



FP7- 257666-eCUTE

(January, 2012)

Education in Cultural Understanding, Technologically-Enhanced

Collaborative Project (ICT-2009.4.2)

Technology-enhanced learning

Start date of project: **01/09/2010**

Duration: **36 months**

(D3.1). (Behavioural Interface)

Due date of deliverable: Month 15

Actual submission date: 19-01-2012



AUTHORS: Nick Degens, John Mc Breen, Gert Jan Hofstede (Wageningen University)

CHECKERS: Ana Paiva, Eva Krumhuber

STATUS: final

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Deliverable <D3.1/final>



FP7- 257666-eCUTE

(January, 2012)

PROJECT COORDINATOR

Name: Ruth Aylett

Address: School of Mathematics and Computer Science, Heriot-Watt University Edinburgh, EH14 4AS

Phone Number: +44 131 4514189 Fax Number: +44 131 451 3327

E-mail: ruth@macs.hw.ac.uk

Table of Contents

Table of Contents	3
1. Purpose of Document	4
2. Executive Overview.....	5
3. Introduction.....	6
3.1. The need for theory-based culturally varying behaviour	6
3.2. Structure of this deliverable.....	7
4. Human Social Behaviour.....	8
4.1. Group membership	8
4.1.1. Social Identity Theory	8
4.1.2. Identity Theory	9
4.1.3. Wrapping up	9
4.2. A pragmatic concept: Moral Circles.....	11
4.2.1. Preliminary definitions	11
4.2.2. Types of Moral Circles.....	12
4.2.3. The Core Moral Circle.....	13
4.2.4. Nested Moral Circles.....	14
4.3. Moral Circle behaviour.....	15
4.3.1. Moral circles are context dependent.....	16
4.3.2. Cultural Meta-Rules and Social Norms.....	16
4.3.3. Moral Circles in action: Rituals.....	17
4.3.4. Verbal and non-verbal behaviour	18
5. Modelling Moral Circles.....	21
5.1. Moral Circle primitives	21
5.1.1. Moral Circle Reputation (MCR)	22
5.1.2. Moral Circle Status (MCS).....	22
5.1.3. Moral Circle Centrality (MCC).....	22
5.2. The Behavioural Interface: Moral Circle dynamics	24
5.2.1. Guideline layer	25
5.2.2. Interpretation layer	28
5.2.3. Context layer	32
5.3. The modifying role of culture.....	33
5.3.1. Dimensions of Culture and MC dynamics	34
5.3.1. Synthetic Cultures	36
6. The Behavioural Interface in practice.....	38
7. Conclusion	41
8. References.....	42

1. Purpose of Document

The purpose of the behavioural interface is to specify the range of behavioural inputs that synthetic characters will have to respond to. Since these agents are social beings, designing this interface requires attention to the social cues they need to be aware of when communicating both with each other and the user; these cues constitute the information that the behavioural interface must be able to convey.

The expositions focusing on the learning goals in previous deliverables (D2.1 and D2.2) are our main theoretical inspiration. However, they allow for considerable freedom in the configuration of the behavioural interface. Technology-related constraints are the subject of later deliverables, but will be taken into account here to the extent that we already know they are inescapable.

Our main focus in this deliverable is to design a conceptual model, based on the theoretical analyses in previous deliverables and on unavoidable interface constraints. The overall goal of this deliverable is to present a model that can be used to create not only social verbal and non-verbal behaviour, but also have that behaviour modified by culture. This calls for an abstract interface that allows for the use of many different types of behaviour.

At this point we would like to thank our colleagues from eCute, as well as those from ComplexityNet project SEMIRA (www.semira.wur.nl), who helped us develop the ideas put forward in this deliverable. A particular word of thanks goes to Frank Dignum and Gennaro di Tosto of Utrecht University.

2. Executive Overview

This deliverable develops a conceptual model of believable social behaviour that

- Takes as its inspiration the learning goals of deliverables 2.1 and 2.2;
- Is aware of technological constraints in the project;
- Enables believable social interaction between agents;
- Explicitly models both the individual and the group ('moral circle')
- Can be used with human users;
- Allows for variation across synthetic cultures.

While the specific scenarios will be dealt with in later deliverables, the main features of MIXER (a trust and deceit game for children, with groups of agents on the scene simultaneously) and TRAVELLER (a tour among many strange peoples, with a lot of greeting and collaborative problem-solving going on) are kept in mind in this deliverable.

The deliverable still leaves many degrees of freedom for the actual development of the scenario materials. It should facilitate constructive design-oriented discussion among members of eCute and serve as a useful foundation for future deliverables. The concepts put forward here will no doubt have to be adapted to both implementation constraints and new insight.

3. Introduction

3.1. The need for theory-based culturally varying behaviour

This deliverable will not just come up with a list of required communication acts; that list will be developed from a fully-fledged, if simplified, model of social behaviour. Why? In order to make this clear, we return to the basic aims of eCute. The following is copied from the eCute description of work, and it speaks for itself.

“Role-play and game-based simulations such as Barnga! (Thiagarajan 1990) are also widely used with the aim of creating safe environments in which participants can be exposed to emotional states such as culture shock and those arising from intercultural conflict and then reflect on their own experience. This is the hallmark of an educational domain in which affect is of vital importance and knowledge is on its own insufficient, so that eCute’s technology development will incorporate a strong affective model. Though role-play and game-based simulation are widely used, there is little or no use of ICT to support or enhance this. eCute to our knowledge will be the first project to systematically examine how ICT can be used with these scenarios.”

So, affect is central. This is why we aim to create an affectively believable agent world for the human learner to interact with.

Yet, we wish to impart real knowledge about cultural differences, which is part of the project’s aims. So our believable agents should be capable of having different cultures. In creating what are effectively cultural story-worlds, we prefer to create synthetic cultures rather than modelling real-world ones. Real-world cultures are much more complex and rich than can be represented in game-based simulations and caricatures could be distracting, play badly with existing prejudices, or offend members of the portrayed culture. Synthetic Cultures avoid these pitfalls and allow exaggeration to be used positively to improve the clarity of the simulation. However constructing such a culture requires a comparative model of culture that provides a solid empirical framework.

A further argument for the use of synthetic rather than real cultures is that they are among the very few scripts for cross-cultural simulations available that are based on theory. Bhawuk & Brislin (1992) compared scripts based on only the individualism-collectivism dimension and found that the basis in theory was defective. The first four synthetic cultures, based on the original dimensions of Hofstede (1980), (Individualism, Power distance, Masculinity, Uncertainty Avoidance) were developed by Pedersen & Ivey (1993) as a tool for counselling.

The successful use of the Synthetic Cultures and the lack of similar theory-based scripts in simulation gaming led to the creation in 1999 of a 360-degree set of synthetic cultures with applications in several

simulations games (G.J. Hofstede & Pedersen 1999). Experiences with these were the basis of the book by Hofstede et al. (2002) that placed Synthetic Cultures within the scope of available practice. Harry Triandis, former president of the American Psychological Association, writes in his foreword to that book: “In fact it has been shown (Bhawuk 1998) that theory-based cross-cultural training is more effective than training that consists of scattered samples of beliefs, attitudes, and experiences. Why? It is easier for the learner to absorb the material and generalize to new situations if the training is based on theory.”

“Reflection by the learner is an important mechanism through which results from synthetic cultures can be generalised to real cultures, and existing educational practice links role-play and simulation to subsequent individual reflection and group discussion for this reason.” (eCute Description of work, 2010)

Note that Hofstede et al.’s Synthetic Culture profiles were not based on a formalized model of believable social behaviour; instead, real human groups played with the synthetic culture scripts, so that there was no need to replace their behaviour with a model’s. In eCute, the agent world requires a model of human social behaviour that can then be augmented with cultural differences.

More information on using a theory based learning approach can be found in eCute deliverable 2.1.

3.2. Structure of this deliverable

All people are moral, but culture modifies that morality. This idea is the basis of this deliverable.

To create behaviour that showcases cultural differences, one must first implement some level of basic social behaviour. We can then add culture to this foundation, by having it vary group dynamics.

In the following sections we will begin by discussing some important group level social theories, then we introduce our ideas for a new way of looking at group dynamics, specify these ideas for use within agent architectures, talk about how culture modifies this model, and provide reflections on the trade-offs that will be involved in further implementing the interface in eCute.

A central concept in this deliverable will be the *moral circle*. This is our operationalization of a set of concepts about human sociality that range from entire societies, or indeed all of humanity, down to small groups. We shall introduce moral circles loosely at first, as implicit in social psychological and sociological theory. Then we come up with a more formal set of concepts and definitions that should serve as a design basis for the eCute agent worlds.

4. Human Social Behaviour

4.1. Group membership

Two important theories that deal with the self from a social perspective are Identity Theory (IT), from a sociological perspective, and Social Identity Theory (SIT), from a social psychology perspective. Central in both theories is the concept of a dynamic representation of your ‘identity’ based on the groups to which you belong. In other words, depending on the group you are in, you will act and appraise differently.

Both theories have a similar mechanism to establish this identity. *“In social identity theory and identity theory, the self is reflexive in that it can take itself as an object and can categorise, classify, or name itself in particular ways in relation to other social categories or classification. This process is called self-categorisation in social identity theory (Turner, Hogg, Oakes, Reicher, and Wetherell 1987); in identity theory it is called identification (McCall and Simmons 1978). Through the process of self-categorisation or identification, an identity is formed.”* (Stets & Burke 2000). The difference is that IT focuses more on an individual’s role related behaviour, while SIT focuses more on the relationships among social categories.

4.1.1. Social Identity Theory

Social Identity Theory (SIT) makes a distinction between social and personal identities. Tajfel & Turner (1986) argue that this underlies the difference between interpersonal situations (where behaviour is largely influenced by personal variables) and group situations (control largely by category-based processes). SIT assumes that social identity is derived largely from group memberships and attempts to elucidate group situations on this basis. Moreover, one’s self-concept is set, in part, by the defining characteristics of the social categories to which one belongs (Hogg et al. 1995). It also posits self-esteem as the core motivation of peoples’ drive to achieve or maintain a positive social identity, which is primarily based on making favourable comparisons between the in-group and relevant out-groups. From Brown (2000), *“In the event of an ‘unsatisfactory’ identity, people may seek to leave their group or find ways of achieving more positive distinctiveness for it.”*

According to Hogg and Abrams (1988), in SIT a social identity is a person’s knowledge that he or she belongs to a social category or group. Those who hold a shared social identification (i.e. see themselves as members of the same social category) are a social group. In-group and out-group perceptions are formed by *“a social comparison process whereby persons who are similar to the self are categorised with the self and are labelled the in-group; persons who differ from the self are categorised as the out-group.”* (Stets and Burke 2000)

Different groups value different attributes, and so their norms of behaviour shall differ also. This is an

aspect that is not yet satisfactorily treated in SIT, where distinctions are made with respect to a limited set of variables such as the status, permeability and stability of groups. Brown (2000) notes the theory's paucity of variation across groups and adds that within SIT:

“all groups – be they small face-to-face units or large scale societal categories – are thought to be psychologically equivalent for their members, at least as far as the operation of social identity processes is concerned.”

He then suggests that *“there is much more to social identification than maintaining positivity through biased intergroup comparisons.”* Moreover Smith & Tyler (1997) noted that *“SIT does not include any variable that represents people's feelings about their position within the group.”*

These observations on the limitations of SIT justify our attempt, in this deliverable, to add greater richness to individuals' perceptions of groups, group processes and inter-group relations through our concept of moral circles.

4.1.2. Identity Theory

In Identity Theory (IT), self-categorization is part of the formation of one's identity and this categorization is dependent upon a named and classified world (Stryker 2000). The categories that are present within a culture include roles that designate the relatively stable, morphological components of social structure (Stets & Burke 2000). Stets and Burke (2000) summarise the essence of IT as follows:

“In identity theory, the core of an identity is the categorization of the self as an occupant of a role, and the incorporation, into the self, of the meanings and expectations associated with that role and its performance (Burke and Tully 1977, Thoits 1986). These expectations and meanings form a set of standards that guide behavior (Burke 1991; Burke and Reitzes 1981).”

4.1.3. Wrapping up

Both SIT and IT recognise that individuals view themselves as part of a structured society, though the forms of self-classification differ for SIT and IT, being in terms of groups and roles respectively (Stet & Burke 2000). Summarising the two theories Stet and Burke (2000) state *“In general, one's identities are composed of the self-views that emerge from the reflexive activity of self-categorization in terms of membership in particular groups or roles.”*

Both theories deal with belonging to groups, however, neither of them focus on the internal dynamics of groups.

FP7- 257666-eCUTE

(January, 2012)

One striking example of social dynamics that can as easily be internal rather than external group dynamics is the following:

“When our reality, which we value and believe in, is questioned or rejected by others, then that brings into immediate evaluation the issue of whether we have constructed the ‘correct’ reality, whether we are on the right track, so to speak. And when someone rebukes or rejects us, we as the centre of that reality similarly come in into question. We find that the rejection often causes us to question ourselves, to be self-searching, and to begin to doubt our own value as an individual. In effect the rejection by someone else often says that we are not the kind of person we believe ourselves to be.” (Villard & Whipple 1976).

This quotation relates to the confirmation or questioning of our identities and self-categorisations that are present in many of our social interactions. The issue is not one of common membership of the group, but it is about the moral quality of that membership. Behaviour that would disqualify one as someone who ‘knows how to behave’ is actually quite common in cross-cultural encounters. This is something that we will try to instantiate in our model of moral circles and their dynamics in the next chapter.

The combination of both identity theory and social identity theory provides support for the moral circle, a concept intertwined with group behaviour in all its facets. Members of a moral circle are those who belong (who identify with other members), and follow the rules of behaviour for that moral circle. This concept will serve as the basis of our parameterisable framework.

The worst thing one can do to another is denying their membership of the moral circle of humanity:

“Perhaps the most destructive of all the communication behaviours that can be launched from the inclusion dimension is disconfirmation. Disconfirmation as defined by Watzlawick is the denial of self’s existence. While rejection communicates the message, “you are wrong” or, “you are no good,” disconfirmation says “You do not exist”. And because it is a denial of one’s existence, there is no opportunity for the individual to modify his behaviour. In fact, if one does not exist, then the question of rightness or wrongness of his behaviour becomes irrelevant. For if one does not exist, he has no behaviour.” (Villard & Whipple 1976).

The following quote from a very different source makes a similar point:

“The denial of full humanness to others, and the cruelty and suffering that accompany it, is an all-too-familiar phenomenon. However, the concept of dehumanization has rarely received systematic theoretical treatment.” (Haslam 2006).

So the combination of both identity theory and social identity theory, backed up by other social psychological literature, provides support for a pragmatic concept called the moral circle. This basic concept will serve as the basis of our parameterisable framework. Based on the synthetic cultural scripts, agents will instantiate moral circle dynamics in different ways, and the ascribed roles will differ as well.

4.2.A pragmatic concept: Moral Circles

We shall now operationalize this concept of Moral Circles. In order to develop a conceptual model and software, we need a formalized, fully specified set of concepts. SIT and IT provide many useful insights. Both are open-ended theories that are still developing, and which cannot be directly used for our design purposes. Both also contain conceptual enrichments that we need to simplify for the sake of feasibility. This is why we introduce the moral circle in a very specific sense.

Why use the concept of a moral circle? To begin with, it is generic. Hofstede et al. (2010) use it as a general indication of a human unit of social agency, ranging from a few people to all of humanity, taking inspiration from evolutionary biologist David Sloan Wilson, who describes humans as a ‘eusocial’ species, i.e. one in which the group has supplanted the individual as the main level of evolution. Now, while in most eusocial species it is rather simple to determine the unit of evolution – it would be the colony of bees, for instance – this is not so in humans. Yet the assumption is that we have a biological propensity, including moral sentiments, to act as group members. And it is this propensity that is the main justification for our concept or Moral Circle – which we shall often abbreviate ‘MC’ from this point onwards. In other words, acting for the survival and prosperity of our moral circles is in our nature.

Note that social animals, including humans, can also defect on their Moral Circle obligations. Agents thus have two potentially conflicting drivers of behaviour: ‘Me drives’ versus ‘we drives’. These two are interdependent; an agent who has lost status or reputation in their moral circle might try and start anew in another moral circle.

An additional advantage is that the MC concept sits well with the two theories we have introduced above, SIT and IT, which also similar group dynamics, such as rejection and disconfirmation, to be at work across different types of human groups. We can use MCs to operationalize concepts used in SIT and IT.

4.2.1.Preliminary definitions

It is clear that one moral circle in the mind of each agent will not be enough; any human can be part of more than one MC, and in the context of cross-cultural training we need to make this distinction, among others.

FP7- 257666-eCUTE

(January, 2012)

A first, informal definition is as follows. MCs exist within the minds of individuals; they delineate their social worlds. A MC is comprised of three elements: the people to whom it applies, their mutual perceptions of social attributes, and the rules that regulate their behaviour. It also has potential emergent attributes: the central tendencies of their perceptions.

Several MCs can apply in an agent's mind at the same time. One will always be there as a backdrop: The Core Moral Circle, containing the set of people that are intrinsically considered part of the social world. Other MCs are much smaller and much more limited to specific contexts. These are termed Nested MCs. A work team, or a travelling band, or a school class, are examples.

It is important at this point to distinguish what we mean by a MC from the much broader term of a *group*. The common usage of the word *group* can sometimes closely resemble what we wish to represent with MCs. That is, collections of individuals who have meaningful social relations with particular rules of behaviour that evolve and continue through time. However, in this document we shall use the term MC exclusively, for such "groups" that fit our MC definition. Moreover, a MC is an abstraction in the mind of an agent, not an actual group of agents.

The term *group* itself shall be reserved for collections of individuals together in a certain place, the group of people at a bus stop, or the group of colleagues together at the coffee machine. Such a group is always an instantiation of *one or more* MCs in the minds of its participants; for instance, some of the agents waiting at the bus stop might be educated in such a way that all the others are included in their CMC – and treat them accordingly when the bus arrives – while others might limit their CMC to a subset of the group of waiting people.

It follows that in any simulation, groups will form and un-form as soon as agents move around, and on condition that there be a concept of space in which it is clear whether agents perceive one another or not. The interactions in these groups will then serve to alter moral circles – for instance, agents might find themselves accepted or refused as new members of a moral circle.

As we've seen before from SIT and IT, groups in the usual sense of the word may differ on any symbolic attribute. So it is expected that within a larger MC there will be many different smaller moral circles. For example, on a country level there might be two "groups" which have the same Core moral circle rules, but have different perceived attributes. These differences in perceived attributes can lead to in-group / out-group dynamics between the two. In our model, the groups would be represented as different moral circles.

4.2.2. Types of Moral Circles

As said, within the mind of an individual, we can differentiate between two types of Moral Circle: a Core Moral Circle, and Nested Moral Circles. The CMC deals with all people counted as 'belonging'

to the social world. We call this the ‘Core’ Moral Circle because it is the core of all social behaviour and acts as the hub in which the dynamics of nested MCs takes place. For instance, it carries cultural meta-rules of behaviour.

The nested MCs deal more with social categories, such as your work environment, your sports team, or a group of friends with whom you are organizing a trip. These Nested MCs are perceived by their members to confer rights and obligations on those who belong to the MC. They can vary hugely in membership, formality and the nature of the rights and obligations they confer. There may be formal roles that come with a certain moral or hierarchical authority.

The members perceive each other to matter within specific social contexts, i.e. the context determines which MC applies. For example, a member of an organisation, such as a private company, does not apply the work MC rules when at home with their family; at home, CMC rules hold. Conversely, that person’s life at work would be quite difficult if work rules were in flagrant conflict with CMC rules.

In the following sections we will discuss both types of MCs in more depth.

4.2.3. The Core Moral Circle

A first distinction to make at this point is between those who apply the same ‘moral circle rules’, and those to whom these ‘moral circle rules’ are applied. In cross-cultural contexts the two might not be remotely the same, and this can be a reason for conflicts. Do note that the concepts introduced below are not taken from the existing body of theory.

Here is a description of both.

- **DeepCulture:** Those who shaped my CMC (core moral circle) rules.

Usually these are my parents, siblings, peers, and wider community, with the national level empirically having proven to be important in most cases. We’ll call these my ‘DeepCulture’ for short, to avoid confusion with e.g. organisational culture, or with group identity.

- **The Core Moral Circle or ‘MyPeople’:** Those to whom one applies one’s CMC rules, which is set by one’s DeepCulture.

Usually these are the same as the members of my DeepCulture, except:

- In cross-cultural contexts, when people with different DeepCulture can be friends in a common group;
- In-group / out-group contexts, when people with identical DeepCulture can be cast as inferior, and dehumanised;

- In individualistic societies, where implicitly foreigners are included in MyPeople (CMC). This value pattern can break down in the face of in-group/out-group casting though.

The CMC can be very wide: Dutch children, for instance, are likely to be taught that all people in the world are included. But it is not homogeneous.

The following quotation from Software of the Mind speaks of learning who is in a MC. It can be read as defining the boundaries of the CMC for a collectivist group. As mentioned before, individualist societies tend to teach their young that everybody should be valued, and at least in theory, valued equally.

“...we do not use the same moral rules for members of our group as we do for others. But who is our group? This turns out to be a key question for any group, and from childhood on we learn who the members of our group are and who are not, as well as what that means. People draw a mental line around those whom they consider to be their group. Only members of the moral circle thus delineated have full rights and full obligations.”
(Hofstede et al. 2010).

The CMC has a set of rules for morally evaluating behaviours that are largely fixed and that are set by one's DeepCulture. Membership of the CMC is highly correlated with DeepCulture, to the extent that people from one society are culturally more alike than those from different societies. The CMC is different from Nested MCs in the sense that it is ever present in the moral evaluation of others, and its rules for appropriate behaviour are always salient.

The CMC is formed primarily by the cultural values inherited from one's DeepCulture. In fact, the cultural values are inherited *through* the guidelines to appropriate behaviour that are inculcated as a child progresses from a CMC novice to a full member. These guidelines to appropriate behaviour touch on the most fundamental aspects of social life: Who is important? Who is admired? How do you show respect? Should you be strong and only respect the powerful? Or should you care for all members of your group, and help them when necessary? Must you be Right? Or is it more important to be wise? Are impulses to be resisted and controlled? Or should they be indulged, when possible?

4.2.4. Nested Moral Circles

Each context shapes its own MC typology, which depends on who is involved and what MCs they perceive to be relevant to the situation. Several MCs can affect the actions of any one person at any time. One MC is usually more salient than others. For instance, in most cultures, leaving work duties to marry or bury a family member would be allowable, or even endorsed. The priority between events is itself symbolic of a prioritisation among MCs. How much nesting of MCs to account for, and how

to prioritise, is an important design issue for our model.

Nested MCs can range from long lasting organised groups, with clear memberships and formal roles, such as companies, to the relatively informal, such as groups of acquaintances who share some minor, mutually understood, rights and obligations. The Nested MC rules that relate to basic values are inherited from the CMC. Social norms are often peculiar to a given organised MC. Nested MCs carry many of the *practices* of a society, while the *values* are carried by the CMC. In terms of Hofstede's work on culture, the CMC carries national culture, while members of nested MCs inherit their CMC culture as far as values are concerned, but where practices are concerned, these latter can adhere to norms that are local to the particular nested MC and to the contexts in which that MC is salient. As a consequence, nested MCs differ in the social norms that define moral behaviour within that MC. How salient each Nested MCs is, in the mind of an individual, depends on the situation in which they find themselves.

A more formal Nested MC has both more specific social norms (rules of appropriate behaviour) and a strong inertia in membership; whether you're in or out is usually being determined by clear attributes e.g. employment or club membership. Membership changes in more formal Nested MCs are usually mediated by formal rituals, often denoting a change in status.

More informal Nested MCs can be, for example, groups of specific friends (some you might know from your studies, others from your sports club). These more informal MCs consist of guides to appropriate behaviour when with these groups of friends. Membership of such an informal MC is often not as clearly defined as in more formal MCs. The relevant social norms for an informal MC will not be stated in any text, and can evolve more freely through an emergent consensual process, than is usual in formal MCs.

The rules of Nested Moral Circles are highly influenced by the cultural background of their members. Culture also has a significant influence of the relative salience of MCs when more than one is present in a person's mind.

4.3.Moral Circle behaviour

In our conception, MCs are essential to group behaviour and therefore fundamental to understanding how best to improve a person's cultural competence. Culture-specific competence can evolve when people are exposed to a different culture and learn that culture's unwritten rules of the social game, which in large part are guidelines about how to behave in MCs. General cultural competence requires understanding the nature of group behaviour and how this can vary. This means acquiring an instinct for the dynamics of MCs. Who matters here? And how should I behave towards them, and others?

Appropriate behaviour in a group context forms the basis of our model of MCs and hence their

implementation in intelligent agent architectures. It should be clear by now that MCs are about more than just morality. They also affect how we organise our societies, how our communities interact, how work organisations function, how friendship networks evolve, indeed all aspects of our social lives. In this section we will describe a few aspects of behaviour that are important within MCs.

4.3.1. Moral circles are context dependent

If a group of agents from one CMC encounter a newcomer who is alien, then they will follow an in-group/out-group pattern of reasoning. Their CMC will be salient during this interaction. If, however, that same group of agents meet a crying baby, they might well split along gender roles, depending on their home culture. If they meet a car accident, and there is a doctor among them, the doctor may start to give orders, and the salient MC for the doctor is his work nested MC. In other words, he will start applying rules that are usually only activated at work.

The message is that it depends on perceived context which MC will take precedence in any agent's mind. This prioritization is itself highly culture dependent.

4.3.2. Cultural Meta-Rules and Social Norms

Cultural meta-rules are guides to social behaviour that are contained within the CMC. They are about the fundamentals of social life, and they are shared within any society that has the same DeepCulture. They deal with the basic question of who should be respected, and how this respect should be shown. They are close to the values of a culture, in the Hofstede sense of shared societal beliefs about what is important. Note that these beliefs are very often not conscious on the part the individuals that hold them. They are part of our basic understanding of how the world works, and what matters. The cultural meta-rules are the practical manifestations of these values in social behaviour.

For example, if a well-dressed elderly stranger is seen to misbehave, by violating a social norm against smoking in front of children, the cultural meta-rules related to this situation would define who could intervene and how that could be done. In some cultures correcting a person of significantly higher status would be strongly repressed, which might leave moving the child away from the smoker as the only feasible solution. In other cultures all persons consider themselves to be equal and even a teenager may take it upon themselves to correct an imposing elderly stranger. Cultural meta-rules are about relationships between people, and the archetypal situations that are an integral part of social life, such as correcting, praising, helping, forgiving, trusting, obeying etc. They are distinguished from the specific social norms, which in the above example was a prohibition of smoking in front of children, by being non-context specific and relational.

Social norms are the practices of a group, they reflect underlying value structures, but are not determined by them. They evolve to be accepted by the larger part of society, or a segment of that society, as a guide to proper moral behaviour. Both the interpretation of the moral quality of behaviour

and the translation of intentions into actions, are mediated by the current social norms. These social norms are the most malleable part of MC rules. They can evolve much more quickly than the core cultural-meta-rules of behaviour, which are the direct reflection of values. A population can come to believe that drink-driving or smoking indoors in the presence of non-smokers are normatively wrong, in a relatively short period of time. However, the underlying value structure, and MC dynamics will not have altered significantly, if at all. The detailed functioning of MCs in practice reflects the underlying cultural values, as the cultural-meta-rules mould the social norms of a society. Social norms are one of the tools for interpreting the moral quality of the actions of others. They also indicate what behaviours are allowed (and effective) for translating social intentions into actions.

4.3.3. Moral Circles in action: Rituals

Hofstede et al. (2010) include rituals among the carriers of culture. This makes them very good candidates for the scenarios to be modelled in MIXER and TRAVELLER. In our models, any ritual action will have to be symbolized in some way – in a manner that is made possible by the set of behaviours that the behavioural interface specifies.

Actions can be purely *instrumental*, e.g. picking up an object that has fallen to the floor. The actions in which eCute is interested are those that serve a *symbolic* purpose, such as greeting actions. Actually, most of our actions contain a symbolic element: for instance, what objects would you pick off the floor, in which places, and with which people present?

Rituals range from the ceremonial and memorable to the mundane and transient. In fact, any group of people (in our sense of the word, as a collection of people gathered in one place) that has a degree of shared attention, can be said to be engaged in a mundane ritual, in the sense of Rothenbühler (1998). We could also say that such a group is an ephemeral MC – it could at some point grow into a group of friends, for instance, that develops its own norms.

Rituals are ordered sets of sequential actions that as a whole have a social meaning, though some or all of the individual actions might serve no instrumental purpose. *Rituals* can be seen as markers of the social order. That social order defines identity, relationships and obligations, or in the words of this article, MC membership, the relationships between MC members, and the MC rules of behaviour. The degree to which these relationships are codified is dependent on culture. The stronger the codification, the stronger the *Ritual* needs to be that either reinforces or changes it. In other words, a MC can only change, or maintain itself, through ritual.

Rituals can have a predefined purpose. They have predefined purposes and attendant roles and rules, and orders of events or sub-rituals. For instance, purification often precedes religious *Rituals* or meals.

Smaller units of action can include adjacency pairs, as developed by INESC-ID for FATiMA; these are

speech acts (or non-verbal acts) that require a reaction. They could e.g. come in handy for modelling breaches of convention; any agent that perceives a breach of convention can start an adjacency pair with the offender or in the current ritual, and engage in some kind of sanctioning – or at least develop the intention of sanctioning.

This deliverable will not go into more detail about rituals. Hofstede (2011) elaborates further. In upcoming deliverables, rituals will be revisited in combination with the synthetic cultures.

4.3.4. Verbal and non-verbal behaviour

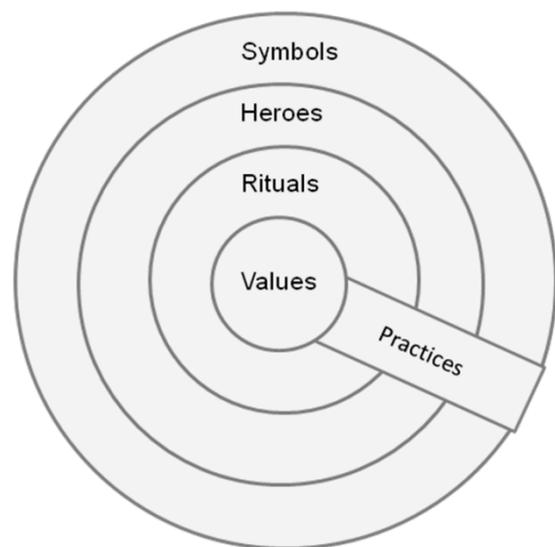
Whatever their level of abstraction, rituals, cultural meta-rules and social norms are all enacted in real life through the actions of individuals. These actions can be separated into two major components: verbal and non-verbal behaviour. Of these two, non-verbal behaviour is probably the most important for intercultural competence; this opinion is also shared by Ting-Toomey (1999), who states in her work that non-verbal behaviour causes many serious misunderstandings when communicating with people from different cultures.

What kinds of non-verbal behaviour can we distinguish? In their new handbook of methods in nonverbal behaviour research, Harrigan et al. (2005) talk about three important domains of nonverbal behaviour: Proxemics, Kinesics and Gaze. In this document we will use their working definitions for these three domains:

- Proxemics is the study of our perception and structuring of interpersonal and environmental space;
- Kinesics refers to actions and positions of body, head and limbs;
- Gaze involves movements and direction of the eyes in visual interaction.

Two elements that we highlight from kinesics are gestures and expressions, as they have a special meaning in a certain environment. The gestures themselves are a manifestation of culture; in Hofstede's Onion (Hofstede et al. 2010) they are considered a part of practices (they can be found among symbols and rituals; see Figure 1). These gestures can range from very direct gestures, such as shaking somebody's hand, or perhaps bowing, to more abstract gestures, such as those of a Japanese tea ceremony.

Many researchers have tried to see if there are



differences in non-verbal behaviour in groups with different cultural backgrounds (for more information, consult the literature review about cultural differences in interpersonal behaviour in deliverable 2.1). Again, our aim is not to provide an in-depth review of the field, but instead we shall present a few examples of the three domains of non-verbal behaviour in action.

Jan et al. (2007) looked into differences in proxemics between Arabs, Americans and Mexicans. They made participants watch two scenarios in which a group of simulated agents are talking while standing in a circle. In the first scenario they gave the agents a Mexican Spanish script, and in the second scenario they gave the agents an Arab script. In practice this means that they had a culturally appropriate level of eye-gaze, turn-taking, and proxemics based on literature (so for example, they used for Mexican Spanish agents the values of 0.45m for intimate, 1.0m for personal and 2.0m for social zones and 0.45m, 0.7m, 1.5m for the respective zones for Arab agents. They found that Arabs significantly prefer the Arab agents with regards to proxemics, while Americans and Mexicans had no specific preference for either. This shows that people from different cultures can be more sensitive to particular forms of non-verbal behaviour. (Jan et al. 2007)

Rehm et al. (2008, 2009) have designed experiments to look at differences in nonverbal behaviour between people from Germany and Japan. In the experiment participants (from both Germany and Japan) were told that they would have to negotiate, or meet, with a representative of a foreign company who was unknown to them (the role of the representative was played out by an actor). They found that Japanese participants used very different gestures from German participants: Japanese persons put their hands more to their wrists (joining hands in front of the body), while Germans put their hands more to their elbows (crossing arms in front of the trunk). There was also a significant difference in the number of gestures that were used: German participants used more than three times as many gestures as the Japanese participants. What is most interesting from the research they have done is that all observed forms of gestural expressivity (Repetition, Fluidity, Power, Speed and Spatial Extent) were significantly different (Rehm et al. 2008).

These are not the only researchers that have studied culturally varying nonverbal behaviour. In their reviews, Ting-Toomey (1999) and Matsumoto (2006) cite numerous studies that have found significant differences in proxemics, kinesics and eye-gaze between different cultures. There are some recurrent basic elements in the literature: expression, gestures, gestural modifiers (referred to in the research of Rehm et al. as gestural expressivity), gaze, and proxemics.

While we do want to implement as many non-verbal actions as possible, we must also consider what is technically feasible. Therefore the small list above, of important non-verbal behaviours, will be used to discuss the possibilities for the eCute scenarios. Besides the specific actions, the design of the MCs within the scenarios (and thus the agents) will be dependent on the synthetic cultures created by Hofstede et al. (2002). In the Annex you can find a summary of these synthetic cultures and the effects

FP7- 257666-eCUTE

(January, 2012)

they have on behaviour. Note that actual agents need not have these extreme behavioural tendencies; instead they will have decision functions that are modified by cultural parameters. These decision functions will be part of the scenario design, and that is the subject of a later deliverable.

5. Modelling Moral Circles

In the previous chapter we have discussed many different desirable properties of Moral Circles. In this chapter we present specified MC primitives based on these properties, and discuss an operational framework in which they can be used. It is clear that both the framework and the variables will never capture the full richness of real life. Instead they are meant as a modelling tool, just rich enough to create emergent behaviour that is representative of real world behaviour.

Moral circles order individuals' understanding of the social world. They do this by assigning the following attributes. The first major attribute is **membership**. This is those who belong to the MC, for whom rules of this MC apply. Within membership there are two possible **roles**, full-member and novice. A novice is forgiven faux-pas against the MC rules more easily, as it is understood that these are still being learnt. A child is a CMC novice, a trainee is a novice within a company or departmental MC.

There are three further variables that define an individual's perception of those they see as fellow MC members. Consider the graphical impressions to the right. If these represent artist's impressions of moral circles, an individual finds itself in a space in which one position is not equivalent to the next. Possible dimensions in this space are **Status, Reputation and Centrality**. These represent the primitives of MCs that we shall now discuss.

5.1.Moral Circle primitives

The CMC can be quite fluid in people's minds and perceptions of who is a member will always be subjective. Each of the MC variables represents an individuals' perceptions, and these can vary greatly. This is of course one way that intercultural problems can arise. We perceive our own actions to be morally right; others might see them as morally reprehensible. The MC is always instantiated differently in the eye of each beholder. In fact, the level of agreement in perceptions among members of any MC is an important emergent variable of that MC.

5.1.1.Moral Circle Reputation (MCR)

Moral Circle Reputation is a measure of how well a person lives up to their MC derived obligations and their respect for the rights of other MC members. It can be conceived of as the ‘moral status’ of a MC member, but we do not use this term for the sake of clarity. Another similar term is ‘standing’, a variable that could be binary or scalar. An agent can be ‘in good standing’ versus ‘in bad standing’ with its fellows (Nowak & Sigmund 2005).

Within our model we want to represent moral behaviour. This means that two important elements need to be present within our model: actions have to be judged as to whether they are moral or not, and members of the moral circle need a perceived level of morality (with unknown people these will be primarily based on exterior attributes). These are the concepts that will be instantiated as MCR within our model.

Each instantiated MC has certain rights and obligations conferred on its members. So if a member of a MC does something that goes against expectations based on an understanding of these right and obligations, it has an effect on their perceived MCR. Each member of the MC has a perception of the MCR of other known members and of their own. So you might think less of yourself if you have done something wrong, and others might also think less of you. This decrease can, depending on the level of MCR change, be attenuated by an appropriate atonement.

5.1.2.Moral Circle Status (MCS)

As well as MCR we also have to deal with differences in *hierarchical* status. An elderly gentleman might not always have a high reputation, but he can still have status due to his age. The same can be said for the CEO of a company. Status in a MC, whether informal or formal, will be instantiated as ‘Moral Circle Status’ (MCS) in our model.

Status is related to dominance as understood in a person’s CMC, which is used to establish the pecking order within a group. Many difficulties between individuals arise because there are differences in perceived MCS (You’re not in charge, I am!) MCS is related to the distribution of formal roles within nested moral circles, which can range from that of the managing director of a multinational company to the most junior trainee. There are generally less misunderstandings in nested MCs regarding MCS, when this is defined by formal roles.

5.1.3.Moral Circle Centrality (MCC)

As we have seen in the previous chapter, one treats close friends and family differently from other people, even though they may all be considered members of the CMC. This means that our model of MCs needs to allow for these kinds of differences.

One option would be to create different levels of ‘centrality’ within a MC. Another is to simply create a new MC for every cohesive subgroup within the larger MC. The choice is largely a matter of

FP7- 257666-eCUTE

(January, 2012)

practical expediency. In this section we shall now assume that MC Centrality (MCC) is used.

MCC deals with “Those who matter to me (within this MC)” and defines the position of a person within that MC. People that are very central matter always, and can almost never be excluded from one’s MC. The most central relationships in the CMC are parent/child, sibling and partner relationships, with close friends and perhaps other close relatives also in one’s “inner” circle. In a nested MC the inner circle will be comprised of people that are closest to you within that MC. These could be the colleagues that you share an office with, or colleagues with whom you work regularly. A MC member with low centrality is someone who matters, but not as much as members of the inner circle; it could be an acquaintance or a fellow queue member. Remember that MCC is about perceptions, and thus may differ across members of the same MC (you might consider somebody within your own Inner Circle, but they might not consider you part of theirs).

Based on our examples above, MCC is probably best represented on a sliding scale; even among friends you can establish a certain hierarchy. Within our initial model, we think it might be better represented as a binary variable to reduce complexity.

- Periphery, you are perceived by somebody as a member of the MC, but you do not matter as much to them as the members of their inner circle;
- Inner Circle, you are perceived as a member of the Inner Circle of the MC, and thus matter a lot to them.

In addition to MCC, we add two roles: novice and member. These apply to both the inner circle and the periphery of MCs. An example illustrating their use:

A daughter brings her boyfriend home to meet her parents. The daughter and her boyfriend have known each other for a while now, and this is the first time that he’s going to meet her parents. The girl’s father knows that his daughter is really happy with him as her boyfriend. So the issue here becomes what position the boyfriend is going to take in the father’s perception of his own CMC.

- He can reject him (I don’t want him in my house anymore!);
- He can accept him very tentatively into the periphery of his MC as a novice (I don’t know who you are, but you are welcome in my house);
- He can accept him tentatively into the Inner Circle as a novice (You are going to be a part of our family, please be at home!).

While he is tentatively accepted into the MC in the last two situations, he is still considered a novice. This means that he can perhaps still afford to err with respect to some of the family’s rules of proper behaviour. This status can last a certain, DeepCulture dependent, time. In this example the novice

status is with regard to certain family or cultural practices, whereas for a child-novice in the CMC, more serious errors can be corrected, and forgiven.

5.2. The Behavioural Interface: Moral Circle dynamics

So far, we discussed the things that can happen in a MC, but not the process according which these things happen. To be able to operationalize the moral circle in a particular context, we have chosen three different levels of group behaviour. These three levels are all equally important and can be used within intelligent agent architectures, together they constitute the behavioural interface (see Figure 3).

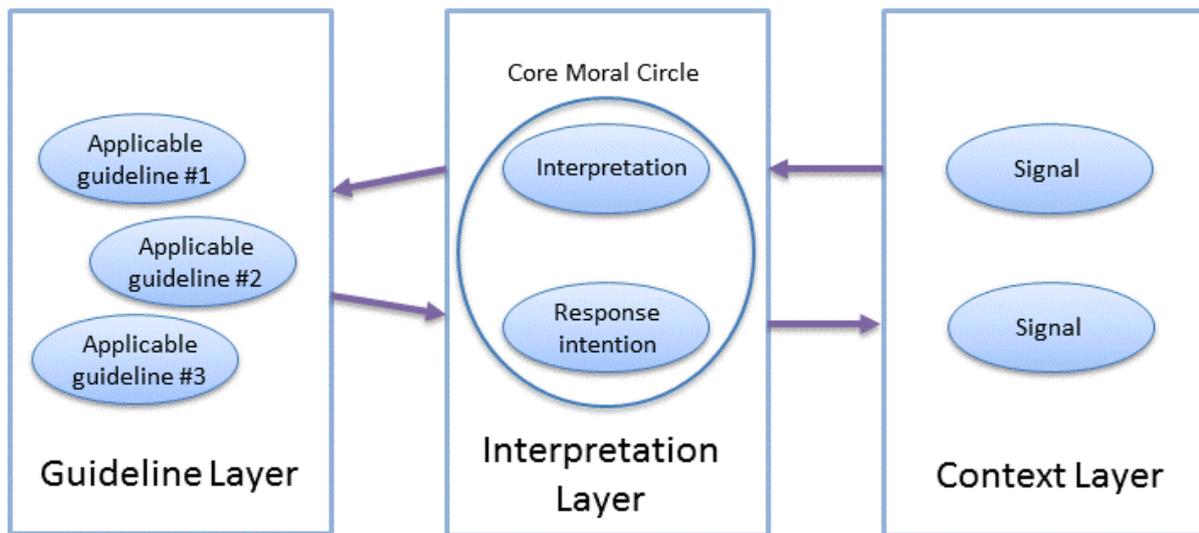


Figure 2: An Agent's Mind in the Behavioural Interface

Within the behavioural interface, a signal is interpreted, then applicable Moral Circle guidelines are sought (moving from right to left in the upper portion of the figure). The interpretation layer then uses the applicable guidelines, along with cultural meta-rules, to select an appropriate response intention. This is then emitted as another signal (which could be an action).

The three levels can be defined in more detail as follows:

Table 1: Three Levels of the Behavioural Interface

Contents

Guideline layer	This level consists of universal group level primitives and dynamics. The primitives are variables that are present for each member of a MC, and the dynamics mainly deal with intentions that are based on the MC primitives.
Interpretation layer	The real world and the group level primitives and dynamics meet each other in the middle layer. Here a filter is applied based on the MC properties (this is where culture enters the equation). On one hand it relates real world actions and relationships to MC primitives, and on the other hand it deals with translating intentions into actions.
Context layer	This layer focuses on different states of the modelled world, where there are interactions between people (verbal, non-verbal, relational)

5.2.1.Guideline layer

The guideline layer deals with two of the three MC primitives: MCR (Reputation), and MCS (Status).

Below you can find Tables 2 & 3, in which MCR and MCS interact in both a formal nested MC and the CMC.

Table 2: Example of MCR and MCS interacting in the CMC:

	Low Status	High Status
Low Reputation	A child stealing candy	An elderly person molesting a child
High Reputation	A child helping an old lady across the street	An adult defending his friend

One issue that arises here is the difficulty of determining what is ‘good’ and what is ‘bad’. These are dependent on your direct environment and your cultural background. However, is it possible to create guidelines that are present within every group? We believe so and consequently we have created a level of reasoning that abstracts from the meanings that are attributed in specific contexts. Hence, this layer will be as context-free as possible.

This framework interprets actions in the context layer into MC Primitives; then translates MC primitives into MC intentions, in order to perform an action in the context layer. This process of converting a given input, into a corresponding output, we call a ‘guideline for social behaviour’.

The input can consist of the following elements:

- Relative differences in the primitives (between yourself and others);
- Fluctuations in the primitives (of yourself or others);
- Do you agree with the proposed changes and fluctuations? (e.g. if somebody wants to increase your reputation, do you think that is fair, or do you accept it anyway even if you doubt its validity).

The output exists out of MC intentions. These intentions are context free messages that signal a change in Reputation and Status *of a member(s)* of the MC. These are not instantiated yet, which means that they do not specify how this change in belief shall affect one’s actions. This will be determined in the interpretation layer. Each MC intention has;

- An intensity which depends on the level of difference or change. This allows for us to work with pivotal actions, which are actions that might fundamentally change the nature of your relationship with somebody else.

An (incomplete) interface can be found on the next page:

Table 3: Incomplete Interface of the Guideline Layer

Change in MC primitives	Do you agree?	Show MC intention	Intensity
Their(MCR) < TheirPrevious(MCR)	Yes	Reproach	Difference(Their(MCR) < TheirPrevious(MCR))
	No	Lift up	Difference(Their(MCR) < TheirPrevious(MCR))
Their(MCR) >= TheirPrevious(MCR)	Yes	Admiration	Difference(Their(MCR) < TheirPrevious(MCR))
	No	Pull down	Difference(Their(MCR) < TheirPrevious(MCR))
Your(MCR) < YourPrevious(MCR)	Yes	Shame	Difference(Your(MCR) < YourPrevious(MCR))
	No	Defend	Difference(Your(MCR) < YourPrevious(MCR))
Your(MCR) >= YourPrevious(MCR)	Yes	Pride	Difference(Your(MCR) < YourPrevious(MCR))
	No	Humility	Difference(Your(MCR) < YourPrevious(MCR))
Relative differences			
Their(MCS) > Your(MCS)	Yes	Obeys	Difference(Their(MCS) > Your(MCS))
	No	Resent	Difference(Their(MCS) > Your(MCS))
Their(MCS) < Your(MCS)	Yes	Command	Difference(Their(MCS) < Your(MCS))
	No	Reject	Difference(Their(MCS) < Your(MCS))

NB. while these MC dynamics already have meaning in our current context, the words used to denote these meanings do not necessarily have their standard meanings here. They are meant as indicators, they still lack a clear and definitive meaning, which would arise from their application in a model. For instance, “command” here might mean disrespect in its purest form (I command you to do whatever I want), or it might mean benevolence (I must make the right decisions for you). These meanings are attributed when generating actions in the interpretation layer.

It might be that multiple guidelines to group behaviour for an individual are triggered simultaneously in a certain situation. Which one is most salient depends on the cultural background of the individual concerned (i.e. his/her DeepCulture). For modelling reasons, we believe that it is best to initially apply only the most salient MC guideline.

5.2.2. Interpretation layer

There can be many MCs circles present in one’s mind at any given time, the CMC, whose rules are established by DeepCulture, always applies. However, it is also possible that multiple nested MCs apply as well. Each of these MCs helps one to give meaning to behaviour when dealing with members of that MC.

Rituals

In our model we use Rituals to represent significant transitions in MCR, MCS and the MCC. We include six Rituals that revolve around either a positive or negative change with one of our three MC primitives. Specifically, for MCR one can change between positive and negative, for MCS between high and low, and for MCC from not in the MC, to the periphery, to the inner circle, and vice versa. These transitions will often involve a very specific set of sequential actions that all members of the MC would understand. After these actions have been completed, all the members present at that moment would understand which primitive has changed.

There are two sides to the interpretation layer. The way up, which consists of interpreting actions into MCR and MCS changes, and the way down, which consists of translating MC intentions into action(s). In this section we will discuss the ‘way up’ first.

Moral Circle Reputation

Those who are within our MC can fall into two categories: those who have high MCR and those with low MCR. All of them are present in one’s mind, and receive moral judgement, but those with low MCR are not perceived to properly uphold the rules of the MC. Significant loss of MCR in the opinion of a sufficient number of MC members could lead to the other member’s wishing to disassociate themselves from the low MCR member.

The most powerful Rituals involve the CMC. Imagine a father who wants to disown his children. They

might not talk to them directly anymore, but they remain present in their mind. The most ‘central’ members of the family MC can rarely be forgotten (removed from the MC) regardless of any perceived lack of MCR. This is a particular case, the exception that proves that rule, and that shows the flexibility of our framework to represent and explore group dynamics.

Moral Circle Status

The clearest examples of MC primitive changes can be found in formal status changes. However, formal status is obviously present for informal nested MCs, which makes it difficult to give clear examples of generic MCS changes in nested MCs. One simple example has to do with the workplace, and formal nested MCs, where your MCS is related to your formal role (and can change by getting a promotion or demotion). In the CMC you can become older (and thus have your MCS increase automatically).

Moral Circle Centrality

When looking at the CMC (do you count as a human being), a striking example of MC Centrality is pre-colonial slavery. These slaves from other countries had to live by different rules than their masters. They were dehumanized; thus not considered to be real humans. In our terminology, they were not members of the CMC in the eyes of most people in the slave holding societies.

The example above highlights the subtle but important distinction between not-in and out. You can also distinguish ‘in’ from ‘out’. It is possible that one does something so bad, that they instantly lose their membership of the MC. An example from real life, in many countries you instantly lose the right to be considered fully human if you molest a small child. These people instantly lose many fundamental rights, and many obligations for normal members of society no longer apply to them.

It is also an important transition to be included in someone’s MC. This can happen for humans, and even for animals and objects (anthropomorphism). Some people consider their dog part of one of their MCs, and there are even situations in which people include robots into a MC.

Transitions from and to the Inner Circle are less clear. However, the people involved will generally always know when such a transition has transpired.

Table 5: Instantiated Rituals

Type of Rituals	Example from the Core Moral Circle	Example from a nested Moral Circle
Becoming a part of the inner circle of a moral circle	Getting married	Becoming a core member
Getting kicked out of the inner moral circle (rejection)	Losing friend status	Getting rejected
Becoming a member of the moral circle	Meeting somebody for the first time	Starting a new job
Getting kicked out of the moral circle (disconfirmation)	Molesting a small child	Getting fired
Getting a higher status	Becoming older	Getting a promotion
Getting a lower status		Getting demoted

‘The way up’

The first step is deciding which MCs are most relevant in the current environment. Culture modifies which MCs are more important in any given context, as well as how easily we can choose our MC of reference to best satisfy our personal drives. The most salient MC may vary based on context. For example, an agent could choose to smoke in a bar because it’s okay for the MC of his friends, while ignoring non- members of this MC also present in the bar. Then the CMC membership of others in the bar would become more salient if a stranger was to sanction his smoking. For modelling reasons we intend that agents will only consider their most salient MC.

The second step is finding Social Norms that apply to the behaviour of the persons involved and determine their MC primitives (MCC, MCR and MCS). Since these primitives and social norms might differ between MCs, they have to be instantiated specifically within the mind of the agent. This means that for every group we include within our scenarios, we have to develop this script manually.

The third step is calculating the MCR change. This will have to be done through an algorithm that contains all the primitives. However, there is no universal formula for this, as culture will modify this algorithm. For example, the influence of MCC might be heavily dependent on whether the agent has a more collectivistic or individualistic script. This would influence both how easily one becomes a close

friend and how you treat friends.

An example: An agent (AgentA) is saying goodbye to another agent (AgentB). AgentA perceives AgentB as a member of the Inner Circle of his CMC; he considers AgentB a close friend. However, agent B does not consider AgentA a close friend (and thus not a part of his Inner Circle). Now let's say they have the same basic rule, that if you say goodbye to someone in the periphery, and there is no MCS and MCR difference, you shake his hand. If you say goodbye to someone who is in your inner circle, then you hug him. Depending on the cultural script of the agent, this might be a huge change in MCR (How dare you think I'm a good friend), or it can end up as a small change in MCR (oh well, you're still a nice guy).

Table 6: Different Behaviours for Different MCC

	Action in moral circle
Neutral or Positive MCR change in Inner Circle (no MCS, MCR diff.)	They hug you
Neutral or Positive MCR change in Periphery (no MCS, MCR diff.)	They shake your hand

'The way down'

Once the 'way up' is completed, the most salient guideline is chosen and an appropriate MC intention with an intensity based on any change in MCR is generated. This all occurs in the guideline layer. This MC intention needs to be translated into an appropriate action on 'the way down' through the interpretation layer.

In the example above, where AgentA says goodbye to AgentB, they need to act in a manner that fits the rules that are appropriate within their MCs. How does one show reproach? It might be that it's not visible at all from an outsider's perspective. However, members of that MC will be better able to interpret any action, or imagine the real feelings hidden therein.

Table7: Different Approaches in Different MCs with regard to Guidelines

	Nested MC 1	Nested MC 2
Admire	Hug them back warmly	Grip them tight
Reproach	Ask them what is wrong with them	Say nothing, look away, and ignore them

FP7- 257666-eCUTE

(January, 2012)

Whether or not a MC intention can be realised in a given situation, is strongly dependent on the social norms that apply to that situation. These social norms will also have an effect on how a MC intention will be realised. Social Norms can apply to specific roles within a moral circle or to persons of a particular status, or when particular configurations of status among those present are realised.

Let's examine a few examples of the social norms of different MCs affecting the translation of MC intentions into appropriate actions, which can of course mean not acting on the MC intention.

When one friend has an intention to reproach another for an act that he perceives to be against a social norm, such as interrupting a high status person during a conversation, whether he can act upon this intention may be determined by whether they are still in interaction with the high status person. If they are, only a non-verbal reproach may be appropriate as a verbal approach might violate the social norm which designates interrupting higher status persons as disrespectful. However, a verbal reproach would be more likely once the higher status person is no longer present, as this social norm would no longer apply 'on the way down' in translating the intention to an appropriate act.

However if the high status person himself were to do something inappropriate that generates the intention to reproach on the part of a lower status person, a social norm of not correcting those of higher would normally prevent this intention materialising into an action.

Let us now look again at the example of two friends in a queue in a shop with strangers when someone tries to jump the queue. The queue jumper naturally generates the intention to reproach by reducing his reputation through breaking a social norm. If the queue jumper is high status (elderly) then this would invoke the "do not correct elderly persons" social norm, and the friends may not react against the queue jumper. They might prefer to find the incident amusing or share their suppressed resentment between themselves.

5.2.3.Context layer

The lowest level focuses on the actions that happen in the real world. 'On the way up' relevant information must be gathered as an input for the middle level. This means that the actions and the relationships must be evaluated and then translated into MC primitives. 'On the way down' actions chosen in the Interpretation Layer must be acted out in through the embodied agent.

As discussed in the section on Moral Circle Behaviour, we want to use the following actions for agents:

- Proxemics
- Eye-gaze
- Expression

- Gestures

A more detailed specification will be a part of deliverable 3.2 and 3.3.

5.3. The modifying role of culture

Our model of the MC, as presented in the previous chapter, should allow a population of computer agents to behave in a believable manner. But what about cultural differences? The MC framework permits modifying agents' behavioural tendencies according to culture. The details of how this is done will be discussed in later deliverables. Here, we shall briefly introduce the concept of culture as we use it.

When we talk about culture, we use the definition found in Hofstede et al. (2010). This definition makes it clear that culture is at the core of affective and cognitive social functioning.

“Every person carries within him- or herself, patterns of thinking, feeling, and potential acting that were learned throughout the person’s lifetime. Much of it was acquired in early childhood, because at that time a person is most susceptible to learning and assimilating. As soon as certain patterns of thinking, feeling, and acting have established themselves within a person’s mind, he or she must unlearn these patterns before being able to learn something different, and unlearning is more difficult than learning for the first time.

Using the analogy of the way computers are programmed, this book will call such patterns of thinking, feeling and acting mental programs, or, as per the book’s subtitle, software of the mind. This does not mean, of course, that people are programmed the way computers are. A person’s behaviour is only partially predetermined by his or her mental programs: he or she has a basic ability to deviate from them and to react in ways that are new, creative, destructive or unexpected. The software of the mind that this book is about only indicates what reactions are likely and understandable, given one’s past.

A customary term for such mental software is culture. This word has several meanings, all derived from its Latin source, which refers to the tilling of the soil. In most Western languages culture commonly means civilisation, or refinement of the mind, such as education, art and literature. This is culture in the narrow sense. Culture as mental software, however, corresponds to a much broader use of the word that is common among sociologists and, especially, anthropologists: this is the meaning that will be used throughout this book.” (Software of the mind)

5.3.1 Dimensions of Culture and MC dynamics

The synthetic cultures will be the core topic of a later deliverable. Here we shall sketch how Hofstede's dimensions of culture could affect the functioning of the moral circle model that we introduce in this deliverable.

Hofstede et al. conceptualize culture as a limited number of major societal issues, to each of which a society finds a shared solution. These issues are conceptualized as continua, as scales with a lower and an upper end. Hofstede et al. (2010) call these 'dimensions' and they describe six of them that vary across nationalities.

The Hofstede model is based on questions about everyday work practices; the dimensions of values were a serendipitous finding. They refer not to convictions or beliefs but to broad tendencies to perceive the social world in a certain way. The model has grown over time, as more sources of data were consulted. The latest model (Hofstede et al. 2010) consists of six dimensions. Each of them is modelled as a continuum running along a scale from 0 to 100. This means that almost all actual values will share characteristics of both extremes.

Here they are, together with some explanation, with an impression of the perceptual capacities that agents need to have in order to accommodate the dimension in a model, and with their effect on moral circle dynamics:

Identity: individualism versus collectivism

Essentially this is the extent to which members of a society feel responsible for themselves, or for the larger group they belong to. In the first case, rights and obligations should be the same for all people, while in the second case, the boundary of the in-group is also a moral boundary beyond which MC rules do not hold. Agents in collectivistic cultures will act differently depending on whether other agents are in-group members or not. In the extreme, out-group agents are simply not considered part of their CMC, and therefore require no social response. On the other hand, their CMC will be quite formally specified, with a large significance attached to an agent's social topology (MCS, MCR, and MCC) within the MC in comparison to personal drives. It might be expedient to define formal roles for the CMC, with associated scripted rituals to effect various MC changes. Comparatively little importance will be attached to nested MCs.

Hierarchy: large power distance versus small power distance

This is the extent to which the less powerful members of a society expect and accept that power and rights are distributed unequally. Large PDI splits up the MC into status levels (MCS) that are not permeable, and depend on position in society. Agents in cultures of large power distance will respond differently to others depending on how they perceive their MCS

relative to their own. Status differences will be effective barriers to communication, and particularly to volitional behaviour travelling upwards.

Aggression and gender: masculinity versus femininity

This dimension is about assertive dominance and emotional gender roles. It contrasts a strong-handed, competitive orientation in ‘masculine’ cultures, in which people in general cannot be assumed to be trustworthy, men are supposed to be tough and women subservient and tender, versus a consensus-seeking and care-taking orientation for both women and men in ‘feminine’ cultures. In masculine cultures, it is desirable that agents be gendered and are able to recognise gender. For our MC primitives this will mean that in masculine MCs, reputation (MCR) will be very unequally divided across the MC, with a tendency to blame the weak or ugly and admire the strong and beautiful. While neither strength nor beauty is currently modelled in our framework, they may sometimes be implied by status (of men, and women, respectively).

Otherness and truth: uncertainty avoidance versus uncertainty tolerance

In uncertainty avoiding societies, anxiety levels are high. In defence against it, strict rules, rituals, and taboos govern life. Distinctions between categories should be sharp and the unknown is considered dangerous. Out-group members and institutions will not be trusted by agents from such cultures. In uncertainty tolerant cultures, relaxation is the rule and actions are results-based rather than anxiety based. The level of fixity of all kinds of rituals increases with uncertainty avoidance. This will mean that MCs with high uncertainty avoidance will have more elaborately scripted rituals and a greater likelihood for newcomers of failing to perform them correctly. This may keep them in a novice status for a much longer time than they would have had in an uncertainty tolerant MC.

Immutability vs. pragmatism: short-term versus long-term orientation

In short-term oriented societies immediate gratification of needs and keeping up social appearance, behaving well and respecting tradition are seen as virtues. In long-term oriented societies, reasoning is pragmatic and planning, foresight and perseverance are valued. Agents from such societies are more likely to learn from experience and to change their norms accordingly. Reputation (MCR) will change more easily as a function of behaviour in a long-term oriented MC.

Gratification of drives: Indulgence vs. Restraint.

Indulgence stands for a society where people feel happy and healthy, and like to enjoy life and to spend time with friends, but could also slip into violence if they feel like it. Restraint stands for a society in which people feel the burden of duties heavily, and keep both positive and negative impulses in check. This dimension will heavily impact the exuberance of greetings.

FP7- 257666-eCUTE

(January, 2012)

More generally, any impulse that comes from an agent's drives is less likely to be restrained by social pressure – such as norms, or other agents' authority – in an indulgent MC than in a restrained one.

Note that the six dimensions are not personality traits, but societal patterns! This means that, unlike personality traits, they will be shared by the persons with the same DeepCulture. Yet culture, as an unconscious set of basic values, should not be confused with conscious group affiliation. Enhancement of the model on the issue of group affiliation is certainly a possible extension. An application would be multi-cultural groups, in which visible group membership is shared but non-visible cultural values are not. Such situations, in which group identity and group culture are not congruent, are a fertile area of research.

Also note that the picture drawn is necessarily simplified. It presents the two caricatured extremes of each dimension. In reality, almost all real world cultures have intermediate positions on almost all dimensions. Furthermore, the dimensions of culture can only be isolated from one another in an artificial way. The six dimensions are no more than abstractions that capture main behavioural trends. In reality, cultures have a recognisable feel to them, a Gestalt that can be described, albeit only roughly, by its combination of dimension scores. So a model of human behaviour will usually require modelling all the dimensions, unless it is very limited in scope.

5.3.2. Synthetic Cultures

Hofstede and Pedersen have operationalized the dimensions of cultures in 10 (now 12) synthetic cultures. These synthetic cultures can be used for experiential learning (in agent-based environments) and are based on the extremes of the dimensions of culture. They contain two parts: what is in the mind of the agents (their cultural meta-rules and social norms) and how they behave (in the form of specific actions and rituals).

Note that actual agents need not have these extreme behavioural tendencies; instead they will have decision functions that are modified by cultural parameters. Synthetic cultures will serve as inspiration for the design of these decision functions. That, however, is part of the scenario design, and will be the subject of a later deliverable. We have included a short description of one of the synthetic cultures and the effect they have on behaviour on the following pages. Incorporating these tendencies in the agent world is a task for later (see Table 8);

Table 4: Large Power Distance (Hipow) vs. Small Power Distance (Lopow)

<p>The Hipow culture is characterized by high power distance. Seven key elements of high-power-distance societies are the following:</p>	<p>The Lopow culture is the opposite of the Hipow culture. It is characterized by extremely low power distance. Seven key elements are the following:</p>
<ul style="list-style-type: none"> • Might makes right and power is good; • Power, status, and privileges go together; • Less powerful people are dependent on more powerful ones; • Centralization is popular; • Subordinates expect direction. They do not speak up without being asked; • The ideal boss is a benevolent autocrat or ‘good father.’; • Style of speech is formal and references hierarchical positions. 	<ul style="list-style-type: none"> • Inequalities among people should be minimized, privileges and status symbols are frowned upon; • There should be, and is, interdependence between less and more powerful people; • Hierarchy in organizations means an inequality of roles only, established for convenience; • Decentralization is popular; • Subordinates expect to be consulted; • In a conversation, anyone can take the word whenever they wish to; • Powerful people try to look less powerful than they are.
<p>Some words that Hipows will use with a positive connotation are: respect, father (as a title), master, servant, older brother, younger brother, wisdom, favor, protect, obey, orders, pleasing.</p>	<p>Some words with a positive connotation among Lopows are rights, complain, negotiate, fairness, task, necessity, codetermination, objectives, question, criticize.</p>
<p>Hipows are restrained and formal in their behavior. They express their intentions in the following manner:</p>	<p>Lopows express their intentions freely, in the following manner:</p>
<ul style="list-style-type: none"> • Friendliness: soft-spoken, polite, listening; • Unfriendliness: quiet, polite, not listening; • Trust: asking for help and direction; • Distrust: not asking for help; • Interest: positive and animated, but no eye contact; • Boredom: expressionless, unanimated, but with eye contact. 	<ul style="list-style-type: none"> • Friendliness: animated, smiling, listening; • Unfriendliness: not listening, talking to others; • Trust: physical closeness, talking freely; • Distrust: physical distance, silence; • Interest: animated, eye contact, with interjections; • Boredom: expressionless, unanimated, no eye contact.

6. The Behavioural Interface in practice

Our ideas, as put forward in the deliverable, will now start a life in the project eCute. This will no doubt induce refinements and changes. Here are some points of discussion, apart from the main thrust of the ideas that should according to us be addressed in the coming months.

The counts_as operator

To use the model presented above within software environments, such as FAtiMA, the counts as operator could be very useful. It can specify that under certain context conditions (e.g. when a particular MC is salient), a certain signal acquires a certain symbolic meaning in terms of MC effects.

Emergent Attributes at Moral Circle level

Working with moral circles allows us to use emergent attributes at moral circle level. For instance, belief updates that one of the members of a MC undergoes can be ‘automatically’ spread among its members, without necessarily having to model all of the individual interactions between members. This belief update mechanism among MC members can be itself subject to the moderating role of Mc parameters, such as status and reputation differences, and of cultural meta-rules about how these MC attributes affect the spread of beliefs.

Controlling the avatar (WP4)

How does the trainee control their avatar? The user should be able to change the non-verbal behaviour of their avatar within the scenarios; they could for example change their proxemics, or their expression based on a certain context. What kind of input modalities the trainee will have to use to interact with the agents is part of Work Package 4.

From the perspective of the behavioural interface there are very few requirements for these input modalities. One thing that needs to be clear for the trainees is that the specific non-verbal behaviours need to follow intuitively from the user selections within the input modalities. Otherwise the user might make unintended actions and misunderstand the intentions of the agents as a result.

Link with existing Intelligent Agent Architecture (WP5)

How our behavioural interface fits in to the FAtiMA architecture that is being used and continually refined, at INESC-ID is an assignment for collaborative design. We believe that the behavioural interface is sufficiently simple and abstract, and compatible with the OCC model (Ortony et al., 1998), that it can easily be incorporated within the architecture. Technical workshops to discuss the exact details will be held in the coming months.

Visual components (WP6)

In order to have the trainee interact with agents that show non-verbal behaviour, the trainee needs to

be able to see both the agents and their behaviour. Within digital scenarios this can be a difficult task, as there are many different ways of representing information in such an environment. It is important to remember that the trainee needs to (visually) see the following elements of other agents:

- the orientation of their head to establish the direction of their gaze;
- the front of their face to establish their expression;
- Their relative location to establish their proxemics;
- Their arms and legs to establish their gestures.

These elements are very dynamic, and may change multiple times during a single conversation (especially for eye-gaze during interactions with multiple characters).

Within digital scenarios this leads to the practical question of camera placement. By camera here we refer to the view that the trainee will see on their computer screen (their viewport into the scenario). In general there are two different placements for the camera: third person, outside of the avatar of the trainee, and first person, through the eyes of their avatar.

There may be some problems when using the third person perspective, unless the camera is quite close to the characters and views the characters from the front (en face), the user might not be able to infer the correct expression, gaze and gestures. However, such a frontal view could lead to confusion with regards to proxemics, as there would be tight restrictions on where other agents can stand with regards to each other.

From the first person perspective the user will be able to view the scenario through the eyes of their avatar. This would lead to a level of freedom in which the user would be able to 'direct' their attention by directly manipulating their view of the scenario. This benefit comes with a price however: the trainee will not have as good an overview of the situation because their viewport will be quite small. This will not be an issue if the trainee only has to interact with a single agent at a time. However, when they interact simultaneously with multiple agents, the trainee would most likely need to 'reorient' their camera as they often would not be able to see all the agents at once.

It is also possible to use multiple cameras, giving the user a third person and first person view at the same time. This has the added problem that it might make the user confused as it becomes less clear where they should focus their attention.

Dynamic vs. static enactment (WP6)

Within digital scenarios non-verbal behaviour can be both dynamic (from one behaviour slowly to another), to very static (directly changing from one to the other). How dynamic does this enactment of non-verbal behaviour need to be within our scenarios?

To give an example of the two extremes of dynamic enactment: the expression of an agent could change directly from angry to sad (without any different visual states of the expression in between), or it could change quite subtly (the expression slowly changing from very angry to very sad), like it would in real life. This of course applies to more than just expressions; gestures and gaze could also change very gradually or directly.

With regards to the learning goals, changing behaviour is not a primary aim of the eCute project, so from the perspective of the project there is no real preference. Completely dynamic behaviour would be the most realistic, but in addition to being more time-consuming it also runs the most risk of entering the uncanny valley, by too nearly mimicking reality. More static behaviour is less life-like, but, in our opinion, still capable of representing the fact that different people show different behaviour.

How to deal with the dynamic vs. static issue is dependent on the (game) engine that will be used and on the amount of time that is available. Besides real animations we can also think of alternatives, such as having icons represent expression and gestures quite statically. This method is used a lot in Japanese media.

Designing characters (WP6)

The context-specific modifiers (i.e. speed and extent) vary the gestures. However, each of the specific gestures that would be part of agents' repertoire needs to be created manually. The range of these specific gestures is dependent on the context of the scenarios, and this leaves considerable freedom. However, we need to keep in mind that the agents should be able to act out these gestures in a realistic manner.

With regards to the appearance of the characters, there are some theoretical constraints to the characters from the perspective of the behavioural interface. Players need to visually make assumptions about the Status of the agents, which means that agents need to have a certain age (for the Core Moral Circle) and certain attributes which denote formal roles and thus a certain amount of MCS. An example of this would be suits in the work environment.

7. Conclusion

This deliverable attempts a major task, which is to derive a theory-based model of generic believable human behaviour, to be used in cross-cultural training scenarios.

We regard the Moral Circle defined above as a theoretically justified representation of the social world, and particularly group behaviour, which is adapted to implementation in agent architectures that can allow for cross-cultural variation in behaviour. Its implementation in agent architectures will help further refine our understanding, as it will provide a platform to test these theoretical ideas using humans' subjects' perceptions of social situations involving simulated agents. We conclude this deliverable with a quotation from Software of the Mind (Hofstede et al. 2010):

"Our mental programs are adapted to life in a moral circle. We take pride in the achievements of our children; we are happy when our favourite sports team wins; many of us sing patriotic or religious songs with feeling and pledge allegiance to our national flag. We are ashamed of the failures of members of our group, and we feel guilty about our crimes. ... Moral, group-related emotions, are universal."SOTM p13

"The moral circle, in many guises and on scales from single marriage to humanity as a whole, is the key determinant of our social lives, and it both creates and carries our future".

In this deliverable we have presented not only a generic model of agent behaviour based on the Moral circle concept, but also a conceptual model for the behavioural interface. While both are linked, changes in either that do not affect the other significantly are still possible. This deliverable presents a set of proposals for elaboration and refinement. Further developments in eCute will be the proof of the pudding.

8. References

- Bhawuk, D. P. S., & Brislin, R. W. (1992). The measurement of intercultural sensitivity using the concepts of individualism and collectivism. *International Journal of Intercultural Relations*, 16, 413-436.
- Bhawuk, D. P. S. (1998). The role of culture theory in cross-cultural training: A multimethod study of culture-specific, culture general, and culture-theory based assimilators. *Journal of Cross-cultural Psychology*, 29, 630-655.
- Brown, R. (2000a). Social identity theory: Past achievements, current problems and future challenges. *European Journal of Social Psychology*, 30, 745-778.
- Harrigan, J. A., Rosenthal, R., Scherer, K. R. (2005). *The New Handbook of Methods in Nonverbal Behavior Research*. Oxford University Press, Oxford, UK.
- Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, 10, 252–26
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.
- Hofstede, G. J., & Pedersen, P. (1999). Synthetic Cultures: Intercultural Learning Through Simulation Gaming. *Simulation & Gaming*, 30(4), 415-440.
- Hofstede, G. J., Pederson, P., Hofstede, G. (2002). *Exploring Culture : Exercises, Stories and Synthetic Cultures*. Intercultural Press
- Hofstede, G. H., Hofstede, G.J., Minkov, M. (2010). *Cultures and Organizations: Software of the Mind* (3rd edition), McGraw-Hill.
- Hofstede, G.J. (2011) Modelling rituals for Homo biologicus. European Social Simulation Association Conference 2011 Proceedings
- Hogg, M.A., and Abrams, D. (1988). *Social Identifications: A Social Psychology of Intergroup Relations and Group Processes*, Routledge, London and New York
- Hogg, M A., Terry, D.J., and White, K.M. (1995). "A Tale of Two Theories: A Critical Comparison of Identity Theory With Social Identity Theory." *Social Psychology Quarterly* 58:255-6
- Jan, D., Herrera, D., Martinovski, B., Novick, D., Traum, D. A. (2007). Computational Model of Culture-Specific Conversational Behavior. In: Pelachaud, C., Martin, J.-C., Andre, E., Chollet, G., Karpouzis, K., Pel' e, D. (eds.) *IVA 2007. LNCS (LNAI)*, vol. 4722, pp. 45–56. Springer, Heidelberg
- Matsumoto, D. (2006). Culture and nonverbal behavior. In Manusov, V., & Patterson, M. (eds.). *Handbook of nonverbal communication* (pp. 219-235). Thousand Oaks, CA: Sage.
- McCall, G. J., & Simmons, J. L. (1978). *Identities and interactions*. Free Press, New York

FP7- 257666-eCUTE

(December, 2011)

- Nowak, M. A. & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature* 437, 1291-1298
- Ortony, A., Clore, G. L., & Collins, A. (1998). The cognitive structure of emotions. Cambridge: Cambridge University Press
- Pedersen, P.B., Ivey, A.E., (1993). *Culture-Centered Counseling and Interviewing Skills: A Practical Guide*. Greenwood Press, Westport CT.
- Rehm, M., Nishida, T., André, E., Nakano, Y.I., (2008). Culture-Specific First Meeting Encounters between Virtual Agents. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) *IWA 2008. LNCS*, vol. 5208, pp. 223–236. Springer, Heidelberg
- Rehm, M., Nakano, Y., Andre, E., Nishida, T., Bee, N., Endrass, B., Huang, H., Lipi, A., Wissner, M. (2009). From Observation to Simulation - Generating Culture Specific Behavior for Interactive Systems. In: *AI & Society* 24 (3)
- Rothenbuhler, E. W. (1998). *Ritual communication: From everyday conversation to mediated ceremony*. Thousand Oaks, CA: Sage
- Smith, H.J., Tyler, T.R. (1997). Choosing the right pond: The impact of group membership on self-esteem and group-oriented behaviour. *Journal of Experimental Social Psychology*, 33, 146-170.
- Stets, J. E., & Burke, P. J. (2000). Identity theory and social identity theory. *Social Psychology Quarterly*, 63, 224–237
- Stryker, S., and Burke, P.J. (2000). The Past, Present, and Future of an Identity Theory. *Social Psychology Quarterly* 63:284-97
- Tajfel, H. & Turner, J. C. (1986). The social identity theory of inter-group behavior. In S. Worchel & L. W. Austin (Eds.), *Psychology of Intergroup Relations*. Chicago: Nelson-Hall
- Thiagarajan, S. (1990). *Barnga: A Simulation Game on Cultural Clashes*. Intercultural Press
- Ting-Toomey, S. (1999). *Communicating across cultures*. Guildford, New York.
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D. & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford: Blackwell
- Villard, K. L. & Whipple, L. J. (1976). *Beginnings in relational communication*. Wiley, New York.