Understanding the impact of culture on cognitive trust-building processes: how to increase the social influence of virtual autonomous agents

Melania Borit
University of Tromsø
Tromsø, Norway
melania.borit@uit.no

Loïs Vanhée
LIRMM,
Montpellier, France
lois.vanhee@lirmm.fr

Petter Olsen
NOFIMA
Tromsø, Norway
petter.olsen@nofima.no

Abstract

Building fine-tuned socially believable autonomous agents interacting with humans in virtual environments is an important aspect of agent design, as humans are influenced more by virtual agents with a high degree of behavioral realism. However, modeling complex psychological processes such as the trust-building between humans and culturally-adaptable agents in a realistic manner is not a trivial task. When designing cultural features of intelligent agents we suggest using our model that integrates cognitive aspects of trust with culture. Certain thinking patterns are involved when a trustor evaluates the trustworthiness of a trustee and thus builds trust in him/her and finally decides whether to trust the person or not. Scientific literature describes five distinct such cognitive patterns called cognitive trust-building processes (CTBPs). Among other factors influencing how trust among parties is built, such as context or personality, culture shapes the preference for certain CTBPs over others (e.g. preferring to build trust by evaluating motivations over assessing ability to fulfill promised duties). National cultures can be evaluated by a combination of scores of cultural dimensions (CD) (e.g. individualism, masculinity). Drawing on theories from organizational management, cross-cultural psychology and social psychology, our model uses CD values to calculate values for each CTBP, Their order indicates the sequence in which each CTBP might be invoked. Agents would behave more realistically if they responded with trust/distrust behavior towards the user according to the agents designed cultural background and if they would display their own trust-building behavior according to their own designed culture. The focus of this paper is on understanding the theoretical underpinnings of the influence of culture on trust-building processes and on explaining how the results of our model can be applied in designing socially believable agents.

In: R. Cohen, R. Falcone and T. J. Norman (eds.): Proceedings of the 17th International Workshop on Trust in Agent Societies, Paris, France, 05-MAY-2014, published at http://ceur-ws.org

1 Introduction

According to the social influence theory of[BLB+02], humans are more likely to be influenced by a virtual entity the more they believe this entity is controlled by another human, i.e. the more they believe this entity is an avatar and not a software agent. In order to increase this social influence in the context of a human interacting with an agent[1], this theory further suggests increasing the agent's visual and behavioral realism. While visual, or photographic realism (i.e. the degree to which the agent resembles a human), is important in some cases, but unnecessary in many others [BM13], behavioral realism seems to be the key of enhancing social influence in human agent interaction [BLB+02, BM13].

Behavioral realism is the extent to which agents behave like humans counterparts in the physical world [BLB+02]. Verbal and non-verbal behaviors are displays of the humans psychological features such as cognitive processes (e.g. perception, belief, awareness, reasoning, judgment), emotional processes or motivational ones. Given the influence of culture on human psychology and thus human behavior [Tri94, MJ12, HHM10], the research on developing agents that are culturally adaptable is growing. Until now, the largest focus has been on modeling specific aspects of conversational behavior, such as language [JVM05], gestures, posture, or proxemics [DHM+07], followed by relational behavior (i.e. the way humans perceive and treat others), such as the status one attributes to another [DHM+13], and psychological processes, such as emotions [BPST07] and motivation [SA11]. Nevertheless, a large palette of psychological features still has to be fine-tuned such that agents become more human-like and among these in this paper we focus on modeling the cognitive trust-building process CTBP with the aim of improving trust-building in human-agent interaction. Many factors influence these CTBPs in a human: life history, personality, temperament, emotions, propensity to trust, culture etc. As for the moment no theoretical framework integrates all these factors, so we chose to focus only on one of them, namely on culture.

People coming from different cultures form trust by involving different cognitive processes and by valuing differently the characteristics of the other party [DCM98]. National cultures can be evaluated by a specific combination of scores in at least four cultural dimensions (CDs): individualism, masculinity, power distance and uncertainty avoidance [HHM10]. Furthermore, culture influences the preference for certain cognitive trust-building process over others (e.g. preferring to build trust in a trustee by evaluating his/her motivations over by assessing his/her ability to fulfill promised duties). In turn, these preferences influence the importance given by the trustor to the factors of perceived trustworthiness (ability, benevolence and integrity) from which to evaluate the trustworthiness of another individual.

In a virtual environment there can be a wide variety of relationship types between the human (i.e. the user) and the agents. These can be neutral, friendly or hostile, and the relationship between the user and the agents can be fixed, or it can change over time, depending on the actions of the user. In human relations, trust is an important aspect of most relationships, and it is one that normally has the potential to change over time. Trust across cultures is particularly problematic, as cultures differ significantly when it comes to how trust should be established and maintained. If person a from culture A wants person b from culture B to trust her, she generally has to choose among several approaches. The most common are either act like a trustworthy person from culture A should and hope that b either recognizes this fact (b then needs to know something of culture A) or that actions inspiring trust in culture A do the same in B). Alternatively, if aware of the differences between cultures A and B, person a could attempt to act like a trustworthy person from culture B (as a views it), hope that b recognizes this fact, and hope that b accepts that this behavior does indeed signify someone worthy of trust. Virtual environments could be symmetrical in the same way as the agent could be programmed to recognize the culture of the user based on the users avatar or behavior. In such an environment, both the user and the agent can take the role of trustor and trustee, as the situation demands. With the agent as (potential) trustor and the user as (potential) trustee, the challenge for the user would be How do I get this agent to trust me?. With the agent as (potential) trustee and the user as (potential) trustor, challenge for the user would be Can I trust this agent? (or possibly In this activity I need to trust an agent; which one out of several alternatives do I choose?).

When designing cultural features of agents we propose using the results of our numerical model that integrates cognitive-trust building processes with culture. With cultural dimensions as input this model calculates appropriate values for each cognitive trust-building process. Arranged in descending order, these values indicate the sequence in which each process might be invoked by individuals of specific cultures. Among possible applications, this model can be used in virtual environment development by providing agents designers with the necessary guidelines for creating realistic cultural characters that behave more like humans when it comes to the issue of

[1]Discussing risks, cautions, problems and moral aspects of social influence of agents in human-agent interaction is outside the scope of this paper.

building trust, both as potential trustors and as potential trustees. The focus of this paper is on understanding the theoretical underpinnings of the relationship between culture and cognitive trust-building processes and on the possible application of this simple numerical model and not any experiments performed with it. We have chosen this because in interdisciplinary studies such as those bringing together social sciences, psychology and computer sciences it is crucial that the theories a model builds on are understood so they are adequately captured in the mechanisms of the model. If this does not happen, the model might run smoothly but its results will not reflect the realities captured by the theory or which have to be proven by it. According to existing theory, humans from different cultures build trust in different ways, thus different agents (e.g. German, Afghan) should be designed to behave like it were building trust in this specific way and our model suggests a possible way to do this. While our model has general application potential, implementation and experimentation depend on the preference of each designer on using a specific agent architecture over another and on the objective of the implementation/experimentation.

The next section outlines the theoretical concepts we use in the model and its application to agents, with the following one touching upon previous approaches to implement virtual agents that reflect culture through their verbal and non-verbal behavior. We then introduce the culture - trust model we developed and discuss how it can be applied in order to enhance social influence in human-agent interaction.

2 Theoretical foundation

2.1 Culture and cultural dimensions

[2] The concept of culture has received many definitions and it was said that almost as many definitions of culture exist as scholars studying the phenomenon [KW88]. One of the dominant culture paradigms in business studies is the one introduced by Geert Hofstede in 1980. In his landmark study, Cultures consequences: international differences in work related values, it is argued that culture is the collective programming of the mind that distinguishes the members of one group or category of people from others; it includes patterns of thinking, feeling, and acting derived from ones social environment rather than from ones genes [HHM10]. This definition is then applied to build categories (or dimensions) of cultures. In particular, investigated groups are nations and organizations. The framework introduced by Hofstede is more widely cited in the Social Science Citation Index than any other competing theories of culture and nearly all cross-cultural management studies are influenced by it [Nak09]. Despite its popularity, Hofstedes approach to culture was criticized by some scholars, many from outside the organizational management field. Some of the criticism refers to simplifying culture to national cultures, being out-dated or having too few dimensions [HJA11, Nak09]. Nevertheless, since it is considered that this paradigm superseded other culture theories [Nak09], we chose to use it ourselves in this study.

In his research on cross-cultural aspects, Hofstede developed the concept of cultural dimension which is a specific aspect of a culture that can be measured relative to other cultures [HHM10]. Six such cultural dimensions (CDs) have been identified: power distance (from small to large), collectivism versus individualism, femininity versus masculinity, uncertainty avoidance (from weak to strong), long-term versus short-term orientation and indulgence versus restraint (for reasons explained in section 4.1. we deal only with four of these dimensions). These dimensions operate together and their operation is influenced by political and economic circumstances [Hof04] and by individual personality factors [HHM10].

- *Power distance (PDI):* it expresses the degree to which the less powerful members of a society accept and expect that power is distributed unequally. Thus, it influences the expectation and importance given to power status. In high PDI societies, leaders are expected to take directions and subordinates to obey and not take initiatives. In low PDI countries, the leader is just a role and subordinates are equal to their superiors. E.g. China, Russia (high PDI) opposed to Scandinavian countries (low PDI).

- *Individualism (IDV):* it influences the definition of individual identity. The lower the IDV, the more one individual's identity is linked to his/her social context (e.g. relatives, colleagues). This context leads to a collective image that has to be pre-served (helping each other within the group, hiding errors, rejecting outsiders). Conversely, in high IDV cultures, individuals expect a treatment independent of any context. E.g. the U.S., Great Britain (high IDV) opposed to South American countries (low IDV).

[2] Due to space limitations, we cannot expand too much here on the values, beliefs, expectations, goals, emotions, motives and behavior underlying cultural dimensions and cognitive trust-building processes. See [HHM10, DCM98] for a comprehensive discussion.

- *Masculinity (MAS)* (alternatively labelled achievement versus cooperation orientation [19]): it indicates preferences on assertiveness, toughness, performance and material success. In high MAS cultures, good performance should be recognized and rewarded, leading to competition. Conversely, low MAS cultures favour modesty, tenderness and high quality of life. Interactions focus on building cooperation and establishing consensus. E.g. Scandinavian countries (low MAS) versus Japan, Hungary (high MAS).

- *Uncertainty avoidance (UAI):* it favors the desire for clear and explicit situations with predictable outcomes. In high UAI cultures, this desire leads to establishment of rules (formal or not), making everything explicit with low ambiguity. Conversely, individuals with low UAI culture dislike the presence of rules. They tend to accept more easily situations with unspecified behavior or unclear outcome. E.g. Greece, Portugal (high UAI) versus Denmark, Vietnam (low UAI).

Partly based on replications and extensions of an IBM survey conducted between 1967 and 1973, the latest database of cultural dimensions values contains scores for 73 countries and three regions [The12]. These values are national averages, which represent the type of behaviors individuals are more likely to perform in a country than in another. These values obviously do not mean that every individual in a given country will perform differently than in another one.

2.2 Trust

Trust has been defined in many different ways by different scientific domains. Only in the period 1960 1990, areas such as management, marketing, psychology and sociology employed more than 70 definitions of trust [CF10]. However, one of the most influential definition is the one proposed by Mayer, Davis and Schoorman in 1995: trust (noun) is the willingness of a party (trustor) to be vulnerable to the actions of another party (trustee/target) based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party.

2.3 Trustworthiness and perceived trustworthiness

The best device for creating trust is to establish and support trustworthiness [Har04] and many authors have theorized why a party will be judged as trustworthy. If to trust is a concept linked with the trustor, then to be trusted (passive form of to trust), to be trustworthy and trustworthiness are concepts linked with the trustee. Trustworthy (adj.) is defined as worthy of being trusted (www.thefreedictionary.com), while trustworthiness (n.) is considered to be the personal state or quality of being trustworthy (www.thefreedictionary.com). Connecting this last definition with the one of trust, we can say that trustworthiness is the trustees personal quality of being worthy to be invested with the willingness of a trustor to be vulnerable to the actions of the trustee (based on the expectation that the trustee will perform a particular action important to the trustor, irrespective of the ability to monitor or control the trustee). This quality is not a general truth, an absolute term about the trustee, but a relative one, connected to the trustors perception of the trustee. This means that while a target can be judged as trustworthy by a trustor, another trustor might consider the same target as untrustworthy. [MDS95] describe three conditions that can give a trustee this quality of trustworthiness in the eyes of a trustor: his/her displayed ability, benevolence and/or integrity (ABI). They frame these conditions within the concept of factors of perceived trustworthiness (FPTs), defined as unique perceptual perspectives from which to consider the trustee.

- *Ability:* it represents the perception of the trustor over a group of skills, competences and characteristics that enable a party (the trustee) to perform in a specific domain (e.g. a patient trusts a certain neurosurgeon well known for her skills for a complicated brain operation).

- *Benevolence:* it is the trustees perceived altruism, his/her disposition to be genuinely interested in the trustors welfare even at his/her own expense (e.g. a young couple accepts the loan requirements from a bank known to help young people seeking to start a family).

- *Integrity:* it is the trustor's perception of the trustees faithfulness to a set of principles that the trustor finds acceptable (e.g. a firm starts doing business with another firm that is known for its business integrity; dealing with Swiss banks might be an obvious example here).

2.4 Cognitive trust-building processes

Some authors consider that trust can be thought of as a continuous mono-dimensional space with trust and distrust as its two extremes [GV10, CE03] theorize that trust develops via a dynamic process that has three evolutionary phases. In the beginning of the relationship, if there is a total lack of information about the trustworthiness of the other, trust starts theoretically around the zero point (the trustor is not willing to make himself/herself vulnerable to the actions of the other party)[3]. During the building-trust phase, the trustor engages in actions meant to acquire knowledge about the trustworthiness of the target (his/her ability, benevolence and/or integrity), while the trustee engages in actions meant to prove his/her trustworthiness. Thus, over time, if these trust-building actions are successful, the general level of trust grows until it begins to level off during the maintenance phase and then stays relatively constant. If during this second stage a trust-destroying event occurs, the overall level of trust drops quickly to the level of distrust and even more intense trust-building actions on behalf of the trustee are required to raise trustors trust from this level. Our culture - trust model deals only with the first phase of this process (i.e. the trust-building one), as it attempts to identify what cognitive processes the trustor is using when evaluating the trustworthiness of the trustee.

During the trust-building phase, certain thinking patterns are involved when the trustor evaluates the trustworthiness of the trustee and thus builds trust in him/her. According to [DC97], trust building relies on the formation of a trustor's expectations about the motives and behaviors of a trustee [DC97]. Scientific literature describes five distinct cognitive patterns which can be employed by the trustor when forming such expectations: calculative, capability, prediction, intentionality, and transference [DCM98]. These cognitive patterns are called cognitive trust building processes (CTBPs).

- Calculative: this process engages the trustor in evaluating the costs and/or rewards of the trustee cheating or staying in the relationship and if the benefits of cheating do not exceed the costs of being caught, the trustor infers that it would be contrary to the trustee's best interest to cheat and therefore the trustee can be trusted. E.g. a research lab wants to collaborate with a competing lab for analyzing a complex chemical. The costs and benefits of the competing lab are analyzed and if the benefits of cheating (patenting the results as own only research) do not exceed the costs of being caught, the competing lab will be trusted in order to be taken in as a partner.

- Prediction: this process engages the trustor in analyzing the trustees behavior in order to decide whether it is consistent or not. E.g. students trust a teacher to be fair with them in the third year of their bachelor degree because they recollect this teacher always being fair since they entered university.

- Intentionality: this process engages the trustor to estimate trustees motivation. E.g. when asked by the leader of the department to undergo a certain evaluation process, the employee trusts the leader to do so because he/she thinks the leader is interested in the employees personal development and not in the profit of the firm performing the evaluation.

- Capability: this process engages the trustor to analyse the trustees ability to act. E.g. in an amusement park, the users of the roller-coaster evaluate the technical competences of the staff serving the facilities to run the roller-coaster in safety conditions.

- Transference: this process engages the trustor in transferring trust from a known entity (proof source) to an unknown one. The proof source can be a person (e.g. friends, family) or an institution (e.g. the government). E.g. the government advices citizens to use a certain bank for specific transactions abroad. Citizens highly trust the government, therefore they will use the services of this particular bank.

In practice, these CTBPs are interrelated, as some factors can invoke multiple CTBPs (e.g. frequent contact with a target may activate both the prediction process and the intentionality one) [DCM98]. Therefore, the evaluation of a trustees trustworthiness may be influenced by several CTBPs, which are often combined and invoked in an unconscious sequenced preference [SSD$^+$10] and triggered sometimes by the same stimulus.

[3]Theory also mentions that the general willingness to trust others (i.e propensity to trust) differs from person to person [MDS95]. Thus, regardless their mental pathway of building trust, some people are more likely to trust others, while some are not. This propensity is also influenced by culture [SMD07]

2.5 Linking trust-building processes with factors of perceived trustworthiness

There is a clear link between how CTBPs manifest themselves and how FPTs are integrated within these processes and this link has an influence over the behavior of both the trustor and the trustee. As said above, the trustor engages in actions meant to acquire knowledge about the trustworthiness of the target, while the trustee engages in actions meant to prove his/her trustworthiness. On the one hand, while acquiring knowledge about the trustworthiness of the target, the trustor will engage in one or more cognitive CTBPs [DCM98]. When engaged in this activity, the trustor will assess the three FPTs of the target from the perspective of the cognitive process he/she uses. For example, if the trustor is using the intentionality CTBP, he/she will look for benevolence signs in the targets behavior while if the trustor is using the capability CTBP he/she will assess the targets ability [DCM98]. Or yet another example, while engaged in the prediction CTBP, the trustor will look at the consistency of the targets past actions or at the extent to which the targets actions are congruent with his/her words [DCM98]. These issues, as well as past actions consistency and actions-words congruency significantly affect the degree to which a party is judged to have integrity [MDS95].

On the other hand, the trustee can choose to display one characteristic or another (e.g. ability, benevolence or integrity) in order to be perceived as trustworthy by the trustor appreciating more that FPT because of the CTBP he/she employs. For in-stance, trust formed via the capability CTBP is based on the trustee being evaluated as able to fulfill his/her promises. Thus, trust is established when the trustor perceives that the trustee has the necessary skills and competences. This suggests at least one strategy for trustees seeking to form trust based on a capability TBP. They can for example increase their perceived abilities by revealing the number of experts hired for performing the specific action the trustor trusts them to accomplish.

2.6 Linking trust with culture

Many scholars consider that among other matters such as affect and emotion, culture impacts trust building through the perceived trustworthiness of a trustee [Hof04, HHM10, MDS95]. For example, more performance oriented societies (high MAS) tend to place a higher value on the ability variable of the ABI framework, while the feminine cultures (low MAS) tend to emphasize more the benevolence variable [SMD07]. The most straightforward connection between trust and culture is done by [DCM98]. These authors describe how national culture, through relation to risk (UAI), authority (PDI) and self (MAS, IDV), impacts TBPs. Based on the idea that the values guiding peoples behavior influence trust building process, the authors formulated 15 propositions about how cultural dimensions facilitate or inhibit the application of CTBPs. For example, one such proposition states that in low UAI cultures trustors are more likely to form trust via a calculative process.

3 Social influence in virtual environments: related work

Humans have certain expectations about communicative signals, patterns, and reactions of interaction partners [SA11]. Such expectations have a crucial influence on how humans unconsciously choose different cognitive pathways of building trust in others [DCM98]. According to the social influence theory, humans are influenced more by virtual agents with a high degree of behavioral realism ([BLB$^+$02, BM13, MdCG13]. Believability of an agent means that humans can apply their usual mental models of communication when interacting with the agent ([SA11]). Thus, the agent is expected to behave as humans would do in a given situation, both in its own initiated behavior and the one generated as a response to the user. This is essential for behavioral realism and thus social influence if an agent looks like a human (or is supposed to represent a human), people expect it to behave like one as well, and will be disturbed by, or misinterpret, discrepancies from human norms [MWR03, NIL00]. Nevertheless, the appearance and behavior of these characters is in most cases based on the cultural background of their designers [KED$^+$12]. Thus, there is the risk that characters developed for a particular culture might not find acceptance when being presented to another culture [KED$^+$12]. Even though a few attempts have been made to create agents that reflect a particular cultural background, interaction with these characters still remains an awkward experience in particular when it comes to non-verbal interaction. As a consequence, agents might not be accepted, and their social influence will be minimal when interacting in a culturally inappropriate manner [KED$^+$12].

Adapting an agent to a different culture involves not only the obvious language translation, but also the careful reconsideration of each of the key characteristics identified in the literature: identity, backstory, appearance, content of speech, manner of speaking, manner of gesturing, emotional dynamics, social interaction patterns, role and role dynamics [SMK$^+$11]. However, usually, cultural/social adaptation of agents has focused on visual ele-

ments relating to non-verbal behaviors, especially gesture, facial expression and inter-personal distance [AVA+09] or on linking only verbal behavior to emotion/motivation/trust/culture (i.e. verbal/non-verbal cues → psychological features/culture) and not so much on the verbal/non-verbal manifestation of cognitive processes in a specific cultural setting (i.e. culture → psychological features → verbal/non-verbal display). For example see the research of [AVA+09, DHM+13, DHM+07, HRMM02, IC07, MHR00, MDA+09, OB97, DPP04, SMK+11, TR09] for culture [HJA11] for culture and emotions; [BPST07, MWR03, NIL00] for emotions [SA11] for motivation [BC01, CS03, MAM97] for trustworthiness of human-like agents and [DLP13] for building social identity of these agents. While personality relates to the level of the individual, culture relates to social groups and it is not enough to modify the expressive behavior of the individual character; they must also be considered as social actors and behave in a socially acceptable manner[AVA+09]. Even as each human personality is unique, each culture tends to evoke specific modes of adjustment and reactions in different situations[SMK+11].

4 Designing realistic cultural agents in relation to cognitive trust-building processes

4.1 Modeling trust-building cognitive processes according to cultural background

In attempting to link culture with trust in a quantitative way, our model draws on the qualitative links between cultural dimensions and cognitive trust-building processes described by [DCM98]. According to the qualitative assessment of these authors, high and low CD values have positive influences on different CTBP. Moreover, the same type of culture (low IDV/MAS/PDI, high UAI) influences similarly the CTBPs of intentionality and transference, while the opposite of this type of culture (high IDV/MAS/PDI, low UAI) influences the calculative CTBP. Doney et al. acknowledge that their research provides insight into the implications of trust developed by alternative processes without specifying a hierarchical approach. Furthermore, they assume that each dimension of culture is of equal importance. Thus, following their reasoning, we also explicitly assume that each cultural dimension has the same weight in the process of trust building (i.e. we transform the qualitative links of [DCM98] in linear formula, assuming that each cultural dimension has an independent and equal impact on the importance given to a TBP). However, we are aware that it may be that trustors assign greater weight to trust developed via one cognitive process compared to another and that context may influence the relevance of a particular cultural dimension. Based on further theoretical and empirical studies our model may be modified in order to reflect these, by using other methods for computing the formula for TBPs (e.g. maximum amongst positive cultural dimensions, mean of positive links, squared cultural dimensions scores). Nevertheless, for the moment, as indicated by [DCM98], questions regarding whether, or how, trustors prioritize these TBP remain unanswered. Because [DCM98] used only four cultural dimensions, our model uses only four, too.

We first normalized each cultural dimension in a [-1;1] space, where -1 corresponds to the lowest cultural value and 1 to the highest one, by using the following formula:

$$n\_value_{CD}(c) = \frac{value_{CD}(c) - min_{CD}}{max_{CD} - minCD} \times 2 - 1$$

Where $n\_value_{CD}(c)$ is the normalized value of the country c in the cultural dimension CD, $value_{CD}$ is the normalized value of c in CD, min_{CD} and max_{CD} are respectively the minimal and maximal values of this cultural dimension. With this representation, a country with a low value for a cultural dimension has a value below 0 and above 0 otherwise, the maximal value is 1 and the minimal is -1. This normalization preserves the distance ratios between the cultural dimensions of various countries.

Using this scale, we evaluate the importance given to each CTBP in using the following formula:

$$value_{Calc}(c) = n\_value_{IDV}(c) + n\_value_{MAS}(c) + n\_value_{PDI}(c) - n\_value_{UAI}(c) \quad (1)$$

$$value_{Pred}(c) = -n\_value_{IDV}(c) - n\_value_{MAS}(c) + n\_value_{PDI}(c) + n\_value_{UAI}(c) \quad (2)$$

$$value_{Capa}(c) = n\_value_{IDV}(c) + n\_value_{MAS}(c) + n\_value_{PDI}(c) + n\_value_{UAI}(c) \quad (3)$$

$$value_{Int}(c) = -n\_value_{IDV}(c) - n\_value_{MAS}(c) - n\_value_{PDI}(c) + n\_value_{UAI}(c) \quad (4)$$

$$value_{Trans}(c) = -n\_value_{IDV}(c) - n\_value_{MAS}(c) - n\_value_{PDI}(c) + n\_value_{UAI}(c) \quad (5)$$

Where $value_{Calc}$ is the importance given to the calculative TBP, $value_{Pred}$ is the importance given to the predictive TBP, $value_{Capa}$ is the importance given to the capability TBP, $value_{Int}$ is the importance given to

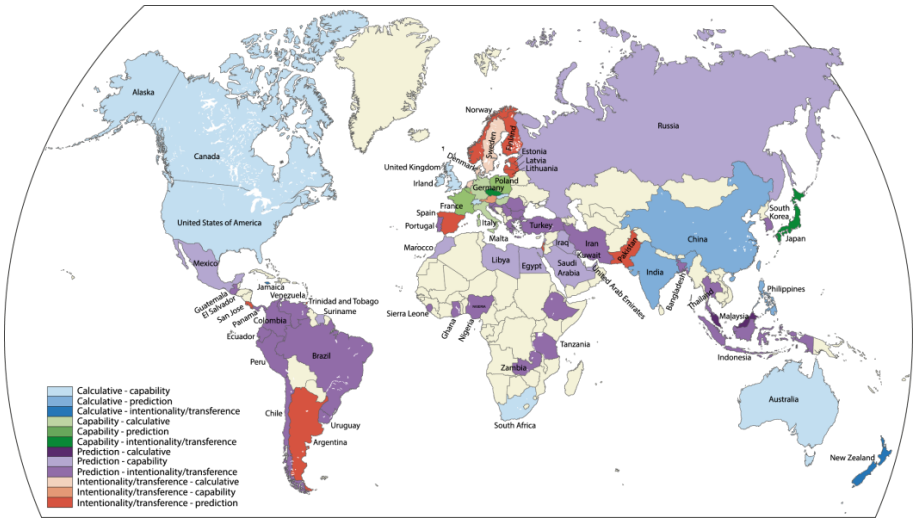

Figure 1: Countries and their respective pairs of the first two cognitive trust-building processes as calculated using our culture-trust model

the intentionality TBP, $value_{Trans}$ is the importance given to the transference TBP. Each formula corresponds to the linear combination of the positive relations between high/low values of cultural dimensions and CTBPs given by [DCM98].

Following the recommendations from [CF10] to look closer at cognitive models of trust, our model focuses on how trust building (from a cognitive perspective) and culture might interact and what could be the outcome of this interaction. At the same time, one has to keep in mind that this is a formalization of the first phase of the trust process only (see Section 2.4) and cannot answer the question of how trust will evolve. For example, if a trustor used the transference process in this first phase, the target might need to maintain trust via another CTBP to ensure the completion of the second phase [DCM98].

When fed with the values of the national cultural dimensions calculated by Hofstede, our culture - trust model calculates appropriate values for each trust-building cognitive process. Arranged in descending order, these values indicate the sequence in which each CTBP might be invoked by individuals of specific cultures. The data-map displayed in Figure 1 presents, for the sake of visualization clarity, only pairs of the first two trust building processes as calculated using our model. A table with all the data calculated for all trust-building processes for all the countries for which cultural dimensions values are available can be provided at request.

The results obtained by this model are coherent with former theoretical descriptions of how people from different cultures build trust from [Hof04, HHM10] and [SMD07]. For example, [SMD07] state that more collaborative, being oriented feminine (thus, low MAS) cultures tend to put more emphasis on the benevolence variable. Hence, for people in these cultures, trust emerges if the trustor perceives target to be genuinely interested in the trustors welfare and motivated to seek joint gain. This way of building trust is specific to the intentionality process [DCM98]. According to our formalization, nine of the most feminine cultures rank the intentionality/transference process first. Moreover, these results seem to be consistent with indirect descriptions of the trust-building phase found in empirical case-studies [CE03, Ide09]. Furthermore, there is a statistically significant relationship between these results and values of CTBPs proxy indicators taken from available world-wide surveys (Borit and Vanhée, work under progress).

This model can be used to infer the CTBP sequence for any individual given that the scores of this individual for the four CDs are known. Applications that calculate these scores are already available on the market (e.g. CultureGPS or ComPass).

4.2 Consequences for agents design

When designing agents, our culture-trust model can be used from a double perspective: 1) considering the user as a trustor (with the agent as a trustee) and 2) considering the user as a trustee (with the agent as a trustor).

In the first case, users would have to consider the trustworthiness of the agent. In order for the user to trust the agent the way the user would trust a human in the role that the agent plays, the agent would have to manifest the characteristic signs of the cognitive trust-building processes that a human would invoke in the respective situation, given the culture of that human. For example, an agent designed to exhibit Nigerian trust-building traits guiding the user through the mountains would have to display characteristics of the prediction and

intentionality/transference CTBPs (e.g. performing certain rituals, revealing its benevolent actions), in order for the user to trust it (if the user has a certain degree of cultural awareness). If the agent would display the characteristics of capability and calculative CTBPs (e.g. talking about its expertise in guiding) then this behavior would mirror a North-American way of thinking, which would be in discrepancy with the cultural background it is designed to represent.

Instructional role-play simulations like the ones developed by Alelo Inc.[Ale14] provide dos and dont's in intercultural encounters. By trial and error the user learns about the practices (rituals, symbols) of specific cultures (the answers to the what and how questions when meeting a new culture). Nevertheless, the simulation does not provide the user with the underlying cognitive premises of the agent behavior (the why question). For example, the simulation teaches the user that she/he has to ask the agent designed to exhibit Taiwanese cultural traits about how the dinner was prepared and provide compliments to build rapport, but it does not offer the explanation that actually this is an action meant to build trust according to the prediction and intentionality/transference CTBPs that are usually firstly invoked by the Taiwanese humans. When the user understands the connection between the agent behavior and the specific way this agent trusts people, the user can handle different kinds of situations than the one explicit in the simulation (for example, how to behave when convincing a Taiwanese human/agent to join him/her to a party, not only how to behave at dinner).

In the second case, the agent would have to evaluate the actions of the user and decide whether to trust him/her or not. In the same example, if the user is North-American, then he/she is more likely to invoke the capability and calculative CTBPs and be more task-oriented than the virtual counterpart. In this case, the user would not engage in trying to know the agent on a more personal basis (e.g. not trying to build a personal relationship by asking different questions about the general wellbeing of the agent and of its family), but would ask directly for expert services. Having a different cognitive pattern of building trust, in order to behave realistically the agent should display a non-trusting behavior until the player changes his/her strategy.

From these two perspectives, our model could be included in the further development of other intercultural training tools such as Traveller [DHM+13], TLTS [JVM05], ELECT BiLAT [HBL+06] or ORIENT [AVA+09].

4.3 Possible other applications in a multiagent systems context

Our formalization could be used in order to replicate the development of trust in the context of artificial societies influenced by their cultures. In particular, this model might be used in multi-agent systems at the agent level in order to reproduce the impact of culture on multi-agent interactions. For example, it could be included in contemporary trust models in order to give them a cultural setting and then observe different outcomes influenced by this cultural setting. That fact that the modeler can simulate the cognitive trust building processes of an agent in a realistic manner could be a powerful tool. Additionally, BDI agents that are capable of abducing other agent's intentions (intentionality CTBP) and behavior (prediction CTBP) could be designed using this model. The impact of trust has a clearly observable importance in organizations. [D'I94] illustrates how three very similar organizations under the influence of different cultures establish trust in a different way. They may rely on a) honour attaching individuals to a status and making them responsible for defending it; b) contracts binding individuals to a performance measure; or c) consensus integrating concerned individuals by decision making, creating a cooperation spirit [D'I94]. Each form of organization can be linked to at least one trust building process: honour with prediction and calculative, contract with calculative and capability, and consensus with intentionality. Agent-based models of culture and organizations (as described in [VDF13]) would benefit from including the importance of trust in organizations and the influence of culture on it through integrating our formalization. Among others, these models could show different organizational dynamics based on how trust between the agents is build according to their cultural background.

Moreover, organizations can display a large panel of behaviors which might inspire trust through different TBPs. To this extent, organizations could specialize in certain types of trustworthiness display, aiming at different targets. For instance, consider an organization favoring costumer service (benevolence, through the prediction and intentionality TBPs) while another proposes quality (ability, through the capability TBP). Our formalization could be used to analyze crucial changes an organization has to make in order to fit a given market. In addition, simulating several organizations in a multi-cultural setting could show what type of national market niches can appear.

5 Conclusions and future work

People coming from different cultures form trust by involving different cognitive processes and by valuing differently the characteristics of the other party [DCM98]. In line with the research related to including cognitive processes (e.g. trust building), emotions, motivations and culture in the agents design, we propose integrating trust and culture theories in the design of agents as a way to enhance their social believability and thus their social influence in a human-agent interaction context. When designing the cognitive characteristics of agents, we propose considering also the trust-building cognitive aspect and suggest using the results of a culture trust model we have built. This numerical model integrates CTBPs with culture: having as input cultural dimensions values the model calculates appropriate values for each CTBP. These values indicate the sequence in which each CTBP might be invoked, thus providing agents designers with guidelines of creating cultural characters that are believable when it comes to how they build trust. Agents could be able to behave more realistically when displaying themselves trustworthy behavior and responding with trust or distrust in the player according to its cultural background. The focus of this paper is on understanding the theoretical underpinnings of the relationship between culture and trust and on explaining how the results of our model can be applied in designing agents in order to improve trust-building in human-agent interaction.

As future work with direct application on modeling the trust-building relationship between humans and agents we intend to create correspondence rules that connect CTBPs to concrete specifications of conversational behavior (e.g. language, gesture expressivity, posture or proxemics), appearance, content of speech, emotional dynamics, social interaction patterns, role and role dynamics. Also, since our culture - trust model deals only with the first phase of the trust cognitive process, we intend to look closer into the time - CTBPs and destroyed trust - CTBPs relationships, an approach that has also direct application in the area of enhancing social believability of agents (e.g. the user might meet the same agent several times during the game; how should the agent maintain trust over time? How should the agent behave if the user behaves in a trust-destroying manner?).

Acknowledgments

The first author wishes to thank to Bruce Edmonds, Jorge Santos, Petter Holm and Emma Norling for their comments on the general aspects of this research. This research was facilitated by the EWMA project (Norwegian Research Council project number 195160).

References

[Ale14] Alelo Inc. Alelo Inc., 2014.

[AVA+09] Ruth Aylett, Natalie Vannini, Elisabeth Andre, Ana Paiva, Lynne Hall, and Sibylle Enz. But that was in another country: agents and intercultural empathy. In *The 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009),*, pages 329–336, 2009.

[BC01] Timothy Bickmore and Justine Cassell. Relational agents: a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, number 3, pages 396–403, 2001.

[BLB+02] Jim Blascovich, Jack Loomis, Andrew C Beall, Kimberly R Swinth, Crystal L Hoyt, and Jeremy N Bailenson. Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, 13(2):103–124, 2002.

[BM13] Jim Blascovich and Cade Mccall. Social Influence in Virtual Environments. In K Dill, editor, *of Media PsychologyThe Oxford Handbook*, number January, pages 305–315. Oxford University Press, New York, NY, 2013.

[BPST07] Tibor Bosse, Matthijs Pontier, Ghazanfar F Siddiqui, and Jan Treur. Incorporating Emotion Regulation into Virtual Stories. In C. et al. Pelachaud, editor, *IVA, LNAI 4722*, pages 339–347, 2007.

[CE03] Steven C Currall and Marc J Epstein. The fragility of organizational trust: Lessons from the rise and fall of Enron. *Organizational Dynamics*, 32(2):193–206, 2003.

[CF10] Christiano Castelfranchi and Rino Falcone. *Trust Theory: A socio-cognitive and computational model*, volume 20. Wiley, 2010.

[CS03] A J Cowell and K M Stanney. Embodiment and interaction guidelines for designing credible, trustworthy embodied conversational agents. *Intelligent Virtual Agents*, pages 301–309, 2003.

[DC97] Patricia M Doney and Joseph P Cannon. An Examination of the Nature of Trust in Buyer-Seller Relationships. *Journal of Marketing*, 61(2):35–51, 1997.

[DCM98] Patricia M Doney, Joseph P Cannon, and Michael R Mullen. Understanding the influence of national culture on the development of trust. *Academy of management review*, pages 601–620, 1998.

[DHM+07] Jan Dusan, David Herrera, Bilyana Martinovski, David Novick, and David Traum. A computational model of culture-specific conversational behavior, 2007.

[DHM+13] Nick Degens, Gert Jan Hofstede, Samuel Mascarenhas, André Silva, Ana Paiva, Felix Kistler, Aleksandra Swiderska, Eva Krumhuber, Arvid Kappas, Colette Hume, and Ruth Aylett. Traveller Intercultural training with intelligent agents for young adults. In *Proceedings of the Eight International Conference on the Foundations of Digital Games (workshop 1st International Workshop on Intelligent Digital Games for Empowerment and Inclusion)*, 2013.

[D'I94] P D'Iribarne. The Honour Principle in the 'Bureaucratic Phenomenon'. *Organization Studies*, 15(1):81–97, 1994.

[DLP13] Joana Dimas, P Lopes, and R Prada. One for all, all for one: Agents with social identities. *Proceedings of the CogSci*, pages 2195–2200, 2013.

[DPP04] Fiorella De Rosis, Catherine Pelachaud, and Isabella Poggi. Transcultural believability in embodied agents: a matter of consistent adaptation. *Agent Culture: Human-Agent Interaction in a Multicultural World*, pages 75–106, 2004.

[GV10] Michelle Greenwood and Harry J Van Buren III. Trust and stakeholder theory: trustworthiness in the organisation-stakeholder relationship. *Journal of business ethics*, 95(3):425–438, 2010.

[Har04] Russell Hardin. *Trust and trustworthiness*, volume 4. Russell Sage Foundation Publications, 2004.

[HBL+06] Randall W Hill Jr., James Belanich, H C Lane, Mark Core, Melissa Dixon, Eric Forbell, Julia Kim, and John Hart. Pedagogically Structured Game-Based Training: Development of the Elect BiLAT Simulation. 2006.

[HHM10] Geert Hofstede, Gert Jan Hofstede, and Michael Minkov. *Cultures and Organizations: Software of the Mind, revised and expanded 3rd ed.* McGraw-Hill, New York, NY, 2010.

[HJA11] Lynne Hall, Susan J Jones, and Ruth Aylett. Fostering empathic behaviour in children and young people: interaction with intelligent characters embodying culturally specific behaviour in virtual world simulations. In *INTED2011 Proceedings*, pages 2804–2814, 2011.

[Hof04] Gert Jan Hofstede. *Hide or confide?: the dilemma of transparency.* Number 248 p. Reed Business Information, 's-Gravenhage, 2004.

[HRMM02] Barbara Hayes-Roth, Heidy Maldonado, and Marcia Moraes. Designing for Diversity: Multi-Cultural Characters for a Multi-Cultural World. *Proceedings of IMAGINA*, 2002.

[IC07] Francisco Iacobelli and Justine Cassell. Ethnic Identity and Engagement in Embodied Conversational Agents. In *IVA 2007, LNAI 4722*, pages 57–63, 2007.

[Ide09] Uwafiokun Idemudia. Assessing corporate-community involvement strategies in the Nigerian oil industry: An empirical analysis. *Resources Policy*, 34(3):133–141, 2009.

[JVM05] W L Johnson, H H Vilhjalmsson, and S Marsella. Serious games for language learning: How much game, how much ai? In C.-K. Looi, G I McCalla, B Bredeweg, and J Breuker, editors, *AIED. Frontiers in Artificial Intelligence and Applications*, pages 306–313. IOS Press, 2005.

[KED+12] Felix Kistler, Birgit Endrass, Ionut Damian, Chi Tai Dang, and Elisabeth André. Natural interaction with culturally adaptive virtual characters. *Journal on Multimodal User Interfaces*, 6(1-2):39–47, 2012.

[KW88] George D Kuh and Elizabeth J Whitt. *The Invisible Tapestry. Culture in American Colleges and Universities. ASHE-ERIC Higher Education, Report No. 1, 1988.* ERIC, 1988.

[MAM97] Susanne Van Mulken, Elisabeth André, and Jochen Müller. An Empirical Study on the Trustworthiness of Life-Like Interface Agents. In *HCI*, volume 2, 1997.

[MDA+09] Samuel Mascarenhas, João Dias, Nuno Afonso, Sibylle Enz, Ana Paiva, and Cavaco Silva. Using Rituals to Express Cultural Differences in Synthetic Categories and Subject Descriptors. In *The 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 305–312, 2009.

[MdCG13] Celso Melo de, Peter Carnevale, and Jonathan Gratch. Peoples biased decisions to trust and cooperate with agents that express emotions. In *The 16th International Workshop on Trust in Agent Societies co-located with the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, St Paul, Minnesota, USA, 2013.

[MDS95] Roger C Mayer, James H Davis, and F David Schoorman. An integrative model of organizational trust. *Academy of management review*, pages 709–734, 1995.

[MHR00] Heidy Maldonado and Barbara Hayes-Roth. Toward Cross-Cultural Believability in Character Design. *Embodied conversational agents*, 2000.

[MJ12] David Matsumoto and Linda Juang. *Culture and Psychology.* Cengage Learning, 2012.

[MWR03] Stacy Marsella, Admiralty Way, and Marina Rey. Modeling Coping Behavior in Virtual Humans: Don't Worry , Be Happy. In *AAMAS*, pages 313–320, 2003.

[Nak09] Cheryl Nakata. *Beyond Hofstede. Culture frameworks for global marketing and management.* 2009.

[NIL00] Clifford Nass, Katherine Isbister, and Eun-ju Lee. Truth Is Beauty: Researching Embodied Conversational Agents. *Embodied conversational agents*, pages 374–402, 2000.

[OB97] Patricia O'Neill-Brown. Setting the Stage for the Culturally Adaptive Agent. *AAAI Technical Report FS-97-02*, 1997.

[SA11] Felix D Schönbrodt and Jens B Asendorpf. The Challenge of Constructing Psychologically Believable Agents. *Journal of Media Psychology: Theories, Methods, and Applications*, 23(2):100–107, 2011.

[SMD07] F David Schoorman, Roger C Mayer, and James H Davis. An integrative model of organizational trust: Past, present, and future. *Academy of management review*, 32(2):344–354, 2007.

[SMK+11] Reid Simmons, Maxim Makatchev, Rachel Kirby, Min Kyung Lee, Imran Fanaswala, Brett Browning, Jodi Forlizzi, and Majd Sakr. Believable Robot Characters. (Stebbins 1886):39–52, 2011.

[SSD+10] Mark N K Saunders, Denise Skinner, Graham Dietz, Nicole Gillespie, and Roy J Lewicki. *Organizational trust: a cultural perspective.* Cambridge University Press, 2010.

[The12] The Hofstede Centre. National cultural dimensions, 2012.

[TR09] David Traum and Marina Rey. Models of Culture for Virtual Human Conversation. In C Stephanidis, editor, *Universal Access in HCI, Part III, HCII 2009, LNCS 5616*, pages 434–440, 2009.

[Tri94] Harry Charalambos Triandis. *Culture and social behavior. McGraw-Hill series in social psychology.* Mcgraw-Hill Book Company, New York, NY, England, 1994.

[VDF13] L. Vanhée, F. Dignum, and J. Ferber. Towards Simulating the Impact of National Culture on Organizations. In *The 14th International Workshop on Multi-Agent-Based Simulation*, Saint Paul, Minnesota, 2013.