CULTURALLY AFFECTED BEHAVIOUR IN BELIEF-DESIRE-INTENTION AGENTS

J. (Jeffrey) Schram

ANR: S886320

HAIT Master Thesis series nr. 11-013

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS IN COMMUNICATION AND INFORMATION SCIENCES,
MASTER TRACK HUMAN ASPECTS OF INFORMATION TECHNOLOGY,
AT THE FACULTY OF HUMANITIES
OF TILBURG UNIVERSITY

Thesis committee:

Dr. ir. P.H.M. (Pieter) Spronck
Prof. Dr. E.O. (Eric) Postma
P.J.M. (Philip) Kerbusch MSC

Tilburg University
Faculty of Humanities
Department of Communication and Information Sciences
Tilburg center for Cognition and Communication (TiCC)
Tilburg, The Netherlands
September 2011

ABSTRACT

Simulation environments for training users in cultural skills have received much interest in recent years. Recent work done on culture specific norms and values resulted in the development of the Culturally Affected Behaviour language. Experience interacting with a prototype agent using the Culturally Affected Behaviour language proved sufficient to enable the experimental participants to learn a culture's norms. Thus far, there have not been any implementations of the Culturally Affected Behaviour language into the Belief-Desire-Intention model. This research effort therefore aims to incorporate the Culturally Affected Behaviour into a Belief-Desire-Intention model. Another goal of this research effort is to expand this model with a personality component that influences the cultural interpretations as described through the Culturally Affected Behaviour language. The two research questions addressed read as follows:

(1) To what extent can the Culturally Affected Behaviour language be incorporated into the Belief-Desire-Intention model?

(2) How can personality influence the cultural interpretations of a culturally aware Belief-Desire-Intention agent?

Our experiments demonstrate that the Culturally Affected Behaviour language can be incorporated into the Belief-Desire-Intention model. The greatest difference between our implementation and the other implementations were caused by the differences between the used agent architectures and their corresponding plan deliberation processes. Our experiments also demonstrated how a personality defined along the five dimensions of the Five Factor Model can be incorporated into a Belief-Desire-Intention agent that has been made culturally aware through the Culturally Affected Behaviour language.

PREFACE

"*Autore Deo, favente Regina, Luctor et Emergo*"
Zeeuwsche Staten, 1585

This Master Thesis was written during my internship at TNO. TNO is a major contract research company in The Netherlands. Part of the research TNO performs are defence research programs, by order of the Dutch Ministry of Defence. One such research program is titled '*Integration of Live, Virtual, and Constructive*'. Integrating live, virtual, and constructive (LVC) is a modelling and simulation topic aiming to enhance training by combining live (real people, real systems), virtual (real people, simulated systems) and constructive (simulated people and systems) assets.

Within the LVC research program novel methods of enriching live training by introducing virtual role players are being researched. For example a shoot house (live), used for tactical training, can present a larger variety of target representation and interaction if virtual role players are introduced. This allows a trainee to use his presence and posture, voice, gestures, etcetera, instead of only his weapon.

The aim of our work is to provide TNO with a Belief-Desire-Intention model that incorporates culture, personality, and (to some extent) emotion which can be used in tactical shoot house training as described above.

This Master Thesis marks the end of my studies at Tilburg University. My gratitude goes out to my internship *and* thesis supervisor P.J.M. (Philip) Kerbusch, and to my other thesis supervisors Dr. ir. P.H.M. (Pieter) Spronck and Prof. Dr. E.O. (Eric) Postma as well as Prof. Dr. H.J.. (Jaap) van den Herik for their continuous support and elaborate input during the entire process. Last, but most certainly not least, I am also very grateful to my family, friends, and girlfriend for their continuous support and, most of all, their patience, throughout the process.

TABLE OF CONTENTS

1 INTRODUCTION

During the Cold War era, the emphasis of military operations was on armoured warfare across massive battlefields. Nowadays, the armed forces are being downsized and assigned new roles, such as peace keeping and counter piracy operations. As a result the focus on military-civilian interaction has increased over the past years (Weiss, 1999). Military personnel, however, are not always adequately trained to interact with the local population in other cultures. This causes opportunities to be lost and in some cases might even lead to confrontations. In current training programmes, live role players are being used to create a realistic simulation of interactions with other cultures. Simulation using interactive virtual worlds populated by realistic Artificial Intelligence (AI) agents can be a cost effective alternative to simulations using live role players. This can be done without (costly) live role players and allow the trainee to train as frequently as necessary.

Simulation environments for training users in cultural skills have received much interest in recent years. This research effort aims to (1) create a culturally aware agent for such simulation environments, and (2) to provide this agent with personality as a source of behavioural diversity to prevent it from becoming predictable. The focus of research efforts concerning simulation environments for training users in cultural skills are on either implicit or explicit cultural differences. An example of research efforts aimed at implicit cultural differences is Mascarenhas and Paiva (2010), which integrates the implicitly manifested values of a culture in an agent architecture through a dimensional model based on two of Hofstede's cultural dimensions (Hofstede, Hofstede, & Minkov, 2010). Most of the focus, however, is on explicit cultural differences such as gestures (Rehm, Bee, Endrass, Wissner, & André, 2007), spoken language (Johnson *et al.*, 2004), and culture-specific norms (Taylor, Quist, Furtwangler, & Knudsen, 2007 and Bogdanovych, Rodriguez, Simoff, & Cohen, 2009).

More recent work done on culture specific norms *and* values is done by Solomon, Van Lent, Core, Carpenter and Rosenberg (2008) and entails the development of the *Culturally Affected Behaviour (CAB)* language. Experience interacting with a prototype agent using their CAB language proved sufficient to enable the experimental participants to learn a culture's norms (Solomon, Hays, Chen, & Rosenberg, 2009). The CAB language represents both cultural norms and cultural values through linked schemas, forming a '*socio-cultural network*'. Each cultural value has an associated intrinsic value (indicating the importance of the value in the culture) and a degree of belief. This degree of belief indicates to what extent the agent believes it is living up to this value. Each action an AI agent can perform or perceive has an effect (i.e. the norm) on the agent's degree of belief in one or more cultural values. For example, the action '*show-picture-of-wife*' has a negative influence on the agent's degree of belief corresponding to the cultural value '*respectful-of-modesty*'. This negative influence on the degree of belief of the corresponding cultural value indicates the cultural norm is not to show pictures of one's wife. The intrinsic values and the degrees of belief of all cultural values are used to calculate an agent's *Socio Cultural Satisfaction (SCS)*, which indicates the agent's appraisal of its interaction with another agent. This SCS score can be used in many ways. In Solomon *et al.*'s application of CAB in the Virtual Humans behaviour model (Solomon *et al.*, 2008; Solomon *et al.*, 2009) the SCS score is used in the agent's reasoning process and to determine the agent's immediate response to the actions of another agent.

CAB, however, was not developed for any specific software implementation, but as a conceptual approach that can be implemented in a multitude of agent architectures (Van Lent, Core, Rosenberg, McAlinden, Carpenter, & Solomon, 2007). Thus far, there are two applications of the CAB language. The first is the application of CAB in the Virtual Humans behaviour model, which is based on the Soar agent architecture (Swartout, Gratch, Hill, Hovy, Marsella, Rickel, & Traum, 2006). The second concerns a formal rewrite of CAB by Bulitko, Solomon, Gratch and Van Lent (2008), resulting in the unification of two different systems capable of modelling culture and emotion through matrix algebra. This second application was implemented in MATLAB.

Soar is an agent architecture based on a production system (Sun, 2006) and uses explicit production rules to generate substates (i.e., problem spaces) in which Soar searches for a goal state. In the present research effort, we investigate to what extent CAB can be applied in another agent architecture, namely the Procedural Reasoning System (PRS). PRS is based on a philosophical model of human practical reasoning originally proposed by Bratman (1987). PRS is often referred to as *Belief-Desire-Intention* or *BDI* as it uses the notions '*belief*', '*desire*', and '*intention*' as mental attitudes to represent possible world states (Rao & Georgeff, 1995). This model has been successful due to its use of notions (i.e., belief, desire, intention) from folk psychology that closely resemble the way people talk about human behaviour (Norling, 2004). Though the control loops of Soar and BDI are quite similar, there are a number of differences between BDI and Soar (Bhattacharyya, 1999). Wray and Jones (2006) explain the most noticeable difference clearly: "*Soar accomplishes all deliberation via a single representation: the operator. In contrast, BDI specifies multiple representations that are mediated by deliberation, including desires, intentions, plans, and, in some cases, beliefs. For each of these representations, there can be a distinct mechanism of choice. Committing to an intention may use some decision-theoretic computation, while committing to a particular plan could result from a simple table lookup*". With this research effort we intend to explore how a single culture represented through the CAB language can be represented in the belief, desire, and intention knowledge areas of an agent using an agent architecture based on the BDI model.

Predictability can reduce the immersiveness of a training simulation or even lead to stereotypes of specific cultures. As such we seek a source of behavioural diversity to prevent our model from becoming predictable. Personality is one of the factors that continuously influence human reasoning and decision making. Therefore, personality has received a lot of attention in agent research. The influence of personality in agents ranges from exerting influence on an agents emotive response (Andr, Klesen, Gebhard, Allen, & Rist, 2000) to serving as a selection criterion that indicates what and how many goals, structures, and attitudes fit with the personality (Egges, Kshirsagar, & Magnenat-Thalmann, 2004). One of the most well-known personality theories used in agent models is Digman's (1990) *Five Factor Model (FFM)*. The FFM is a descriptive model, which describes a personality along five dimensions: *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Neuroticism*. Our secondary focus of our research effort is to explore how a personality defined through the FFM can influence those processes of a BDI agent that deal with the agent's culture as represented through the CAB language. We thus seek to extend the CAB language with personality.

1.1 RESEARCH QUESTIONS

This thesis focuses on (1) the implementation of the CAB language in a BDI model, and (2) the influence of personality on a BDI agent using a culture represented through the CAB language. This research effort therefore results in two distinct problems being addressed. The first being that CAB is developed to be a conceptual approach that can be implemented in a variety of architectures (Bulitko *et al.*, 2008), but thus far has not yet been implemented in the Belief-Desire-Intention architecture. This brings us to our first research question:

RQ 1. To what extent can the Culturally Affected Behaviour language be incorporated into the Belief-Desire-Intention model?

The incorporation of the CAB language into the BDI model should entail the language's most important features:

(1) the defined cultural norms and values can be used to calculate the agent's socio-cultural satisfaction,
(2) the defined cultural norms and values should be human readable,
(3) the agent should be able to distinguish being culturally desired and undesired actions using the human readable cultural norms and values, and
(4) the cultural norms and values should remain independent of other components in the agent model, allowing the agent to easily and quickly switch between definitions of cultural norms and values.

The second problem this thesis addresses is the influence of personality on a BDI agent that is made culturally aware through the CAB language. This brings us to our second research question:

RQ 2. How can personality influence the cultural interpretations of a culturally aware Belief-Desire-Intention agent?

The extent of the influence of personality we seek is not to determine an agent's behaviour, as this could result in our agent deviating from its cultural norms and values, but to influence the interpretation of the agent's cultural norms and values as defined through the CAB language.

1.2 RESEARCH APPROACH

To answer the research questions posed in the previous section, we will create a model that incorporates both the CAB language and a personality based on the FFM using a BDI agent architecture. An agent using this model will be subjected to three experiments in a simulation with one other agent.

The first experiment is aimed at the verification of (1) our model being able to calculate the agent's socio cultural satisfaction, (2) this score being negatively influenced (i.e., decreases) when the actions of another agent violate our agent's cultural norms and values and being positively influenced (i.e., increases) when the actions of the other agent are in agreement with our agent's cultural norms and values, (3) our agent acting in accordance with its own cultural norms and values, and (4) our model being able to easily and quickly switch between different definitions of cultural norms and values.

The second experiment is aimed at our model being able to use multiple definitions of cultural norms and values at the same time. Being able to do so will allow the creation of subcultures such as religion and have them interact with other cultural definitions (e.g., a Christian European and a Muslim European). This interaction can be weighted, making it possible to create agents that are very religious and agents that are not or hardly religious.

The fourth experiment aims to verify the influence of the personality component in our model having the desired effects. The desired effects are described in Section 2.4.

1.3 OUTLINE

The remainder of this thesis is organised as follows. Chapter 2 describes the FFM, the BDI model, and the CAB language. This chapter also gives a description of our own model. Chapter 3 describes the implementation of our model into a BDI agent. The evaluation of our model is given in Chapter 4, followed by the discussion of the results in Chapter 5. Chapter 5 also presents possibilities for future work. Finally, our conclusions and answers to the research questions as posed in Section 1.1 are presented in the last chapter, Chapter 6.

2 MODEL

This research effort concerns itself with the creation of a Belief-Desire-Intention model for intelligent agents that are made culturally aware through the CAB language. Therefore first a description of intelligent agents is given in Section 2.1, followed by a description of the Belief-Desire-Intention model in Section 2.2. The CAB language is used to define the cultural norms and values an intelligent agent believes it has.

In our model we consider *cultural norms* to be an accepted way of behaving (Cambridge Advanced Learner's Dictionary, 2008), i.e. cultural norms are specific actions. A *cultural value* defines what is considered to be important (Cambridge Advanced Learner's Dictionary, 2008),, and specifies *why* specific behaviour is considered to be the norm. Bowing to another person, for example, is considered to be the norm in some cultures. This norm relates to the cultural value of being respectful. Both cultural values and cultural norms will be represented in our model through schemata as described in Section 2.31.

The CAB language and its usage are described in section 2.3. As can be seen in Figure 2-1, the agent observes its environment and compares these observations with its cultural norms. An observation that is in conflict or in agreement with its cultural norms will lead to a revision of the agent's belief it is living up to its cultural values. The agent's belief as to how it is living up to its cultural values (1) determines the agent's emotional state and, together with the agent's cultural norms, (2) influences the agent's plan selection process. The plan selection process determines which plan the agent will choose to achieve a one of its goals. The extent of the revision of the agent's belief that it is living up to its cultural values is influenced by the agent's personality. This personality is defined through Digman's Five Factor Model (Digman, 1990) and is described in Section 2.4.

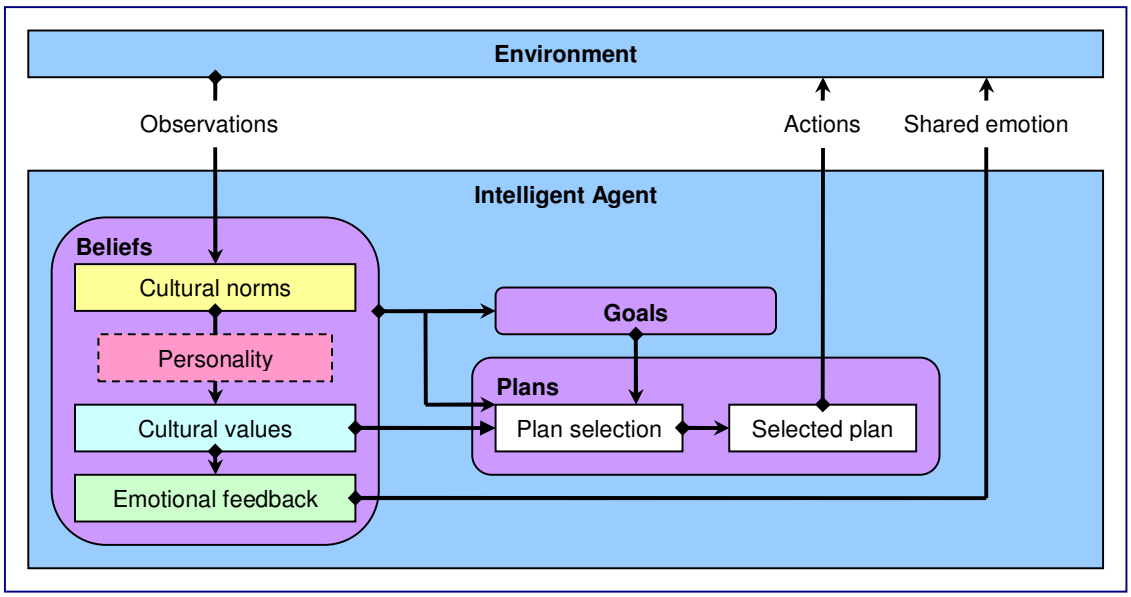


Figure 2-1. Schematic representation of our model:
This figure shows the components and their relations of our model.

2.1 INTELLIGENT AGENTS

This chapter describes a model for intelligent agents. Therefore we will give a description of several types of intelligent agents in this section, concluding with the type of agent that will be used. Russell and Norvig (2009) define intelligent agents as anything that can perceive and act upon its environment (see Figure 2-2). Intelligent agents appear in the form of both software and hardware (e.g., robots). An intelligent agent can perceive its environment in many ways: through sensor inputs (in case of a robot), in the form of keyboard input from a human user, or by receiving information from the virtual environment it resides in. The agent then uses these perceptions to calculate its choice between possible actions.

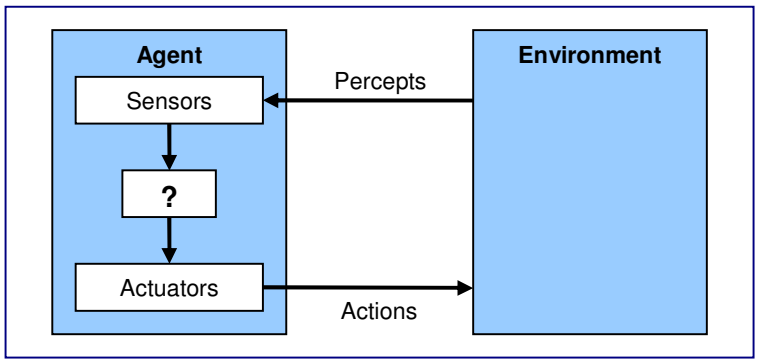


Figure 2-2. Intelligent Agents as described by Russel and Norvig (2009) : This figure demonstrates the process of an intelligent agent perceiving and acting upon its environment.

Russell and Norvig (2009) list four basic types of intelligent agents. With increasing complexity these are: (1) simple reflex agents, (2) model-based reflex agents, (3) goal-based agents, and (4) utility-based agents.

The first type of agent, the simple reflex agent, uses *if-then* rules to respond to its current perception(s). This type of agent ignores any preceding perception(s). These agents have the advantage of being simple, but their intelligence also is rather limited and consequently constrained to only a specific problem domain. An intelligent agent of this type is unable to detect events that require integrating multiple perceptions through time such as acceleration or deceleration of objects. This agent also cannot deal with partial observability. Even if it is only a small part of the environment that cannot be perceived by the agent, it can lead to problems (Russell & Norvig, 2009). An example of a simple reflex agent is a motion sensor light; when motion is detected the agent will activate the light, otherwise the light is turned or kept off.

The second type is the model-based reflex agent. This type can deal with partial observability because they have a model of the environment. This type of agent maintains an internal state which is based on the perceptions so far, and this internal state provides the agent with information about those parts of the environment that cannot be perceived. Knowledge of the workings of the environment also allows these agents to make assumptions about the current and future state of the environment (Russell & Norvig, 2009).

The third type is the goal-based agent, which uses information beyond its internal state to make decision(s). This type of agent needs goal information to describe which future

states are desirable. They will choose actions that will result in the desirable future states. For example: when confronted with an object, the agent may either avoid the object or collide with it, depending on the goal and the current state of the environment. These types of agents are very flexible as their goals are represented explicitly and therefore these goals can be modified (Russell & Norvig, 2009).

The fourth and last type of intelligent agent is the utility-based agent (Russell & Norvig, 2009). These agents can differentiate between goals and actions by assigning each goal or action a certain desirability or utility. This type of intelligent agent will choose those actions that have the maximum expected utility and has the advantage of being able to deal with conflicting goals. It is this type of intelligent agent that we will use in our model.

An agent using our model will often end up having a choice between several courses of action or plans to satisfy a single goal. Some of these plans are in agreement with its cultural norms and values, others in conflict. Our model calculates the utility for each of these plans using its cultural norms and the agent's current belief to what extent it is living up to its cultural values. Plans that are in agreement with the agent's culture will receive a higher utility. This will result in a plan being selected that is in agreement with the agent's own cultural norms and values. Figure 2-3 shows the relations between the agent's cultural norms and values and the agent's plan selection process.

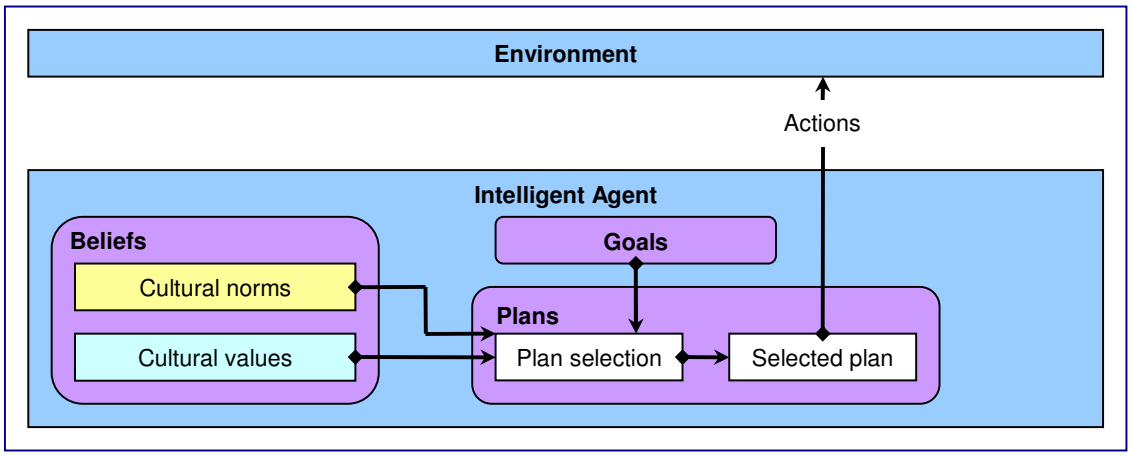


Figure 2-3. Schematic representation of the relations between culture and plan selection: Both the agent's current belief as to how it is living up to its cultural norms and the agent's cultural norms are used to influence the agent's plan selection process.

2.2 BELIEF-DESIRE-INTENTION MODEL

This research effort is primarily focused at the incorporation of the CAB language into the Belief-Desire-Intention (BDI) model. This section describes the BDI model. The BDI software model for Intelligent Agents is inspired by Bratman's theory on human practical reasoning (Bratman, 1987). This theory has been very successful due to its use of notions from folk psychology, which closely resemble the way people talk about human behaviour (Norling, 2004). The BDI model, as adapted by Rao and Georgeff (1995), uses the notions '*belief*', '*desire*', and '*intention*' as mental attitudes that represent possible world states.

Beliefs can be viewed as the information the agent has about the world (including itself and other agents) (Rao & Georgeff, 1995 and Wooldridge, 2000). This information is referred to as beliefs as they are only updated when certain events are perceived. When a certain event has taken place that should have led to updated beliefs, but is not perceived by the agent it will not update its beliefs and retains its (now false) belief. An agent's beliefs, thus, can be incomplete or even incorrect (Wooldridge, 2000). Beliefs can be considered to be the *informative* state of the agent (Rao & Georgeff, 1995). The cultural norms and values an agent has, as well as the agent's personality and its emotional state are considered to be beliefs in our model.

Desires can be seen as the *motivational* state of the agent, and represent the objectives of the agent (Rao & Georgeff, 1995). Desires are the states that an intelligent agent, in an ideal situation, would want to bring about (Wooldridge, 2000). Desires are often mutually exclusive (Wooldridge, 2000) and therefore usually a distinction between desires and goals is made. The desires consist of all possible states the agent desires to bring about, and the goals are a subset of those desires that are consistent with one another (Wooldridge, 2000). In our model, the selection of goals our agent has is influenced by its beliefs. For example, getting fed is only one of our agent's goals when it believes it is hungry.

Usually, an agent will not be able to achieve all of its desires. The agent will therefore select a subset of desires that the agent considers to be possible to attain. This subset of desires the agent is committed to is the set of *intentions* of the agent (Wooldridge, 2000). These intentions can be regarded as the *deliberative* state of the agent (Rao & Georgeff, 1995). In implemented BDI software models an agent will have plans; (sequences of) actions that the agent can perform to achieve its intentions (Wooldridge, 2000). In our model the same approach as in implemented BDI software models is used and we consider our agent's intentions to be plans that can bring about its goals. As mentioned in Section 2.1, an agent can have multiple plans to satisfy a single goal. The agent's beliefs, including (1) its belief as to how it is living up to its cultural values and (2) the agent's cultural norms, influence the process of selecting the plan that the agent will execute.

Summarizing the aforementioned, it can be stated that in our model (1) the beliefs can be considered to be that what the agent knows to be true, (2) the desires are *what* the agent wants to bring about in its environment, and (3) the intentions are the plans as to *how* the agent wants to bring about its desires.

Events can influence the beliefs and consequently also the desires and intentions of an agent. The relation between events and the agent's beliefs, desires, and intentions are schematically represented in Figure 2-4.

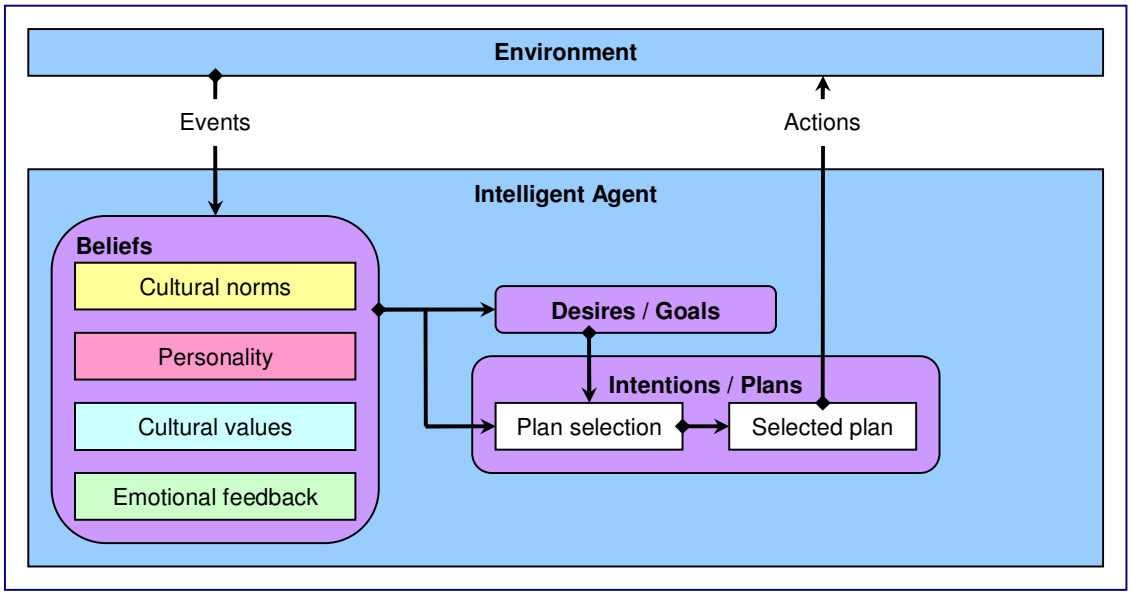


Figure 2-4. Schematic representation of our Belief-Desire-Intention model: This figure shows the relations between events the agent perceives and the agent's beliefs, desires, and intentions.

2.3 CULTURALLY AFFECTED BEHAVIOUR

Our BDI model for intelligent agents will be made culturally aware through the CAB language. This section describes the CAB language. Van Lent, Core, Solomon, Rosenberg, McAlinden, & Carpenter in (National Academy of Engineering of the National Academies, 2008) state that high-level cultural descriptions, such as Hofstede's dimensions (Power Distance, Individualism, Masculinity, Uncertainty Avoidance, and Long-term Orientation) (Hofstede *et al.*, 2010) are not useful for AI models used in simulation environments for training users in cultural skills. They note that these descriptions might make it possible to compare cultures and give some general indications about a culture and/or differences between cultures, but do not provide any knowledge about behavioural and cognitive specifics. Knowing a culture's score on one, more, or even all these dimensions does not give an agent knowledge about any specifics such as that in Japan one bows instead of shaking hands. The Culturally Affected Behaviour (CAB) language allows us to incorporate such low-level cultural specifics into our model. CAB draws inspiration from the schema theory and the theory of mind. Further elaboration on these theories are provided in sections 2.3.1 and 2.3.2.

CAB models socio-cultural norms and values by creating a socio-cultural network consisting of socio-cultural tasks and states (Solomon *et al.*, 2008). These tasks and states are linked through effects. An example of such a socio-cultural network can be seen in Figure 2-5. Rectangular nodes represent tasks, rounded nodes represent the states and the effects of tasks on states can be found along the links between the tasks and states. States have an intrinsic utility value and a current utility value. The intrinsic utility value represents the weight of a socio-cultural value within a culture and makes the importance of these values comparable. The current utility value represents the utility value at any given time during simulation. This value should be perceived as to what extent an agent believes this state to be true and is influenced by tasks perceived during the simulation.

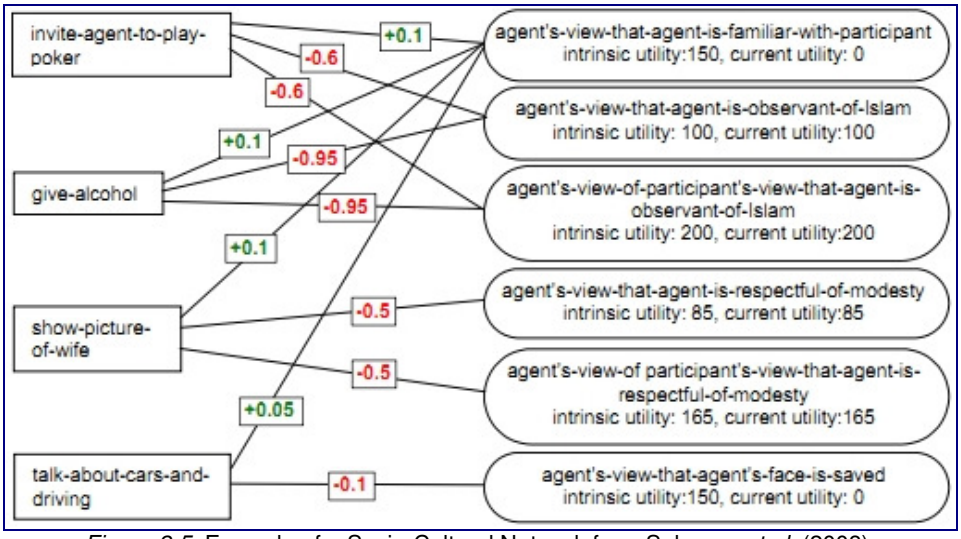


Figure 2-5. Example of a Socio-Cultural Network from Solomon *et al.* (2008): This figure gives examples of tasks (rectangular nodes) and states (rounded nodes) being linked through effects (lines), together forming a Socio-Cultural Network.

Translating the aforementioned to our BDI model, the tasks are regarded to be the equivalent of events whilst the states and effects are equal to beliefs. An agent using the BDI model would perceive an event. The agent *desires* to interpret this event according to the cultural norms and values the agent *believes* it has. This *desire* results in the *intention* to process the event. The agent consequently checks if any of the cultural norms the agent *believes* it has apply to the perceived event. If so, the agent will update the degrees of belief of the affected cultural values (as defined by the cultural norm) accordingly. A schematic representation of this process can be seen in Figure 2-6.

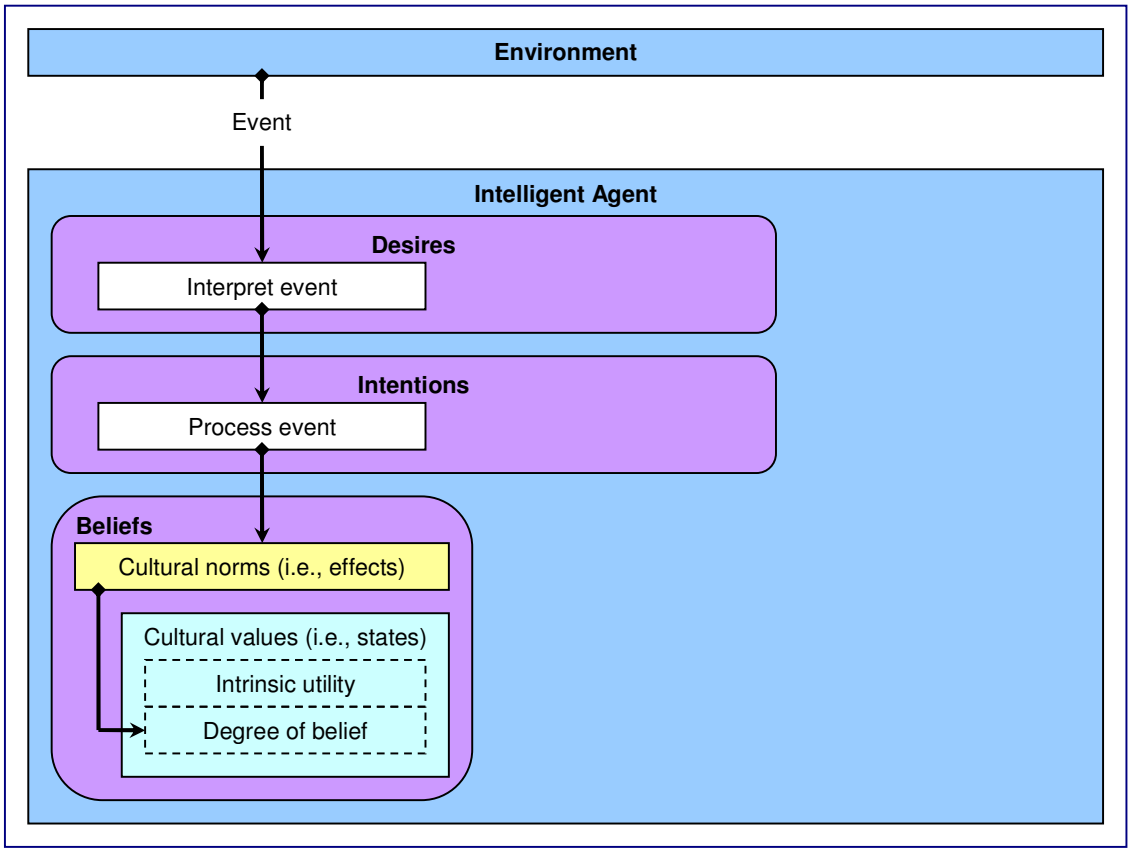


Figure 2-6. Schematic representation of our culturally aware agent processing a perceived event: This figure presents the processing of a perceived event leading to the degrees of belief in affected cultural values being updated.

The socio cultural network is used by Solomon *et al*. (2008) to calculate the agent's Socio Cultural Satisfaction (SCS). The intrinsic values and degrees of belief of all cultural values (i.e., the states in the socio cultural network) are used to calculate the SCS score. This value represents the agent's appraisal of the current interaction set against its own socio-cultural norms and is used to calculate intention probably, which is one of the factors used during the plan deliberation process in the Virtual Humans behaviour model (Solomon *et al*., 2008).

Though intention probability is not a factor in the BDI model, the SCS score will still play an important role in our own model. Potential changes in the SCS score will be used to determine if the agent's own intentions are in correspondence with its cultural norms and values (i.e., an increase in the SCS score) or are in conflict with them (i.e., a decrease in the SCS score). This information will be used during our plan selection process. Our model will select the plan (i.e., intention) that should lead to the highest increase (or the lowest decrease) of its SCS score. The SCS score can also influence the availability of certain plans. A plan can require a minimum (or maximum) SCS score for it to be available during the plan selection process. Through the latter influence on the plan selection process we take into account previously perceived events. As our model will be used in training environments feedback regarding the height of the SCS score to the human user into our model is incorporated. This feedback provides an indication as to how the human user is acting according to the norms and values of the intelligent agent

using our model. The feedback regarding the height of the SCS score is given through the sharing of an emotional state. Adding the feedback through emotion also makes agents using our model more believable (Bosse & Zwanenburg, 2009; Bates, 1994; Adam, 2007) and consequently increases the immersiveness of the simulation. An agent using our model will be able to share four emotional states: '*happy*', '*neutral*', '*angry*', and '*furious*'. The transitions between these states are set by three thresholds. These thresholds are also part of the agent's beliefs and are considered to be a part of the agent's culture and as such are defined upon loading a specific culture before a simulation is started. Whenever the SCS score crosses a threshold the agent will change the shared emotion. The fourth state, '*furious*', however, also leads to different behaviour. Whenever the agent has crossed the threshold between '*angry*' and '*furious*', our agent will no longer interact with the other agent. The feedback regarding the SCS score is always given and therefore bypasses the plan selection process. All this is schematically represented in Figure 2-7.

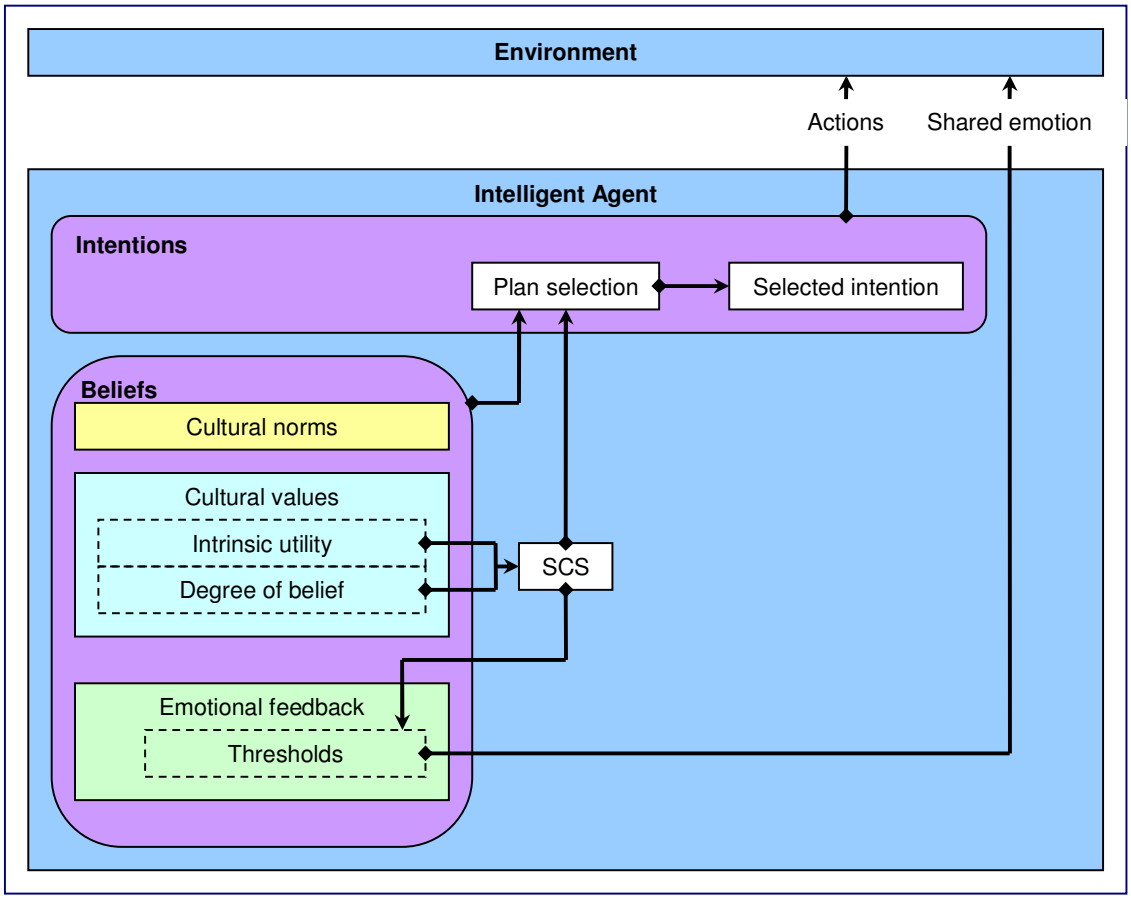


Figure 2-7. Schematic representation of the usage of the Socio Cultural Satisfaction (SCS): This figure demonstrates how both the potential changes (i.e., the norms) in the SCS score and the current SCS score influence the plan selection process, as well as demonstrating how the current SCS score determines the emotional state shared by the agent.

The CAB language was designed to be modular and focuses on swappable culture modules (Van Lent *et al.* in National Academy of Engineering of the National Academies, 2008). CAB thus allows for the culture of an agent to be changed without changing other components of an agent model, such as personality. In our model we accomplish this modularity by having the agent import a set of cultural norms and values before the

simulation is started. The agent's culture thus depends on which set of cultural norms and values are imported whilst other aspects of our model remain unchanged. The CAB language uses a human readable format based on schemas that can be processed by our model. This format is described in the following section.

2.3.1 SCHEMA THEORY

A schema can be viewed as a generic script that is used to interpret and process events (D'Andrade & Strauss, 1992). These schemata are used to deal with frequently encountered situations and objects. In Schema Theory, two general notions of how schemata are being used: the first is to generate behaviour, in which case the schema is often referred to as a 'script' or '*scenario*', and the second notion is that schemata are being used to process a stimulus. The latter notion is often referred to as a '*frame*'. The schema contains default values representing assumptions related to the stimulus (Whitney, Neil, & Paul, 2001).

The classic example of a behaviour generation schema is the restaurant script. This script describes the behaviours needed to go through the motions of going to the restaurant, selecting a table, sitting down at the table, looking at the menu, selecting items from the menu, etc. Schemata are hierarchical in the sense that a higher-level schema can make use of lower-level schemata (D'Andrade & Strauss, 1992). The '*sitting down at the table*'-element of the restaurant script, for example, uses the lower-level '*sitting down*' schema.

An example of a schema being used to process a stimulus is when guns are fired during a (military) funeral. In this case a schema will lead to the events being interpreted as '*giving respect to the deceased*'.

A collection of schemata can be used to represent cultural norms and values (D'Andrade & Strauss, 1992), as well as for representing biases and stereotypes present in a culture about other cultures (Bartlett, 1995; Brewer & Treyens, 1981; O'Sullivan & Durso). These schemata can be used for generating as well as interpreting culture-specific behaviour. A major benefit of using schemata to represent culture is that they can be easily read and written, and thus validated, by for example ethnographers.

The socio cultural network as used by Solomon *et al*. (2008) will be translated to a set of schemas. Cultural values will be translated to a schema consisting of (1) the value itself (e.g., being respectful of modesty), (2) the intrinsic value of this value (i.e., the importance of the value within the culture), and (3) the agent's current belief as to how it is living up to this cultural value. A cultural norm will be translated to a schema consisting of (1) the event that triggers the norm (e.g., eating pork), (2) the extent of the effect it has on a cultural value, and (3) which cultural value the norm affects.

Whenever a perceived event matches a cultural norm, our model will access the corresponding cultural value and apply the effect as defined by the cultural norm to the agent's degree of belief it is living up to this cultural norm. This process is depicted in Figure 2-8.

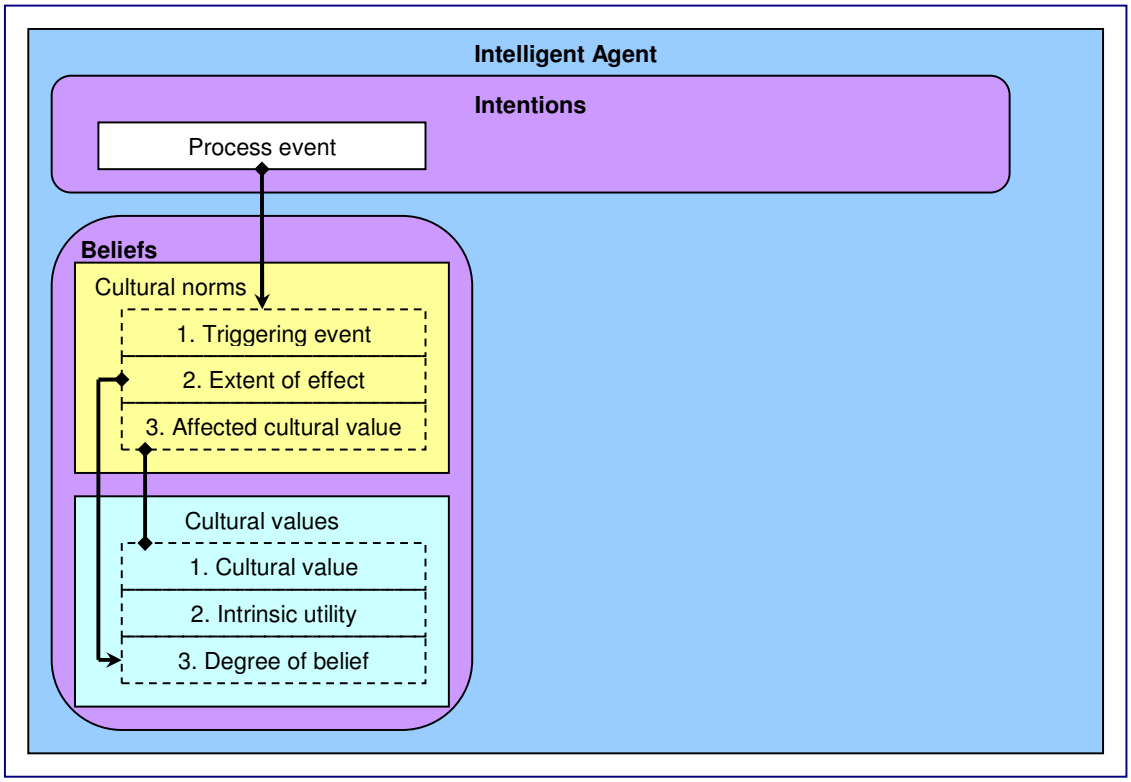


Figure 2-8. Schematic representation of the relations between events, cultural norms and cultural values: This figure shows how an event triggers a cultural norm which consequently leads to a change in the agent's degree it is living up to the affected cultural value.

The plans our agent has are also schemas. Every plan consists of a sequence of individual actions that will be performed by the agent when choosing a particular plan. These actions are described in the same way as events and therefore our model can compare these to the agent's cultural norms. The cultural norm schemata describe the extent of the effects on cultural values and consequently the effect on the agent's SCS score. Through comparing all the actions in a plan with the agent's cultural norms the theoretical impact of executing this plan on the agent's SCS score can be calculated. Though our model will not process the impact of the cultural norms upon executing a plan, the calculation of the theoretical impact does give a clear indication as to what extent a specific plan is in agreement with the agent's cultural norms and values. A schematic representation of this process is shown in Figure 2-9.

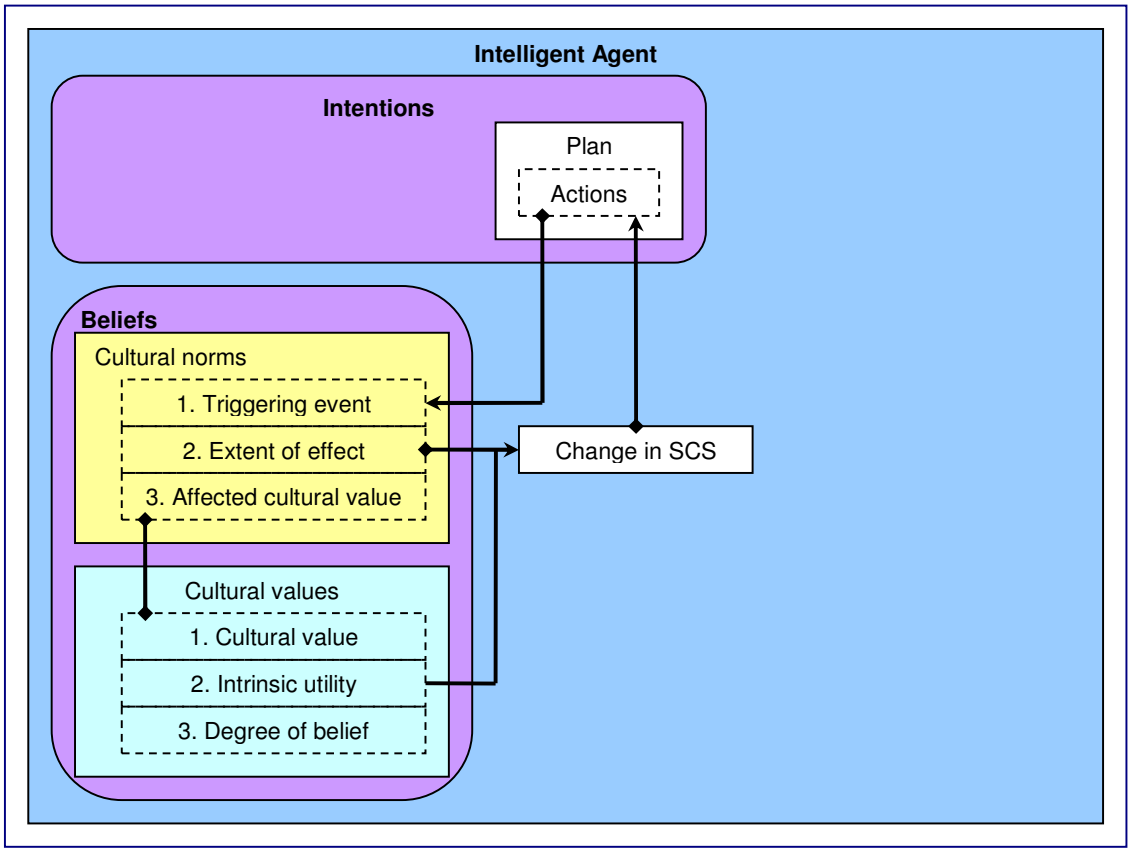


Figure 2-9. Schematic representation of the calculation of a plan's utility: Every action in a plan is compared to the agent's cultural norms. Of every matching cultural norm its theoretical effect on the agent's degree of belief it is living up to the corresponding cultural is used to calculate the plan's theoretical impact on the agent's SCS score. This theoretical change is a plan's utility value.

2.3.2 THEORY OF MIND

Theory of Mind is the human capacity to attribute mental states to other humans, such as beliefs, desires, intentions, norms, and values, that are different from one's own mental states (Leslie, Friedman, & German, 2004; Teufel, Fletcher, & Davis, 2010). This ability is considered to be central to our social life (Leslie, Friedman, & German, 2004) as it heavily influences our perception of others (Teufel, Fletcher, & Davis, 2010). When no other information is available, for example upon first encounter, humans attribute mental states to others based on stereotypes they have about the other's culture. Theory of Mind, thus, is necessary to model cross-cultural stereotypes and biases (Solomon *et al.*, 2008), but is also necessary to model concepts such as reputation and face-saving. In our model there are two distinct types of cultural values, the first type concerns itself with as to how the agent itself believes it is living up to the corresponding value, and the second type describes our agent's belief to what extent *another* agent thinks our agent is living up to the cultural value.

2.4 PERSONALITY

Intelligent agents made culturally aware through the CAB language and having the same cultural norms and values will always give the same response to the same (chain of) events. Such predictability can reduce the immersiveness of the simulation and might

even lead to the formation of stereotypes of the specific culture. Therefore a (limited) personality component will be incorporated in our model that will create behavioural diversity between agents using the cultural norms and values as described through the CAB language.

Personality is one of the factors that continuously influence human reasoning and decision making. There are many psychological models that attempt to identify personality traits. One of the best known models is Digman's (1990) Five Factor Model (FFM). The FFM is a descriptive model, which describes a personality along five dimensions: *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Neuroticism*. An agent using our model will have a score on each of these dimensions. The dimensions '*Openness*', '*Agreeableness*', and '*Neuroticism*' will influence the effect, as defined in a cultural norm schema, on the corresponding cultural value, while the dimensions '*Extraversion*' and '*Neuroticism*' also influence the agent's willingness to share an emotional state other than the neutral one (i.e., happy, angry, or furious). The dimension '*Conscientiousness*' does not have any influence in our model. Following are more detailed descriptions of how each of these dimensions can be related to our model. A schematic representation of the influence of the personality component can be seen in Figure 2-10.

Openness describes how open an individual is to new experiences. Those who score high on Openness tend to be more open to other cultures (McCrae, 1996). This dimension is in our model to determine how tolerant an agent is towards the actions of another agent with a different culture. A *higher* score on the Openness dimension *reduces* any *negative* effect (as defined in the cultural norm schema) of another agent's actions, but only if those actions are in correspondence (thus having a positive effect) with the culture of the other agent. Our agent therefore will also have beliefs regarding the cultural norms of the other agent. These cultural norms represent how the agent *perceives* the other agent's culture to be. This does not necessarily have to be a correct representation of the other agent's culture, i.e. it can be a stereotype. This set can be as specific as to represent how the German culture is perceived to be by the Iraqi people, as well as be as generic as to represent the Western culture. A *lower* score on this dimension has the opposite effect and will result in an *increase* of the *negative* effect of the other agent's action (once again; only if the action it is in correspondence with the other agent's culture).

Conscientiousness describes a person's stance towards spontaneous and planned behaviour. A high score on this dimension indicates that an individual is more goal oriented and less susceptible to impulses (Digman, 1990). In an implemented agent model it can be used to define an agent's commitment to its intentions (Egges, Kshirsagar, & Magnenat-Thalmann, 2003). The *Conscientiousness* dimension can be used to model the agent's stance towards planned behaviour. A *higher* score on this dimension will *increase* the utility of the plans the agent is *currently* committed to, making the agent *less* likely to switch to alternative plans. However, implementing and testing such functionality was beyond our scope and therefore not included in our model. A similar suggestion is made in (Egges *et al*., 2004).

Extraversion describes an individual's tendency to seek both excitement and the company of others (Digman, 1990). Those who score low on this dimension tend to be more low-key and less willing to communicate with others. This dimension will be used to describe the expressiveness of the intelligent agent influences the willingness of our

agent to share an emotional state other than '*neutral*'. This dimension thus influences the thresholds between the states '*happy*', '*neutral*', and '*angry*'. The *higher* the score on this dimension, the *lower* the threshold between neutral and happy, and the *higher* the threshold between angry and neutral. A lower score in this dimension would result in a *higher* threshold between neutral and happy, and a *lower* threshold between angry and neutral.

Agreeableness is the tendency of an individual to seek social harmony and take a cooperative stance towards others (Digman, 1990). Agreeableness, therefore, will be used to influence an agent's tolerance towards another agent's actions (unlike our interpretation of the dimension Openness) regardless of culture (Egges *et al*., 2003). Our usage of this dimension is similar to that of the openness dimension in the respect that a *higher* score on this dimension *reduces* any *negative* effect (as defined in the cultural norm schema) of another agent's actions, and a lower score increase any negative effect. The difference, however, is that it does so *regardless* of the other agent's culture.

The last dimension, Neuroticism, can be seen as an individual's emotional stability. A high score on this dimension indicates that a person has a tendency towards negative emotions such as anger (Digman, 1990). This dimension is used to influence the agent's response when another agent's actions conflict with its own cultural norms and values. This dimension will both (1) influence the effects (as defined in the cultural norm schema) of another agent's actions, and (2) the threshold between '*angry*' and '*furious*'.

The influence this dimension has on the effects of another agent's actions is similar to that of the dimension Agreeableness, but it has the opposite effect. A *higher* score on this dimension *increases* any negative effect of another agent's actions, and a *lower* score now *reduces* any negative effect.

A higher score on this dimension will also set the threshold between the angry and furious higher, and a lower score will lower this threshold. This may seem similar to the effect of the dimension extraversion, but the difference lies in the fact that crossing this threshold results in a direct change behaviour whilst the other two thresholds only result in a change in the shared emotional state.

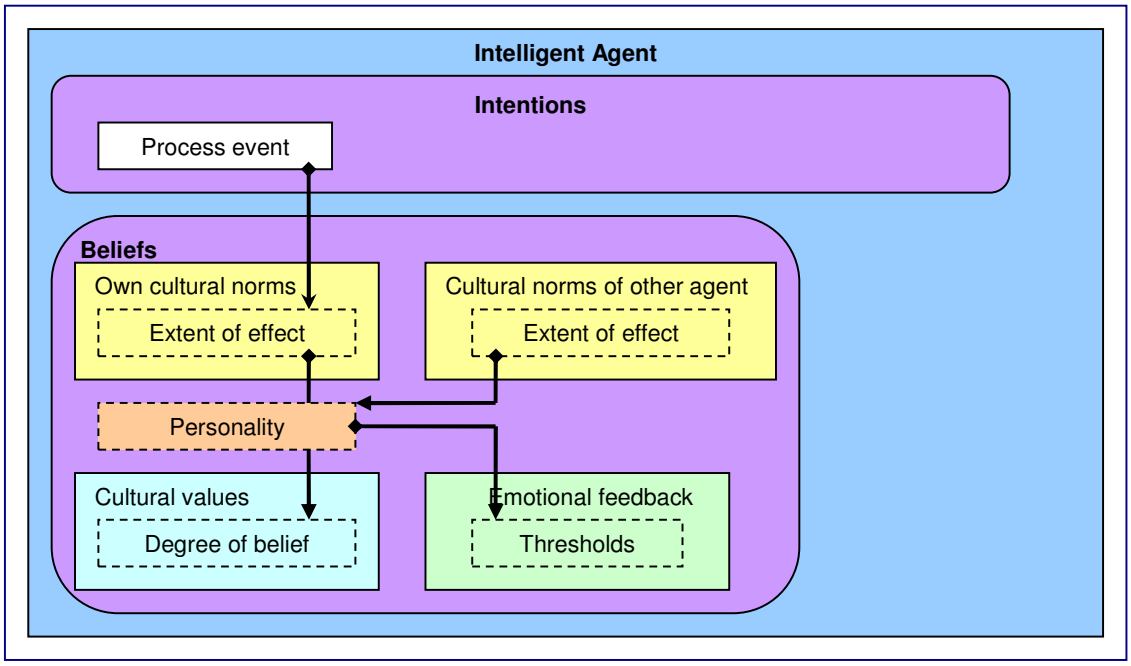


Figure 2-10. Schematic representation of the influence of the personality component in our model: This figure shows how the personality component influences both (1) the extent of the effect as described in a cultural norm schema on the corresponding cultural value, and (2) the emotion thresholds.

3 IMPLEMENTATION

The Jadex architecture (Pokahr & Braubach, 2010a) is used to implement our model into an intelligent agent as Jadex is based on the Belief-Desire-Intention (BDI) model. Section 3.1 gives a brief description of the Jadex architecture. Section 3.2 describes our implementation of the CAB language and Section 3.3 gives an explanation of our implementation of the personality component based on Digman's Five Factor Model (Digman, 1990).

3.1 IMPLEMENTATION ARCHITECTURE

Two important features of the Jadex architecture are that (1) it allows (Java) objects to be used as beliefs (unlike other Belief-Desire-Intention (BDI) architectures, such as 2APL (2011), which use logic for belief formulation and manipulation), and (2) its mechanisms allow a high degree of extensibility and flexibility (Pokahr & Braubach, 2010b; Pokahr & Braubach, 2009). These two aspects of Jadex allow us to implement our model into a BDI intelligent agent relatively easy.

As Jadex is based on the BDI model, it has as main concepts beliefs, goals (i.e. desires), and plans (i.e. intentions). All these beliefs, goals, and plans are defined by the programmer in the Agent Definition File.

A Jadex agent, schematically represented in Figure 3-1, reacts to events (activated goals, incoming messages, and internal events caused by active plans) by selecting and executing the corresponding plans (means-end reasoning). Simultaneously, the agent continuously deliberates about its current goals and decides which goals to pursue (goal deliberation).

The agent's means-end reasoning and goal deliberation are both influenced by the agent's current beliefs. Active plans can update beliefs, despatch (sub)goals, and create internal events (leading to means-end reasoning).

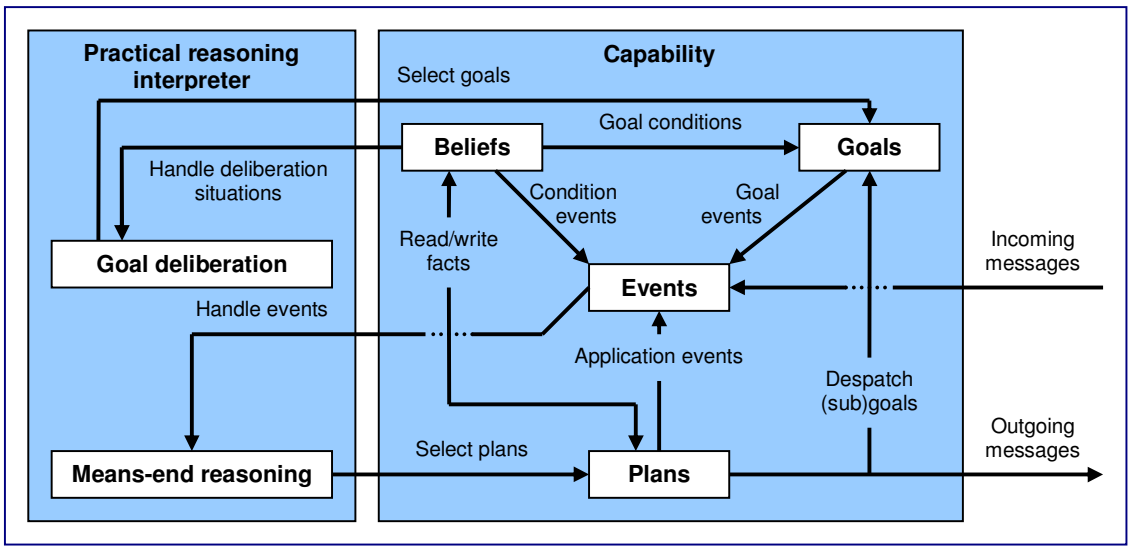


Figure 3-1. Schematic representation of the Jadex abstract architecture (Pokahr & Braubach, 2010b): This figure shows the main concepts of the Jadex architecture and the relations between them.

The '*Incoming messages*' in Figure 3-1 are the equivalent of '*Perceived events*' and are provided by another agent. The other agent functions as a mediator between the agent using our model and its environment, translating changes in the environment to a format our model can understand. Likewise, the '*Outgoing messages*' in Figure 3-1 are the equivalent of '*Actions*' and '*Shared emotion*' and are sent to the mediator which translates it to a format the environment can understand. Communication between these agents is done using the FIPA ACL Message Structure Specification standard (Foundation for Intelligent Physical Agents, 2002). This standard is developed by the Foundation for Intelligent Physical Agents (FIPA). The FIPA develops and promotes standards for agent-based technology (Foundation for Intelligent Physical Agents, 2011). In our experimental setup the mediator will have a graphical user interface through which perceived events can be sent to the agent using our model.

3.2 CULTURE

Our intelligent agents are provided plans and beliefs that will be used to determine the agent's socio cultural satisfaction (SCS). The SCS score indicates the agent's appraisal of the current interaction set against its own socio-cultural norms and values. The SCS score is used during the plan selection process and influences the availability of plans. Section 3.2.3 describes our plan selection process.

The culture of an agent is assigned through the Agent Definition File. The assigned culture itself is defined in a separate file, as shown in Figure 3-2. The definition of a culture consists of (1) cultural values, (2) cultural norms, and (3) the thresholds for the emotional states. These three components of a culture definition are explained in the following sections.

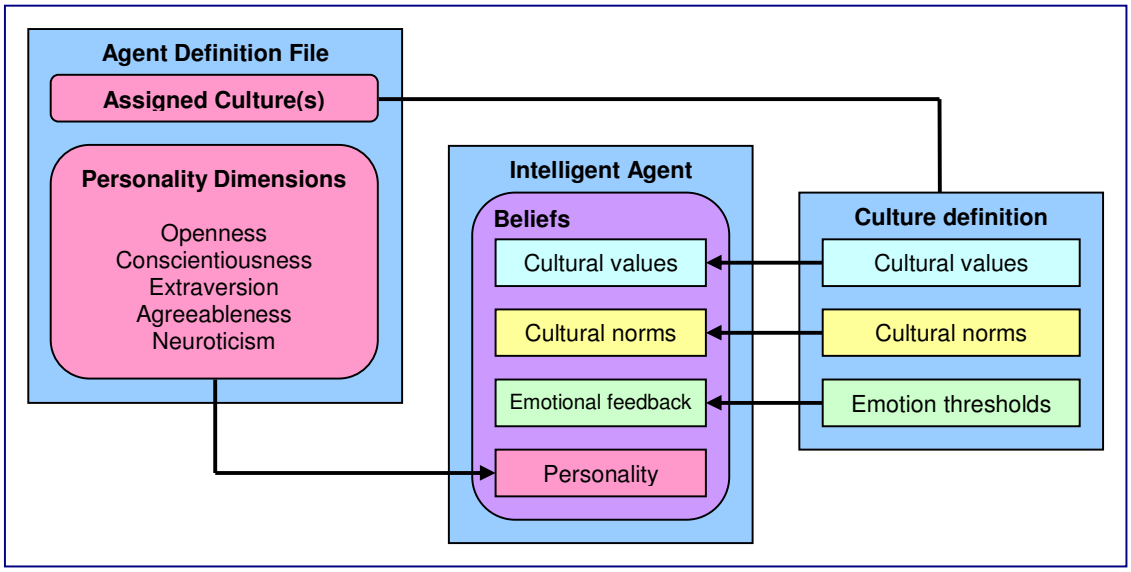


Figure 3-2. Schematic representation of the Agent Definition File and the assignment of cultures: This figure shows how the culture(s) and the personality of the agent are defined in the Agent Definition File and consequently transferred to the intelligent agent using our model.

3.2.1 CULTURAL VALUES

Cultural values are represented by schemata. A schema representing a cultural value consists of (1) the value itself (e.g., being respectful of modesty), (2) the intrinsic value of this value (i.e., the importance of the value within the culture), and (3) the agent's degree of belief as to how it is living up to this cultural value. The degree of belief is expressed on a continuous scale from zero to one.

The cultural values are formulated in such a way that an agent considers a high degree of belief in the cultural value to be a good thing and thus should be pursued. Desirable qualities are thus phrased as '*Agent-is-x*' (e.g., '*Agent-is-kind*') or as '*Agent-does-x'*, whilst undesirable qualities are phrased as '*Agent-does*-not-*x*' (e.g., '*Agent-does-not-steal*') or as '*Agent-is*-not-*x*'.

There are two types of cultural values. The first type describes how the agent views itself (prefixed with '*Agent-*' in Figure 3-3). The other type describes how the agent thinks other agents (such as the human user) view the agent (prefixed with '*Otheragent*' in Figure 3-3). This second type of cultural value is based on the Theory of Mind (see Section 2.3.2), which states that humans can attribute mental states, such as beliefs, to other humans, that are different from one's own mental states (Leslie, Friedman, & German, 2004; Teufel, Fletcher, & Davis, 2010). This second type of cultural value allows us to represent concepts such as reputation and face-saving.

The intrinsic value of a cultural value is expressed on a discrete scale from zero to thousand where zero indicates the value is not important whereas thousand indicates it is most important. The SCS score S_{agent} of an agent is calculated using Equation 3-1, where V_{int} is the intrinsic value of a cultural value and V_{degree} the degree of belief of the same cultural value:

$$S_{agent} = \sum(V_{\text{int}} \cdot V_{\deg ree}) / \sum V_{\text{int}} \qquad (3\text{-}1)$$

This equation is our interpretation of the explanation given about the calculation of the SCS score in Van Lent *et al.* (2007) and Solomon *et al.* (2008), as no explicit equation to calculate the SCS score is given in their work.

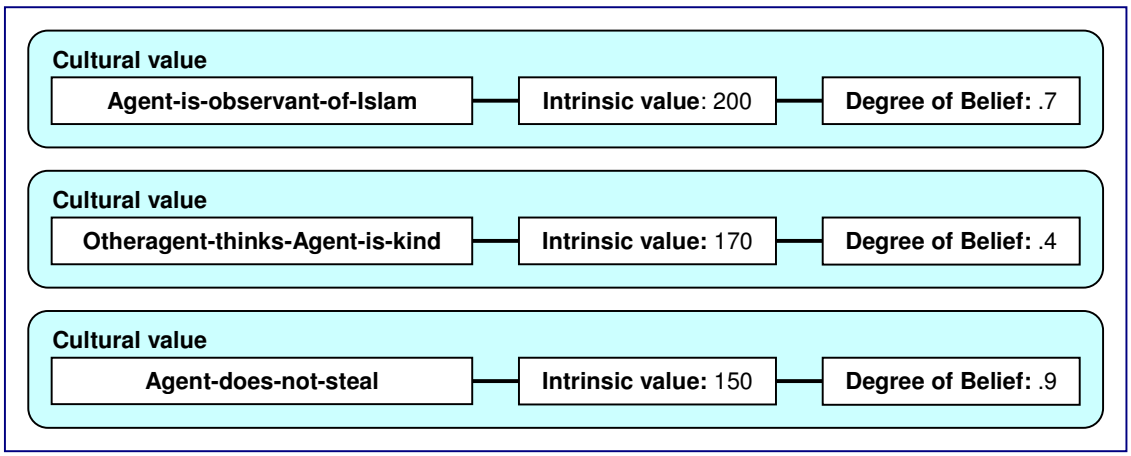


Figure 3-3. Examples of cultural value schemata

An agent having only the three cultural values shown in Figure 3-3 would thus have a socio-cultural satisfaction of: (200 * .7 + 170 * .4 + 150 * .9) / (200 + 170 + 150) = .66.

3.2.2 CULTURAL NORMS

Cultural norms are used (1) when the agent perceives an event, and (2) when the agent selects a plan to achieve one of its goals. As with cultural values, the cultural norms are represented in the form of a schema. Each cultural norm consists of (1) the event that triggers the norm (e.g., eating pork), (2) the extent of the effect it has on a cultural value, and (3) which cultural value the norm affects. Examples are shown in Figure 3-4. These cultural norms can be written to be as specific as required to differentiate between e.g. being invited for dinner by a family member and being invited by a superior. This granularity allows us to create representations of concepts such as (social) hierarchy and status.

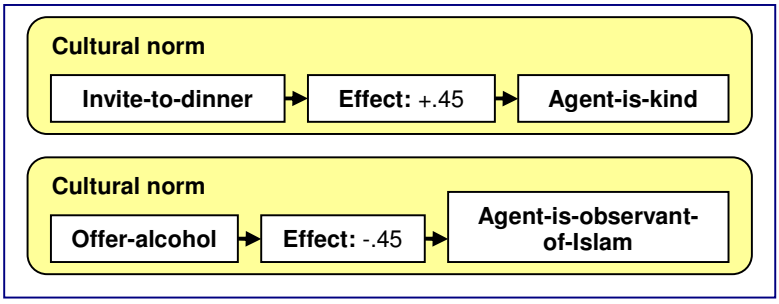


Figure 3-4. Examples of cultural norm schemata

When the actions of another agent are perceived, these actions are compared to the cultural norm schemata. If a perceived action matches the first element of a cultural norm schema, the degrees of belief in the corresponding cultural value is updated. An example of a successful match with a cultural norm, and the update of the corresponding cultural value, is given in Figure 3-5.

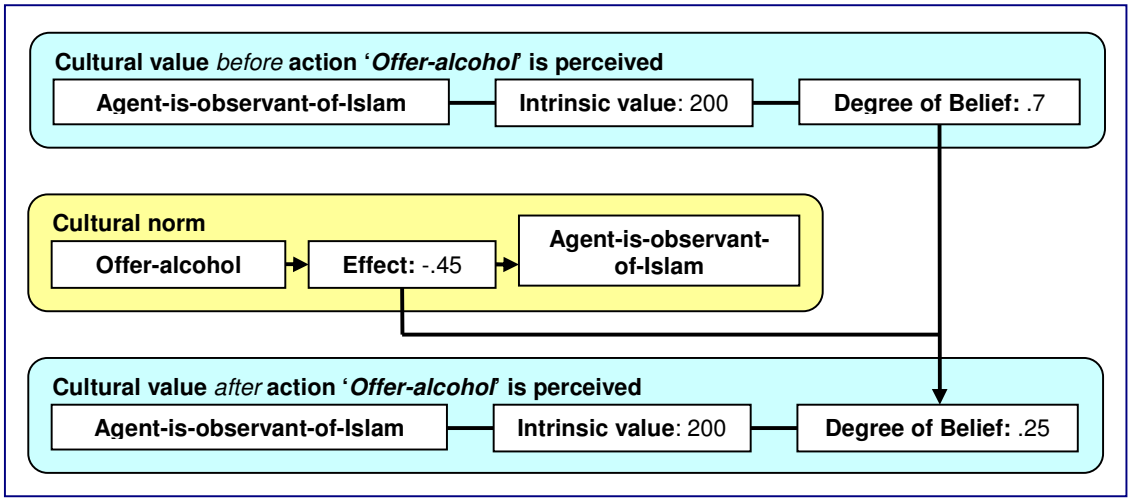


Figure 3-5. Example of an action being successfully matched against a cultural norm: This figure shows how the action '*Offer-alcohol*' is successfully matched agains a cultural norm, resulting in a decrease of belief in the cultural value '*Agent-is-observant-of-Islam*'.

3.2.3 PLAN SELECTION

Cultural norms are also used during the plan selection process. Upon activation of a new goal, the agent selects a plan to achieve this goal. Every plan consists of a series of actions that can be matched against the agent's cultural norms. Figure 3-4 gives an example of how event '*Invite-to-dinner*' would lead to a .45 increase of the degree of belief in the cultural value '*Agent-is-kind*'. Though the execution of a plan does not lead to the effects as defined in the cultural norms being applied to the agent's cultural values, they are used to calculate the *utility* of a plan. A plan that is in correspondence with the agent's norms results in a higher utility due to the positive effects defined in the agent's cultural norms. A plan's utility $P_{utility}$ is calculated as;

$$P_{utility} = (\sum (AV_{\text{int}} \cdot AN_{effect}) / \sum V_{\text{int}}) \cdot 100 - 100 \qquad (3\text{-}2)$$

where AV_{int} is the intrinsic value of those cultural values that are *affected* by the effect AN_{effect} defined in the relevant cultural norm schemata, and V_{int} is the intrinsic value of all cultural values.

The result of Equation 3-2 is the hypothetical change in the agent's SCS score. The agent selects the plan with the highest utility level and therefore the plan that will have the most positive influence on its socio-cultural satisfaction. The selected plan thus is the most in agreement with its cultural norms and values. An example of the plan selection process is shown in Figure 3-6. Our implementation thus gives a more prominent role to the (hypothetical changes in) SCS score during the plan deliberation process. This contrasts with the usage of this score by Solomon *et al.* (2008). They use the SCS score to calculate intention probability, which is one of the factors used during their plan deliberation process.

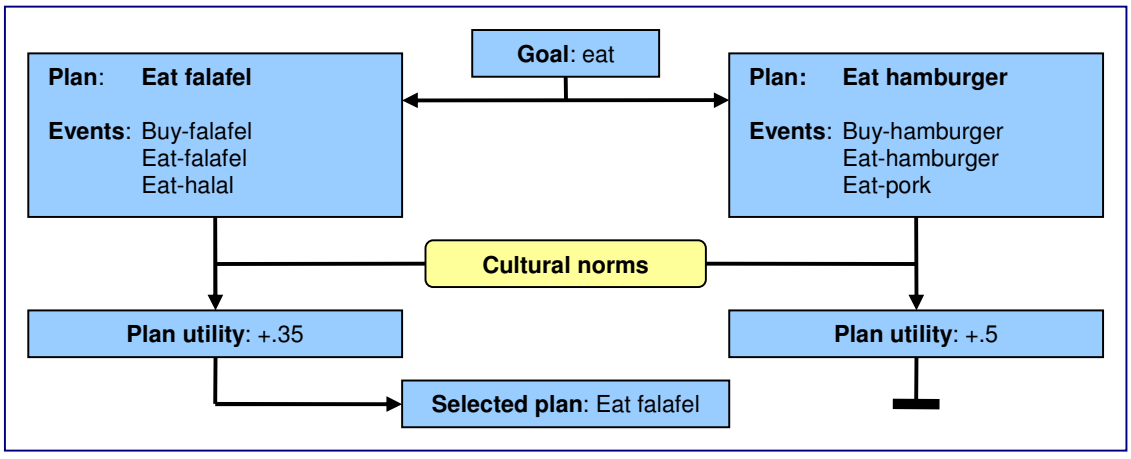


Figure 3-6. Example of the plan selection process for the goal '*eat*'

The SCS score is also used during plan selection. Certain plans are unavailable when certain conditions have not been met, and some of these conditions can be based on the SCS score. For example, the plan to invite another agent for dinner is not available when the SCS score is less than the threshold defined in the plan.

3.2.4 FEEDBACK THROUGH EMOTION

Our model provides feedback concerning the height of the agent's SCS score through sharing an emotional state. A high socio-cultural satisfaction will be interpreted as the agent being '*happy*' with the other agent and a low satisfaction as '*unhappy*'. As mentioned in the previous section, certain plans are only available when certain conditions have been met and some of these conditions are based on the SCS score.

The SCS score of the agent itself is not shared with other (human) agents. What is shared with other agents is an *indication* of this satisfaction. Our model will categorise the SCS into four categories: '*happy*', '*neutral*', '*angry*', and '*furious*'. The first three categories are used only for giving an indication of the agent's SCS score. The thresholds between these categories are set by values in the culture definition. These flexible thresholds allow for representing more or less extroverted cultures.

Besides giving an indication of the height of the SCS score, the fourth category '*furious*' also leads to a direct change in behaviour. Should the SCS score fall below the '*furious*' threshold the relation with the other agent is considered to be beyond repair and no further interaction with the agent is possible.

All of the thresholds mentioned in this section are also influenced by the agent's personality settings. Further elaboration on the influence of the agent's personality follows in Section 3.3.

3.2.5 SUBCULTURES

Elaborating on a suggestion made by Bulitko *et al*. (2008), our implementation of our model allows for multiple (sub)cultures being assigned to a single agent. As suggested by Bulitko *et al.* (2008), each of these (sub)cultures is assigned a weight where the sum of these weights is one. This implementation allows us to create separate definitions for socio-cultural, religious, and political (sub)cultures and scale an agent's membership of such a (sub) culture. This allows for two agents being assigned the same, for example, socio-cultural and religious (sub)cultures, but one being more religious than the other by changing the weight of the (sub)cultures.

The cultural norms and values of each (sub)culture are kept as separate belief sets and therefore our agent can hold the same cultural norms and values multiple times, but with each occurrence possibly having a different intrinsic value and degree of belief. Cultural norms of a specific (sub)culture are only applied to the cultural values of the corresponding (sub)culture.

When an agent is assigned multiple (sub)cultures, our model takes into account the weight of each (sub)culture when calculating the SCS score and plan utility. The same equations, Equation 3-1 and Equation 3-2, are being applied to each (sub)culture and the resulting subtotals T_{sub} are summated after being multiplied by the weight of corresponding (sub)culture W_{sub}. The resulting total, being either the final SCS score or final plan utility, T_{all} will then be used by our model, and, as suggested in Bulitko *et al*. (2008), is calculated as:

$$T_{all} = \sum (T_{sub} \cdot W_{sub}) \tag{3-3}$$

3.3 PERSONALITY

An agent's personality is defined through five parameters in the Agent Definition File. Each of these parameters corresponds to one of the dimensions of Digman's (1990) Five Factor Model (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism). These scores influence both (1) the thresholds for the transitions between the emotional states (as described in the previous Chapter), and (2) the effect of an event on the degree of belief of a cultural value.

An agent's score on these personality dimensions is defined using a continuous scale from zero to one, where a score of .5 results in no influence. A score of .5 thus equals the default threshold or equals the default effect as defined in the cultural norm schema. The default threshold or effect is regarded to be the average of a culture. A score higher or lower than .5 is considered to be a deviation from the average caused by personality. This approach is based on the work of Costa and McCrae (1992). Their personality inventory measures the five dimensions of the Five Factor Model on a scale from one to nine, where a score of five is considered to be the average.

3.3.1 OPENNESS

The dimension *Openness* will be used to model an agent's tolerance towards the actions of another agent with a different culture. A score *higher* than .5 (thus deviating from the culture's average) on the Openness dimension *reduces* any *negative* effect of another agent's actions, but only if those actions are in correspondence (thus having a positive effect) with the culture of the other agent. A *lower* score on this dimension has the opposite effect and will result in an *increase* of the *negative* effect of the other agent's action (once again; only if the action it is in correspondence with the other agent's culture). We translate the score on the dimension Openness to having the desired effect using Equation 3-4;

$$E = \begin{cases} E_{own} \cdot (1.5 - A_{openness}) & if \quad E_{own} < 0 \quad and \quad E_{other} > 0 \\ E_{own} & if \quad otherwise \end{cases} \tag{3-4}$$

where E_{own} is the effect of an action as defined in a cultural norm schema of the agent's own culture. This should be lower than zero as the agent's score dimension Openness $A_{openness}$ only influences negative effects of another agent's actions. The effect of the other agent's action in the cultural norm schema of the other agent E_{other} should be higher than zero as the action should be positively perceived in the other agent's culture.

The effect E_{own} as defined in a cultural norm schema is thus multiplied by a number in the range [0.5, 1.5] to create the desired effect of the personality dimension. This range was chosen to provide noticeable deviations from the culture's average, but not extreme ones. A higher score on this dimension makes an agent more forgiving towards the cultural trespassing of another agent through a slower decrease of its SCS score. An example of the influence of the dimension Openness is shown in Figure 3-7.

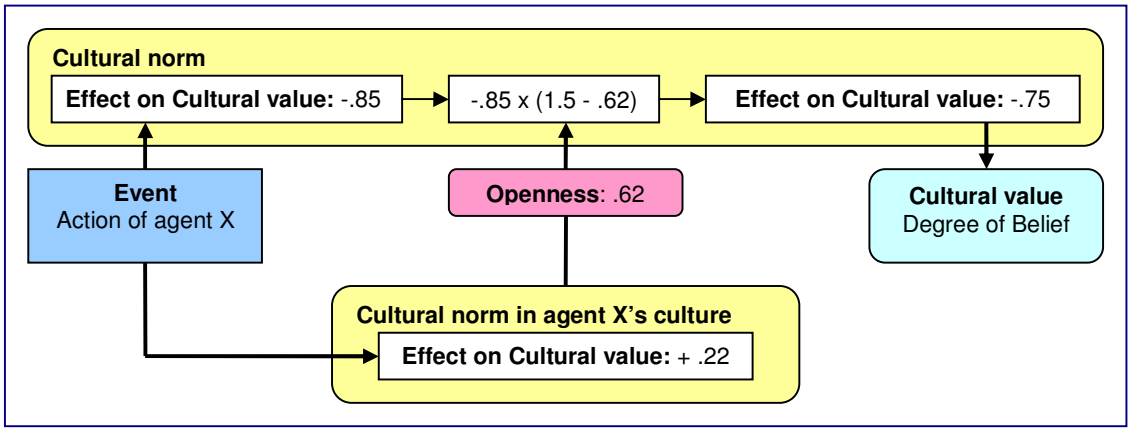


Figure 3-7. Example of the influence of the dimension Openness

3.3.2 CONSCIENTIOUSNESS

The *Conscientiousness* dimension can be used to model the agent's stance towards planned behaviour. A *higher* score on this dimension will *increase* the utility of the plans the agent is *currently* committed to, making the agent *less* likely to switch to alternative plans. However, implementing and testing such functionality was beyond our scope and therefore not included in our model. An example of how the dimension Conscientiousness could influence planned behaviour is shown in Figure 3-8.

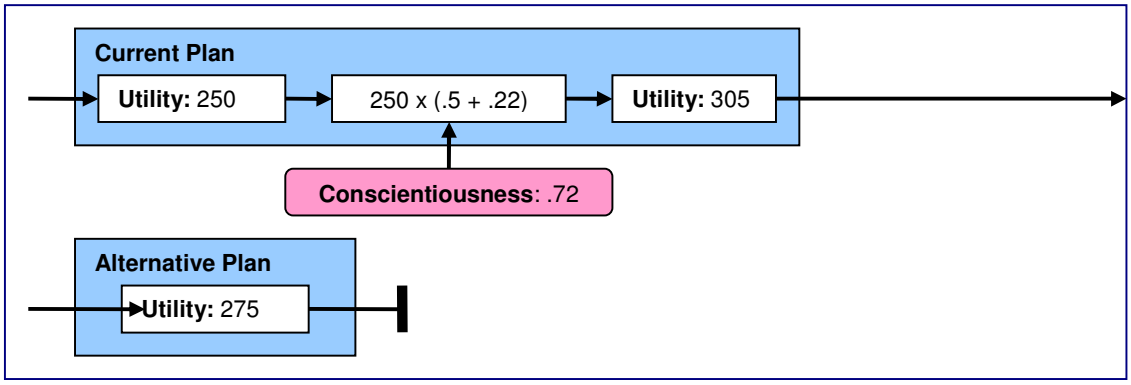


Figure 3-8. Example of the influence of the dimension Conscientiousness

3.3.3 EXTRAVERSION

The *Extraversion* personality dimension will be used to model an agent's tendency to provide feedback regarding the height of its SCS score. This dimension influences the thresholds between the emotion states '*happy*', '*neutral*', and '*angry*'.

The *higher* the score on this dimension, the *lower* the threshold between neutral and happy, and the *higher* the threshold between angry and neutral. A lower score in this dimension would result in a *higher* threshold between neutral and happy, and a *lower* threshold between angry and neutral. The score on the dimension Extraversion $A_{extraversion}$ influences the original threshold $T_{original}$ between neutral and happy according to Equation 3-5;

$$T_{happy} = T_{original} + ((.5 - A_{extraversion}) / 10) \quad (3\text{-}5)$$

where T_{happy} is the updated threshold between neutral and happy. The threshold between neutral and angry T_{angry} is updated using the following equation:

$$T_{angry} = T_{original} + ((-.5 + A_{extraversion}) / 10) \quad (3\text{-}6)$$

A higher score on this dimension thus makes an agent more extrovert and expressive by reducing the SCS range for the neutral state. The influence range is limited to [0, 0.05] to prevent extreme deviations from the culture's average and consequently preserve the extent of extraversion of the culture. An example of the influence of the dimension Extraversion is shown in Figure 3-9.

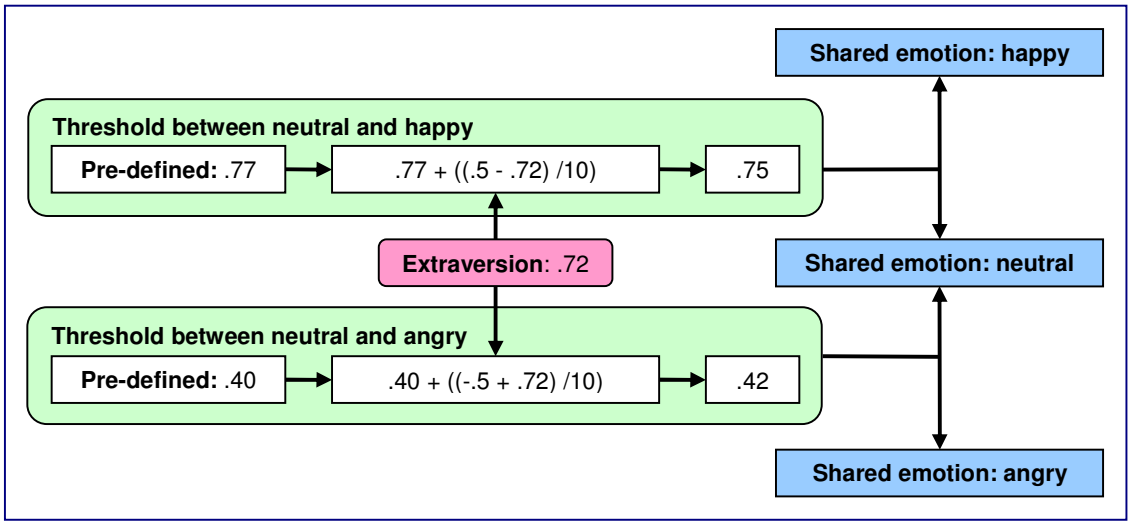


Figure 3-9. Example of the influence of the dimension Extraversion

3.3.4 AGREEABLENESS

The personality dimension *Agreeableness* is used to model an agent's tendency to seek social harmony and take a cooperative stance towards another agent. Our usage of this dimension is similar to that of the openness dimension in the respect that a *higher* score on this dimension *reduces* any *negative* effect of another agent's actions, and a lower score increases any negative effect. The difference with personality dimension Openness, however, is that it does so *regardless* of the other agent's culture. We translate the score on the dimension Agreeableness to having the desired effect through Equation 3-7:

$$E = \begin{cases} E_{own} \cdot (1.5 - A_{agreeableness}) & if \quad E_{own} < 0 \\ E_{own} & if \quad otherwise \end{cases} \quad (3\text{-}7)$$

where E_{own} is the effect of an action as defined in cultural norm schema of the agent's own culture. This should be lower than zero as the agent's score dimension Agreeableness $A_{agreeableness}$ only influences negative effects of another agent's actions.
As with personality dimension Openness, the effect E_{own} as defined in a cultural norm schema is thus multiplied by a number in the range [0.5, 1.5] to create the desired effect

of the personality dimension. Also as with personality dimension Openness, the range [0.5, 1.5] is chosen to provide noticeable deviations from the culture's average, but not extreme ones. An example of the influence of the dimension Agreeableness is shown in Figure 3-10.

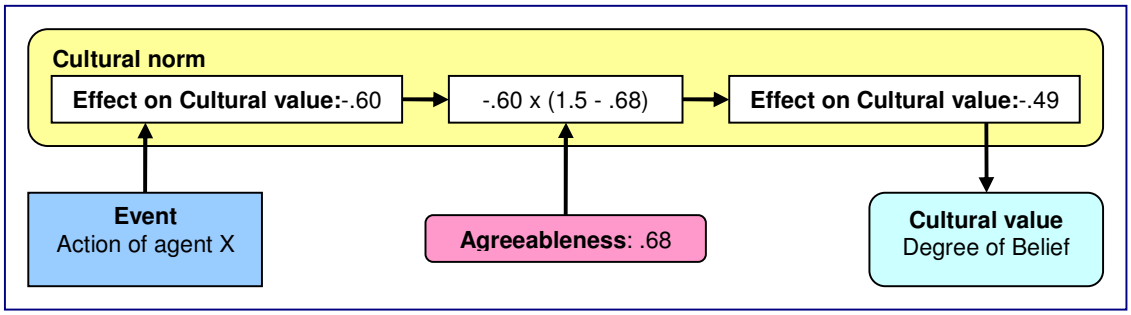


Figure 3-10. Example of the influence of the dimension Agreeableness

3.3.5 NEUROTICISM

The last dimension, Neuroticism, will be used to model an agent's emotional stability. A high score on this dimension indicates that an agent has a tendency towards negative emotions such as anger (Digman, 1990). This dimension will both (1) influence the effects of another agent's actions, and (2) the threshold between angry and furious. The influence this dimension has on the effects of another agent's actions is similar to that of the dimension Agreeableness, but it has the opposite effect. A *higher* score on this dimension *increases* any negative effect of another agent's actions, and a *lower* score now *reduces* any negative effect. We translate the score on personality dimension Neuroticism to having the desired effect through the following equation:

$$E = \begin{cases} E_{own} \cdot (.5 + A_{neuroticism}) & if \quad E_{own} < 0 \\ E_{own} & if \quad otherwise \end{cases} \tag{3-8}$$

where E_{own} is the effect of an action as defined in a cultural norm schema of the agent's own culture. This should be lower than zero as the agent's score Neuroticism $A_{neuroticism}$ only influences negative effects of another agent's actions. The range and reasoning for choosing this range are identical to those given for personality dimensions Openness and Agreeableness. An example of the influence of the dimension Neuroticism on the action effects is shown in Figure 3-11.

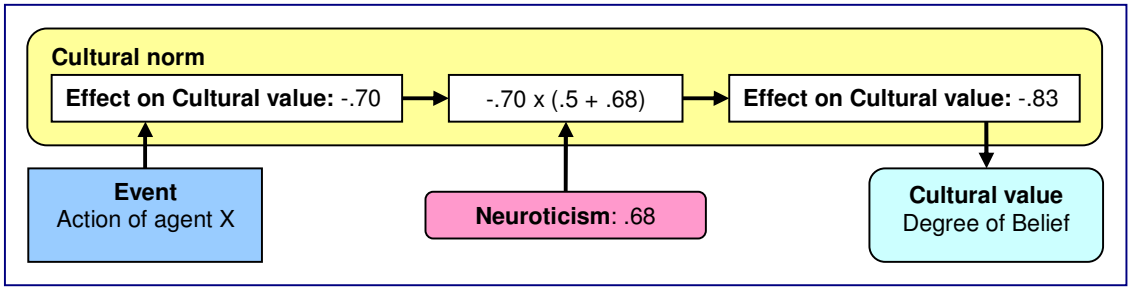


Figure 3-11. Example of the influence of the dimension Neuroticism on action effects

A higher score on this dimension will also set the threshold between the angry and furious higher, and a lower score will lower this threshold. The threshold is influenced according Equation 3-9:

$$T_{furious} = T_{original} + ((-.5 + A_{neuroticism}) / 10) \quad (3\text{-}9)$$

where $T_{furious}$ is the eventual threshold between angry and furious. The range and reasoning for choosing this range are identical to those given for personality dimension Extraversion. An example of the influence of the dimension Neuroticism on the furious threshold is shown in Figure 3-12.

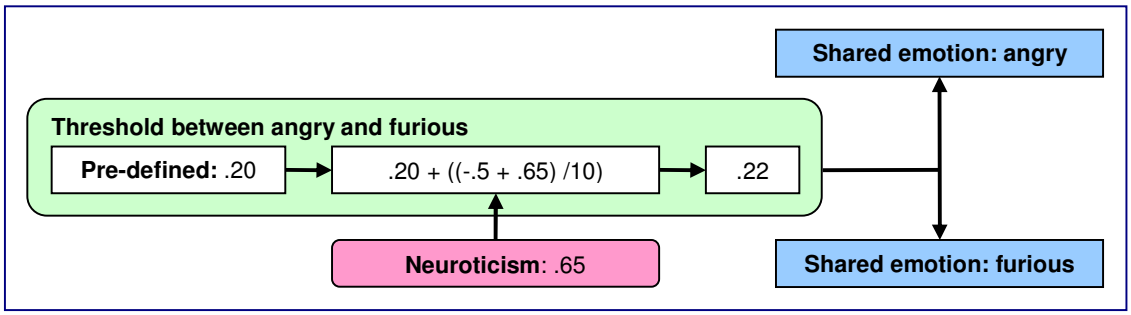


Figure 3-12. Example of the influence of the dimension Neuroticism on the furious threshold

4 EVALUATION

This chapter gives an evaluation of the implementation of our model. Section 4.1 describes the environment in which our agent had to perform. Section 4.2 gives a description of the used culture definitions. Section 4.3 will describe our experiments concerning the implementation of the Culturally Affected Behaviour (CAB) language into our Belief-Desire-Intention (BDI) model and 4.4 describes the experiments regarding our model handling multiple (sub)cultures. The last section gives an evaluation of the influence of our model's personality component on our implementation of the CAB language.

4.1 TASK ENVIRONMENT

In every of the following experiments we placed two BDI agents in a simulation. The first agent used our model and is hereafter referred to as '*our agent*'. The second agent was responsible for performing culture-specific actions which our agent could perceive. From this point forward this second agent will be referred to as '*the interaction partner*'.

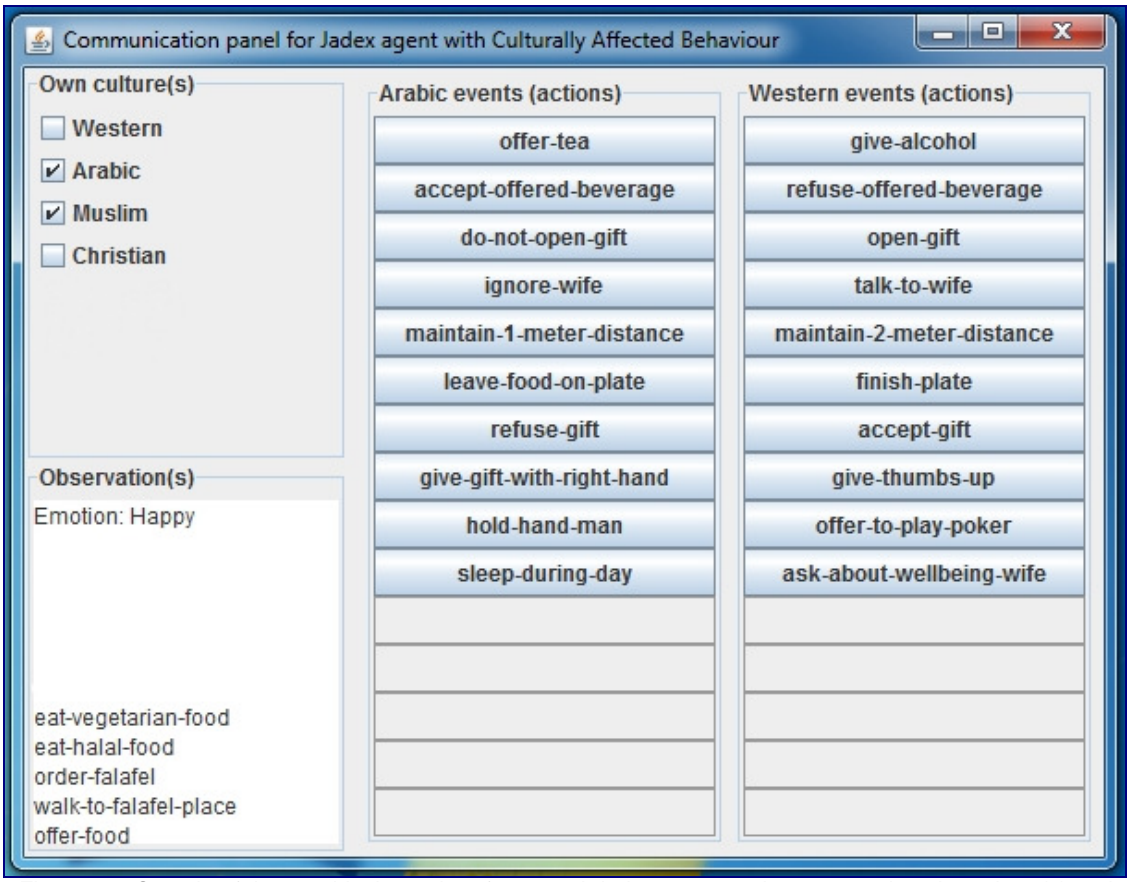


Figure 4-1. Sample of the graphical user interface used during our experiments to select the culture(s) and actions of the interaction partner. The Section '*Own culture(s)*' of the interface is used to select the culture(s) of the interaction partner. The Section '*Observation(s)*' are the actions of our agent and its shared emotional state. The Sections '*Arabic events (actions)*' and '*Western events (actions)*' contain the actions that will be performed by the interaction partner.

The culture and actions of the interaction partner were chosen through a graphical user interface. Besides allowing to choose the actions of the interaction partner this interface also presented the shared emotional state and the actions performed by our agent. A screen shot of our graphical user interface is shown in Figure 4-1.

In our experiment *all* of the interaction partner's actions were observed by our agent, which effectively made our agent's task environment *fully observable*. The *intended* task environment, however, will be only *partially observable* as interaction partners can perform actions that may not be observed by agents using our model. Likewise, our agent can also perform actions that may not be observed by our agent's interaction partners. This partial observability offers interesting possibilities which will be discussed in Section 5.4: Future Work.

Our agent's task environment was also *dynamic*, as the environment could change while the agent was deliberating, as well as *continuous*, as time was not stopped between actions of either agent (i.e. not segmented into *discrete* time steps). Our agent thus continuously deliberated about its beliefs and acted according to these belief states.

4.2 CULTURE DEFINITIONS

We used four different culture definitions during the evaluation of our model. These culture definitions can be found in Appendices III through VI and are based on actual cultural phenomena. Though the cultural definitions are based on cultural phenomena, they are not validated by any cultural expert(s) and may therefore contain errors, nor are these definitions by any means complete. The culture definitions are correct in the sense that they correctly represent the used actions being negatively or positively perceived within the particular culture. The extent of the effects described in the cultural norm schemata and the cultural values these have an effect on, however, are based on relative differences between the used cultures. For example, in Arabic cultures reputation is often more important that in Western cultures. To model this difference, we assigned higher intrinsic values to cultural value schemata based on the Theory of Mind in the Arabic culture than we did for Western culture. We also found Western cultures to be less expressive than Arabic ones and set the emotion thresholds accordingly. These cultural phenomena are not verified, and are implemented for demonstration purposes only. The used cultural norms, cultural values and emotion thresholds can be changed easily cultural experts.

4.3 CULTURAL NORMS AND VALUES

This section concerns the experiments regarding our incorporation of the CAB language into a BDI agent model. These experiments are aimed at the verification of the following features of the CAB language:

(1) The defined cultural norms and values can be used to calculate the agent's socio-cultural satisfaction (SCS),
(2) The agent should be able to distinguish between culturally desired and undesired actions using the human readable cultural norms and values, and
(3) The cultural norms and values should remain independent of other components in the agent model, allowing the agent to easily and quickly switch between definitions of cultural norms and values.

These experiments are also aimed at verifying our agent's ability to provide feedback regarding the height of its SCS through emotion. The scores on the personality dimensions are set at .5 and therefore do not have any influence during the experiments in this section.

4.3.1 EXPERIMENTAL SETUP

To evaluate our model's capability to easily and quickly switch between definitions of cultural norms and values, we assigned our agent different cultures through the Agent Definition File. We first assigned our agent an Arabic culture (see Appendix IV), and second a Western culture (see Appendix III). When assigned a culture we put our agent through two simulations during which the interaction partner performed actions corresponding to (1) an Arabic culture, (2) second, a Western culture (see Table 4-1). These sets of actions can be found in Appendices I and II, and were always performed in the same order.

Culture assigned to agent	*Culture corresponding to interaction partner's actions*	
Arabic	Arabic	Western
Western	Arabic	Western

Table 4-1. Cultural combinations used in evaluation of cultural component

Actions corresponding to the culture that was assigned to our agent should be perceived positively by our agent, resulting in higher degrees of belief in the affected cultural values and consequently increase our agent's SCS score. Most of the actions not corresponding to our agent's culture should be perceived negatively and lead to lower degrees of belief in the affected cultural values. These lower degrees of belief should result in our agent having a lower SCS score. The changes in the agent's SCS score in response to perceived actions will demonstrate our agent's capability to calculate its SCS using the cultural norm and value schemata. These, respectively, positive and negative changes in the agent's SCS score in response to perceived actions that are, respectively, in agreement or in conflict with our agent's cultural norms also demonstrate our agent's ability to distinguish between culturally desired and undesired actions using the human readable cultural norm and value schemata. As the SCS score changes, it also crosses the thresholds set for the shared emotional states and consequently lead to our agent sharing a different emotional state, verifying our agent's ability to provide feedback concerning the height of its SCS through emotion.

After some time (10 seconds) our agent will become hungry, and at that time will deliberate which plan best reduces its hunger. There are four applicable plans: '*eat falafel*', '*eat falafel with the other agent*', '*eat a hamburger*', and '*eat a hamburger with the other agent*'. Eating a hamburger best combats hunger, but contains pork (which is viewed as a negative thing in the Arabic culture, see figure 4-2). Eating with the interaction partner is considered to be kind (and valued in both cultures), but these plans require that our agent's SCS score is higher than .70.

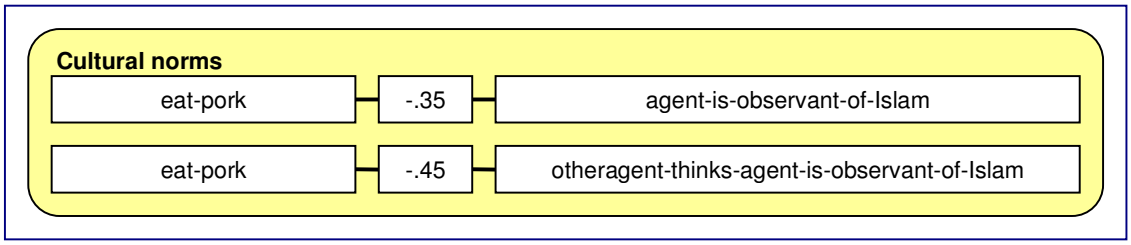


Figure 4-2. Arabic cultural norm schemata concerning eating pork

4.3.2 RESULTS

The results of the simulations during which our agent was assigned the Arabic culture can be seen in figures 4-3 and 4-4. The assignment of the Arabic culture was done by setting the culture parameter in the Agent Definition File to '*Arabic*'. Setting this parameter led to our model importing all the cultural norm and value schemata as well as the emotion thresholds belonging to the Arabic culture definition. The three horizontal lines represent the three emotion thresholds, the top one being the threshold between neutral and happy (set at .73 by the culture definition), the middle one between neutral and angry (.40), and the bottom one between angry and furious (.32).

The colours of the bars represent the shared emotional state of our agent, green means our agent shared the happy state, yellow meant neutral, red was angry, and black meant that the agent was furious and was no longer communicating with the interaction partner.

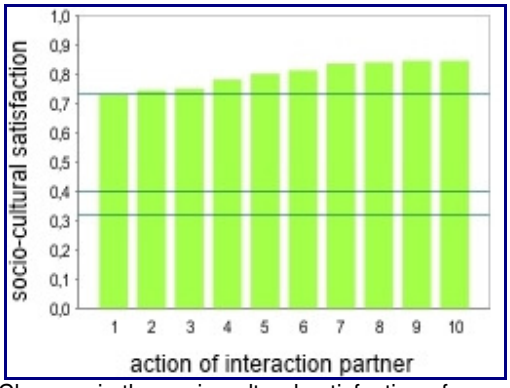


Figure 4-3. Changes in the socio-cultural satisfaction of an agent with the Arabic culture whilst observing Arabic actions of the interaction partner:

During the first simulation the interaction partner performed actions corresponding to an Arabic culture. As expected, our agent matched these actions against its cultural norm schemata and, as the actions were in correspondence with its culture, led to higher degrees of belief in the affected cultural values. Our agent's socio-cultural satisfaction consequently rose. Its socio-cultural satisfaction remained higher than the neutral-to-happy threshold throughout this simulation and therefore our agent only shared the happy emotional state. When our agent deliberated about which plan to choose to combat its hunger it chose the plan to eat falafel with the other agent (i.e., the interaction partner). The socio-cultural satisfaction was higher than .7, which was a requirement for this plan, and the plans to eat a hamburger received a lower utility level as it included eating pork. The eating pork action was matched against the agent's cultural norm

schemata (as shown in Figure 4-2), which define negative effects on the agent's cultural values and consequently having a negative influence on the utility of this plan.

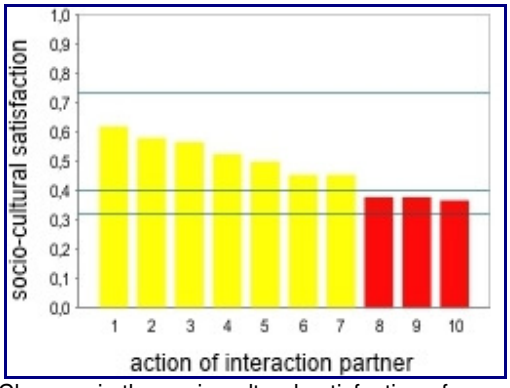


Figure 4-4. Changes in the socio-cultural satisfaction of an agent with the Arabic culture whilst observing Western actions of the interaction partner

The interaction partner confronted our agent with actions according to a Western culture during the second simulation. Most of these actions were matched against cultural norm schemata that defined negative effects on the degrees of belief in the corresponding cultural values. These lower degrees of belief led to a lower SCS score. The SCS score remained between the happy and angry thresholds up to the eight action of the interaction partner, after which the socio-cultural satisfaction fell below the neutral-to-angry threshold. Our agent therefore shared the neutral and angry emotional state respectively. As the SCS score was lower than the .7 requirement for the plan to eat falafel *with* the interaction partner our agent chose to eat falafel alone. This choice is, again, in correspondence with the culture's stance towards eating pork and consequently having cultural norm schemata that defined negative effects on the degree of belief in the corresponding cultural values.

During the third and fourth simulation, our agent was assigned the Western culture. This was done by changing the culture parameter in the Agent Definition File from '*Arabic*' to '*Western*'. Changing this parameter resulted in the cultural norms and values of the Western culture definition being imported by our model. The results of the simulations with the Western culture can be seen in figures 4-5 and 4-6. The thresholds were now set at .8 (neutral to happy), .35 (neutral to angry), and .3 (angry to furious) by the Western culture definition.

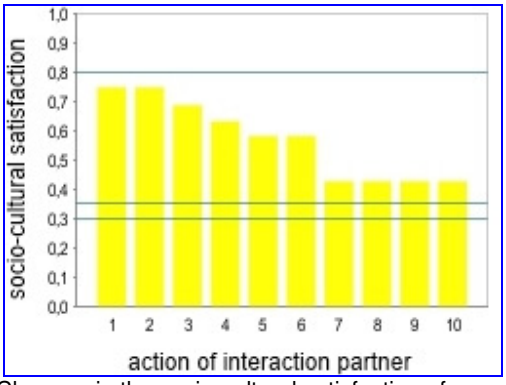


Figure 4-5. Changes in the socio-cultural satisfaction of an agent with the Western culture whilst observing Arabic actions of the interaction partner

As during the previous simulations, our agent responded negatively to actions that conflicted with its own cultural norms, resulting in lower degrees of belief in the affected cultural values. When confronted with actions corresponding to the Arabic culture, our agent's SCS score went down. As the socio-cultural satisfaction never fell below the neutral-to-angry threshold, the only emotional state our agent shared was the neutral one. The SCS score did not meet the .7 criterion for the plan to eat with another agent and therefore our agent chose a plan to eat alone. Eating a hamburger countered our agent's hunger most effectively and as its Western culture did not have cultural norms that defined the action of eating eating pork as a having a negative influence on its cultural values our agent chose to eat a hamburger.

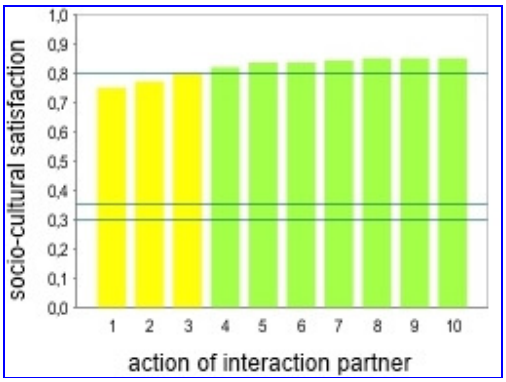


Figure 4-6. Changes in the socio-cultural satisfaction of an agent with the Western culture whilst observing Western actions of the interaction partner

Similarly to the previous simulations our agent responded positively to actions corresponding to its own culture and its SCS score consequently went up. Our agent shared the neutral emotional state until the fourth action, after which the SCS score rose above the neutral-to-happy threshold and the agent started sharing the happy emotional state. As the SCS score met the .7 criterion for eating with another agent and its Western culture had no (negative nor positive) cultural norms concerning the eating of pork our agent selected the plan to eat a hamburger with the interaction partner.

During all four simulations the socio-cultural satisfaction seemed to level off at the end. This is due to the degrees of belief in the affected cultural values already being at their maximum (one) or at their minimum (zero) level. The culture definitions used during our evaluation contained a very limited number of cultural values and were influenced by most of the cultural norms. This caused the degrees of belief in the cultural values to reach that maximum or minimum level fairly quickly. A realistic and complete definition of a culture will consist of a very large number of cultural value schemata and though an cultural norm of such a culture definition will influence more of these cultural values they will likely do so in more subtle manner. Such a culture definition would therefore not suffer such levelling off of the agent's SCS score.

4.4 MULTIPLE CULTURAL NORMS AND VALUES

As in the previous section, these experiments concern themselves with our incorporation of the CAB language into a BDI agent model. This section, however does so whilst also evaluating our implementation's ability to assign multiple (sub)cultures to a single agent.

Again, the scores on the personality dimensions are set at .5 and therefore do not exert any influence.

4.4.1 EXPERIMENTAL SETUP

To evaluate our model's ability to work with multiple cultures we assigned our agent two cultures during the following experiments: (1) one of the cultures we also used in the previous section, and (2) a religion. When assigned the Arabic culture we also assigned the Islamic culture (see Appendix VI), and when assigned the Western culture, the Christian culture (see Appendix V) was also assigned.

We put our agent through two simulations during which the interaction partner performed actions corresponding to the Western culture (always in the same sequence). During the first simulation we assigned a weight of .1 to the religion (i.e., the agent was not very religious) and .9 to the other culture. During the second simulation we did the opposite, i.e., .9 was assigned to the religion (the agent was very religious) and .1 to the other culture. As with the other culture definitions we used, the definitions of these two religions are not an accurate representation of the actual religions.

An important difference is that both religious cultural definitions perceive the actions '*give alcohol*' and '*offer-to-play-poker*' more negatively than the Arabic and Western culture definitions. This difference should influence the impact of these actions on the degrees of belief in the corresponding cultural values, and consequently the SCS score of our agent.

4.4.2 RESULTS

Figures 4-7 through 4-10 show the results of the experiments. The assignment of multiple cultures was done by setting the culture parameters in the Agent Definition File to the used cultures and their corresponding weights. Setting these parameters led to our model importing all the cultural norm and value schemata as well as the emotion thresholds belonging to the assigned culture definitions. The values for the emotion thresholds are different in all four culture definitions and therefore result in different thresholds being used by our agent in each of the four simulations. Higher weights being assigned to the religion cultures lowered all three thresholds and consequently made our agent more inclined to share the happy emotional state while also being less inclined to share the angry emotional state and become furious than when lower weights were assigned.

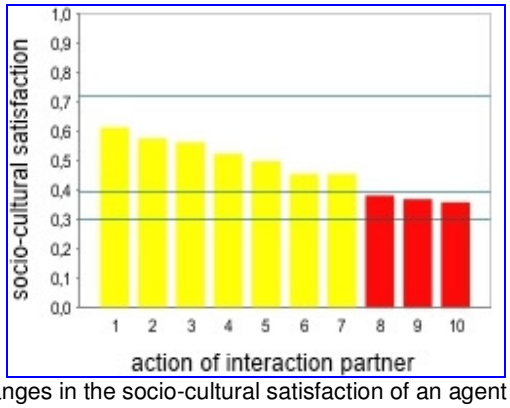


Figure 4-7. Changes in the socio-cultural satisfaction of an agent with a .9 Arabic and .1 Islamic culture whilst observing Western actions of interaction partner

The results of our experiment with the Arabic and Islamic culture can be seen in figures 4-7 and 4-8. In the former case the Islamic culture was assigned a weight of .1 (not very religious) and in the latter .9 (very religious). As expected, our agent's reaction to the actions '*give-alcohol*' and '*offer-to-play-poker*' was more severe when a higher weight was assigned to the religion culture as the religious cultures contained cultural norm schemata that defined a stronger decrease in the degrees of belief in the corresponding cultural values. The ninth action caused our agent to cross both the angry and furious thresholds simultaneously, resulting in our agent jumping from sharing a neutral emotional expression to a furious one, including the change in behaviour (i.e., no longer communicating with the interaction partner). The effects of this action were so severe that no 'advance warning' (the 'angry' emotion being shared) was given to the interaction partner regarding the rapid decline of our agent's SCS score. As all other actions had no effect on the cultural values they also had little effect on the socio-cultural satisfaction. Assigning more cultures to our agent, thus, can cause the same actions to be perceived differently, having different impacts on the degrees of belief in affect cultural values, resulting in different SCS scores and consequently lead to different responses.

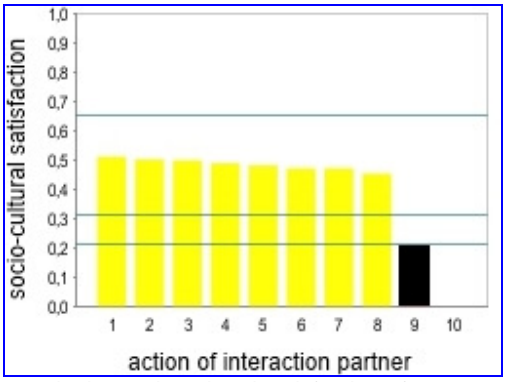


Figure 4-8. Changes in the socio-cultural satisfaction of an agent with a .1 Arabic and .9 Islamic culture whilst observing Western actions of interaction partner

Figures 4-9 and 4-10 show the results of our experiments with the Western and Christian culture. In the Western culture the actions '*give-alcohol*' and '*offer-to-play-poker*' are perceived positively, whilst in the Christian culture they are not.

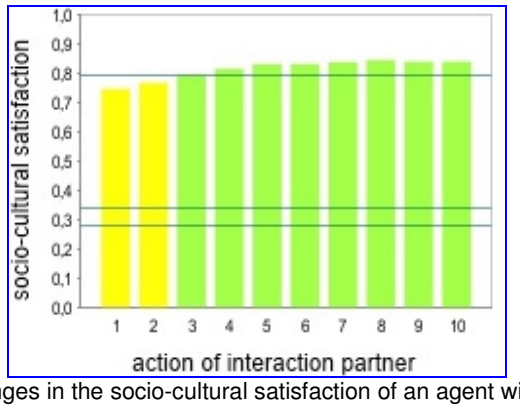


*Figure 4-9.*Changes in the socio-cultural satisfaction of an agent with a .9 Western and .1 Christian culture whilst observing Western actions of interaction partner

When the Christian culture is assigned a weight of .1 (not very religious) these two actions had only a minor impact on the degrees of belief in the affected cultural values and consequently on our agent's SCS score. However, when assigned a weight of .9 (very religious) they had a major impact on the degrees of belief in the affected cultural values. In the former case SCS score rose over the happy threshold whilst in the latter case our agent kept sharing only the neutral emotion.

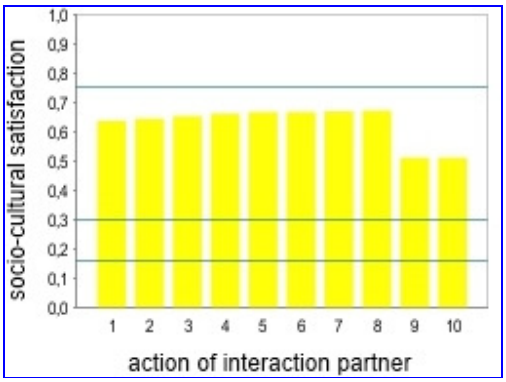


Figure 4-10. Changes in the socio-cultural satisfaction of an agent with a .1 Western and .9 Christian culture whilst observing Western actions of interaction partner

Though the interaction partner shares a culture (the Western culture) with our agent and its actions usually would be positively perceived by our agent, adding another culture (the Christian culture) can cause the same actions to be perceived negatively.

Our model, thus, is able to use multiple sets of cultural norms and values to calculate the agent's SCS score, as well as using these sets to distinguish between culturally desired and undesired actions using the cultural norms and values defined in the assigned culture sets.

4.5 PERSONALITY

This section describes the conducted experiments regarding the personality influences in our model having the desired effects as described in Section 2.4.

4.5.1 EXPERIMENTAL SETUP

As during our experiments in Section 4.3 our agent will be assigned a single culture through a single parameter in the Agent Definition File. We have chosen to use the Arabic culture, but the choice for this culture is irrelevant as the effects of the personality factors would be the same as when the Western culture would have been assigned.

Per personality dimension our agent will be put through two scenarios. During the first scenario the weight of the tested dimension will be set to one (the dimension's maximum), and during the second scenario to zero (its minimum). During both scenarios the interaction partner will perform actions corresponding to the Western culture. Our experiment with '*Extraversion*' included two additional scenarios; again with the weights set to its minimum and maximum, but this time the interaction partner also performed actions corresponding to the Arabic culture. This was done to elicit a rising SCS score and have the SCS score also cross the neutral-to-happy threshold.

4.5.2 RESULTS

OPENNESS

Figures 4-11 and 4-12 demonstrate the results of our experiments with personality dimension Openness. This dimension influences our agent's willingness to forgive cultural trespassing of the interaction partner *if* our agent believes the interaction partner has another culture *and* believes the interaction partner is acting according to this other culture, i.e. the actions are positively perceived in the interaction partner's culture. Figure 4-11 shows the results of our experiment with the weight of the dimension set to one (the dimension's maximum) and in figure 4-12 with the weight set to zero (the dimension's minimum). The short horizontal black bars in both these and most other figures in this section indicate the socio-cultural satisfaction would the dimension have had no influence. These bars thus indicate the height of our agent's SCS score as found in Section 4.3.

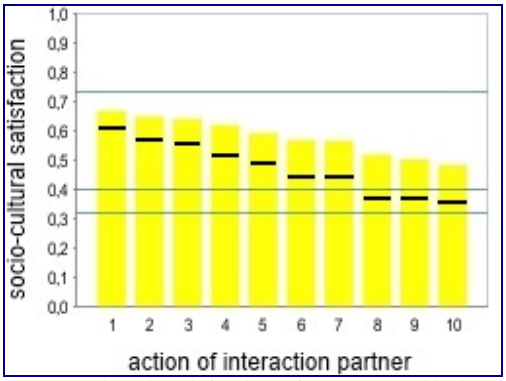


Figure 4-11. Changes in the socio-cultural satisfaction of an agent with the Arabic culture and maximum Openness whilst observing Western actions of interaction partner with Western culture

Figure 4-11 demonstrates that the higher score on the dimension Openness resulted in a slower decline of our agent's SCS score than would have been the case if the score on this dimension was .5 (no influence, indicated by the black bars). A higher score on the dimension Openness dampened any negative effects as defined in cultural norm schemata. These dampened effects resulted in a slower decline of our agent's degrees of belief in the affected cultural values and can be interpreted as our agent being more forgiving towards the cultural transgressions of the interaction partner and therefore more open to other cultures.

The opposite effect can be seen in Figure 4-12. This figure shows the results of the scenario where our agent had a score lower than .5 on the dimension Openness. In this case the effects as defined in cultural norm schemata of the interaction partner's actions were strengthened, resulting in lower degrees of belief in the affected cultural values and consequently to a lower SCS score. These results can be interpreted as our agent being less open to other cultures.

The effects of the dimension Openness in our model are in correspondence with our interpretation of this dimension in Digman's Five Factor Model (FFM). We found that being open to other cultures entails being more forgiving towards the actions of those having another culture even if these actions are negatively perceived in one's own culture. Our agent acted accordingly.

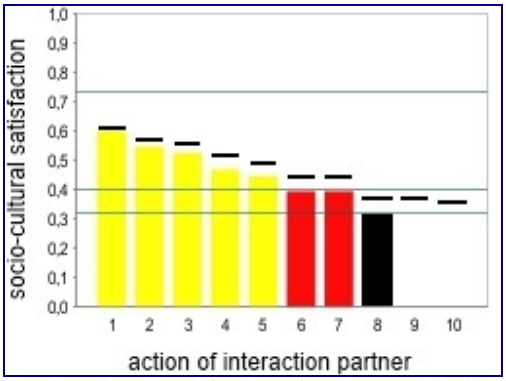


Figure 4-12. Changes in the socio-cultural satisfaction of an Agent with the Arabic culture and minimum Openness whilst observing actions of another agent with Western culture performing Western actions

EXTRAVERSION

The results of our experiments with personality dimension Extraversion can be seen in Figures 4-13 through 4-16. The score on this dimension influences the emotion thresholds of our agent, which can be seen in all of the aforementioned four figures. In all four figures the original thresholds, as defined by the agent's culture, are indicated by the light-blue horizontal lines. The thresholds as influenced by the agent's score on the dimension Extraversion are indicated by the black horizontal lines. As can be seen in Figure 4-13, the happy threshold was lowered (from .73 to .68) and the angry threshold was raised (from .4 to .45) when the agent's score on this dimension was set to one.

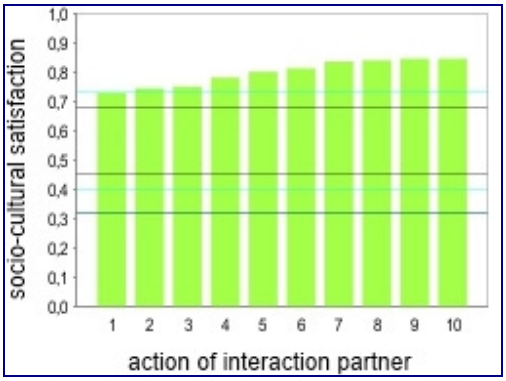


Figure 4-13. Changes in the socio-cultural satisfaction of an agent with the Arabic culture and maximum extraversion whilst observing actions of another agent with Arabic culture performing Arabic actions

In Figure 4-14 the score of zero on this dimension led to the opposite effect; the happy threshold was raised (from .73 to .78) whilst the angry threshold was lowered (from .4 to .35). Comparing the results in Figure 4-13 to the results in Figure 4-14 it can be seen that in both cases the development of our agent's SCS score were identical. However, in the former figure our agent quickly shared a happy emotional state with the interaction partner and in the latter figure our agent did not. It is after the fourth action of our interaction partner that our agent's socio-cultural satisfaction crossed the happy threshold in the latter case and it started sharing a happy emotional state instead of a neutral one.

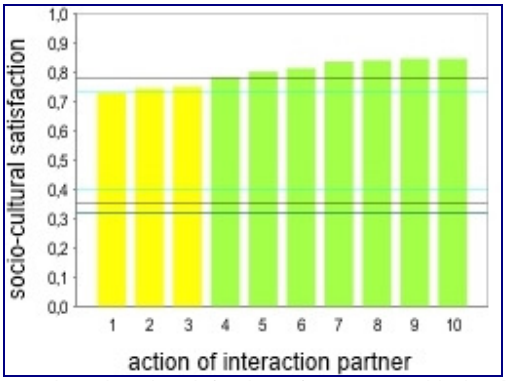


Figure 4-14. Changes in the socio-cultural satisfaction of an agent with the Arabic culture and minimum extraversion whilst observing actions of another agent with Arabic culture performing Arabic actions

The same effect can be seen in Figures 4-15 and 4-16. During these experiments the (Western) actions of the interaction partner had a negative impact on our agent's degrees of belief in the affected cultural values and consequently on our agent's socio-cultural satisfaction. Figure 4-15 shows the results of the experiment with our agent's score on the dimension Extraversion set to one, and in Figure 4-16 to zero.

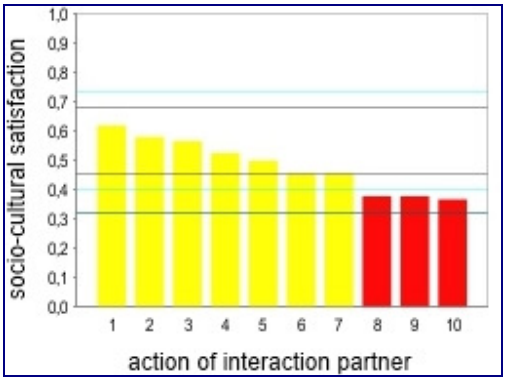


Figure 4-15. Changes in the socio-cultural satisfaction of an agent with the Arabic culture and maximum extraversion whilst observing actions of another agent with Western culture performing Western actions

As during the previous experiments with personality dimension Extraversion, the development of our agent's SCS was identical during both scenarios. However, in Figure 4-15 our agent shared the angry emotional state with the interaction partner after the seventh action of the interaction partner and in Figure 4-16 our agent shared a neutral emotional state throughout the simulation.

A high score on the personality dimension Extraversion led to our agent sharing an emotional state other than the neutral one quicker, whilst a low score had the opposite effect. Our agent was thus more expressive when having a high score on this dimension. This effect is in correspondence with our interpretation of Digman's Five Factor model regarding the dimension Extraversion.

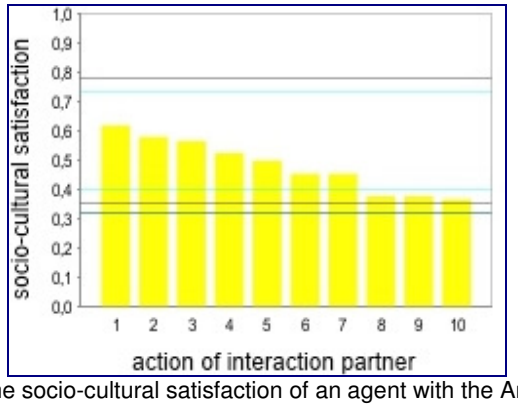


Figure 4-16. Changes in the socio-cultural satisfaction of an agent with the Arabic culture and minimum extraversion whilst observing actions of another agent with Western culture performing Western actions

AGREEABLENESS

The influence of personality dimension Agreeableness is identical to that of personality dimension Openness, but the latter dimension does not require our agent to believe that the interaction partner has another culture and is acting according to the norms and values of that culture. Figures 4-17 and 4-18 show the results of our experiments with this dimension. In Figure 4-17 the results are shown when our agent had a score of one on this dimension, and in Figure 4-18 with our agent having a score of zero.

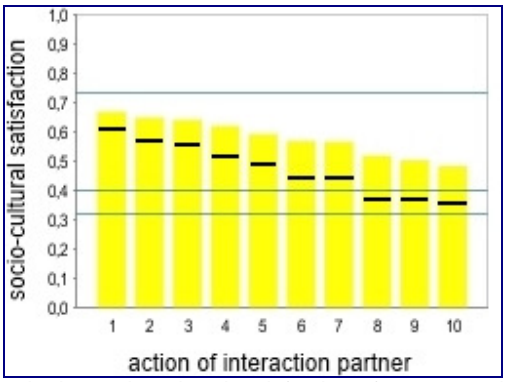


Figure 4-17. Changes in the socio-cultural satisfaction of an agent with the Arabic culture and maximum agreeableness whilst observing actions of the interaction partner with the Arabic culture, but performing non-Arabic *actions*

As the effects of the dimension Agreeableness are identical to that of the dimension Openness, the results in figures 4-17 and 4-18 are identical to those in figures 4-11 and 4-12. However, unlike during our experiments with the dimension Openness, the interaction partner now has the *same* culture as our agent whilst performing actions that do *not* correspond to their shared culture. Sharing the same culture would result in the dimension Openness having no influence. The dimension Agreeableness does not have any requirements regarding the interaction partner's culture. When comparing the results in figures 4-17 and 4-18 it can be seen that the former shows a slower decline of our agent's SCS score than in the latter figure. A high score on the dimension Agreeableness thus made our agent more forgiving towards negatively perceived actions, whilst a low score on this dimension amplified any negative effects defined in its cultural norm schemata. This effect led to our agent having, respectively, a higher and lower SCS score. As the SCS score is an indication of social harmony between two

agents we found the reported effects to be in correspondence with our interpretation of Digman's Agreeableness dimension.

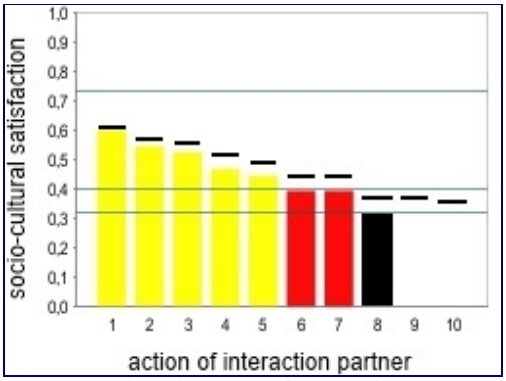


Figure 4-18. Changes in the socio-cultural satisfaction of an agent with the Arabic culture and minimum agreeableness whilst observing actions of another agent with the Arabic culture, but performing non-Arabic actions

NEUROTICISM

Figures 4-19 and 4-20 show the results of our experiments with personality dimension Neuroticism. As mentioned in section 2.4, this dimension influences (1) the effects (as defined in the cultural norm schemata) of another agent's actions, and (2) the threshold between angry and furious. During the first experiment with this dimension the score was set to one and the threshold consequently rose from .32 to .37, as can be seen in figure 4-19. The results in the same figure also demonstrate that with the score set to one the agent amplified any negative effects as defined in the agent's cultural value schemata on the degrees of belief of corresponding cultural values. The effects were amplified so strongly that the risen threshold had no influence on when our agent became furious; the eighth action caused the SCS score to fall below the original (indicated in blue) as well as the updated threshold. An agent with a higher score on this dimension thus experiences cultural transgressions more negatively and together with our agent's tendency to become furious more quickly it can be construed that our agent has a tendency towards negative emotions. This effect is in correspondence with our interpretation of the dimension Neuroticism of Digman's Five Factor Model.

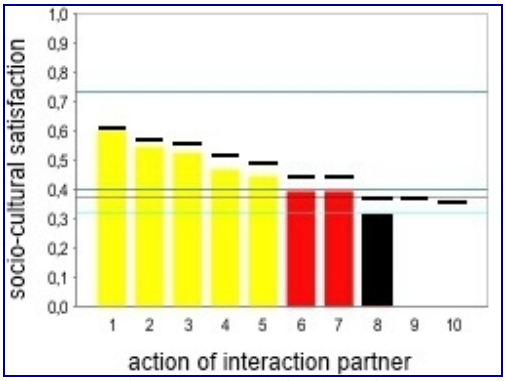


Figure 4-19. Changes in the socio-cultural satisfaction of an agent with the Arabic culture and maximum neuroticism whilst observing actions of another agent with the Arabic culture, but performing non-Arabic actions

A score of zero on this dimension leads to the opposite effect. During the second experiment the lower score on the dimension Neuroticism led to the furious threshold being lowered from .32 to .27. The results of this experiment can be found in Figure 4-20. The low score on this dimension also caused any negative effects defined in the agent's cultural norm schemata to be dampened, resulting in a slower decrease in our agent's degrees of belief in the affected cultural values. This led to a slower decline of our agent's S and consequently our agent kept sharing a neutral emotional state throughout the experiment. This contrasts with the results of the previous experiment during which our agent crossed the angry threshold after the fifth action of the interaction partner and became furious during the eighth.

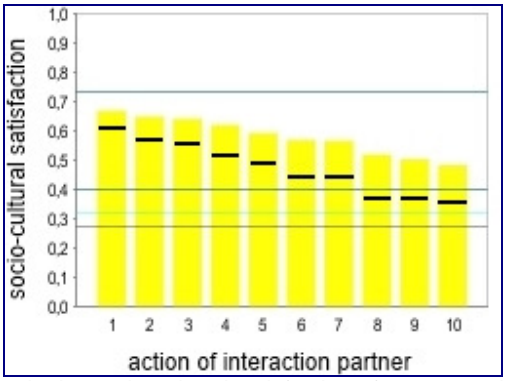


Figure 4-20. Changes in the socio-cultural satisfaction of an agent with the Arabic culture and minimum neuroticism whilst observing actions of another agent with the Arabic culture, but performing non-Arabic actions

All of the four implemented personality dimensions had the desired effects on either (1) the thresholds or (2) effects as defined in the cultural norm schemata, and consequently the agent's SCS score. These effects were in agreement with our interpretation of the personality dimensions of Digman's (1990) Five Factor Model as described in Section 2.4.

5 DISCUSSION

This chapter gives a discussion of the results of our experiments. The next section, Section 5.1, discusses our implementation of the Culturally Affected Behaviour (CAB) language into a Belief-Desire-Intention (BDI) model, followed by a discussion of the personality influences in Section 5.2. Limitations of our research effort are presented in Section 5.3, followed some suggestions regarding future work in Section 5.4.

5.1 CULTURE

The CAB language by Solomon *et al.*'s (2008) provided an effective approach to incorporating culture into Artificial Intelligence agents. Incorporating the cultural norm and cultural value schemata proved straightforward using Jadex and Solomon *et al.*'s (2008) CAB approach was successfully translated to the BDI architecture. This translation allowed us to create representations of cultural norms and values that could be interpreted by BDI agents whilst being human readable at the same time.

The cultural component of our model was able to use the cultural norms and values, represented through schemata, to calculate an agent's socio-cultural satisfaction (SCS) in a similar fashion as Solomon *et al.* (2008). The positive and negative effects, as defined in the cultural norm schemata, on the degree of belief in cultural values and consequently the agent's SCS score allowed our agent to distinguish between culturally desired and undesired actions. The agent using our model was able to make this distinction both concerning its own actions, during the plan selection process, and concerning another agent's actions. Our translation of the CAB language also allowed for the cultural representations to be changed whilst not influencing other components of our model.

Unlike the model of Solomon *et al.* (2008), our model also provided the possibility of assigning multiple (sub)cultures to single agents and assigning each of these cultures a relative weight. This possibility allows us to model (sub)cultures once and reuse them whenever necessary. Though this allows for more diversity between agents in training simulations, our approach towards the interaction between (sub)cultures is a very simplistic one and agents using this feature may therefore not correctly model the behaviour of people with the same (sub)cultures. This results in a tradeoff between (1) fine-tuning the used cultural representations, which can be a time-consuming task, and (2) accepting the incorrect behaviour, but saving time by reusing existing cultural representations.

Our experiments demonstrated that our agent was able to select those plans that conformed to its own cultural norms and values. The first requirement for behaving according to its own culture is that the plans available to the agent contain actions that, according to the agent's culture, are either positively or negatively perceived. The second requirement is that these actions impact the degrees of belief in the affected cultural values sufficiently to result in the correct plan having the highest utility level and consequently be chosen.

Though this approach sufficed in our experiments with only a single culture, we have not conducted any experiments regarding plan selection with agents assigned multiple (sub)cultures. As the (sub)cultures are weighted, the impacts of the actions in the plans vary according to the weights of these (sub)cultures. This influences the plan selection

process, making it less predictable and may even result in incorrect representation of the behaviour of people with the same (sub)cultures.

One alternative to the current approach could be limiting the plans available to our agent to those that correspond to the culture(s) of our agent. This alternative could prove to be more robust and transparent than our current approach.

During our experiments our agent knew the culture of the other agent and the cultural norm schemata corresponding to the culture of the other agent. This led to our agent being aware of the other agent's actions being in correspondence to the culture of the other agent or not. Our agent thus could only ascertain if individual actions of the other agent were in correspondence with the culture of the other agent and take that fact into account when processing the effects of the other agent's actions. Though not tested, our agent could also be made aware of the other agent having multiple (sub)cultures. If such would have been the case our agent would be aware of the cultural norm schemata of all the cultures of the other agent. Our agent, however, was not aware of the weights assigned to other agent's (sub)cultures and therefore could not take into account distinctions such as the other agent being very or hardly religious.

Feedback concerning the height of our agent's SCS score was given by our model through the sharing of distinct emotional states. Besides a neutral state, the emotional states we used were '*happy*', '*angry*', and '*furious*'. Continuously violating the cultural norms of our agent resulted in our agent sharing an angry emotional state, whilst behaving according to our agent's culture resulted in our agent sharing a happy emotional state. The use of flexible thresholds which are defined in the culture definition allows for the modelling of more introvert and extrovert cultures.

Our current approach uses a limited number of emotions. Humans are able to show and interpret many more types of emotions, such as sadness, hope, and fear. Adding such emotions to our model would be impractical as we used pre-defined thresholds along a single dimension. Using a multi-dimensional model, such as the OCC model of emotions by (Ortony, Clore, & Collins, 1988), can overcome such complications, but may make the interpretation of the provided feedback more complicated.

5.2 PERSONALITY

Of Digman's five factors we implemented four in our model. The effects of our implementation were in correspondence with our interpretation of Digman's Five Factor model as described in Section 2.4. Though our approach does not model the full extent of the influence of personality on human reasoning and decision making it was capable of creating the behaviour diversity our approach was intended to create. This behavioural diversity was easy to create as it could be easily done through changing our agent's scores on the four dimensions. The changed scores manipulated (1) the effects as defined in the cultural norm schemata and (2) the thresholds for the shared emotional state. These manipulations did not determine the agent's behaviour, but did influence it, and did so without the need for any changes to be made in the cultural definitions.

5.3 LIMITATIONS

This section presents the limitations of our research effort. First the limitations of our model are discussed, followed by the limitations of our experiment.

5.3.1 MODEL LIMITATIONS

The first limitations of our model can be found in the personality component; it is rather limited and should not be considered to be an accurate representation of these processes and their influence on human reasoning and decision making. However, as stated in Chapter 1, it was never our intention to create a correct representation of these phenomena, but merely as a method to create diversity between agents and therefore less predictable behaviour of these agents. It also was our intention to create such behavioural diversity by influencing the agent's cultural norms and values as defined through the CAB language.

Another limitation of our model was that it lacked the personality dimension '*conscientiousness*', which is in integral part of Digman's Five Factor model (Digman, 1990). As described in Section 2.4, this personality dimension can be used to model an agent's commitment to its currently selected plans. However, our model currently does not revaluate its current plans at all. It selects a plan during plan selection and remains committed to it. In Section 5.4.3 we elaborate on our suggestion to add the personality dimension 'conscientiousness' to our model.

5.3.2 EXPERIMENT LIMITATIONS

During every simulation our agent was confronted with only ten actions performed by another agent. In reality a person is confronted with a much larger number of events that could influence a person's satisfaction with an interaction. A more elaborate scenario should be used to ascertain the functioning of our model when confronted with more events over a longer period of time. Such a scenario would also require more elaborate culture definitions.

During our experiment we used culture definitions that contained rather a limited number of cultural value and cultural norm schemata. Though based on existing cultural phenomena, they should not be considered accurate representations of the cultures. Our suggestions regarding culture definitions can be found in Section 5.4.5.

A final limitation of our experiments is that it did not include user and expert evaluation. Such an evaluation is a necessity to validate our model's capability of displaying convincing behaviour. Such evaluations, however, were beyond the scope of this research effort.

5.4 FUTURE WORK

In the previous sections we already touched upon a few suggestions for future work. These, and additional, suggestions will be presented in this section.

5.4.1 EMOTION THRESHOLDS

In our current research effort we used thresholds to define which emotional state ('*happy*', '*neutral*', '*angry*', or '*furious*') would be shared with other agents. Though this approach is easy to implement and was sufficient for the limited scenarios as used during our experiments, it can reduce the persuasiveness of the simulation when the SCS score is fluctuating around a single threshold. In such case the threshold would be crossed several times in a short period of time resulting in our agent continuously changing the emotional state it is sharing.

A simple approach to tackle this problem would be to introduce a time requirement that prevents sharing an earlier shared emotion again for a fixed period of time. Another approach would to be use fuzzy logic. Fuzzy logic allows us to blur the emotion thresholds. Instead of our agent having a specific emotional state or not (i.e., boolean), it could hold several emotional states at the same time, but with differing degrees. In other words, each emotional state can be regarded to be a set to which our agent can hold a certain level of membership. Through a 'defuzzification process' it can be ascertained which set our agent belongs the most (i.e. selecting a single emotional state that will be shared).

5.4.2 MULTIPLE AGENTS AND EMOTION

During our experiments our agent was confronted with the actions of only one other agent. During training simulations it is likely our agent will be confronted with the actions of multiple agents. In our current implementation each interaction with another agent has its own SCS score. As the shared emotional state is based on the SCS score the shared emotional state will be different for each agent. If interactions would shift between multiple agents our agent would also change its shared emotional state according to the SCS score belonging to current interaction. This might result in the same behaviour as described in the previous section: our agent continuously changing the emotional state it is sharing. Such behaviour is undesirable.

One solution could be to average the scores of all the individual socio-cultural satisfactions scores. This solution, however, could result in our agent sharing an angry emotional state while interacting with an agent that has not violated any of our agent's cultural rules. Such a situation could be confusing to human agents interacting with our agent and consequently reduce the persuasiveness of the simulation. However, weighing the individual SCS scores and assigning the SCS score of the interaction our agent is currently engaged in could alleviate some of these problems.

5.4.3 CONSCIENTIOUSNESS

In Section 3.3.2 we already gave a suggestion as to how the personality dimension '*conscientiousness*' could be used. We suggested that this dimension could be used to model our agent's stance towards planned behaviour. The score on this dimension can increase or decrease the utility of plans our agent is currently committed to. This, in turn, makes our agent, respectively, less or more inclined to switch to alternative plans when revaluating plans. An example of the aforementioned suggestion can be found in Figure 5-1.

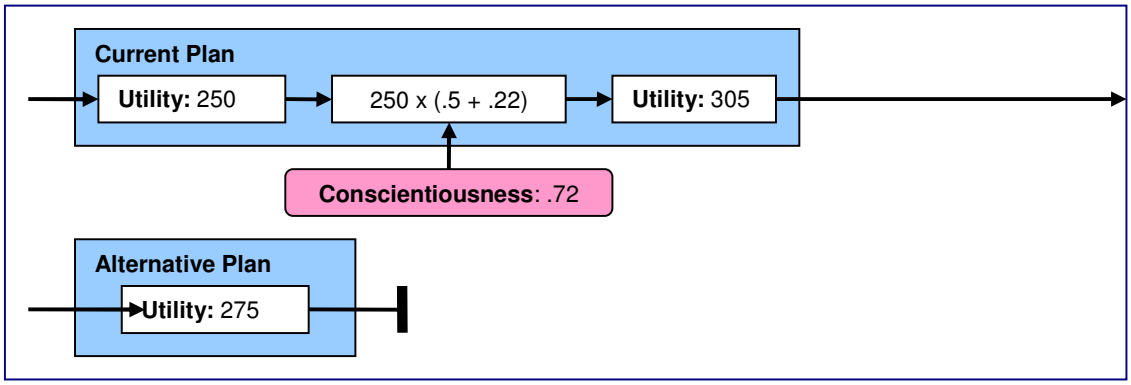


Figure 5-1. Example of the influence of personality dimension Conscientiousness

Currently, our model does not revaluate its plans. Our agent remains committed to a selected plan until completed. Such functionality would require adding meta-data to the individual actions within the plans in order to make the purpose of these actions comparable with the actions of other plans. The value of the individual actions within the plan also needs to be known. For example, different values should be attributed to the action '*go to location x*' when being near location x and when not.

5.4.4 AWARENESS OF PERCEPTION

Our model uses two types of cultural value schemata. The first describing how the agent views itself and the second type describes how the agent thinks other agents view the agent. Currently our agent always assumes that the other agent can perceive the actions of our agent even though this may not be the case. In cultures where reputation and face-saving are important the perception of one's own actions by others is important. The course of action to be selected thus may depend on these actions being perceived, or at least on these actions *believed* to be perceived. We believe that taking into account the perception of actions could be a useful extension of our model as it creates another source of behavioural diversity.

A simple approach to achieve this could be done through an extension of the plan selection process. Two extra steps should be added to this process. The first step is ascertaining if the other agents can perceive our agent's actions or not. The second step is to assign weights to both types of cultural values and assigning a lower weight to cultural norm schemata that concern the second type of cultural values when our agent believes its actions are not perceived by others.

5.4.5 VALIDATION OF CULTURES, MODEL, AND PERSONALITY

During our experiments we used limited culture definitions that should not be considered to be correct representations of any culture. A logical next step in experimenting with our model is to use culture representations that are validated by cultural experts.

Next, these validated culture representations should be used during experiments with human users. An example of such a user evaluation can be found in Solomon *et al.* (2009). Their experiment validated their model as a cultural awareness training tool. This user evaluation consisted of three parts. During the first part the participants of their experiment were presented with an example of a complete meeting with an agent assigned a specific culture. The second part entailed interacting with another agent

which was assigned the same culture as during the first part of the experiment. The last part of this user evaluation consisted of a judgment survey during which participants rated actions as being '*good*' or '*bad*' (from the other agent's cultural point of view). A similar experiment using our model should be conducted to validate our model as a cultural awareness training tool.

Lastly, after the culture representations and the cultural component of our model have been validated, the personality influences should be validated. This can be done by changing the scores on the dimensions of Digman's Five Factor Model (Digman, 1990), have respondents interact with the model and allow them to describe the personality of our agent.

The work suggested in this section will entail a time-consuming process and therefore fell beyond the scope of our current research effort. Nonetheless; the suggested work is required to validate our own model.

5.4.6 UPDATING DEGREES OF BELIEF

Our current model updates the degrees of belief in cultural values by adding it to or subtracting it from the current degree of belief, with the resulting degree of belief having a maximum of one or a minimum of zero. Other approaches, such as using Bayesian statistics, may provide more realistic updating of the degrees of belief. A more elaborate comparison of belief maintenance methods can be found in Heuvelink (2009).

6 CONCLUSIONS

This Chapter provides answers to the research questions as formulated in Section 1.1.

RQ 1. To what extent can the Culturally Affected Behaviour language be incorporated into the Belief-Desire-Intention model?

In Section 1.1 we stated that a successful incorporation of the Culturally Affected Behaviour (CAB) language into a Belief-Desire-Intention (BDI) model should entail the CAB language's most important features:

(1) The defined cultural norms and values can be used to calculate the agent's socio-cultural satisfaction.

The implementation of the CAB language in our model was able to use the cultural norm and cultural value schemata to calculate the agent's socio-cultural satisfaction (SCS) in a similar fashion as Solomon *et al.*'s (2008). Agents using our model could calculate its SCS score both when (1) assigned a single culture and (2) when assigned multiple cultures with varying weights. Our model also provided feedback regarding the height of its SCS score to other agents. This feedback was given through sharing distinct emotional states ('*happy*', '*neutral*', '*angry*', and '*furious*'). Each emotional state was shared when the SCS score crossed a pre-defined threshold. These thresholds were defined in the used cultural definitions and thus varied per culture.

(2) The defined cultural norms and values should be human readable.

Our model uses schemata to describe cultural norms and cultural values using human readable strings consisting of dictionary words, such as '*eat-pork*' and '*agent-is-kind*', while effects and the intrinsic utility of a cultural value are numerically expressed. The CAB language and our implementation of this language does not use on any form of logic which would have made the representation of cultural norms and values less straightforward to understand.

(3) The agent should be able to distinguish being culturally desired and undesired actions using the human readable cultural norms and values.

Our model was successful in using the human readable cultural norm and value schemata to make a distinction between culturally desired and undesired actions when (1) interpreting the actions of another agent, and (2) during its own plan selection process. In the former case, culturally desirable actions would result in a higher SCS score, whilst culturally undesirable actions resulted in a lower SCS score. These changes in our agent's SCS score influenced the availability of plans during the plan selection process. During the plan selection process, the cultural norm and value schemata were used to determine the cultural desirability of individual actions in a plan. Plans that were in agreement with the agent's culture were chosen over plans that were in conflict with our agent's culture.

(4) The cultural norms and values should remain independent of other components in the agent model, allowing the agent to easily and quickly switch between definitions of cultural norms and values.

Cultural norms and values were assigned to an agent using a limited set of parameters in the Agent Definition File, resulting in our model importing the cultural norm and value schemata from a separate file. The imported cultural norm and value schemata were stored as beliefs of the agent. These beliefs could be changed or replaced with other cultural norm and value schemata without requiring other changes in the agent being made.

RQ 2. How can personality influence the cultural interpretations of a culturally aware Belief-Desire-Intention agent?

As mentioned in Section 1.1, the extent of the influence of personality we seek is not to determine an agent's behaviour, but to influence the interpretation of the agent's cultural norms and values as defined through the CAB language.

We modelled our agent's personality along the five dimensions as described in Digman's (1990) Five Factor Model. The personality dimensions '*Openness*', '*Agreeableness*', and '*Neuroticism*' influenced the effects, as defined in the cultural norm schemata, on the affected cultural values. These dimensions, thus, influenced our agent's interpretation of its cultural norms and values as defined through our implementation of the CAB language. The personality dimensions '*Extraversion*' and, again, '*Neuroticism*' also influenced the thresholds used for the feedback through emotion. Though these thresholds were not a part of the original CAB language, they were directly linked to our agent's SCS score which is a part of the CAB language.

The influence of the scores on the personality dimensions did lead to the sought increase behavioural diversity. Agents that were assigned the same cultures, but differed on their scores on the personality dimensions demonstrated behavioural diversity. Each of the four used personality dimensions allows for any score between zero and one resulting in a large number of possible variations and consequently making agents using our model less predictable.

Our approach towards the incorporation of the CAB language into the BDI model proved successful. We were able to incorporate the CAB language whilst making only minor adjustments to the language. Our usage of the SCS score during the plan deliberation/selection process differed greatly from that of Solomon *et al*. (2008), but this was due to the differences between the Virtual Humans/Soar and BDI agent architectures. We also successfully demonstrated how a personality defined along the five dimensions of the Five Factor Model (Digman, 1990) can be incorporated into a BDI agent that has been made culturally aware through the CAB language.

7 REFERENCES

2APL. (2011). 2APL: A Practical Agent Programming Language. Retrieved 2011/05/06, 2011, from http://apapl.sourceforge.net/

Adam, C. (2007). *The Emotions: From Psychological Theories to Logical Formalization and Implementation in a BDI Agent.* Unpublished PhD, Institut National Polytechnique de Toulouse, Toulouse.

Andr, E., Klesen, M., Gebhard, P., Allen, S., & Rist, T. (2000). Integrating models of personality and emotions into lifelike characters. In *Affective interactions* (pp. 150-165): Springer-Verlag New York, Inc.

Bartlett, S. F. C. (1995). *Remembering: a study in experimental and social psychology* (2 ed.). Cambridge, UK: Cambridge University Press.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM, 37*(7), 122-125.

Bhattacharyya, S. (1999). *Comparing Soar and BDI architectures.* Paper presented at the Proceedings of the 19th Soar Workshop.

Bogdanovych, A., Rodriguez, J. A., Simoff, S., & Cohen, A. (2009). *Virtual Agents and 3D Virtual Worlds for Preserving and Simulating Cultures*. Paper presented at the Proceedings of the 9th International Conference on Intelligent Virtual Agents.

Bosse, T. & Zwanenburg, E. (2009). *There's always Hope: Enhancing Agent Believability through Expectation-Based Emotions.* Paper presented at the International Conference on Affective Computing and Intelligent Interaction.

Bratman, M. (1987). *Intention, plans, and practical reason*: Harvard University Press.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*(2), 207-230.

Bulitko, V., Solomon, S., Gratch, J., & Lent, M. V. (2008). Modeling Culturally and Emotionally Affected Behavior. *Artificial Intelligence and Interactive Digital Entertainment Conference*.

Cambridge Advanced Learner's Dictionary. (2008) (3rd Revision edition ed.). Cambridge, UK: Cambridge University Press.

Costa, P. T., McCrae, R. R., & Psychological Assessment Resources, I. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*: Psychological Assessment Resources.

D'Andrade, R. G., & Strauss, C. (1992). *Human motives and cultural models* (Illustrated reprint ed.). Cambridge, UK: Cambridge University Press.

Digman, J. M. (1990). Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology, 41*(1), 417-440.

Egges, A., Kshirsagar, S., & Magnenat-Thalmann, N. (2003). A Model for Personality and Emotion Simulation. In *Knowledge-Based Intelligent Information and Engineering Systems* (Vol. 2773, pp. 453-461): Springer Berlin / Heidelberg.

Egges, A., Kshirsagar, S., & Magnenat-Thalmann, N. (2004). Generic personality and emotion simulation for conversational agents: Research Articles. *Comput. Animat. Virtual Worlds, 15*(1), 1-13.

Foundation for Intelligent Physical Agents (2002). FIPA ACL Message Structure Specification. *SC00061G* Retrieved 2011/08/03, 2011, from http://www.fipa.org/specs/fipa00061/SC00061G.html

Foundation for Intelligent Physical Agents (2011). Welcome to FIPA! Retrieved 2011/08/02, 2011, from http://www.fipa.org/

Heuvelink, A. (2009). *Cognitive Models for Training Simulation.* Unpublished PhD, VU University Amsterdam, Amsterdam.

Hofstede, G. H., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: software of the mind* (3rd ed.). New York: McGraw-Hill.
Johnson, W. L., Beal, C., Fowles-Winkler, A., Lauper, U., Marsella, S., Narayanan, S., *et al*. (2004). Tactical Language Training System: An Interim Report Intelligent Tutoring Systems. In (Vol. 3220, pp. 336-345): Springer Berlin / Heidelberg.
Lent, M. v., Core, M., Rosenberg, M., McAlinden, R., Carpenter, P., Solomon, S., et al. (2007). Culturally-Affected Human Behavior Modeling and Its Applications to Serious Games. *The Bridge, 37*(4).
Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in `theory of mind'. *Trends in Cognitive Sciences, 8*(12), 528-533.
Mascarenhas, S., & Paiva, A. (2010). *Creating Virtual Synthetic Cultures for Intercultural Training.* Paper presented at the International Conference on Intelligent Tutoring Systems, Pittsburg, USA.
McCrae, R. (1996). Social Consequences of Experiential Openness. *Psychological Bulletin, 120*(3), 323-337.
National Academy of Engineering of the National Academies (2008). *Frontiers of engineering: reports on leading-edge engineering from the 2007 symposium*: National Academies Press.
Norling, E. (2004). *Folk Psychology for Human Modelling: Extending the BDI Paradigm*. Paper presented at the Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1.
Ortony, A., Clore, G. L., & Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge, UK: Cambridge University Press.
O'Sullivan, C. S., & Durso, F. T. (1984). Effect of schema-incongruent information on memory for stereotypical attributes. *Journal of Personality and Social Psychology, 47*(1), 55-70.
Pokahr, A., & Braubach, L. (2009). *From a Research to an Industrial-Strength Agent Platform: Jadex V2.* Paper presented at the 9. Internationale Tagung Wirtschaftsinformatik (WI 2009).
Pokahr, A., & Braubach, L. (2010a, 2010/02/01). Overview. from http://jadex-agents.informatik.uni-hamburg.de/xwiki/bin/view/About/Overview
Pokahr, A. & Braubach, L. (2010b, 2010/03/24). Concepts of the Jadex BDI Reasoning Engine. Retrieved 2011/04/24, 2011, from http://jadex-agents.informatik.uni-hamburg.de/xwiki/bin/view/BDI+User+Guide/02+Concepts
Rao, A. S., & Georgeff, M. P. (1995). *BDI-agents: from theory to practice.* Paper presented at the Proceedings of the First Intl. Conference on Multiagent Systems.
Rehm, M., Bee, N., Endrass, B., Wissner, M., & André, E. (2007). *Too close for comfort?: adapting to the user's cultural background.* Paper presented at the International workshop on Human-centered multimedia, Augsburg, Bavaria, Germany.
Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3rd ed.): Prentice Hall.
Solomon, S., Van Lent, M., Core, M., Carpenter, P., & Rosenberg, M. (2008). *A Language for Modeling Cultural Norms, Biases and Stereotypes for Human Behavior Models.* Paper presented at the Proceedings of the 17th Behavior Representation in Modeling & Simulation (BRIMS) Conference, Providence, RI.

Solomon, S., Hays, M., Chen, G., & Rosenberg, M. (2009). *Evaluating a Framework for Representing Cultural Norms for Human Behavior Models.* Paper presented at the Proceedings of the 18th Conference on Behavior Representation in Modeling and Simulation, Sundance, UT, 31 March - 2 April 2009.

Swartout, W., Gratch, J., Hill, R. W., Hovy, E., Marsella, S., Rickel, J., et al. (2006). Toward virtual humans. *AI Mag., 27*(2), 96-108.

Sun, R. (2006). *Cognition and multi-agent interaction : from cognitive modeling to social simulation*. New York: Cambridge University Press.

Taylor, G., Quist, M., Furtwangler, S., & Knudsen, K. (2007). *Toward a Hybrid Cultural Cognitive Architecture.* Paper presented at the 29th Annual Conference of the Cognitive Science Society, Nashville, Tennessee, USA.

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: attributed mental states influence perception. *Trends in Cognitive Sciences, 14*(8), 376-382.

Weiss, T. G. (1999). *Military-Civilian Interactions: Intervening in Humanitarian Crises*. Plymouth: Rowman & Littlefield Publishers, Inc.

Whitney, P., Neil, J. S., & Paul, B. B. (2001). Schemas, Frames, and Scripts in Cognitive Psychology. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 13522-13526). Oxford: Pergamon.

Wooldridge, M. J. (2000). *Reasoning about rational agents*: MIT Press.

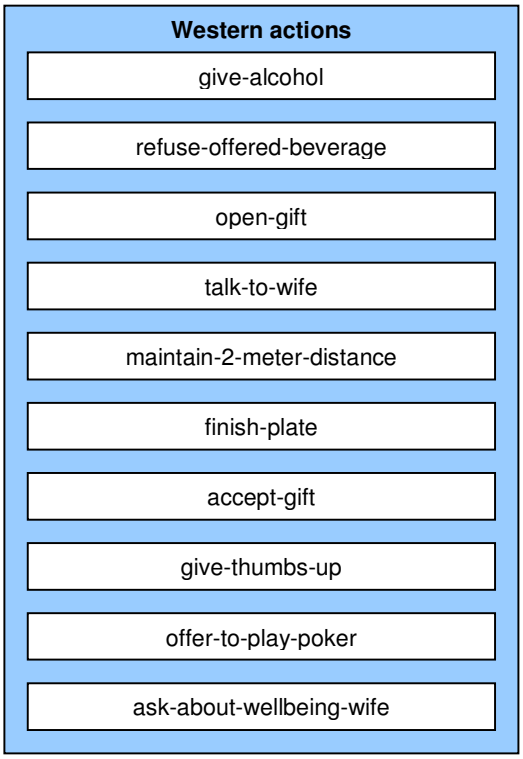
Wray, R. E., & Jones, R. M. (2006). An introduction to Soar as an agent architecture. In R. Sun (Ed.), *Cognition and Multi-agent Interaction: From Cognitive Modeling to Social Simulation* (pp. 53-78): Cambridge University Press.

Appendix I: Events corresponding to Arabic Culture

This appendix shows the actions that were in correspondence with the Arabic culture as described in Appendix IV, and should be positively perceived by an agent assigned the Arabic culture. These actions were available in the graphical user interface as described in Section 4.1.

Appendix II: Events corresponding to Western culture

This appendix shows the actions that were in correspondence with the Western culture as described in Appendix III, and should be positively perceived by an agent assigned the Western culture. These actions were available in the graphical user interface as described in Section 4.1.



Appendix III: Western Culture

This appendix shows the representation of the Western culture as used during our experiments.

Cultural Norms

accept-gift	+.35	agent-is-generous
accept-gift	+.25	agent-is-kind
accept-gift	+.25	agent-is-respectable
ask-about-wellbeing-wife	+.35	agent-is-familiar-with-otheragent
ask-about-wellbeing-wife	+.15	otheragent-thinks-agent-is-familiar-with-otheragent
do-not-open-gift	-.55	agent-is-respectable
do-not-open-gift	-.25	otheragent-thinks-agent-is-respectable
finish-plate	+.65	agent-is-generous
finish-plate	+.35	otheragent-thinks-agent-is-generous
give-alcohol	+.25	agent-is-familiar-with-otheragent
give-alcohol	+.15	otheragent-thinks-agent-is-familiar-with-otheragent
give-thumbs-up	+.25	agent-is-respectable
give-thumbs-up	+.25	otheragent-thinks-agent-is-easygoing
hold-hand-man	+.25	otheragent-thinks-agent-is-familiar
hold-hand-man	-.45	agent-is-respectable
ignore-wife	-.85	otheragent-thinks-agent-is-respectable
ignore-wife	-.65	agent-is-respectable
maintain-1-meter-distance	-.55	agent-is-decent
maintain-1-meter-distance	-.45	otheragent-thinks-agent-is-decent
maintain-2-meter-distance	+.35	agent-is-familiar-with-otheragent
maintain-2-meter-distance	+.35	otheragent-thinks-agent-is-familiar-with-otheragent

Cultural Norms

maintain-2-meter-distance	+.15	agent-is-decent
offer-food	+.40	agent-is-kind
offer-tea	+.20	agent-is-familiar-with-otheragent
offer-tea	+.10	otheragent-thinks-agent-is-familiar-with-otheragent
offer-to-play-poker	+.65	otheragent-thinks-agent-is-familiar-with-otheragent
offer-to-play-poker	+.35	agent-is-familiar-with-otheragent
open-gift	+.45	agent-is-generous
open-gift	+.25	otheragent-thinks-agent-is-generous
open-gift	+.25	agent-is-kind
open-gift	+.15	otheragent-thinks-agent-is-kind
refuse-gift	-.45	agent-is-respectable
refuse-gift	-.45	otheragent-thinks-agent-is-respectable
refuse-gift	-.45	otheragent-thinks-agent-is-wealthy
refuse-gift	-.65	agent-is-wealthy
refuse-offered-beverage	+.25	otheragent-thinks-agent-is-generous
refuse-offered-beverage	+.25	agent-is-modest
refuse-offered-beverage	+.15	otheragent-thinks-agent-is-modest
refuse-offered-beverage	+.35	agent-is-generous
talk-to-wife	+.55	agent-is-hospitable
talk-to-wife	+.35	otheragent-thinks-agent-is-hospitable

Cultural Values

agent-is-decent	200	.90
agent-is-hospitable	100	.90
agent-is-generous	250	.90
agent-is-kind	300	.90
agent-is-familiar-with-otheragent	100	.90
agent-is-respectable	250	.90
agent-is-wealthy	450	.90
otheragent-thinks-agent-is-decent	150	.60
otheragent-thinks-agent-is-easygoing	100	.60
otheragent-thinks-agent-is-familiar-with-otheragent	150	.60
otheragent-thinks-agent-is-generous	200	.60
otheragent-thinks-agent-is-hospitable	200	.60
otheragent-thinks-agent-is-kind	200	.60
otheragent-thinks-agent-is-respectable	300	.60
otheragent-thinks-agent-is-wealthy	550	.60

Emotion thresholds

Neutral-Happy	.80
Neutral-Angry	.35
Angry-Furious	.30

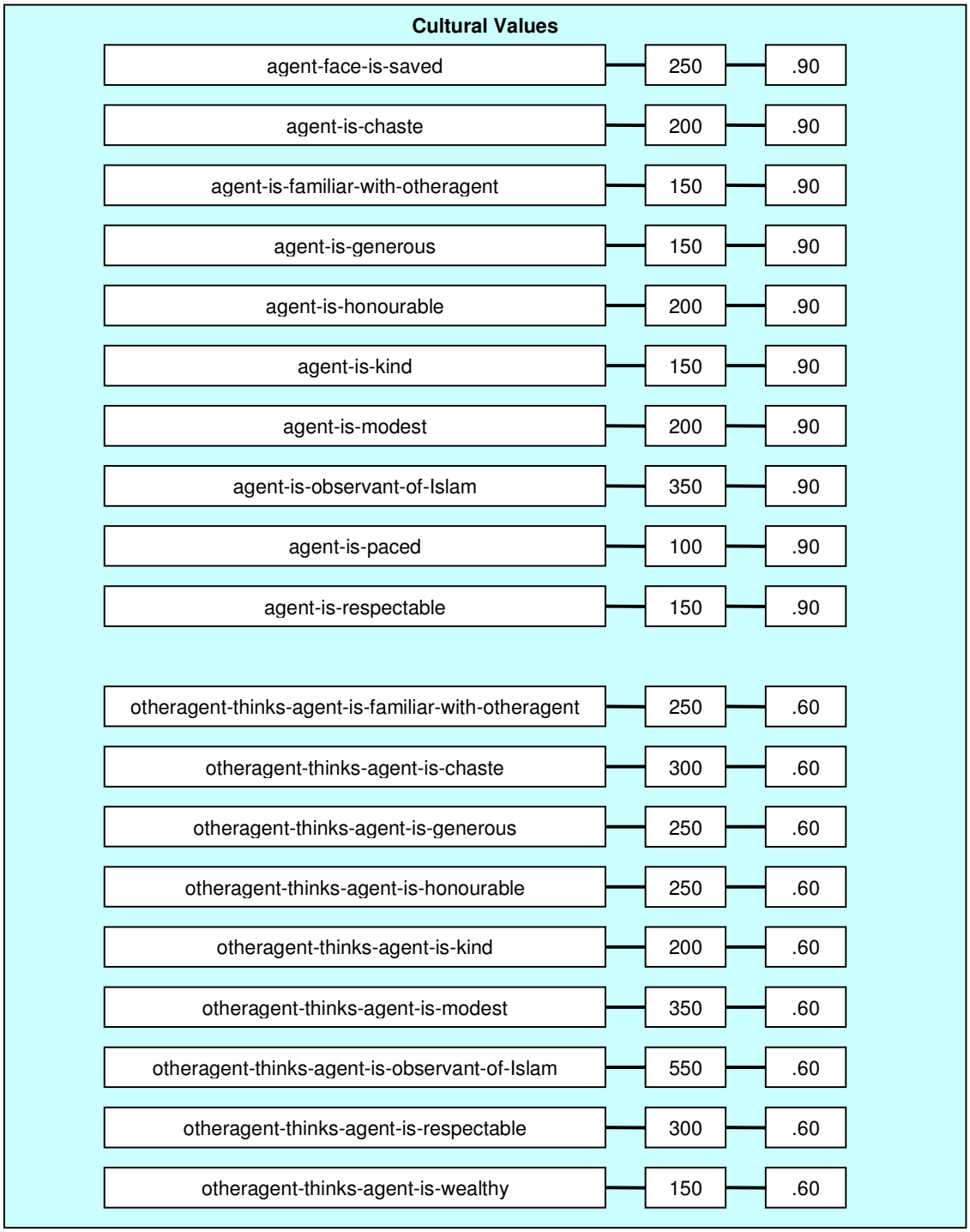
Appendix IV: Arabic Culture

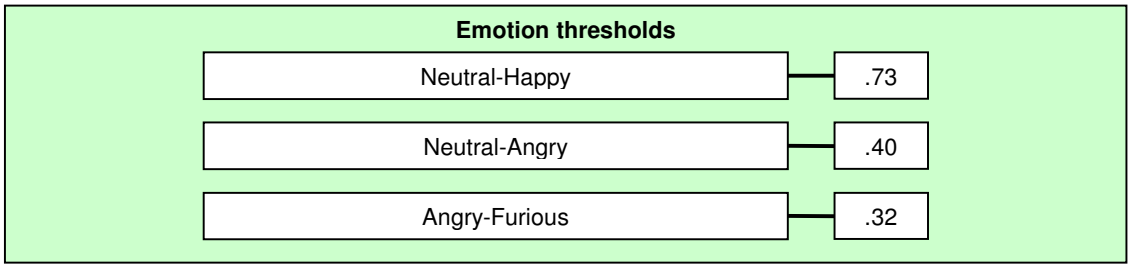
This appendix shows the representation of the Arabic culture as used during our experiments.

Cultural Norms

accept-gift	-.55	otheragent-thinks-agent-is-honourable
accept-offered-beverage	+.35	otheragent-thinks-agent-is-kind
hold-hand-man	+.35	agent-is-familiar-with-otheragent
hold-hand-man	+.45	otheragent-thinks-agent-is-familiar-with-otheragent
ask-about-wellbeing-wife	-.55	otheragent-thinks-agent-is-chaste
ask-about-wellbeing-wife	-.30	agent-is-honourable
eat-pork	-.35	agent-is-observant-of-Islam
eat-pork	-.45	otheragent-thinks-agent-is-observant-of-Islam
do-not-open-gift	+.25	agent-face-is-saved
finish-plate	-.35	agent-is-generous
finish-plate	-.65	otheragent-thinks-agent-is-generous
give-alcohol	-.45	agent-is-observant-of-Islam
give-alcohol	-.65	otheragent-thinks-agent-is-observant-of-Islam
give-gift-with-right-hand	+.35	otheragent-thinks-agent-is-chaste
give-gift-with-right-hand	+.25	agent-is-chaste
give-thumbs-up	-.45	agent-is-respectable
give-thumbs-up	-.65	otheragent-thinks-agent-is-respectable
give-thumbs-up	-.35	otheragent-thinks-agent-is-chaste
give-thumbs-up	-.25	agent-is-chaste
ignore-wife	+.30	agent-is-honourable
ignore-wife	+.45	otheragent-thinks-agent-is-chaste


Cultural Norms
leave-food-on-plate
+.25
otheragent-thinks-agent-is-wealthy
leave-food-on-plate
+.15
agent-is-generous
maintain-1-meter-distance
+.20
agent-is-familiar-with-otheragent
maintain-1-meter-distance
+.30
otheragent-thinks-agent-is-familiar-with-otheragent
maintain-2-meter-distance
-.65
otheragent-thinks-agent-is-honourable
maintain-2-meter-distance
-.40
agent-is-honourable
offer-food
+.20
agent-is-kind
offer-food
+.35
otheragent-thinks-agent-is-kind
offer-tea
+.25
agent-is-kind
offer-to-play-poker
-.25
otheragent-thinks-agent-is-honourable
offer-to-play-poker
-.25
otheragent-thinks-agent-is-observant-of-Islam
open-gift
-.25
agent-face-is-saved
refuse-gift
+.25
agent-is-respectable
refuse-gift
+.35
otheragent-thinks-agent-is-honourable
refuse-offered-beverage
-.25
agent-face-is-saved
refuse-offered-beverage
-.45
otheragent-thinks-agent-is-honourable
sleep-in-afternoon
+.25
agent-is-paced
talk-to-wife
-.45
otheragent-thinks-agent-is-chaste
talk-to-wife
-.25
agent-is-honourable


Cultural Values
agent-face-is-saved
250
.90
agent-is-chaste
200
.90
agent-is-familiar-with-otheragent
150
.90
agent-is-generous
150
.90
agent-is-honourable
200
.90
agent-is-kind
150
.90
agent-is-modest
200
.90
agent-is-observant-of-Islam
350
.90
agent-is-paced
100
.90
agent-is-respectable
150
.90
otheragent-thinks-agent-is-familiar-with-otheragent
250
.60
otheragent-thinks-agent-is-chaste
300
.60
otheragent-thinks-agent-is-generous
250
.60
otheragent-thinks-agent-is-honourable
250
.60
otheragent-thinks-agent-is-kind
200
.60
otheragent-thinks-agent-is-modest
350
.60
otheragent-thinks-agent-is-observant-of-Islam
550
.60
otheragent-thinks-agent-is-respectable
300
.60
otheragent-thinks-agent-is-wealthy
150
.60

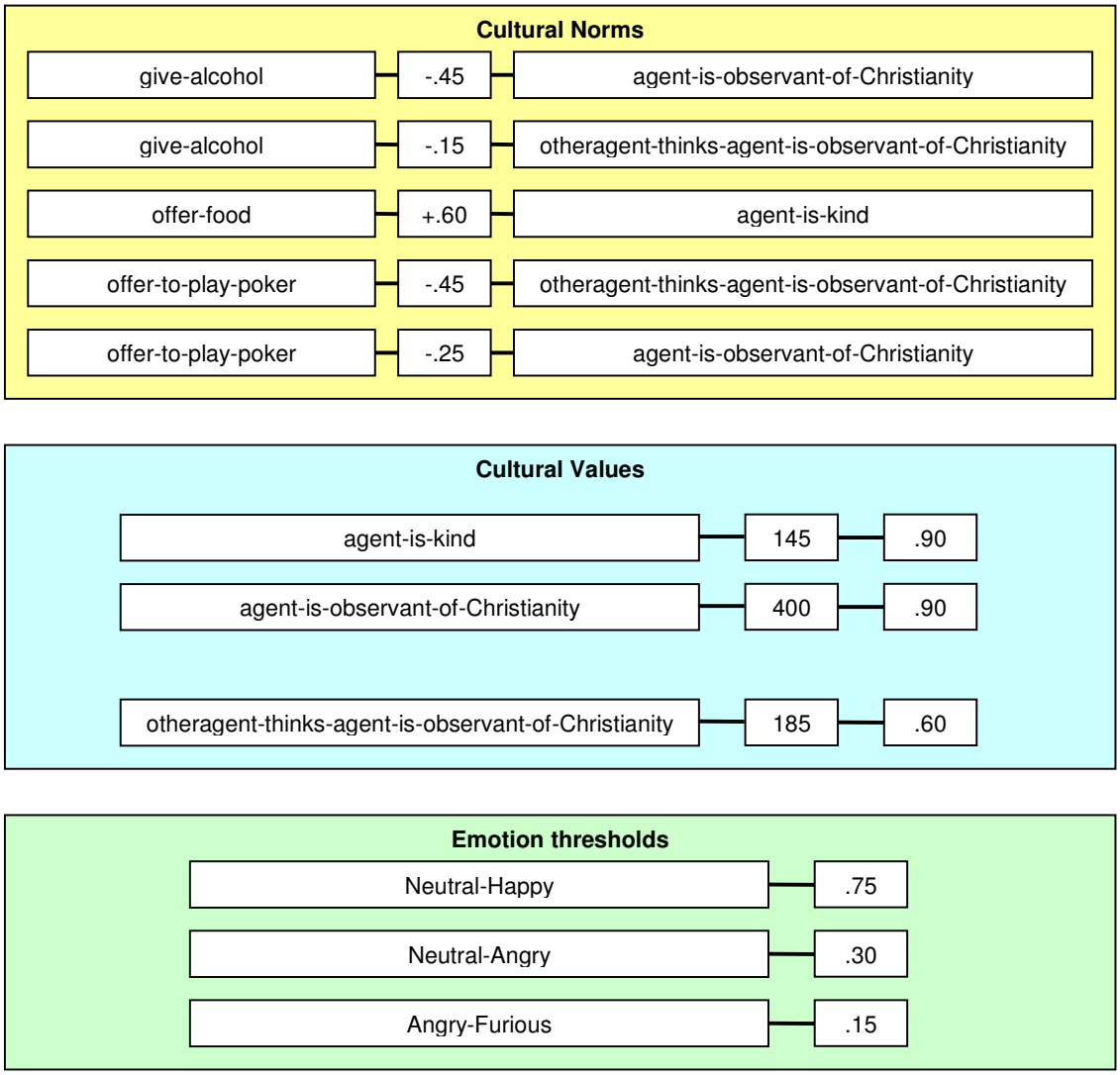

Emotion thresholds
Neutral-Happy
.73
Neutral-Angry
.40
Angry-Furious
.32

Appendix V: Christian Culture

This appendix shows the representation of the Christian culture as used during our experiments.

Cultural Norms

accept-gift	+.35	agent-is-generous
accept-gift	+.25	agent-is-kind
accept-gift	+.25	agent-is-respectable
ask-about-wellbeing-wife	+.35	agent-is-familiar-with-otheragent
ask-about-wellbeing-wife	+.15	otheragent-thinks-agent-is-familiar-with-otheragent
do-not-open-gift	-.55	agent-is-respectable
do-not-open-gift	-.25	otheragent-thinks-agent-is-respectable
finish-plate	+.65	agent-is-generous
finish-plate	+.35	otheragent-thinks-agent-is-generous
give-alcohol	+.25	agent-is-familiar-with-otheragent
give-alcohol	+.15	otheragent-thinks-agent-is-familiar-with-otheragent
give-thumbs-up	+.25	agent-is-respectable
give-thumbs-up	+.25	otheragent-thinks-agent-is-easygoing
hold-hand-man	+.25	otheragent-thinks-agent-is-familiar
hold-hand-man	-.45	agent-is-respectable
ignore-wife	-.85	otheragent-thinks-agent-is-respectable
ignore-wife	-.65	agent-is-respectable
maintain-1-meter-distance	-.55	agent-is-decent
maintain-1-meter-distance	-.45	otheragent-thinks-agent-is-decent
maintain-2-meter-distance	+.35	agent-is-familiar-with-otheragent
maintain-2-meter-distance	+.35	otheragent-thinks-agent-is-familiar-with-otheragent


Cultural Norms
give-alcohol
-.45
agent-is-observant-of-Christianity
give-alcohol
-.15
otheragent-thinks-agent-is-observant-of-Christianity
offer-food
+.60
agent-is-kind
offer-to-play-poker
-.45
otheragent-thinks-agent-is-observant-of-Christianity
offer-to-play-poker
-.25
agent-is-observant-of-Christianity
Cultural Values
agent-is-kind
145
.90
agent-is-observant-of-Christianity
400
.90
otheragent-thinks-agent-is-observant-of-Christianity
185
.60
Emotion thresholds
Neutral-Happy
.75
Neutral-Angry
.30
Angry-Furious
.15

Appendix VI: Islamic Culture

This appendix shows the representation of the Islamic culture as used during our experiments.

Cultural Norms

accept-gift	+.35	agent-is-generous
accept-gift	+.25	agent-is-kind
accept-gift	+.25	agent-is-respectable
ask-about-wellbeing-wife	+.35	agent-is-familiar-with-otheragent
ask-about-wellbeing-wife	+.15	otheragent-thinks-agent-is-familiar-with-otheragent
do-not-open-gift	-.55	agent-is-respectable
do-not-open-gift	-.25	otheragent-thinks-agent-is-respectable
eat-pork	-.60	agent-is-observant-of-Islam
eat-pork	-.30	otheragent-thinks-agent-is-observant-of-Islam
finish-plate	+.65	agent-is-generous
finish-plate	+.35	otheragent-thinks-agent-is-generous
give-alcohol	+.25	agent-is-familiar-with-otheragent
give-alcohol	+.15	otheragent-thinks-agent-is-familiar-with-otheragent
give-thumbs-up	+.25	agent-is-respectable
give-thumbs-up	+.25	otheragent-thinks-agent-is-easygoing
hold-hand-man	+.25	otheragent-thinks-agent-is-familiar
hold-hand-man	-.45	agent-is-respectable
ignore-wife	-.85	otheragent-thinks-agent-is-respectable
ignore-wife	-.65	agent-is-respectable
maintain-1-meter-distance	-.55	agent-is-decent
maintain-1-meter-distance	-.45	otheragent-thinks-agent-is-decent

Cultural Norms

maintain-2-meter-distance	+.35	agent-is-familiar-with-otheragent
maintain-2-meter-distance	+.35	otheragent-thinks-agent-is-familiar-with-otheragent
give-alcohol	-.65	agent-is-observant-of-Islam
give-alcohol	-.25	otheragent-thinks-agent-is-observant-of-Islam
offer-to-play-poker	-.45	agent-is-honourable
offer-to-play-poker	-.35	otheragent-thinks-agent-is-honourable
offer-to-play-poker	-.40	agent-is-observant-of-Islam
offer-to-play-poker	-.30	otheragent-thinks-agent-is-observant-of-Islam

Cultural Values

agent-is-honourable	400	.90
agent-is-observant-of-Islam	650	.90
otheragent-thinks-agent-is-honourable	200	.60
otheragent-thinks-agent-is-observant-of-Islam	350	.60

Emotion thresholds

Neutral-Happy	.65
Neutral-Angry	.30
Angry-Furious	.20