion, except that the interior has a look of a typical (university) teacher room. All the assets used in both scenes are free to use and were downloaded from Unity Asset Store and Mixamo [50, 51].

It is also important to mention that both dialogues are linear with a small addition of the 'illusion of choice'. At first, we wanted to make the scenarios with several dialogue branches, so that each participant could follow the path that they think is more appropriate to their culture and the virtual characters would response accordingly. However, due to the nature of our evaluation, and because such an approach would provide different experience for the participants, we decided to make these virtual conversations linear. Appendix B provides a couple of examples of dialogue snippets with different conditions for the scenarios.

6.1.2 Challenges

As culture is a very complex phenomenon by itself and yet conceals a lot of unknown, evaluation of the computational cultural model can bump into several obstacles. These obstacles (i.e. challenges) that we could face during the evaluation of our model must be identified in advance in order to avoid them.

The first challenge is the language barrier. Despite that our participants originate from three different countries, each of which has its own unique language, we put all of them on equal terms in our evaluation. In particular, the experiment is made in English. So the first problem here is the participants' knowledge and proficiency of English language. The participants from the same country but with different language competence can experience the same scenarios differently, and therefore - the results will suffer. As it is clearly impossible (and unnecessary) to adjust the scenarios for each participants, we simply ask them how they assess their level of English before virtual conversation sessions. Naturally, this simple approach cannot influence the results, but at least it helps us to understand what went wrong if the participants cannot recognize culturally appropriate behavior.

The second issue of this challenge is that some particular words and phrases in English can be understood differently if the observers have different languages of origin. For example, a single word in Russian that Russian participant understands in one way, the same word in English can be understood completely different, even if the participant's levels of both languages are equal. Moreover, a Dutch participant can grasp the same word in English with a third meaning, and the meaning of the same word in Dutch can also be different. Unfortunately, this issue cannot be fully avoided; it is beyond our topic and is probably related to personality. However, it does not mean that it should not be considered.

Another challenge is the unique features that are related to cultural symbols and rituals. In other words, well-known stereotypes that distinguishes one culture from the other. One may think that it is not an issue at all, that addition of such features can only benefit for the observer's cultural recognition. However, as the participants should recognize the overall culturally-influenced behavior, these features become an issue. For example, it is an absolutely common thing in Russia to gratitude the doctor materially (i.e. grant a gift, such as a pack of chocolates) for any help provided. This ritual is absolute nonsense for people from the Netherlands and Mexico. So if such ritual is added to the doctor-patients scenario with parameters set for Russian culture, the participants from Russia and those who are well familiar with it, recognize it only by this ritual and not the overall actors' behavior. Fortunately, we identified the most of these features in our pre-evaluation user study and do not include them in our scenarios.

In order to minimize these and other challenges that may arise during evaluation, we decided to conduct two experimental runs. The first run is a test, it requires minimum amount of participants and is needed to correct and enhance our experiment based on its results and participants' feedback. The second run is the actual evaluation of our model.

6.1.3 Participants

9 volunteer participants for the first run and 27 volunteer participants for the second run (36 in total), equally distributed in three groups per country of origin (Netherlands, Mexico, Russia). The age of the participants vary between 20 and 30 years, gender does not matter. The participants are either students or employed young adults. The experiment can be recorded with the consent of its participants.

As our experiment aims to verify cultural model, some of the participants can be located in their country. Therefore, the user study with these people is conducted online. In particular, throughout the whole experiment there will be a constant video communication via Skype. In addition, the Skype session can be recorded with the consent of the participants.

6.1.4 Independent and Dependent Variables

According to the above, the independent variables of the experiment are the participants' culture, the degree their familiarity with each of chosen three cultures, and the culture defined in virtual characters (and dimension scores that correspond to that culture). Thereby, the dependent variables are the participants' ability to recognize that culture, as well as their perception of the following aspects of virtual conversation: attitude between actors, respect shown by virtual characters to the participant, group behavior of virtual characters, and the length of the dialogue itself.

6.1.5 Hypotheses

At the beginning of the thesis we proposed our research question that can be summarized as: *how can we develop a cultural model based Hofstede's dimensions, such that the resulting behavior of virtual characters is culturally appropriate?*

In short, the answer to this question lies in the effectiveness of our cultural model and there are multiple ways to define it. Therefore, it is important to specify what exactly will be measured in form of hypotheses.

In case of our research, the most important factor to evaluate is whether the external observers (i.e. the users of the dialogue system) can recognize appropriate to their culture behavior of the virtual characters. If evaluation of our model shows that the participants can easily distinguish a particular culture, this will be the best indication of success of our model.

Thereby, in order to be precise in the answer to our research question and to assess the success of the cultural model, we propose the first and main hypothesis:

- **H.M**: The use of cultural module allow external observers (i.e. participants) to ***successfully recognize a specific culture*** in behavior of virtual characters, given that they are familiar with that culture.

Another factor worth evaluation is the behavior related to specific dimensions. In other words, the evaluation of specific dialogue moves that are based on concrete dimensions and how we expect them to be perceived by the external observers.

Therefore, we propose the following hypotheses related to each of the four chosen dimensions:

- **H.PDI**: In a PDI-dependent virtual conversation, actors with ***high pdi parameter*** are perceived paying respect to and treat actors with ***higher status parameter*** as superiors. On the contrary, actors with ***low pdi parameter*** are perceived to treat other actors with ***higher status parameter*** as equals (i.e. power status does not matter).

- **H.IDV**: In an IDV-dependent virtual conversation, actors with ***high idv parameter*** are perceived to behave individually ***regardless of group id parameter*** (i.e. group id does not matter). On the contrary, actors with ***low idv parameter*** are perceived to behave as a group with actors that have ***the same group id parameter***, and in the group an actor with ***the highest status parameter*** is considered a leader that converse with actors that have ***different group id parameter***.

- **H.LTO**: In a LTO-dependent virtual conversation, actors with ***low lto parameter*** are perceived to converse longer with other actors and frequently participate in non-subject discussions. On the contrary, actors with ***high lto parameter*** are perceived to converse shorter with other actors and do not participate in non-subject discussions.

- **H.IVR**: In a IVR-dependent virtual conversation, actors with ***high ivr parameter*** are perceived to have more positive, friendly and personal attitude. On the contrary, actors with ***low ivr parameter*** are perceived to have more neutral and impersonal attitude.

6.1.6 Questionnaire Design

Throughout the experiment, each participant is asked to answer three sets of questions: a preliminary set, and two post-session sets. It takes no more than 10 minutes in total to answer all three sets.

The purpose of a preliminary set is to gather basic information of the participants, such as country of origin (culture), age, how do they assess their level of English, how often do they usually visit a doctor, how often do they usually have a meeting with a teacher (supervisor), and how familiar they are with each of the three chosen cultures.

Both post-session sets of questions take place right after each of the scenario sessions. The purpose of those is to gather scenario-specific information, such as to which culture the behavior of the actors is more appropriate and how appropriate it is to their culture (grades from 1 to 5, where 1 is "not at all" and 5 is "everything was culturally appropriate"). Also these questionnaire hold the questions related to hypotheses evaluation, whether each of them was complied accordingly to the participant's culture.

The full questionnaire can be found in Appendix C. Note that this questionnaire only meant to show questions that were asked.

6.2 Results Analysis

As mentioned previously, we conducted two experimental runs. The first run is a test run, its purpose is to indicate and exclude the aforementioned challenges and fix any other issues. In other words, the most valuable results of this run is the participants' feedback. The second run is the main evaluation process. Based on its results we comprehend whether we succeeded or not.

6.2.1 First Run

For this experimental run we invited 9 volunteers from three chosen countries (3 per country) to participate in our experiment. In this case, we provided only the doctor-patients scenario and two questionnaires (preliminary and one post-session). While this setting is obviously insufficient for good results, it is enough to get the participants' feedback and understand what can be wrong with the experiment.

Overall, the results are surprisingly positive. [Tab. 6.4] illustrates whether the participants recognized a culture in the overall actors' behavior correctly. As it may be seen, 7 of 9 participants were correct. However, even those were wrong indicated in the questionnaire that they are not very experienced with human-machine interaction and the visits to the doctor are uncommon for them. But what is more important, they are very unfamiliar with the cultures that were set in the scenarios they participated.

Cultural Setting	Mexico	Netherlands	Russia
Netherlands	Correct	Correct	Correct
Mexico	Correct	Incorrect	Correct
Russia	Incorrect	Correct	Correct

Tab. 6.4. The results of the first run doctor-patients scenario; 7 of 9 participants could recognize defined in the scenario cultures (based on their experience, knowledge and perception).

As for the feedback, it was mostly positive but also very constructive and useful. The most common issue was that despite the conversation is linear, the participants had to press the button each time they should make a dialogue move. Indeed, this mechanic might be a bit annoying, but it was made for a reason. If there will be no need to choose a dialogue move (even if in most cases there is a single move possible), there will be no interaction, and therefore the participants will not be interested in the whole process and can miss some of the important clues. To avoid this issue in the final run, we decided to lower down the linearity of the scenarios and add even more of the 'illusion of choice'.

Other issues were related to little obscurity of instructions and several specific lines of speech. As for the latter, with some lines the participants did not agree, even if they could successfully recognize the culture. All the aforementioned and some other issues were fixed before the second experimental run.

6.2.2 Second Run

For the second and final experimental run we invited 27 volunteer participants that were equally distributed in three groups per country of origin (Netherlands, Mexico, Russia). For this run, as it was intended, we provided all three questionnaires and both scenarios.

Fig. 6.2.A student from Russia participating in the experiment (Doctor-patients scenario).

Before proceeding to the results, it is important to provide some clarity. When we finished the experiment and started the results analysis, it became clear to us that 27 participants for our type of evaluation is not enough for statistically significant results. We conducted several statistical tests and did not receive any significant dependencies or correlations. Therefore we made a decision to manually look through the data and find possible confirmations of proposed hypotheses.

The overall results this time are less positive than it was expected. [Tab. 6.5, Tab. 6.6, Tab. 6.7] illustrate whether the participants recognized cultures in the overall actors' behavior correctly in both scenarios. Only 7 of 27 participants could successfully recognize cultures in both scenarios, 18 participants could recognize culture in one of two scenarios, and 2 participants could not recognize cultures in any of the scenarios.

Participants	Familiarity with Dutch Culture (from 1 to 5)	Doctor-patients Scenario Answer	Teacher-students Scenario Answer
D-P.1	5	Netherlands	Russia
D-P.2	5	Netherlands	Netherlands
D-P.3	5	Netherlands	Netherlands
M-P.1	1	Mexico	Russia
M-P.2	3	Mexico	Netherlands
M-P.3	1	Netherlands	Mexico
R-P.1	2	Netherlands	Russia
R-P.2	2	Netherlands	Russia
R-P.3	2	Netherlands	Russia

Tab. 6.5. The overall results of the second experimental run for both scenarios with parameters set for Dutch culture.

Participants	Familiarity with Mexican Culture (from 1 to 5)	Doctor-patients Scenario Answer	Teacher-students Scenario Answer
D-P.4	1	Netherlands	Netherlands
D-P.5	1	Mexico	Russia
D-P.6	2	Mexico	Netherlands
M-P.4	5	Russia	Mexico
M-P.5	5	Mexico	Mexico
M-P.6	5	Mexico	Mexico
R-P.4	1	Netherlands	Russia
R-P.5	1	Netherlands	Russia
R-P.6	3	Netherlands	Russia

Tab. 6.6. The overall results of the second experimental run for both scenarios with parameters set for Mexican culture.

Participants	Familiarity with Russian Culture (from 1 to 5)	Doctor-patients Scenario Answer	Teacher-students Scenario Answer
D-P.7	2	Netherlands	Russia
D-P.8	1	Netherlands	Russia
D-P.9	2	Russia	Russia
M-P.7	1	Netherlands	Mexico
M-P.8	2	Mexico	Russia
M-P.9	1	Mexico	Mexico
R-P.7	5	Netherlands	Russia
R-P.8	5	Netherlands	Russia
R-P.9	5	Russia	Russia

Tab. 6.7. The overall results of the second experimental run for both scenarios with parameters set for Russian culture.

Important thing to notice here is that in cases of Dutch and Mexican defined cultures [Tab. 6.5, Tab. 6.6] participants who indicated the highest score (5) in familiarity with that culture (natives of these countries) successfully recognized their cultures in 5 of 6 cases, while participants with lower familiarity score (1-3) recognized these cultures at most in 3 of 6 cases. Normally, this kind of results are obviously insufficient, but as we had a very small number of participants, we think there is at least a positive tendency.

In case of Russian defined culture [Tab. 6.7], Russian participants (with 5 familiarity score) successfully recognized their culture in 4 of 6 cases, while Dutch and Mexican participants (with 1-2 familiarity score) recognized Russian culture in 4 of 6 and in 3 of 6 cases respectively. In other words, there is no clear tendency. It is also clear that all participants recognized Russian culture in teacher-students scenario (Russian and Mexican scenarios are identical). It is possible that this scenario turned out to be too stereotyped in Russian case. Anyway, we decided to go deeper and look for dependencies in other types of data.

Participants	English Language Assessment	Experience with Human-machine Interaction	Doctor Visits for the Past Year	Teacher Meetings for the Past Year
D-P.7	5	3	2	3
D-P.8	5	3	1	0
D-P.9	5	4	1	2
M-P.7	1	3	2	2
M-P.8	3	2	4+	4+
M-P.9	3	3	3	2
R-P.7	4	4	1	0
R-P.8	3	2	4+	0
R-P.9	4	5	4+	4+

Tab. 6.8. English language assessment, experience with human-machine interaction, and the amount of doctor visits and meetings with a teacher for the past year of the participants who had both scenarios with parameters set for Russian culture.

As it can be seen, there is no clear dependencies between the participants' ability to recognize culture in the overall behavior of virtual characters and data provided by them [Tab. 6.8] (please note that English level and experience with human-machine interaction data were not obtained from the results of any tests, the participants assessed themselves). So that could signify the failure of our approach.

However, perception of human behavior itself is a very complex phenomena, not even to mention the influence of culture, personality and many other aspects. So feedback of the participants helped us to get some interesting insights. For example, one Russian participant perceived the behavior of virtual characters in teacher-students scenario as totally appropriate to Russian culture (while the parameters in that particular case were set for Dutch culture) according to his personal experience. Therefore, we came to the very important conclusion that influence of personality and personal experience on perception of behavior in conversations is at least not less than influence of culture.

In addition, it is important to keep in mind that the developers of these scenarios are also human beings, with their own culture, personality and personal experiences. This fact undoubtedly influenced the results of the experiment to some extent.

Based on the above, it is necessary to verify our main hypothesis, which is as follows:

H.M: *The use of cultural module allow external observers (i.e. participants) to* ***successfully recognize a specific culture*** *in behavior of virtual characters, given that they are familiar with that culture.*

It is definitely difficult to state whether we succeeded or not in this term given the amount of people participated in our experiment. However, relying only on the data and feedback we have, as well as conclusions about personality and personal experiences we made, the proof of this hypothesis tends more to ***success***. Dutch and Mexican participants recognized their culture in more cases than other participants. Russian participants successfully recognized their culture in most cases, even if there is no significant difference with successful recognitions of other participants.

Nevertheless, these overall results do can not show the full image, and thus it is impossible to evaluate our approach based on those only and verify other hypotheses. As we collected a lot of useful data from our experiment, it is worth looking at the results in more details. Therefore, we divide our analysis into two major sections, each of which is based on a particular scenario.

6.2.3 Doctor-patients Scenario Results

As it is given earlier in this Chapter, doctor-patients scenario is based on IVR, LTO and PDI related norms (dimensions are listed in order of priority), so in this subsection we verify relevant hypotheses.

The first and most important hypothesis here is the following one:

H.IVR: *In a IVR-dependent virtual conversation, actors with* ***high ivr parameter*** *are perceived to have more positive, friendly and personal attitude. On the contrary, actors with* ***low ivr parameter*** *are perceived to have more neutral and impersonal attitude.*

To understand how the participants perceive the attitude between actors in this scenario, we asked them to indicate their opinion and answer the following linear scale question:

- "*What do you think of the attitude between actors in this virtual conversation?*",

where answering 1 means "Very (too) neutral and impersonal attitude", 5 means "Very (too) positive and personal attitude", and 3 means that the attitude between

actors was totally appropriate based on their personal opinion. [Tab. 6.9, Tab. 6.10, Tab. 6.11] depict the participants opinion on the attitude between actors in doctor-patients scenario.

Participants	D-P.1	D-P.2	D-P.3	M-P.1	M-P.2	M-P.3	R-P.1	R-P.2	R-P.3
Attitude	4	3	3	1	5	2	4	2	3

Tab. 6.9. The participants opinion on the attitude between actors in doctor-patients scenario with parameters set for Dutch culture.

Participants	D-P.4	D-P.5	D-P.6	M-P.4	M-P.5	M-P.6	R-P.4	R-P.5	R-P.6
Attitude	4	4	2	5	5	4	3	4	1

Tab. 6.10. The participants opinion on the attitude between actors in doctor-patients scenario with parameters set for Mexican culture.

Participants	D-P.7	D-P.8	D-P.9	M-P.7	M-P.8	M-P.9	R-P.7	R-P.8	R-P.9
Attitude	3	3	4	4	4	4	3	4	3

Tab. 6.11. The participants opinion on the attitude between actors in doctor-patients scenario with parameters set for Russian culture.

In case of Dutch and Russian doctor-patients scenarios the participants from these countries found the attitude between actors mostly appropriate($\approx 3,3$ on average for D-P.1, D-P.2, D-P.3, R-P.7, R-P.8, R-P.9), while in case of Mexican scenario Mexican participant found the attitude too positive and personal ($\approx 4,7$ on average for M-P.4, M-P.5, M-P.6). The latter could mean that we put too much emphasis on indulgence dimension for a such scenario. It is also interesting to notice that all nine participants found the attitude in Russian scenario nearly appropriate, a bit more positive and personal ($\approx 3,6$ on average), which is quite strange in our opinion. Other opinions greatly vary and there is no correlation with the regularity of participants' doctor visits, meaning that they had different experiences, probably with doctors who have strong personalities.

Nonetheless, we state that the proof of hypothesis **H.IVR** tends to be **successful**. Dutch and Russian scenarios have mostly appropriate attitude between actors. As for Mexican scenario, although Mexican participants found the attitude extremely positive and personal, they had almost the same opinion on this matter. One should

keep in mind that any opinion is a subjective thing by definition, and recalling one of the challenges proposed in this Chapter, the statement "too positive and personal" could mean different degrees of attitude for different individuals.

The second hypothesis to verify here is the following one:

H.LTO: *In a LTO-dependent virtual conversation, actors with **low lto parameter** are perceived to converse longer with other actors and frequently participate in non-subject discussions. On the contrary, actors with **high lto parameter** are perceived to converse shorter with other actors and do not participate in non-subject discussions.*

To understand how the participants perceive the duration of the conversation in this scenario, we asked them to indicate their opinion and answer the following linear scale question:

- "*What do you think of the length of this virtual conversation?*",

where answering 1 means "Too short", 5 means "Too long", and 3 means that the duration of the conversation was totally appropriate based on their personal opinion. [Tab. 6.12, Tab. 6.13, Tab. 6.14] depict the participants opinion on the length of doctor-patients scenario.

Partici-pants	**D-P.1**	**D-P.2**	**D-P.3**	**M-P.1**	**M-P.2**	**M-P.3**	**R-P.1**	**R-P.2**	**R-P.3**
Conversa-tion Duration	3	3	4	3	4	3	3	3	3

Tab. 6.12. The participants opinion on the duration of the conversation in doctor-patients scenario with parameters set for Dutch culture.

Partici-pants	**D-P.4**	**D-P.5**	**D-P.6**	**M-P.4**	**M-P.5**	**M-P.6**	**R-P.4**	**R-P.5**	**R-P.6**
Conversa-tion Duration	5	4	2	3	5	3	3	3	4

Tab. 6.13. The participants opinion on the duration of the conversation in doctor-patients scenario with parameters set for Mexican culture.

Participants	D-P.7	D-P.8	D-P.9	M-P.7	M-P.8	M-P.9	R-P.7	R-P.8	R-P.9
Conversation Duration	3	3	3	3	3	3	3	3	3

Tab. 6.14. The participants opinion on the duration of the conversation in doctor-patients scenario with parameters set for Russian culture.

All participants that had Russian scenario found its duration totally appropriate (3 on average), and the participants who had Dutch scenario found it almost appropriate ($\approx 3,2$ on average). At the same time, the participants who had Mexican scenario had a rather different opinion on the its duration ($\approx 3,6$ on average, grades vary between 2 and 5). Moreover, two of those participants found it too long (surprisingly, one of them is Mexican). Indeed, Mexican scenario is about twice as long than the other two and contains two non-subject discussions. But it was developed in such a way based on the data received from pre-experimental user study results.

This leads us to a conclusion that based on the data we have, the length of conversations has no big cultural influence (at least in a typical doctor-patients setting) and relies more on interlocutors' personalities. From which follows that hypothesis **H.LTO** has **failed**. Despite that short-term oriented virtual characters conversed longer than long-term oriented ones, all participants found the durations of all three scenarios appropriate (with some exceptions, which most probably are based on personal experiences).

The next hypothesis is the following one:

H.PDI: *In a PDI-dependent virtual conversation, actors with **high pdi parameter** are perceived paying respect to and treat actors with **higher status parameter** as superiors. On the contrary, actors with **low pdi parameter** are perceived to treat other actors with **higher status parameter** as equals (i.e. power status does not matter)*.

To understand how the participants perceive the respect shown by the patients in this scenario, we asked them to indicate their opinion and answer the following linear scale question:

- "*What do you think of the respect shown by the patients to the doctor?*",

where answering 1 means "The patients had to show far more respect", 5 means "The patients showed too much respect", and 3 means that the respect shown was

totally appropriate based on their personal opinion. [Tab. 6.15, Tab. 6.16, Tab. 6.17] depict the participants opinion on the respect shown by the patients to the doctor in doctor-patients scenario.

Participants	**D-P.1**	**D-P.2**	**D-P.3**	**M-P.1**	**M-P.2**	**M-P.3**	**R-P.1**	**R-P.2**	**R-P.3**
Respect Shown	4	4	4	3	5	5	3	3	3

Tab. 6.15. The participants opinion on the respect shown by the patients to the doctor in doctor-patients scenario with parameters set for Dutch culture.

Participants	**D-P.4**	**D-P.5**	**D-P.6**	**M-P.4**	**M-P.5**	**M-P.6**	**R-P.4**	**R-P.5**	**R-P.6**
Respect Shown	3	3	4	5	5	3	4	3	5

Tab. 6.16. The participants opinion on the respect shown by the patients to the doctor in doctor-patients scenario with parameters set for Mexican culture.

Participants	**D-P.7**	**D-P.8**	**D-P.9**	**M-P.7**	**M-P.8**	**M-P.9**	**R-P.7**	**R-P.8**	**R-P.9**
Respect Shown	4	4	4	5	5	4	3	4	3

Tab. 6.17. The participants opinion on the respect shown by the patients to the doctor in doctor-patients scenario with parameters set for Russian culture.

In this scenario PDI related norms played significantly smaller role than IVR and LTO ones, so it makes little sense to rely on the data provided in the tables above. To reinforce this statement, it is enough to look at the data itself and understand that it is not of great interest. The only thing worth mentioning here is that all participants from each of three cultures (D-P.1 - D-P.9, M-P.1 - M-P.9, R-P.1 - R-P.9) found the respect shown by the patients approximately the same (or at least they thought in the same direction), no matter which culture was set in the scenario: from $\approx 3,3$ to 4 on average for Dutch participants, from $\approx 4,3$ to $\approx 4,7$ on average for Mexican participants, and from 3 to 4 on average for Russian participants. In other words, participants thought that the patients overrespected the doctor.

However, we are not yet trying to verify hypothesis **H.PDI** as it is of far greater significance in the second scenario.

6.2.4 Teacher-students Scenario Results

Teacher-students scenario is based on PDI and IDV related norms only (dimensions are listed in order of priority). Therefore, in this subsection we verify hypotheses, relevant to those cultural dimensions.

The first hypothesis here is **H.PDI**, which is given in the previous subsection. To understand how the participants perceive the respect shown by the patients in this scenario, we asked them exactly the same question as for the previous scenario, namely:

- "*What do you think of the respect shown by the students to the teacher?*",

where answering 1 means "The students had to show far more respect", 5 means "The students showed too much respect", and 3 means that the respect shown was totally appropriate based on their personal opinion. [Tab. 6.18, Tab. 6.19, Tab. 6.20] depict the participants opinion on the respect shown by the students to the teacher in teacher-students scenario.

Participants	**D-P.1**	**D-P.2**	**D-P.3**	**M-P.1**	**M-P.2**	**M-P.3**	**R-P.1**	**R-P.2**	**R-P.3**
Respect Shown	5	4	3	3	4	5	3	3	2

Tab. 6.18. The participants opinion on the respect shown by the students to the teacher in teacher-students scenario with parameters set for Dutch culture.

Participants	**D-P.4**	**D-P.5**	**D-P.6**	**M-P.4**	**M-P.5**	**M-P.6**	**R-P.4**	**R-P.5**	**R-P.6**
Respect Shown	3	4	4	3	3	4	3	3	2

Tab. 6.19. The participants opinion on the respect shown by the students to the teacher in teacher-students scenario with parameters set for Mexican culture.

Participants	**D-P.7**	**D-P.8**	**D-P.9**	**M-P.7**	**M-P.8**	**M-P.9**	**R-P.7**	**R-P.8**	**R-P.9**
Respect Shown	2	1	3	5	5	3	2	3	3

Tab. 6.20. The participants opinion on the respect shown by the students to the teacher in teacher-students scenario with parameters set for Russian culture.

Unfortunately, this data shows some really unexpected results. First, Dutch participants found that the students in Dutch scenario paid too much respect (4 on average), and that the students in Russian scenario had to pay far more respect (2 on average), while we expected it to be other way around. Second, Russian participants found it equally almost appropriate ($\approx 2,7$ on average in all three cases), which is also unexpected (at the beginning of this section we also indicated that all Russian participants recognized their own culture in all three cases of this scenario). And finally, one group of Mexican participants found the respect shown in the scenario with their culture defined almost appropriate ($\approx 3,3$ on average) and in case of Russian scenario another group of Mexican participants found that the students paid too much respect ($\approx 4,3$ on average), while in both cases the scenarios are identical.

These and the results from the previous subsection show that either power status aspect is more related to personality and personal experience, or could not avoid the challenges proposed previously, or even the question could be badly formulated. In either case, we can state with confidence (taking into account only the received data) that hypothesis **H.PDI** has **failed**. It is also possible that we basically failed in understanding the cultural features related to power distance for three chosen countries.

Nevertheless, our last proposed hypothesis is as follows:

H.IDV: *In an IDV-dependent virtual conversation, actors with* ***high idv parameter*** *are perceived to behave individually* ***regardless of group id parameter*** *(i.e. group id does not matter). On the contrary, actors with* ***low idv parameter*** *are perceived to behave as a group with actors that have* ***the same group id parameter****, and in the group an actor with* ***the highest status parameter*** *is considered a leader that converse with actors that have* ***different group id parameter****.*

To understand how the participants perceive the group behavior between students in this scenario, we asked them the following question:

- "*What do you think of group behavior of the students?*",

where answering 1 means "The students had too individualistic behavior", 5 means "The students had too collectivistic behavior", and 3 means that the group behavior was totally appropriate based on their personal opinion. [Tab. 6.21, Tab. 6.22, Tab. 6.23] depict the participants opinion on the group behavior of the students in teacher-students scenario.

Participants	D-P.1	D-P.2	D-P.3	M-P.1	M-P.2	M-P.3	R-P.1	R-P.2	R-P.3
Group Behavior	4	3	3	2	2	3	3	1	3

Tab. 6.21. The participants opinion on the group behavior of the students in teacher-students scenario with parameters set for Dutch culture.

Participants	D-P.4	D-P.5	D-P.6	M-P.4	M-P.5	M-P.6	R-P.4	R-P.5	R-P.6
Group Behavior	4	4	5	1	1	3	3	4	4

Tab. 6.22. The participants opinion on the group behavior of the students in teacher-students scenario with parameters set for Mexican culture.

Participants	D-P.7	D-P.8	D-P.9	M-P.7	M-P.8	M-P.9	R-P.7	R-P.8	R-P.9
Group Behavior	3	4	5	1	4	5	3	2	3

Tab. 6.23. The participants opinion on the group behavior of the students in teacher-students scenario with parameters set for Russian culture.

In case of Dutch scenario, Dutch participants found the behavior of students almost appropriate ($\approx 3,3$ on average for D-P.1, D-P.2, D-P.3), while Russian and Mexican participants found the behavior slightly more individualistic ($\approx 2,3$ on average for M-P.1, M-P.2, M-P.3, R-P.1, R-P.2, R-P.3), which is an expected tendency. In case of Mexican scenario, Dutch participants found the behavior too collectivistic ($\approx 4,3$ on average for D-P.4, D-P.5, D-P.6), Mexican participants surprisingly found it too individualistic ($\approx 1,7$ on average for M-P.4, M-P.5, M-P.6), and Russian participants found it nearly appropriate ($\approx 3,7$ on average for R-P.4, R-P.5, R-P.6). In case of Russian scenario, Dutch participants found the behavior more collectivistic (4 on average for D-P.7, D-P.8, D-P.9), Mexican participants found it almost appropriate ($\approx 3,3$ on average for M-P.7, M-P.8, M-P.9), and Russian participants also found it almost appropriate ($\approx 2,7$ on average for R-P.7, R-P.8, R-P.9).

With the exception of that Mexican participants found the behavior of students in Mexican scenario too individualistic (in addition to other possibilities, they could simply misunderstand the question), these results are very good. We can with most confidence state that the proof of hypothesis **H.IDV** tends to be **successful**.

6.3 Discussion

Overall, the results are just average. While the main hypothesis **H.M** has been satisfied (with several limitations), only two of four dimension-specific hypotheses, namely **H.IVR** and **H.IDV**, have been satisfied (also with some limitations and exceptions). Taking also into account the fact that we had a very few participants than we had to have, we cannot state with certainty whether it is a success or not. A lot more future work should be done in terms of sociological and anthropological research, implementation and evaluation approach.

The most valuable conclusion we got from these results is that a lot depends on personality and personal experience when it comes to perception of social behavior. In other words, all three levels of human uniqueness have approximately equal influence. So it is very unlikely possible to develop a decent and realistic cultural multi-party dialogue system without taking into account those aspects.

It is also possible that cultural decomposition approach, such as Hofstede's Software of the Mind, is simply not suitable for a computational model of culture. Possibly, culture should be considered as a whole integrated with the aspects mentioned above.

As for the evaluation process, there are numerous things to consider. In this thesis we proposed several challenges that we tried to avoid, but undoubtedly there are many more due to complexity of the matter.

It is important to note once again that these hypotheses are verified in two specific social contexts proposed earlier in this Chapter (doctor-patients and teacher-students scenarios). However, in other contexts these hypotheses could have different tendencies due to the complexity of cultural phenomena.

7 Conclusion

In this Chapter we enclose our thesis. In section 7.1, we reintroduce our research question and evaluate whether we succeeded or not. In section 7.2, we share our afterthoughts and ideas for future work.

7.1 Approach Assessment

Before assessing the success of our approach, it is necessary to recall the research question of this thesis:

How can we add culture in a form of four Hofstede's cultural dimensions to human-machine conversation in multi-party setting, such that the result will be a culturally appropriate behavior of virtual characters?

The main goal here was to verify whether it is possible to develop a simple but elaborated computational model of culture solely based on Hofstede's cultural concept, such that the behavior of virtual characters will be perceived culturally appropriate. In short, it is difficult explicitly state that we succeeded on this matter. The first reason is that we had insufficient number of participants in the experiment, so we could not conduct a statistical analysis of the data. Another reason is that the results we obtained with a manual data analysis are just average. With several exceptions and limitations, only three of five hypotheses can be assumed as successfully confirmed.

However, as we confirmed our main hypothesis, which states that the use of cultural module allow external observers to successfully recognize a specific culture if they are familiar with it, our contribution tends to be successful. Therefore, we can conclude that adding culture in a form of four Hofstede's cultural dimensions to a human-machine conversation in multi-party setting is possible, but it is necessary to consider aspects of human behavior in addition to culture.

Our secondary goal was to develop this model in a form of additional module to the existing dialogue system. In order to verify this matter, we proposed three requirements for the implementation of the module, all of which were met. There

are still a lot of ways to improve our approach, and some of those are considered for the future work.

Despite the lack of significant results, we managed to draw several essential and useful conclusions, such as the importance of personality and personal experience. The field of realistic virtual characters still holds a lot mysteries to revel, but we hope our approach made at least a small step in the right direction. That being said, we hope our contribution is sufficient enough for the progress of human-machine interaction.

7.2 Afterthoughts and Future Work

Certainly, our approach is not perfect at least because it is on its early stage and had quite short development period. Work in this direction with the involvement of new resources (primarily time resources) could lead to a more significant outcome. In addition, there are several particular possible solutions to consider.

First, it is still possible to appropriately evaluate this approach in its current state by greatly increasing the number of participants. With at least 900 participants (300 per culture in the given setting) it is possible conduct valuable statistical tests to get solid results. Unfortunately, we were not able to find that many participants with time and other resources we had.

As for our cultural model, the first and simplest thing to do is to invent new and more complex cultural norms, which would probably reflect those from real life more realistic. The second improvement here could be the expansion of the module as a whole. For example, combining Hofstede's approach with other concepts (some of which were presented here) could make our model more robust and comprehensive.

In Chapter 3 we discussed that we made a decision to leave interruptions and overlapping out of scope in the current state of our approach, despite that it is possible. There are several reasons behind that decision, one of which is that interruption on a cultural basis is a vast topic by itself and its implementation could not help us to answer the research question. Therefore, one of the next steps for our model development could be the actual implementation of those aspects.

Another way to improve our approach is to further develop it in parallel with dialogue system development process, and do it so that the cultural module will be considered from the very core of the system. Certainly, the development of extensions is possible

and our approach is a proof. But there is an important thing to consider: the dialogue system itself is still on its early development stage, which means that its structure and code will constantly change. That will lead to the point when our approach become outdated quite soon. Therefore, a consolidated development will help to prevent this issue.

References

[1] Gunnar Johannsen. „Human-machine interaction“. In: *Control Systems, Robotics and Automation* 21 (2009), pp. 132–62 (cit. on p. 1).

[2] Jonathan Gratch, Jeff Rickel, Elisabeth André, et al. „Creating interactive virtual humans: Some assembly required“. In: *IEEE Intelligent systems* 17.4 (2002), pp. 54–63 (cit. on p. 1).

[3] Nadia Magnenat Thalmann, Prem Kalra, and Marc Escher. „Communicating with virtual characters“. In: (1998) (cit. on p. 1).

[4] Terrence Fong, Charles Thorpe, and Charles Baur. „Collaboration, dialogue, human-robot interaction“. In: *Robotics Research*. Springer, 2003, pp. 255–266 (cit. on p. 1).

[5] Zerrin Yumak, Jianfeng Ren, Nadia Magnenat Thalmann, and Junsong Yuan. „Multi-party interaction with a virtual character and human-like robot: A case study“. In: (2014) (cit. on p. 1).

[6] David Traum and Jeff Rickel. „Embodied agents for multi-party dialogue in immersive virtual worlds“. In: *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*. ACM. 2002, pp. 766–773 (cit. on p. 1).

[7] David Traum. „Issues in multiparty dialogues“. In: *Workshop on Agent Communication Languages*. Springer. 2003, pp. 201–211 (cit. on p. 1).

[8] Frank Dignum and Gerard Vreeswijk. „Towards a testbed for multi-party dialogues“. In: *Workshop on Agent Communication Languages*. Vol. 2922. Springer. 2003, pp. 212–230 (cit. on p. 1).

[9] David Traum. „Models of culture for virtual human conversation“. In: *Universal Access in Human-Computer Interaction. Applications and Services* (2009), pp. 434–440 (cit. on pp. 2, 13).

[10] Activision. *Next Generation Graphics - Human Face Simulation.* 2013. URL: https://www.youtube.com/watch?v=z0cZin2xDmQ&list=PLXj1dKQF6cvSF12Bs92FGfhRbicZU4Fak (cit. on pp. 2, 14).

[11] Unity. *Unity Adam Demo - The Full Film.* 2016. URL: https://www.youtube.com/watch?v=GXI0l3yqBrA (cit. on pp. 2, 14).

[12] Unreal Engine. *The Vineyard Challenge Winners Sizzle Reel.* 2015. URL: https://www.youtube.com/watch?v=M5nWHzH1nDA (cit. on pp. 2, 14).

[13] Dušan Jan and David R Traum. „Dynamic movement and positioning of embodied agents in multiparty conversations“. In: *Proceedings of the Workshop on Embodied Language Processing*. Association for Computational Linguistics. 2007, pp. 59–66 (cit. on pp. 2, 15).

[14] Francisco Iacobelli and Justine Cassell. „Ethnic identity and engagement in embodied conversational agents“. In: *Intelligent virtual agents*. Springer. 2007, pp. 57–63 (cit. on pp. 2, 15).

[15] Rachael E Jack, Oliver GB Garrod, Hui Yu, Roberto Caldara, and Philippe G Schyns. „Facial expressions of emotion are not culturally universal“. In: *Proceedings of the National Academy of Sciences* 109.19 (2012), pp. 7241–7244 (cit. on pp. 2, 15).

[16] Dušan Jan, David Herrera, Bilyana Martinovski, David Novick, and David Traum. „A computational model of culture-specific conversational behavior“. In: *International Workshop on Intelligent Virtual Agents*. Springer. 2007, pp. 45–56 (cit. on pp. 2, 15).

[17] Samuel Mascarenhas, Rui Prada, Ana Paiva, and Gert Jan Hofstede. „Social importance dynamics: A model for culturally-adaptive agents“. In: *International Workshop on Intelligent Virtual Agents*. Springer. 2013, pp. 325–338 (cit. on pp. 2, 8, 15).

[18] Samuel Mascarenhas, Nick Degens, Ana Paiva, et al. „Modeling culture in intelligent virtual agents“. In: *Autonomous Agents and Multi-Agent Systems* 30.5 (2016), pp. 931–962 (cit. on pp. 2, 8, 15, 16).

[19] Steven Solomon, Michael van Lent, Mark Core, Paul Carpenter, and Milton Rosenberg. *A language for modeling cultural norms, biases and stereotypes for human behavior models*. Tech. rep. UNIVERSITY OF SOUTHERN CALIFORNIA MARINA DEL REY CA INST FOR CREATIVE TECHNOLOGIES, 2008 (cit. on pp. 2, 8, 9, 17, 18).

[20] Gert Jan Hofstede, Catholijn M Jonker, and Tim Verwaart. „Cultural differentiation of negotiating agents“. In: *Group Decision and Negotiation* 21.1 (2012), pp. 79–98 (cit. on pp. 2, 18).

[21] Elnaz Nouri, Kallirroi Georgila, and David Traum. „Culture-specific models of negotiation for virtual characters: multi-attribute decision-making based on culture-specific values“. In: *AI & society* 32.1 (2017), pp. 51–63 (cit. on pp. 2, 19).

[22] Alfred Louis Kroeber and Clyde Kluckhohn. „Culture: A critical review of concepts and definitions.“ In: *Papers. Peabody Museum of Archaeology & Ethnology, Harvard University* (1952) (cit. on pp. 2, 4).

[23] Theodore D Kemper. *Status, power and ritual interaction: A relational reading of Durkheim, Goffman and Collins*. Routledge, 2016 (cit. on pp. 2, 8).

[24] Roy G D'Andrade. „Schemas and motivation“. In: *Human motives and cultural models* 23 (1992), p. 44 (cit. on pp. 2, 8).

[25] Roy G D'andrade. „Cultural meaning systems“. In: *Document Resume* 197 (1984) (cit. on pp. 2, 8).

[26] Alfred Kieser. „Book Reviews: Geert Hofstede: Cultures and Organizations. Software of the Mind: 1991, Maidenhead, UK: McGraw-Hill. 279 pages“. In: *Organization Studies* 15.3 (1994), pp. 457–460 (cit. on pp. 2, 4, 5, 11).

[27] G Hofstede, GJ Hofstede, and M Minkov. „Cultures and Organizations: Software of the Mind. Revised and expanded third edition. New York“. In: (2010) (cit. on pp. 2, 4, 5, 7, 9, 11, 24, 36).

[28] Anne Berry. „Spanish and American Turn-Taking Styles: A Comparative Study.“ In: (1994) (cit. on pp. 2, 10).

[29] Raoul Harel. „Towards multi-modal, multi-party virtual dialogue systems.“ In: (2017) (cit. on pp. 3, 12, 13, 26, 29, 67).

[30] Frank Dignum and Virginia Dignum. „Integrating Cultures: An Introduction“. In: *Perspectives on Culture and Agent-based Simulations*. Springer, 2014, pp. 1–10 (cit. on p. 4).

[31] SH Schwartz and W Bilsky. „(1987). Toward a universal psychological structure of human values“. In: (1987) (cit. on p. 6).

[32] Shalom H Schwartz. „A theory of cultural values and some implications for work“. In: *Applied psychology* 48.1 (1999), pp. 23–47 (cit. on p. 6).

[33] Shalom H Schwartz and Anat Bardi. „Value hierarchies across cultures: Taking a similarities perspective“. In: *Journal of cross-cultural Psychology* 32.3 (2001), pp. 268–290 (cit. on p. 6).

[34] Anat Bardi and Shalom H Schwartz. „Values and behavior: Strength and structure of relations“. In: *Personality and social psychology bulletin* 29.10 (2003), pp. 1207–1220 (cit. on p. 6).

[35] Geert Hofstede and Robert R McCrae. „Personality and culture revisited: Linking traits and dimensions of culture“. In: *Cross-cultural research* 38.1 (2004), pp. 52–88 (cit. on p. 7).

[36] Shaun Nichols and Stephen P Stich. *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds.* Clarendon Press/Oxford University Press, 2003 (cit. on pp. 9, 11).

[37] Tanya Stivers, Nicholas J Enfield, Penelope Brown, et al. „Universals and cultural variation in turn-taking in conversation“. In: *Proceedings of the National Academy of Sciences* 106.26 (2009), pp. 10587–10592 (cit. on p. 10).

[38] Angelo Cafaro, Nadine Glas, and Catherine Pelachaud. „The effects of interrupting behavior on interpersonal attitude and engagement in dyadic interactions“. In: *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems. 2016, pp. 911–920 (cit. on pp. 10, 11).

[39] Michael Argyle. „Bodily Communication. 2nd“. In: *London: Methuen* (1988) (cit. on p. 10).

[40] Mashael Al-Saleh and Daniela M Romano. „Culturally Appropriate Behavior in Virtual Agents: A Review“. In: *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*. 2015 (cit. on p. 14).

[41] Gert Jan Hofstede and Paul Pedersen. „Synthetic cultures: Intercultural learning through simulation games“. In: *Simulation & Gaming* 30.4 (1999), pp. 415–440 (cit. on p. 14).

[42] Paul B Pedersen and Allen E Ivey. *Culture-centered counseling and interviewing skills: A practical guide.* Praeger Publishers/Greenwood Publishing Group, 1993 (cit. on p. 14).

[43] Gert Jan Hofstede. „Role playing with synthetic cultures: the evasive rules of the game“. In: *Experimental Interactive Learning in Industrial Management: New approaches to Learning, Studying and Teaching* 49 (2005) (cit. on p. 14).

[44] Nick Degens, Gert Jan Hofstede, Samuel Mascarenhas, et al. „Traveller–intercultural training with intelligent agents for young adults“. In: *Proceedings of the 8th international conference on foundations of digital games 2013*. 2013 (cit. on p. 15).

[45] Joao Dias, Samuel Mascarenhas, and Ana Paiva. „Fatima modular: Towards an agent architecture with a generic appraisal framework“. In: *Emotion Modeling*. Springer, 2014, pp. 44–56 (cit. on p. 15).

[46] Julia M Kim, Randall W Hill Jr, Paula J Durlach, et al. „BiLAT: A game-based environment for practicing negotiation in a cultural context“. In: *International Journal of Artificial Intelligence in Education* 19.3 (2009), pp. 289–308 (cit. on p. 17).

[47] David V Pynadath and Stacy C Marsella. „PsychSim: Modeling theory of mind with decision-theoretic agents“. In: *IJCAI*. Vol. 5. 2005, pp. 1181–1186 (cit. on p. 17).

[48] Catholijn M Jonker and Jan Treur. „An agent architecture for multi-attribute negotiation“. In: *International joint conference on artificial intelligence*. Vol. 17. 1. LAWRENCE ERLBAUM ASSOCIATES LTD. 2001, pp. 1195–1201 (cit. on p. 18).

[49] Cormas. *Cormas – Natural resources and agent-based simulations*. URL: `http://cormas.cirad.fr/indexeng.html` (cit. on p. 18).

[50] Unity. *Unity Asset Store*. URL: `https://www.assetstore.unity3d.com/en/` (cit. on p. 40).

[51] Mixamo. *Mixamo: Free 3D models and animations*. URL: `https://www.mixamo.com/` (cit. on p. 40).

Appendices

Cultural Module Manual

A

This manual provides important information and instructions for how to install the cultural module (described in Chapter 3 and Chapter 4). Section A.1 contains a list of prerequisites, section A.2 holds installation instructions, and section A.3 briefly describes how to use cultural module.

Note! It is very important to first get acquainted with the dialogue system and manuals for its main components [29], as this cultural module is the part of this project.

A.1 Prerequisites

- **C Sharp** is the programming language used across the project.
- **Microsoft Visual Studio** (version 2015 or above) is the integrated development environment used to manage source files, configure builds and develop scenarios.

A.2 Installation

At this stage of the project there is no place to download pre-compiled DLLs for cultural module or even the dialogue system components. Therefore, it is needed to build the library and its dependencies from source. The relevant directory in "Cultural Multi-party Dialogue" folder is '$core/dialog - culture/$'. This directory contains cultural module itself and dependencies to other components of the system, which can also be found in '$core/$'.

A.2.1 Building from Source

In order to build the module, it is necessary to perform the following steps:

1. Open the module's project solution **dialog-culture.sln** located under 'core/dialog-culture/library/dialog-culture' in Microsoft Visual Studio.

2. Build the solution (F7).

3. A **dialog-culture.dll**, as well as the other dll's of the dialogue system used, should now be available under 'dialog-culture/bin/Debug' or 'dialog-culture/bin/Release' (depending on the specified configuration).

B

Doctor-patients Scenarios Fragments Examples

This Appendix contains two examples of dialogue fragments of both scenarios. Please keep in mind that these fragments are taken out of the full context and thus cannot convey the whole point of differences in culturally-dependent behavior. Full virtual conversation breakdown for this, as well as for the teacher-students scenario, can be found in "Cultural Multi-party Dialogue" folder.

B.1 Doctor-patients Conversation Fragment Example

In short, the story is that a young man, accompanied by his girlfriend, visits the doctor because his leg hurts. This example illustrates the part of the conversation, in which Doctor (i.e. the participant) asks Patient A (i.e. virtual character) the following question:

- **Doctor**: "*What is your height and weight?*" -> **Patient A**.

The response of Patients (i.e. virtual characters) depends on two cultural norms: `NonchalanceNorm` and `ShortTermOrientationNorm`. The conditions of these norms are described in Chapter 5.

Response 1. Based on the parameters, if:

- `NonchalanceNorm = true`;
- `ShortTermOrientationNorm = true`.

Patients perform the following dialogue moves:

- **Patient A**: "*My height is about 180, and my weight... Patient B, do you remember?*" -> **Patient B**;

- **Patient B**: "*His weight is almost 80 kilos. Got it quite recently. Completely gave up going to the gym.*" -> **Doctor**.

In this case, Patients have high IVR score (i.e. behave more positive, personal and relaxed with others) and low LTO score (i.e. converse longer and participate in non-subject discussions). As it may be seen, Patient A behaves more relaxed, and addresses the question to Patient B (who is defined in the scenario as a girlfriend of Patient A). In turn, Patient B gives an answer and concerns a non-subject topic, which also makes the whole response longer.

Response 2. Based on the parameters, if:

- `NonchalanceNorm = true`;
- `ShortTermOrientationNorm = false`.

Patient A performs the following dialogue move:

- **Patient A**: "*My height is about 180, and my weight... Hmm... I think it is almost 80 kilos.*" -> **Doctor**.

In this case, Patient A has high IVR (i.e. behaves more positive, personal and relaxed with others) and high LTO (i.e. converses shorter and does not participate in non-subject discussions). Patient A behaves more relaxed, but does not address Patient B and does not concern any non-subject topics.

Response 3. Based on the parameters, if:

- `NonchalanceNorm = false`;
- `ShortTermOrientationNorm = false`.

Patient A performs the following dialogue move:

- **Patient A**: "*My height is about 180, and my weigh is almost 80 kilos.*" -> **Doctor**.

In this case, Patient A has low IVR (i.e. behaves more neutral and less relaxed with others) and high LTO (i.e. converses shorter and does not participate in non-subject discussions). Patient A gives a short and straight answer (with some tension), and does not concern any non-subject topics.

B.2 Teacher-students Conversation Fragment Example

In short, the story is that two students visit the teacher to ask several study-related questions. In this scenario, the teacher has higher status and his group id is different to the students group id. This example illustrates the part of the conversation, in which Students (i.e. virtual characters) ask Teacher (i.e. the participant) to discuss their essays.

Question 1. Based on the parameters, if:

- `StatusLowToHighNorm = false`.
- `InGroupNorm = false`.

Students perform the following dialogue moves:

- **Student A**: "*Teacher, I also wanted to discuss my essay.*" -> **Teacher**;
- **Student B**: "*Oh, yeah, me too!*" -> **Teacher**.

In this case, Students have low PDI score (i.e. treat others as equals regardless of their status) and high IDV score (i.e. behave individually). Both Students approach Teacher in an easy manner as if they are talking to someone equal, and each consider only their own work (they use pronouns like "I" and "me").

Question 2. Based on the parameters, if:

- `StatusLowToHighNorm = true`.
- `InGroupNorm = true`.

Student B performs the following dialogue move:

- **Student B**: "*Teacher, we also wondered if you have a few minutes more to discuss our essays.*" -> **Teacher**.

In this case, Student B has high PDI score (i.e. pays respect to and treats others with higher status as superiors) and low IDV score (i.e. behaves collectivistic). Student B

respectfully asks Teacher if he has a few more minutes to discuss their essays (she uses "we" and "our" pronouns).

Experiment User Study

C

This Appendix contains the questionnaire for the final experiment (second run).

C.1 Step 1. General Questions

This section of the questionnaire is to be filled at the very beginning of the experiment. Steps 2 and 4 are intended for virtual conversations.

Q1: Where are you from?

☐ Netherlands
☐ Mexico
☐ Russia

Q2: What is your gender?

☐ Male
☐ Female

Q3: What is your age? _____

Q4: How do you assess your level of English language?

Beginner/Elementary ☐ ☐ ☐ ☐ ☐ *Advanced/Fluent*

Q5: How experienced you are with visualized human-machine interaction through virtual characters? Think of any systems with which you could interact with virtual characters, like virtual dialogues in computer games or information terminals.

Not experienced at all ☐ ☐ ☐ ☐ ☐ *Very experienced*

Q6: How many times have you visited a doctor (for any health-related reasons) over the past year?

☐ 0
☐ 1
☐ 2
☐ 3
☐ 4+

Q7: How regularly do you usually visit a doctor (for any health-related reasons)?

☐ Once every few years
☐ Once a year
☐ Two times a year
☐ Three times a year
☐ More

Q8: How many times have you had a meeting with a teacher or a supervisor to have a conversation (for any study-related reasons) over the past year?

☐ 0
☐ 1
☐ 2
☐ 3
☐ 4+

Q9: How regularly do you usually have a meeting with a teacher or a supervisor (for any study-related reasons)?

☐ Once every few years
☐ Once a year
☐ Two times a year
☐ Three times a year
☐ More

Q10: How familiar you are with Dutch culture?

Not familiar at all ☐ ☐ ☐ ☐ ☐ *Very familiar*

Q11: How familiar you are with Mexican culture?

Not familiar at all ☐ ☐ ☐ ☐ ☐ *Very familiar*

Q12: How familiar you are with Russian culture?

Not familiar at all ☐ ☐ ☐ ☐ ☐ *Very familiar*

C.2 Step 3. Doctor-Patients Scenario Questions

This section of the questionnaire is to be filled after participating in doctor-patients scenario (Step 2).

Q13: What do you think of the respect shown by the patients (virtual characters) to the doctor (you)?

They had to show far more respect ☐ ☐ ☐ ☐ ☐ *They showed too much respect*

Q14: What do you think of the attitude between actors in this virtual conversation?

Too neutral and impersonal ☐ ☐ ☐ ☐ ☐ *Too positive and personal*

Q15: What do you think of the length of this virtual conversation?

Too short ☐ ☐ ☐ ☐ ☐ *Too long*

Q16: Based on your experience and/or perception, to which of the following three cultures the overall behavior of virtual characters was more appropriate in this scenario?

☐ Netherlands
☐ Mexico
☐ Russia

Q17: How appropriate to your culture was the overall behavior of virtual characters in that scenario? Please consider dialogue moves only, not appearance and animations of the characters.

Behavior was completely wrong ☐ ☐ ☐ ☐ ☐ *Behavior was totally appropr.*

C.3 Step 5. Teacher-Students Scenario Questions

This section of the questionnaire is to be filled after participating in teacher-students scenario (Step 4).

Q18: What do you think of the respect shown by the students (virtual characters) to the teacher (you)?

They had to show far more respect ☐ ☐ ☐ ☐ ☐ *They showed too much respect*

Q19: What do you think of the attitude between actors in this virtual conversation?

Too neutral and impersonal ☐ ☐ ☐ ☐ ☐ *Too positive and personal*

Q20: What do you think of group behavior of the students?

Behavior was too individualistic ☐ ☐ ☐ ☐ ☐ *Behavior was too collectivistic*

Q21: Based on your experience and/or perception, to which of the following three cultures the overall behavior of virtual characters was more appropriate in this scenario?

☐ Netherlands
☐ Mexico
☐ Russia

Q22: How appropriate to your culture was the overall behavior of virtual characters in that scenario? Please consider dialogue moves only, not appearance and animations of the characters.

Behavior was completely wrong ☐ ☐ ☐ ☐ ☐ *Behavior was totally appropr.*