Don't Be a Stranger-Designing a Digital Intercultural Sensitivity Training Tool that is Culture General

Nick Degens, Gert Jan Hofstede, Adrie Beulens, Eva Krumhuber, and Arvid Kappas

Abstract—Digital intercultural training tools play an important role in helping people to mediate cultural misunderstandings. In recent years, these tools were made to teach about specific cultures, but there has been little attention for the design of a tool to teach about differences across a wide range of cultures. In this work, we take the first steps to create a digital self-contained culture-general training tool. In the first part of the article, we focus on different aspects and methods of intercultural training. This information is then used in the second part to evaluate the effect of these different methods on the perception of behaviour in misunderstandings. We found that experiential and story-based approaches may lead to different perceptions of participants. In the third part, we expanded on these critical incidents, and incorporated virtual characters, to evaluate if experiential incidents in an embedded story can lead to an attribution of perceived differences in behaviour to specific differences in culture and to users becoming less judgemental of inappropriate behaviours by people from different cultures. The results suggest that the tool had some effect, but that a debriefing relating the general differences to specific instances would be beneficial.

Index Terms—Social and behavioural sciences, social issues, computer-assisted instruction

1 Introduction

It is becoming increasingly important that people from over the globe are able to work and live together. In 2010 alone, out of the 501 million people living in Europe, 47 million were born in a different country than the one they are staying in [1]. If you are from Europe, sooner or later you will have to interact with people from other countries.

Interacting with people from different cultures may broaden your horizon, but it may also lead to social stress; integration is not always a smooth process and cultural differences may lead to misunderstandings or even conflicts. If such problems keep occurring, they may eventually lead to the creation of negative stereotypes, which may then influence future interactions with people from those cultures.

Adept management of intercultural encounters poses great demands on people who are to interact with people from other cultures. A psychological framework for behavioural components of such socio-cultural interactions is provided by culture learning theory. This theory draws on social psychology, social skills, and interpersonal behaviours [2], to describe social processes taking place between people that are new to a given culture and members of that culture.

According to culture learning theory, intercultural interactions closely resemble any other type of social interactions in that they can be easily disrupted when the parties engaged fail to regulate the interaction appropriately. Generally, this occurs when they are not familiar with conventions that guide the other's behaviour; when people from different cultural backgrounds violate each other's expectations, effective intercultural interactions may be endangered.

The term 'conventions' in this context is closely related to our interpretation of the concept of 'culture', which, in this work, we consider to be the "collective programming of the mind distinguishing the members of one group or category of people from others" [3]. This 'programming of the mind' can take on different manifestations [4], which Hofstede et al. [3] divide into practices and values.

Practices are the visible manifestation of culture, e.g. the way people from a certain culture behave. These may entail norms concerning the expression of emotions, proper use of posture, and performance and appropriateness of various routines, for example greetings [5].

Values are a less visible manifestation of culture, i.e., the beliefs of a group of people. Hofstede et al. [3] mention that values are "broad tendencies to prefer certain states of affairs over others". These values help groups to determine the difference between right and wrong, and, as a result, help to determine which behaviours are considered appropriate in different social contexts. In contrast to practices, values do not vary per culture, as they are the "core of the culture".

It is difficult to determine the precise relationship between values and practices as they are artificial constructs used to describe differences among groups of people. In a sense, values influence which practices are considered

- *N. Degens is with the School of Communication, Media & IT, Hanze University of Applied Sciences, Zernikeplein 11, 9747 AS Groningen, The Netherlands. E-mail: d.m.degens@pl.hanze.nl.*
- *G. J. Hofstede and A. Beulens are with the Information Technology Department, University of Wageningen, Hollandseweg 1, 6706 KN Wageningen, The Netherlands. E-mail: {gertjan.hofstede, adrie.beulens}@wur.nl.*
- *E. Krumhuber is with the Department of Experimental Psychology, University College London, Gower Street, WC1E 6BT London, United Kingdom. E-mail: e.krumhuber@ucl.ac.uk.*
- *A. Kappas is with the Department of Psychology and Methods, Jacobs University, Campus Ring 128759, Bremen, Germany. E-mail: a.kappas@jacobs-university.de.*

Manuscript received 3 July 2014; revised 12 June 2015; accepted 5 July 2015. Date of publication 15 July 2015; date of current version 22 June 2016.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TLT.2015.2456912

appropriate, but they do not determine the specifics of those practices; based on the same underlying values, different practices may arise (even within the same group).

In the past few years, intercultural training has played an important role in presenting and explaining the practices and values of certain cultures; best practices have shown short-term and long-term benefits [6], enabling people to become more competent at interacting with people from different cultures. Often, the trainees' interpretations of behaviours in intercultural situations are challenged, which help them to understand the differences between another culture and their own [7].

This process usually happens as part of what Fiedler refers to as so-called critical incidents [8]. Learning from these incidents usually involves two steps 1) trainees need to become or made aware of misunderstandings or conflicts that occurred during an interaction between people from different cultures, and 2) they need to understand the underlying reasons for these misunderstandings, i.e., relate the differences in behaviour or interpretation to specific differences in culture (a so-called debriefing).

Traditionally, this form of intercultural training requires professional trainers to lead training sessions, and professional actors, or fellow trainees, to participate. Those involved are required to be in the same location at the same time. This can lead to training sessions being very expensive, time-consuming, and difficult to organise.

To deal with these potential problems, researchers have been looking into creating digital intercultural training tools that enable interested people to train their ability to deal with misunderstandings due to differences in culture without having to attend a potentially expensive on-site training session. Such a tool would need to be self-contained, as interested people should be able to run these digital tools on their own (for example from their own personal computer or smartphone).

Since one cannot employ actors or recruit other participants in a self-contained digital tool, there is a need for these tools to simulate social interaction to ensure that trainees can experience situations in which expectations are violated due to differences in culture. Usually so-called virtual characters are employed, which are able to behave according to a set of pre-defined rules. The advantage of using virtual characters over for instance videos of real-life actors is that it is more scalable; for each new situation videos would have to be made, involving real-life actors (preferably the same as before), while virtual characters can be re-used relatively cheaply. As the training should be validated, there are likely to be a large set of iterations in the design of such a tool, and using professional actors can get very expensive.

There are several examples of digital intercultural training tools incorporating virtual characters, and they have very different applications: learning specific foreign languages in an embedded cultural context [9], for example TLCTS/OLCTS [10]; teaching about the cultural heritage of specific countries [11]; integrating a specific cultural influence for different educational purposes [12]; adapting to the user's specific cultural background [13]; teaching about the behaviours that are considered appropriate in a specific culture, for example ELECT BiLAT [14] (for a more detailed overview of the developments in this field, see work by Lane and Ogan [15]).

The strength of these tools is that they are quite effective at preparing trainees to interact with people from a specific culture. However, most of the knowledge and skills acquired from these tools would not apply to different cultural contexts. In other words, if a trainee had to interact with people from another culture later on, they would have to find and use another training tool for that culture.

At the moment, there are no digital training tools that take a so-called culture-general approach, i.e., training to interact with people from a wide range of cultures. Such a tool would not focus on teaching about practices particular for a certain culture, but rather teach people how to better deal with misunderstandings due to culture that occur during interactions with people from a wide range of cultures.

At the moment, there is little knowledge on how to create a digital culture-general training tool. As such, in this article, we will take the first steps in the design of such a tool. In the first section, the theoretical background, we discuss the aims of our training tool, compile a set of requirements based on these aims, and discuss the theory relevant to those requirements. In the second section, we discuss our first experiment, in which we create a set of critical incidents based on the requirements, and compare different presentation styles of critical incidents to see which is the most effective at achieving our aims. In the third section, we discuss our follow-up experiment, in which we created a digital self-contained prototype, based on the results of the first experiment, and evaluated it to determine whether the prototype can be used as-is, to achieve our aims.

2 Theoretical Background

The goal of our training tool is to create an overall improvement of people's performance in intercultural situations. In this section, we will first define what this desired 'improvement' actually is.

As a starting point for our research we took the Developmental Model of Intercultural Sensitivity (DMIS) [16], [17]. In this model, a traveller in a foreign country will go quite linearly through six stages: denial, defence, minimization, acceptance, adaptation, and integration. These stages range from not being aware of differences due to culture, to being aware of these differences and considering the other culture to be inferior, to accepting that there are differences and that the other culture isn't inferior or superior, to fully embracing the world view of another culture.

Although the model is strong in modelling systemic stages of change, and helps to classify attitudes towards people from other cultures, there are drawbacks to using it to define the learning outcomes of an intercultural training tool. First, it does not specify the learning outcomes that facilitate or moderate the course of such a change [18]. Second, the DMIS assumes that people progress from one cultural learning stage to the next in a predetermined order. At this point, there is no clear empirical evidence that the acquisition of intercultural sensitivity follows such a linear path.

To deal with these two drawbacks, we considered it important to specify the learning outcomes that we would need to target. In their work, Kraiger et al. [19] mention that there are three different types of learning outcomes, namely cognitive, affective, and behavioural outcomes.

Cognitive learning outcomes relate to increasing a trainee's knowledge or understanding about a certain topic. In the literature on intercultural training, this process is sometimes referred to as increasing one's 'intercultural awareness' [20], [21]. *Awareness* in the context of intercultural training indicates that people need to be prepared for self-reflection as they try to understand the influence of their cultural background on the way they perceive, interpret, and act out behaviour. This will require trainees to 'become aware of' differences and similarities between their own cultural background and that of other people [22]. An important first step in this process is that trainees actually notice differences in behaviours between people from different cultures [17]. Typically, a culture-specific tool would focus on increasing an understanding of the practices that are considered appropriate in a specific culture. Instead, we would like to help trainees to understand how people from different cultures perceive and interpret similar social situations (for instance, how people from different cultures would treat strangers).

Affective learning outcomes relate to making trainees aware of their emotions and those of others. In the literature on intercultural training, this process is sometimes referred to as increasing one's 'intercultural sensitivity' [20], [21]. Knowledge of differences in culture does not necessarily prevent conflicts and misunderstandings from affecting trainees emotionally. Trainees will also need to become less judgemental towards the behaviour of people from other cultures, and try to be more empathic and attentive, or *sensitive*, to the others' perspective [22], [4], [23]. If we aim to increase the competence of people to deal with misunderstandings due to culture, it would be important to ensure that our trainees are able to handle the negative emotions that arise during these misunderstandings.

Behavioural learning outcomes relate to acquiring physical skills. In the literature on intercultural training, this process is sometimes referred to as increasing one's 'intercultural competence' [20], [21]. As this learning outcome focuses on teaching specific behaviours, i.e., practices, it falls outside of the scope in this work.

Based on these findings, we have determined the learning outcomes that we will target in our training tool:

- Awareness - Trainees should become aware of differences across cultures;
- Sensitivity - Trainees should be able to observe the behaviour of another group without feeling prejudice.

2.1 Designing a Digital Culture-General Training Tool

It is important to establish which elements need to be present within the tool to meet these outcomes. The first is that trainees should be confronted in some manner with an interaction between one or more characters, involving a misunderstanding or conflict due to culture. To ensure learning accurate details about other cultures, it important that these interactions are not based on hypothetical situations, or based on anecdotal experiences, but are instead based on theoretical frameworks of culture (as also argued by Fowler and Pusch [24]). By doing so we can strengthen the relationship between cultural differences in the training tool and cultural differences in real-life.

As such, these interactions should help users to become aware that a misunderstanding of conflict has occurred; become aware that the misunderstanding or conflict occurred due to difference in perception, interpretation, or agency rooted in cultural differences; and, understand that the misunderstanding or conflict is based on a theoretically valid manifestation of culture.

To ensure that trainees are able to do so, we need to determine...

1) ...*the theoretical framework of culture* to use, to help ensure that a trainee is confronted with believable and realistic interactions in which misunderstandings due to culture occur.
2) ...the way we present these interactions to the user; what *method of intercultural training* can be used to achieve our aims.

We will tackle these two problems in the coming sections.

2.1.1 Theoretical Framework of Culture

It is important to use a theoretical framework of culture for the design of an intercultural training tool. Through this framework, we can ensure that trainees are confronted with believable and realistic interactions in which misunderstandings due to culture occur, which will then strengthen the transfer of acquired skills to the real world.

As mentioned in the beginning of this article, we do not intend to focus on teaching aspects of a single culture; rather, we want to focus on differences between a wide range of cultures. In other words, we do not want to focus on the range of practices that are considered appropriate in specific cultures, but on the range of differences between cultures in general (do note, to highlight these differences, we would still need to instantiate the behaviours in practices; however, these practices are not the focus of our work). In a review of intercultural simulation games [24] the authors suggest that the synthetic cultures created by Hofstede and Pedersen [25] hold a lot of promise for the future of intercultural training tools because of their theoretical value.

These synthetic cultures are based on the extremes of Hofstede's dimensions of cultures, which describe societal issues, to which each society has found a shared solution [3].[1] Hofstede et al. have identified six dimensions of culture, namely (1) power distance, (2) individualism versus collectivism, (3) masculinity versus femininity, (4) long-term versus short-term orientation, (5) uncertainty avoidance, and (6) indulgence versus restraint. Each culture has been assigned a score on each of these dimensions based on data of people from those cultures. As a result, each culture can be classified according to their score, e.g. a culture can be very masculine, very feminine, or, most likely, somewhere in between.

1. It is important to mention that there are some controversies surrounding the model of Hofstede. However, its thorough emperical validation, its extensive use in cultural analysis, and its instantiation through the synthetic cultures lend the model well for our work. For more information on these controversies, and a rebuttal, see work by Minkov and Hofstede [26].

As a first step, we have only decided to incorporate the first three dimensions into our training tool. We will discuss these in more detail.

Power distance. deals with the extent to which less powerful members of a society expect and accept that power and rights are distributed equally or unequally. For large power distance, group membership is dependent on position in society, and people from a 'lesser' group are not consulted, while for small power distance, people of a society can expect to belong to any group, regardless of their position in society.

Individualism versus collectivism. deals with the extent to which members of a society feel responsible for themselves or for the larger group they belong to. For individualism, rights and obligations should be the same for all people, while for collectivism, the boundary of the in-group is considered a moral boundary beyond which typical in-group norms do not hold.

Masculinity versus femininity. deals with the extent to which members of a society focus on performance and winning or taking care of the weak. For masculinity, people in general cannot be assumed to be trustworthy, men are supposed to be tough, and women subservient, while for femininity, there is a lot of focus on seeking consensus and taking care for those who cannot take care of themselves.

The synthetic cultures have primarily been used in non-digital intercultural training games [27]. We build on previous work involving the use of synthetic cultures to create critical incidents that feature differences across cultures according to the dimensions of culture by Hofstede et al. [28], [29].

2.1.2 Methods of Intercultural Training

The next step to creating an intercultural training tool is to determine how the learning outcomes can be achieved.

To help users become aware of misunderstandings or conflicts due to differences in cultures, trainers have developed different approaches for intercultural training, and they mainly revolve around 'learning by doing' as opposed to 'learning by telling/showing' or 'experientially' & 'didactically'. This dichotomy has been very central to the discussion of effective intercultural training tools [30].

The aim of experiential intercultural training is to have the user learn about different cultures through life-like interaction with (simulated) people from those cultures. The aim of didactic intercultural training is to have the user learn more about different cultures by providing them with factual information about these cultures. We will now describe these two approaches in greater detail.

An experiential approach to intercultural training rests on the assumption that the best way to teach the trainees about another culture is to make them experience this culture directly or through simulation [30]. Experiential methods include role plays, in which people act as if they were engaged in a real cross-cultural encounter, simulation games, most popular of which is BaFa BaFa [31], and other intercultural exercises that are based on the content of a specific training session [7]. These activities are designed to give the trainees a set of concrete skills that may be applied in unfamiliar situations, and to practice them in a safe environment that provides feedback [7].

Some intercultural training tools may focus on the interactions between people with different cultural rules, but others sometimes include a story-like environment. For example, in ELECT BiLAT [14], American army soldiers are prepared for conducting bilateral negotiations in Iraq. They are guided through a story that might be similar to what they would expect to find in Iraq, and during that story, they have to interact with virtual characters that show believable Iraqi behaviour. By experimenting with behaviours and perceiving what the virtual characters do, trainees become more aware of the specific practices that are considered appropriate in Iraq.

Methods that are primarily didactic, aim at achieving a cognitive understanding of a given topic by giving trainees factual information about another culture [30]. This knowledge is ordinarily distributed via lectures, written materials, for instance books and videos, and field trips, but also through self-assessment tools that allow for exploration of the trainees' private attitudes, case studies requiring the trainees to find a solution to a cross-cultural issue, and critical incidents presenting conflicts stemming from intercultural differences [7].

One example of a didactic training tool involving interactions between people from different cultures is the Culture Assimilator [8]. These assimilators have certain characteristics (based on the description of critical incidents from Flanagan's work [32]):

"For the purpose of developing culture assimilators, the ideal incident must describe (a) common occurrence in which a [trainee] and a [person from a another culture] interact, (b) a situation which the [trainee] finds conflicting, puzzling, or which is likely to misinterpret, and (c) a situation which can be interpreted in a fairly unequivocal manner, given sufficient knowledge about the culture." [8]

After a user has observed the relevant context, they can then choose between four different responses, of which one is considered correct. They then get feedback based on their choice. Learning happens in three stages: first, when the user becomes aware that their assumptions about the behaviour within the critical incident are incorrect, second, when the user understands what other assumptions they should have made, and third, when the user understands on which observations these assumptions can be based.

These Culture Assimilators may also be used in culture-general training, for example through culture-general assimilators [33]. The first of which was the Culture-general assimilator created by Brislin [34]. Culture-general assimilators are very similar to culture-specific assimilators, in so far that they also feature critical incidents, and are presented in a similar manner. The difference is in the fact that they focus on themes that can be applied to interactions with people from any culture, such as Anxiety or Prejudice.

In the last few decades, researchers have been trying to decide whether an experiential or didactic approach yields more effective intercultural training, but it seems there are no clear answers.

Earley [35] was one of the first researchers to empirically compare an experiential approach to a didactic approach. He contrasted two groups, one group that participated in role-play simulations for intercultural training, and another group that received written materials,

comparing the US to South Korea. Earley found that both forms of training had a beneficial effect, but he did not find any significant differences in effectiveness. Participants did prefer the experiential approach.

Pruegger and Rogers' [36] research led to a different set of conclusions. They contrasted four groups; the first two groups participated in a role-play simulation for intercultural training; the third group attended a lecture about differences in culture; and a comparison group. The authors concluded that experiential learning is significantly more effective concerning attitude change in intercultural training experiences than didactic learning.

2.2 Conclusion

It is clear that the different methods of intercultural training may lead to different results in different contexts, and as such, it is important to understand which method is most useful in the design of our tool.

In the first experiment, we have used the synthetic culture descriptions to create a set of believable critical incidents that are then presented to users in different styles (that vary based on the methods of intercultural training). While it is important that these incidents are able to make trainees become aware of misunderstandings and conflicts due to differences in culture, we first focus on determining the impact of the different methods of intercultural training.

In the second experiment, we will create additional incidents in a virtual prototype, again using the synthetic culture descriptions, and evaluate whether the prototype can be used to achieve the intended learning outcomes.

3 Experiment 1

In the first experiment, we take the first steps in the design of a digital culture-general training tool. We have already determined that it is important that trainees become aware of misunderstandings between people from different cultures. That way, we can teach trainees about differences across cultures, and make them understand how people from those cultures respond to similar social situations.

As a first step, we need to determine the effect of different methods of intercultural training on how participants perceive and interpret interactions between people from different cultures. To do so, we have created two critical incidents in which a misunderstanding may occur due to a difference in culture, adapted these incidents to represent the methods of intercultural training described in the theoretical background, and evaluated the effect of these different incidents on the perception of participants. These critical incidents do not yet involve virtual characters, but are instead scripted and text-based.

We have decided to model the experiential approach similar to existing intercultural training tools using role-play (but smaller in scope). In experiential critical incidents, users should feel like they are participating in real interactions, and feel that they are in charge of their own behaviour. As such, these critical incidents are presented in a first-person perspective, i.e., 'you' versus 'he', and users can select one of four possible actions, which lead to different outcomes for each action.

TABLE 1
Different Components in Each Condition

	Type of Critical Incident	Story
Condition 1	Experiential	No Story
Condition 2	Experiential	Story
Condition 3	Non-experiential	Story
Condition 4	Non-experiential	No Story
Condition 5	No critical incident	Story

We have decided to model the didactic approach similar to Culture Assimilators. In didactic critical incidents, users should not feel like they are the one participating in an interaction. As such, these critical incidents are presented in a third-person perspective, i.e., 'he' versus 'you', and users cannot change the outcome; instead, the users are presented the outcome that was chosen by the designers to be the most conflicting out of the four actions available in the experiential incidents. Do note, in contrast to regular didactic approaches, we are not offering additional reflective questions after the incident. This is why we will refer to these incidents as non-experiential in this article.

An additional component that we have chosen to vary is story. We found that existing intercultural training tools sometimes include an overarching story that is used to connect critical incidents, for example in ELECT BiLAT [14], while this is not included in other tools, such as in the Culture Assimilators. It may be that including such a story also influences the perception of trainees, so it is important to check the impact of story components as well. In the story conditions, the users are told that they are searching for their grandfather's long lost treasure. Before and after each incident, they will read how that incident relates to their treasure hunt and what they do in-between. In the non-story conditions, participants only deal with the incidents; there is no story before and after each incident.

Based on these variations, we created a between-group design in which participants are assigned to different experimental conditions. There are five different conditions, and they can be found in Table 1. Condition 1 to 4 feature comparable incidents; they have similar content, but just vary in style and interaction. In condition 5, the user is only presented in-between texts of the story conditions; whenever a critical incident would normally occur, it is skipped.

3.1 Modelling the Incidents

In the first critical incident, we have decided to model one aspect of individualism, namely the boundary between in and out-group members. The main character, who is lost and needs directions, comes from an individualistic culture, and goes up to some strangers and asks them for directions. The characters he will have to interact with are from a collectivistic culture, and do not feel it is their obligation to help.

In the second critical incident, we have decided to model one aspect of power distance, namely the influence of status on behaviour. The main character, who needs permission from someone with a high status, comes from low power distance culture, and asks for permission as if he were an equal. The characters he will have to interact with are from a large power distance culture, and feel that they should be treated with some more respect.

TABLE 2
Significant Differences on the Comparison between Experiential versus Non-Experiential Incidents and Story versus Non-Story Incidents for the First Critical Incident (Table is Written from the Perspective of Experiential and Story Conditions)

Question	Exp. (versus non-exp.)	Story (versus non-story)
How friendly did the people sitting at the table appear to be?	-	More ($p < 0.05$)
Did the barman make you feel a part of the group in the bar?	-	More ($p < 0.01$)
Did the people at the table make you feel a part of the group in the bar?	More ($p < 0.05$)	-
Do you think the people sitting at the table behaved the way they did because of their personality and character?	-	Less ($p < 0.05$)
Do you think the people sitting at the table behaved the way they did because in their culture they come from that is the way people behave?	More ($p < 0.05$)	-
Did the people sitting at the table seem to be offended by your behaviour?	-	More ($p < 0.01$)
Do you think you behaved appropriately in the bar?	Less ($p < 0.01$)	-
Would you choose the same behaviour again if you were in the bar again?	-	Less ($p < 0.01$)
Do you think the people sitting at the table responded appropriately?	More ($p < 0.01$)	More ($p < 0.05$)

In the experiential incidents, the user can select one action per incident, and, depending on that action, a corresponding outcome. For more information on the contents, see the online experiment.[2]

3.2 Evaluation

Since we are interested in evaluating the perception of the participants, we have included certain types of questions that measure different aspects of the interaction. In particular we are interested to understand the impact of the different methods of intercultural training on the perception of the participants in the context of misunderstandings due to culture. Questions asked ranged from 'not at all' to 'very much' (five-point Likert scale).

Do note, in the non-experiential conditions, the word 'you' is replaced by 'Bob', the name of the main character of the non-experiential conditions. Further information about the evaluation can be found in the online link for the experiment.

- Questions related to the cultural aspects used to model the incident; this would be group membership for the first critical incident, and status differences for the second critical incident;
- Questions related to the participant's perception of the appropriateness of their own behaviour, and that of the behaviour of other characters in the incident;
- Questions related to the affective relationship with the other characters, and the underlying reasons for their behaviour.

3.3 Results

3.3.1 Participants

To attract a broad range of participants, we used social media and online experimental databases to gather participants. In total, 228 participants between 15 and 69 years old took part in our online study (142 females; mean age: 27.6; SD age: 11.5). There were 55 participants in Condition 1, 35 participants in Condition 2, 31 participants in Condition 3, 52 participants in Condition 4, and 55 participants in Condition 5. Participants were randomly assigned to these conditions (there was a higher drop-out rate in condition 2 and 3; we believe this is due to the length of these conditions).

The participants came from 29 different countries, ranging from the Netherlands, Turkey, United Kingdom, United States, Germany, etc. (around 40 percent of the participants came from the Netherlands, and there were no more than 15 participants from each other country). While the group is culturally diverse, it is not as large or as representative as we would like; the sample size is not large enough to deduce generalizable conclusions (for a global population).

3.3.2 Comparison of Conditions

To test for significant differences between the story and non-story conditions (1 versus 2 & 3 versus 4) and the experiential and non-experiential conditions (1 versus 4 & 2 versus 3) in participants' responses, Mann-Whitney U tests were conducted to find if there were significant differences in participants' responses as a function of the experiential and story components. We also looked at the impact of the critical incidents (Condition 2 versus 5 & 3 versus 5), but we did not find any significant results. For the first set of comparisons, we found many results, and have chosen to summarise the findings in Tables 2 and 3. More specific results, including the p-values, can be found in the following sections.

Story versus Non-Story

Critical Incident 1

The participants of the story condition (condition 2, choice 1) . . .

- found that the barman made them feel more a part of the group in the bar ($p = 0.005$, $U = 86.5$, $z = -2.83$, $r = -0.46$);
- thought the people at the table were more offended by their actions ($p < 0.001$, $U = 56$, $z = -3.92$, $r = -0.64$);
- would be less likely to choose the same behaviour again ($p = 0.006$, $U = 88$, $z = -2.94$, $r = -0.48$);

2. goo.gl/aQIQZy

TABLE 3
Significant Differences on the Comparison between Experiential versus Non-Experiential Incidents and Story versus Non-Story Incidents for the Second Critical Incident (Table is Written from the Perspective of Experiential and Story Conditions)

Question	Exp. (versus non-exp.)	Story (versus non-story)
I felt that the park supervisor had a [much higher/much lower] social status than me.	Higher ($p < 0.05$)	-
How trustworthy did the park supervisor appear to be?	-	More ($p < 0.05$)
Do you think you behaved appropriately with the park supervisor?	More ($p < 0.05$)	-
Would you choose the same behaviour again if you were interacting with the park supervisor again?	Less ($p < 0.01$)	Less ($p < 0.01$)

. . .than in the non-story condition (condition 1, choice 1).

The participants of the story condition (condition 2, choice 2) . . .

- thought the people at the table were friendlier ($p = 0.023$, $U = 22.5$, $z = -2.36$, $r = -0.51$);
- thought the people at the table behaved the way they did because of their personality and character less ($p = 0.034$, $U = 24$, $z = -2.21$, $r = -0.48$)

. . .than in the non-story condition (condition 1, choice 2).

The participants of the story condition (condition 3). . .

- thought the people at the table acted more appropriate ($p = 0.014$, $U = 558$, $z = -2.46$, $r = -0.27$)

. . .than in the non-story condition (condition 4).

Critical Incident 2

The participants of the story condition (condition 2, choice 1). . .

- found the park supervisor more trustworthy ($p = 0.028$, $U = 67.5$, $z = -2.36$, $r = -0.40$)

. . .than in the non-story condition (condition 1, choice 1).

The participants of the story condition (condition 3). . .

- were less likely to choose the same behaviour if they were to ask the supervisor again ($p = 0.005$, $U = 516$, $z = -2.83$, $r = -0.31$)

. . .than in the non-story condition (condition 4).

Experiential versus Non-Experiential

Critical Incident 1

The participants of the experiential condition (condition 1, choice 3). . .

- felt that the people at the table made them feel more a part of the group ($p = 0.025$, $U = 412.5$, $z = -2.24$, $r = -0.26$);
- thought that the characters sitting at the table behaved the way they did because in the culture they come from that's the way people behave more ($p = 0.024$, $U = 390$, $z = -2.26$, $r = -0.26$);
- thought that their behaviour in the bar was less appropriate ($p = 0.005$, $U = 355$, $z = -2.82$, $r = -0.33$);
- thought the people at the table acted more appropriate ($p = 0.001$, $U = 294$, $z = -3.43$, $r = -0.4$)

. . .than in the non-experiential condition (condition 4).

The participants of the experiential condition (condition 2, choice 3). . .

- felt that the people at the table made them feel more a part of the group ($p = 0.017$, $U = 294$, $z = -3.43$, $r = -0.4$)
- thought the people at the table acted more appropriate ($p = 0.031$, $U = 41$, $z = -2.27$, $r = -0.37$)

. . .than in the non-experiential condition (condition 3).

Critical Incident 2

The participants of the experiential condition (condition 1, choice 2). . .

- felt that the park supervisor had a higher social status than their own ($p = 0.005$, $U = 143$, $z = -2.78$, $r = -0.35$)

. . .than in the non-experiential condition (condition 4).

The participants of the experiential condition (condition 2, choice 2). . .

- felt that the park supervisor had a higher social status than their own ($p = 0.011$, $U = 62$, $z = -2.71$, $r = -0.43$);
- thought that they behaved more appropriately with the park supervisor ($p = 0.028$, $U = 72$, $z = -2.23$, $r = -0.35$);
- were less likely to choose the same behaviour if they were to ask the supervisor again ($p = 0.003$, $U = 52$, $z = -2.92$, $r = -0.46$)

. . .than in the non-experiential condition (condition 3).

3.4 Discussion

As can be seen from the results, we were able to find differences between the experiential and non-experiential conditions and the story and non-story conditions. While it is hard to conclude the specific influence, we can see there is a difference in how participants perceive the behaviour of the characters in the scenario and how they perceive their own behaviour. This is important, as it means that the different intercultural training methods can lead to different perceptions of similar social interactions.

With regard to the behaviour of the characters in the scenario in the story conditions, participants found the characters to be more friendly and trustworthy, that they behaved more appropriate, made them feel more part of the group, and attributed their behaviour less to due to personality and character. Concerning the perception of their own behaviour, the participants considered the characters to be more offended by their actions, and were

less likely to choose the same behaviour again in the story conditions.

With regard to the behaviour of the characters in the scenario in the experiential conditions, participants found that the characters made them feel more a part of the group, that they behaved more appropriate, and attributed their behaviour more to the culture they're from. Concerning the perception of their own behaviour, the participants considered their behaviour both less and more appropriate, and were less likely to choose the same behaviour again in the experiential conditions.

Again, it is hard to conclude the specific influence, as the results vary across different scenarios and different choices. In general, the results may be interpreted as follows: 1) the characters were perceived more positively (more friendly, welcoming, appropriate) and 2) the participants were more critical of their own behaviour (less likely to choose again, less appropriate) in some of the story and experiential conditions. Do note, these conclusions cannot be made for experiential and story conditions in general; additional evaluations, with larger groups of participants and different critical incidents, would be required to determine the exact effect of these intercultural training methods. For now, we can conclude that it is important for other researchers and practitioners to do similar evaluations for their intercultural training tools.

From the perspective of theory, these results are consistent with the work of Moreno and Mayer [37], who found that deeper learning, and accordingly more impact, happens when "the student is addressed as a participant, rather than as an observer". They also conclude that incorporating self-referencing into a training tool helps to increase the mental interaction of the learner. A high level of self-reflection is important for intercultural training, as this awareness is vital for recognizing whether there has been a conflict or misunderstanding.

Some limitations need to be mentioned about our experiment. The first is, it was conducted with text-based critical incidents, to avoid the effects of visualisation on the interpretation of the participants. As a result, the results might not apply to a virtual environment. The second is, we only looked at the relative values of the user's interpretation; we did not measure if participants were actually aware of a misunderstanding. This might yield interesting insights into the effectiveness of different methods of intercultural training.

4 Experiment 2

We used the results from the first experiment to design a digital prototype for culture-general training, henceforth called Traveller. Traveller was developed as part of the European project eCute,[3] and its content is partly based on expanded versions of the incidents presented in the first experiment and a collection of new incidents. There is now also a larger back-story, giving the playing more context and an additional drive to progress through the incidents. Traveller includes so-called virtual agents that are able to make decisions autonomously depending on the context and the selected cultural profile (see Fig. 1). There are many

Fig. 1. Example of an interaction with virtual characters in a critical incident.

components in Traveller, for more information about Traveller as a whole, see previous work [38]; for now we will only focus on the scenario of Traveller.[4]

In Traveller, the user take a role of a young person, who has to travel the world in search of his grandfather's treasure, by interacting with people from three different countries/cultural backgrounds.

There are two goals of this experiment: the first is to evaluate whether going through Traveller can lead to users attributing perceived differences in behaviour or interpretation to differences in values across cultures, and the second is to evaluate if this awareness of differences may lead to users becoming less judgemental of inappropriate behaviours (the two learning outcomes established in chapter 2).

Usually, intercultural training includes a debriefing, helping the user to bridge this gap between the perception and interpretation of behaviour and the actual intentions behind that behaviour. Since our training tool is meant to be self-contained, it is not yet clear how such a debriefing should be structured. Therefore we are focusing on the tool as-is, to evaluate its effect on users; this will help us in future work to establish the requirements of a debriefing method.

To evaluate Traveller, we need to understand how Traveller affects the perception of users. We have done so by randomly assigning participants to one of two groups: the first group answered the evaluation questions without going through Traveller, and the other group went through Traveller and answered the evaluation questions afterwards. In the rest of this article, we will refer to the first group as the 'group without Traveller', and the second group as the 'group with Traveller'. This was chosen instead of a typical pre- and post-test experiment, because the questions would notify the users of our experimental intent, and may have led to users attributing all the behaviour in the scenario to cultural differences.

We conducted the experiment at a university in Germany with international students and at two universities in the Netherlands with Dutch students. Our expectations were that, since we are targeting basic elements of intercultural training, the international students would not show as much change with or without Traveller as the Dutch

3. http://traveller.ecute.eu/

4. Traveller can be downloaded at: goo.gl/kgn5Ib

students would show. As such, we would expect there to be more differences between both groups in the condition without Traveller than in the condition with Traveller. We defined the hypotheses of this research as follows:

Hypothesis 1 (H1) There will be no significant difference between the results of the test with and without Traveller for the international students.

Hypothesis 2 (H2) The Dutch students will score significantly higher on the results from the test with Traveller than on the test without Traveller.

4.1 Modelling the Incidents

The user will go through three different countries within Traveller, and the people from each of these countries have a different 'culture'. By this, we mean that the characters from a country will perceive, interpret, and behave differently than characters from the other countries. The differences between the cultures are based on the three dimensions described in Section 2.2: Individualism versus Collectivism, Masculinity versus Femininity, and Large Power Distance versus Smaller Power Distance.

There are two ways that we have instantiated the dimensions in Traveller. The first is through the behaviour of the virtual characters, with which users will interact during their travels, and the primary reason for a misunderstanding or conflict in each incident. Due to the scope of this article, we have chosen not to focus on the design of these virtual characters, and instead refer to our other work for more information [39], [40].

In short, the characters behave in a certain manner according to their cultural background. Individualistic agents focus more on the task at hand than personal relationships, and treat strangers differently from acquaintances. Masculine agents resolve conflicts by fighting them out, and are less likely to forgive a person who has behaved inappropriately. Large power distant agents consider people that are more powerful as more important than less powerful people are, and they are less likely to consult the opinion of subordinates.

The dimensions of culture and the synthetic cultures were instantiated in the critical incidents as follows:

4.1.1 First Country (Collectivistic, Masculine, Small Power Distance)

First incident: The user is lost, and needs directions to find his hotel, so he enters a bar in which two strangers are deep in conversation. This incident is about task orientation, does the user go directly to the characters in the back, to ask them for directions, or does he wait at the bar for the barman; and about the boundary between in and out group. In this case, the collectivistic characters are wary of strangers, and believe that they are not responsible for helping them out.

Second incident: The user needs permission to enter a wildlife park, and has to ask permission from the supervisor of the park at a formal event (which is held in a museum). This incident is primarily about the effect of status on behaviour: do you approach the supervisor casually, and ask him for permission, or do you approach his assistant, who can introduce you to the supervisor. In this case, the small power distant characters pay much attention to differences in status: the user should just walk up to the supervisor and ask him for help.

Third incident: The park supervisor accidentally knocks over an artefact. This incident is primarily about forgiving versus blaming: does the supervisor blame an employee of the museum, or does he accept responsibility. In this case, the masculine characters try to shift blame to others.

4.1.2 Second Country (Individualistic, Masculine, Large Power Distance)

First incident: The user is on a train, and after a visit from a train conductor, discovers that he has bought a wrong ticket. This incident is primarily about forgiving versus blaming: is he blamed for not having a proper ticket, and told he has to pay a fine, or is he forgiven, and all is well. In this case, the masculine characters confront the user, and fine them for their behaviour.

Second incident: While the user is working in a restaurant, he will have to decide which customer to help first. This incident is primarily about the effect of status on behaviour and task orientation: does one help the elder people, who entered the restaurant last, before the younger people, who entered the bar first, or the other way around. In this case, the characters prefer that the elder people should be served first.

4.1.3 Third Country (Individualistic, Feminine, Small Power Distance)

First incident: While climbing a mountain, his guide has an accident, and is unable to continue. This incident is primarily about winning versus caring and task orientation: does the user leave the guide behind, and travel up the mountain alone, or does he take care of the guide. In this case, the character believes that it is more important to help those in need, instead of purely focusing on the goal.

Second incident: After trying to climb the mountain, the user is invited to a feast, and has to make a toast. This incident is primarily about the effect of status on behaviour and winning versus caring and task orientation: does he wait for the elder to make a toast or does he make a toast himself, and, if so, does he toast to success, or the people he has met in his adventures. In this case, he is expected to make a toast to his friends.

Most of the cultural profiles used in the critical incidents are not yet validated. We have recently evaluated the first critical incident between participants from an individualistic and collectivistic culture, and the results suggest that the behaviour of the characters is representative of individualism versus collectivism (Mascarenhas et al., n.d.). In future work we will validate the cultural profiles used in the other incidents.

4.2 Evaluation

As described in the beginning of this article, we are interested in evaluating if Traveller contributes to users attributing the perceived differences in behaviour or interpretation during the tool to the specific differences in culture that we have used to model the incidents & if users become less judgemental of the 'strange behaviours' of another group.

To measure whether we succeeded at these two aims, we have included two types of evaluation. The first focuses on

the intercultural awareness learning outcome, involving questions related to general differences in culture, and the second focuses on the intercultural sensitivity learning outcome, involving questions related to the affective aspects of norm violations. After these questions, participants were then also asked about their experiences in a short interview.

With regards to the norm violations, we used cultural vignettes. These vignettes, short stories about people interacting, use a norm violation as a probe to measure the affective stance of people towards 'deviant' behaviour. Since we did not find any significant results for this evaluation, we have not included additional information on the vignettes. For more information on the vignettes, see the work by Kappas et al. [41].

4.2.1 General Differences in Culture

During their run through Traveller, trainees are not told in abstract forms about general differences due to cultures, but they do encounter them in the game. Encountering these differences could then lead to a stronger representation of general differences in culture, as determined by a change in the probability/frequency that these behaviour could happen in real-life; a so-called availability heuristic. Do note, there are no 'correct' answers, rather, we are primarily interested in the comparison between the two conditions, not the absolute values.

The questions were based on the description of the synthetic cultures (and correspondingly, the dimensions of culture) that were used in the creation of the incidents.

The following questions were rated on a scale from 0 to 100 percent:

- CQ1 - How often do people from other cultures focus on the task-at-hand rather than personal relations?
- CQ2 - How often do people from other cultures treat strangers differently from acquaintances?
- CQ3 - How often do people from other cultures resolve conflicts by fighting them out?
- CQ4 - How often do people from other cultures forgive a person who has just behaved inappropriately?
- CQ5 - How often do people from other cultures consider more powerful people as more important than less powerful people?
- CQ6 - How often do people from other cultures consult the opinion of subordinates?

These questions are based on the primary descriptors of the synthetic cultures that are also present in Traveller (for more information, see the synthetic culture descriptions [27]).

4.3 Results

4.3.1 Participants

In total, 137 participants between 17 and 28 years old took part in our study. There were two groups, international students from Jacobs University in Bremen, Germany (n = 74; 51 females; mean age 19.89; SD age: 1.60) and Dutch students from both Universities in Groningen, the Netherlands (n = 63; 20 females; mean age 21.89; SD age: 2.65). There were 43 students from Jacobs University in the 'without Traveller' condition and 31 students in the 'with Traveller' condition; there were 37 students from the Universities of Groningen in the 'without Traveller' condition and 26 students in the 'with Traveller' condition. Students joined voluntarily, and received a small monetary incentive for their effort. One participant was removed from the data due to the participant not belonging to the target group.

4.3.2 General Differences in Culture

International students. To test for significant differences between the results with and without Traveller for the international students, both the t-test (CQ1 and 6) and the Mann-Whitney U test (CQ2, 3, 4 and 5) were conducted.[5] We found no significant differences.

Dutch students. To test for significant differences between the results with and without Traveller for the Dutch students, both the t-test (CQ1 and 6) and the Mann-Whitney U test (CQ2, 3, 4 and 5) were conducted. We only found a significant difference for CQ1 "How often do people from other cultures focus on the task at hand rather than personal relations?" Dutch participants rated this question significantly higher after having gone through Traveller condition (p: 0.020; mean without: 47.97, mean with: 58.31).

Comparison international and dutch students. To test for significant differences between the international students from Jacobs University and the Dutch students on the test with Traveller (comparison 1) and without Traveller (comparison 2), we used the Mann-Whitney U test. We found significant differences for both question CQ3 ($p = 0.005$; mean Dutch students: 36.35, mean international students 25,78; $U = 505.5$, $z = -2.82$, $r = -0.32a$) and CQ4 ($p = 0.006$; mean Dutch students: 42.03, mean international students: 30.35; $U = 509.5$, $z = -2.78$, $r = -0.31$) in the test without Traveller, but not in the test with Traveller.

4.4 Discussion

We did not find significant differences for students from the international university between the results from the test with and without Traveller, thus confirming hypothesis 1. We found a significant difference for one question between the results of the test with and without for the Dutch students for the questions about general differences in culture, thus partly confirming hypothesis 2.

When comparing the international students with the Dutch students we found a significant difference for two questions for the results from the test without Traveller. These questions had to do with 'resolving conflicts by fighting them out' and 'forgiving a person for behaving inappropriately'. This difference between the two groups was not present in the test with Traveller, which seems to suggest that the groups were more homogeneous after going through Traveller.

While the lack of significant results on the comparison between the test with Traveller and the test without may seem to imply that the tool was not effective, we did find some interesting results in the interviews conducted after the experiment.

When asked about the goal of the experiment, many of the participants responded that we wanted to see how

5. Some of the results were non-normally distributed, for those the Mann-Whitney tests were used. Others were normally distributed, for those the t-tests were used.

people would respond in certain social and cultural situations. It was however quite difficult for them to link the abstract concepts in the questions about general differences in culture to specific instances of behaviour that happened during Traveller. After explaining the link between these two, participants became aware of other instances of behaviour: "Oh! That would explain why the people in the bar were so distant!" This a-ha moment was also present in other participants "I probably should have been more polite to the elderly man"; "I feel as if I'm now more conscious about behaving appropriately in social situations". This may actually reinforce the need for a proper debriefing, helping the users to bridge the gap between what they saw and what it meant.

Another reason for the lack of results may have been the groups that we have selected. Our initial assumption was that the international students have more a-priori knowledge of intercultural differences than the Dutch students. We tried to verify this by having two conditions (with or without Traveller), and not a typical pre/post-test, as we did not want to inform them about our experimental intent. However, due to small sample sizes, we may have ended up with a more 'knowledgeable' group in the 'without Traveller' group. In future work, we will focus on four parts: 1) determining a good way to measure the level of intercultural competence in potential users in such a way that they are not informed about our research intent, 2) involve different groups based on their a-priori knowledge on cultural differences, 3) validating the questionnaires used in our work, and 4) ensuring a larger sample size to understand the specific influence of the cultural background of participants on their experience.

Another aspect that may have influenced our results, that we discovered through the interviews, is that the transition between the different cultures was not noticed by a large majority of the participants. They were aware of some visual differences, such as palm trees in one country, and a hut in another, but they felt that the behaviour of the characters did not change noticeably. This would imply that different cultures were not noticeable enough, possibly requiring a larger amount of incidents alongside clear transitions between the countries.

5 General Discussion and Conclusion

In this article, we take the first steps to creating a digital culture-general training tool using critical incidents. In the theoretical background section we identified a set of requirements, the most important being that users should be aware that a misunderstanding or conflict has occurred, and that they should relate these misunderstandings to differences in behaviour and interpretation across cultures.

Based on these aims, we established and discussed relevant theories, which we then used to create two prototypes, which were evaluated in two experiments. We found in the first experiment that an experiential approach, embedded in a story, may lead to different perceptions on the appropriateness of behaviour (e.g. more friendly characters, less sure of their own behaviour). This is an important finding, because different methods of intercultural training may influence the effectiveness of a training tool.

In the second experiment, our results suggest that our training tool could lead to users becoming more aware of general differences in culture. While the quantitative data only shows a significant improvement for one set of cultural differences, we found in interviews afterwards that, with a small explanation of these aspects, participants had a greater understanding of cultural differences as a whole.

While the results show that we have created a good basis for a culture-general training tool, there are some limitations to the current version of our tool, experimental setup, and method.

The first of which is the validation of individual incidents. In the second experiment, we evaluated the effect of the entire collection of critical incidents. We did so to evaluate whether our training tool can help trainees progress through our learning outcomes, which is more likely to happen if trainees encounter multiple interactions. For future work, it is important to validate invididual incidents in terms of perception and presence of misunderstandings (see our previous work on the evaluation of incidents [39]).

The second of which is the lack of a debriefing in the current setup. By combining quantitative and qualitative data, we found that users are able to relate general differences in culture to the specific actions of the characters, if they were told about the specific differences. This shows that users need to help to translate implicit to explicit knowledge. For future work, we will evaluate the effect of the tool when embedded as part of a larger intercultural training package.

The third of which relates to the evaluation. The questionnaires have not been validated and the groups we have used do not necessarily yield generalizable results. There is also an unspoken question that relates to how the amount of knowledge that people have of differences across cultures should be measures. For these three issues we will need to find an answer in future work.

Regardless of the limitations, we have also made important contributions to the field: 1) we described the design of a digital self-contained culture-general tool (from learning outcomes to working prototype); 2) we described the use of Hofstede's dimensions of culture to create digital scenarios for culture-general training; 3) we looked at different ways to evaluate the effectiveness of culture-general training tools. Future work will continue the current line of research to further expand and validate Traveller.

Acknowledgments

This work was partially supported by European Community (EC) and was funded by eCute (ICT-5-4.2 257666). The authors are solely responsible for the content of this publication; as such, it does not represent the opinion of the EC or the FCT.

References

[1] K. Vasileva, "6.5% of the EU population are foreigners and 9.4% are born abroad," *EUROSTAT, Statist. Focus*, vol. 34, 2011, http://ec.europa.eu/eurostat/en/web/products-statistics-in-focus/-/KS-SF-11-034
[2] M. Argyle, *Social Interaction*. London, U.K.: Methuen, 1969.
[3] G. Hofstede, G. J. Hofstede, and M. Minkov, *Cultures and Organizations : Software of the Mind*, 3rd ed. New York, NY, USA: McGraw-Hill, 2010.

[4] S. Ting-Toomey, *Communicating Across Cultures*. New York, NY, USA: The Guilford Press, 1999.
[5] P. Trower, B. Bryant, and M. Argyle, *Social Skills and Mental Health*. London, U.K.: Methuen, 1978.
[6] J. S. Black and M. Mendenhall, "Cross-cultural training effectiveness: A review and a theoretical framework for future research," *The Acad. Manage. Rev.*, vol. 15, no. 1, pp. 113–136, 1990.
[7] S. M. Fowler and J. Blohm, "An analysis of methods for intercultural training," *Handbook of Intercultural Training*, D. Landis, J. M. Bennett, and M. J. Bennett, eds., 3rd ed. Thousand Oaks, CA, USA: Sage, 2004.
[8] F. E. Fiedler, T. R. Mitchell, and H. C. Triandis, "The culture assimilator: An approach to cross-cultural training," *J. Appl. Psychol.*, vol. 55, no. 2, pp. 95–102, 1971.
[9] A. Sagae, W. L. Johnson, and R. Row, "Serious game environments for language and culture education," in *Proc. North Am. Chapter Assoc. Comput. Linguistics-Human Lang. Technol.*, 2010, pp. 29–32.
[10] W. L. Johnson and A. Valente, "Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures," in *Proc. 20th Nat. Conf. Innovative Appl. Artif. Intell.*, 2008, pp. 1632–1639.
[11] J. Froschauer, M. Arends, D. Goldfarb, M. Weingartner, and D. Merkl, "Designing socio-cultural learning games-challenges and lessons learned," in *Proc. Int. Conf. Inf. Soc.*, 2011, pp. 56–61.
[12] P. Mohammed and P. Mohan, "Integrating culture into digital learning environments: Studies using cultural educational games," *The Caribbean Teaching Scholar*, vol. 1, no. 1, pp. 21–33, 2011.
[13] E. G. Blanchard and R. Mizoguichi, "Designing culturally-aware tutoring systems with mauoc, the more advanced upper ontology of culture," *Res. Practice Technol. Enhanced Learn.*, vol. 9, no. 1, pp. 41–69, 2014.
[14] R. W. Hill, J. Belanich, H. C. Lane, M. G. Core, M. Dixon, E. Forbell, and J. Hart, "Pedagogically structured game-based training: Development of the ELECT BiLAT simulation," in *Proc. 25th Army Sci. Conf.*, 2006, pp. 1–8.
[15] H. C. Lane and A. E. Ogan, "Virtual environments for cultural learning," in *Proc. 2nd Workshop Culturally-Aware Tut. Syst.*, 2009.
[16] M. J. Bennett, "A developmental approach to training for intercultural sensitivity," *Int. J. Intercultural Relations*, vol. 10, no. 2, pp. 179–196, 1986.
[17] M. J. Bennett, "Towards ethnorelativism: A developmental model of intercultural sensitivity," *Education for the Intercultural Experience*, R. M. Paige, eds. Yarmouth, ME, USA: Intercultural Press, 1993.
[18] B. H. Spitzberg and G. Changnon, "Conceptualizing intercultural competence," *The Sage Handbook of Intercultural Competence*, D. K. Deardorff, eds. Thousand Oaks, CA, USA: Sage, 2009.
[19] K. Kraiger, J. K. Ford, and E. Salas, "Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation," *J. Appl. Psychol.*, vol. 78, no. 2, pp. 311–328, 1993.
[20] G. M. Chen and W. J. Starosta, "A review of the concept of intercultural sensitivity," *Human Communication*, vol. 1, pp. 1–16, 1997.
[21] W. B. Gudykunst, R. M. Guzley, and M. R. Hammer, "Designing intercultural training," *Handbook of Intercultural Competence*, D. Landis, and R. Bhagat, eds., 2nd ed. Thousand Oaks, USA, CA: Sage, 1996.
[22] R. W. Brislin and T. Yoshida, *Intercultural Communication Training: An Introduction*. Thousand Oaks, CA, USA: Sage, 1994.
[23] D. J. Cegala, "Interaction involvement: A cognitive dimension of communicative competence," *Commun. Edu.*, vol. 30, no. 2, pp. 109–121, 1981.
[24] S. M. Fowler and M. D. Pusch, "Intercultural simulation games: A review (of the united states and beyond)," *Simul. Gaming*, vol. 41, no. 1, pp. 94–115, 2004.
[25] G. J. Hofstede and P. Pedersen, "Synthetic cultures: Intercultural learning through simulation games," *Simul. Gaming*, vol. 30, no. 4, pp. 415–440, 1999.
[26] M. Minkov and G. Hofstede, "The evolution of Hofstede's doctrine," *Cross Cultural Manage.: An Int. J.*, vol. 18, no. 1, pp 10–20, 2011.
[27] G. J. Hofstede, P. Pedersen, and G. Hofstede, *Exploring Culture: Exercises, Stories and Synthetic Cultures*. Yarmouth, ME, USA: Intercultural Press, 2002.
[28] N. Degens, B. Endrass, G. J. Hofstede, A. Beulens, and E. Andre, "'What i see is not what you get': Why culture-specific behaviours for virtual characters should be user-tested across cultures," in *Proc. AI Soc*, 2014, http://link.springer.com/article/10.1007%2Fs00146-014-0567-2
[29] S. F. Mascarenhas, J. Dias, R. Prada, and A. Paiva, "A dimensional model for cultural behaviour in virtual agents," *Appl. Artif. Intell.*, vol. 24, no. 6, pp. 552–574, 2010.
[30] K. Cushner and R. W. Brislin, "Key concepts in the field of cross-cultural training: An introduction," *Improving Intercultural Interactions: Modules for Cross-Cultural Training Programs*, K. Cushner and R. W. Brislin, eds. Thousand Oaks, CA, USA: Sage Publications, 1997.
[31] R. G. Shirts, "Beyond ethnocentrism: Promoting cross-cultural with Bafa Bafa," *Intercultural Sourcebook: Cross-cultural Training Methods*, S. M. Fowler, eds. Yarmouth, ME, USA: Intercultural Press, 1995.
[32] J. C. Flanagan, "The critical incident technique," *Psychol. Bull.*, vol. 51, no. 4, pp. 327–58, 1954.
[33] D. P. Bhawuk, "The role of culture theory in cross-cultural training: A multimethod study of culture-specific, culture-general, and culture theory-based assimilators," *J. Cross-Cultural Psychol.*, vol. 29, no. 5, pp. 630–655, 1998.
[34] R. W. Brislin, "A culture general assimilator: Preparation for various types of sojourns," *Int. J. Intercultural Relations*, vol. 10, no. 2, pp. 215–234, 1986.
[35] P. C. Earley, "Intercultural training for managers: A comparison of documentary and interpersonal methods," *The Acad. Manage. J.*, vol. 30, no. 4, pp. 685–698, 1987.
[36] V. J. Pruegger and T. B. Rogers, "Cross-cultural sensitivity training: Methods and assessment," *Int. J. Intercultural Relations*, vol. 18, no. 3, pp. 369–387, 1994.
[37] R. Moreno and R. Mayer, "Engaging students in active learning: The case for personalized multimedia messages," *J. Edu. Psychol.*, vol. 92, no. 4, pp. 724–733, 2000.
[38] N. Degens, G. J. Hofstede, S. F. Mascarenhas, A. Silva, A. Paiva, F. Kistler, and R. Aylett, "Traveller-intercultural training with intelligent agents for young adults," in *Proc. 8th Int. Conf. Found. Digital Games*, 2013.
[39] N. Degens, G. J. Hofstede, J. Mc Breen, A. Beulens, S. F. Mascarenhas, N. Ferreira, and F. Dignum, "Creating a world for socio-cultural agents," *Emotion Modeling*, T. Bosse, J. Broekens, J. Dias, J. van der Zwaan, eds. New York, NY. USA: Springer, 2014, pp. 27–43.
[40] S. F. Mascarenhas, N. Degens, A. Paiva, G. J. Hofstede, A. Beulens, R. Prada, and R. Aylett, "Modelling culture in virtual agents-from theory to implementation," unpublished.
[41] A. Kappas, E. Tsankova, and E. Krumhuber, "Facial responses to cultural norm violations: A vignette approach to probe for changes in ethnocultural empathy," in *Proc. 53rd Annu. Meeting Soc. Psychophysiological Res.*, 2013.

Nick Degens received his doctoral degree on the design of an engaging intercultural agent-based training tool. He is currently an acting professor of user-centered design at the Hanze University of Applied Sciences, the Netherlands. His current research focuses on the (optimal) design of innovative adaptive games and virtual reality applications for real-world cases.

Gert Jan Hofstede is currently an associate professor at Wageningen University, the Netherlands. He creates simulations and animates group simulation games with ambiguous incentive systems. Both methods allow natural group behaviours to occur and to be analyzed in areas such as culture, leadership, negotiation, and trust.

Adrie Beulens is an emeritus professor of information technology at Wageningen University, the Netherlands. His research focuses on the design of state-of-the-art of smart systems and system of systems engineering to support innovations in the life sciences application domains.

Eva G. Krumhuber received her doctoral degree in social psychology from Cardiff University. She is an assistant professor at University College London, United Kingdom. Her research focuses on the processes that lead to the attribution and denial of human-like traits and emotions in humans and virtual agents.

Arvid Kappas is a professor of psychology and a dean at Jacobs University. He is an emotion researcher in different fields, such as psychophysiology or affective computing. Currently, he is the president of the International Society for Research on Emotion.