Article

Inter- and Transcultural Learning in Social Virtual Reality: A Proposal for an Inter- and Transcultural Virtual Object Database to be Used in the Implementation, Reflection, and Evaluation of Virtual Encounters

Rebecca M. Hein [1,2,*] , Marc Erich Latoschik [1] and Carolin Wienrich [2]

1 Human-Computer Interaction, Computer Science 9, University of Würzburg, 97074 Würzburg, Germany; marc.latoschik@uni-wuerzburg.de
2 Psychology of Intelligent Interactive Systems, Institute Human-Computer Media, University of Würzburg, 97074 Würzburg, Germany; carolin.wienrich@uni-wuerzburg.de
* Correspondence: rebecca.hein@uni-wuerzburg.de

Abstract: Visual stimuli are frequently used to improve memory, language learning or perception, and understanding of metacognitive processes. However, in virtual reality (VR), there are few systematically and empirically derived databases. This paper proposes the first collection of virtual objects based on empirical evaluation for inter-and transcultural encounters between English- and German-speaking learners. We used explicit and implicit measurement methods to identify cultural associations and the degree of stereotypical perception for each virtual stimuli ($n = 293$) through two online studies, including native German and English-speaking participants. The analysis resulted in a final well-describable database of 128 objects (called *InteractionSuitcase*). In future applications, the objects can be used as a great interaction or conversation asset and behavioral measurement tool in social VR applications, especially in the field of foreign language education. For example, encounters can use the objects to describe their culture, or teachers can intuitively assess stereotyped attitudes of the encounters.

Keywords: virtual stimuli; *InteractionSuitcase*; implicit association test; virtual reality; social VR

1. Introduction

Due to the global COVID-19 pandemic, inter- and transcultural encounters are limited to tile size, either through social media or video conferencing. Multimodal interactive technologies, such as social virtual reality (VR), provide many potentials to make inter- and transcultural encounters come alive, even under restricted travel regulations. People can meet independently from spatial and temporal distances [1]. The rapid technical progress of VR technology is opening up more and more potential uses for learning applications. VR can significantly contribute to education by allowing a direct and self-determined experience of environments, a targeted replication of situations that are difficult to replicate using traditional teaching methods, or by allowing manipulation of virtual objects and digital artifacts [2–4]. Especially, virtual objects and digital artifacts can be shared and manipulated with each other. They can be initiators and facilitators for conversations and inter- and transcultural encounters. Virtual objects can be used to foster creativity, bridge linguistic barriers, or simply learn vocabulary [5]. Object integration in immersive applications is almost limitless and can be beautifully integrated into foreign language learning [3]. Yang and Liao's augmented reality (AR) application, for instance, empowered students to translate, rotate, scale, and modify virtual objects in 3D using intuitive hand gestures. Thus, cultural content could be easily and quickly visualized and paraphrased [3,6]. Nevertheless, the technical possibility should never be the criterion of immersive applications in education, but always the didactic added value, since the novelty effect fades the effectiveness of VR [7]. Furthermore, there is a "vast design and impact space on human

behavior, including predictable impact paths and manipulable variables" such as virtual objects [1]. Only when we can systematically manipulate variables and only when we can assess the manipulations' impact, we acquire knowledge with practical impact [8]. Hence, the connection between virtual learning materials and didactic approaches possesses great importance in the field of inter- and transcultural learning processes and results. Inter- and transcultural sensitivity is a learning goal in today's foreign language teaching. Getting to know a foreign culture, especially in the beginning, lives on the fascination towards it and uses stereotypes and clichés for motivation and attractiveness [9]. Especially in textbooks, stereotypical images repeatedly occur, such as a *cup of tea* or the *Queen* for the UK chapter and *burgers* and the *Statue of Liberty* for USA-related content. Stereotypes can initiate learning processes, on the one hand, they can quickly transform into negative prejudices or even into discriminatory experiences on the other hand. Similarly, virtual objects might probably support the intercultural and transcultural learning processes, but they might also foster stereotypical thinking. The visual stimuli of textbooks are often very specific, while the visual stimuli used for psychological studies are often rather generic for good reason. Therefore, on the one hand, the question arises how must virtual objects for inter- and transcultural encounters via social VR be designed to be supportive and stimulating? Second, there is no database of virtual stimuli that addresses this particular domain. We would like to provide herewith a first proposal for exchange formats with English and German speakers. To explore the design space of virtual objects for intercultural and transcultural encounters in a social VR learning environment, we first created an extensive database of visual stimuli [5]. In this paper, we present this database of virtual objects for EFL teaching and learning. To be able to use virtual objects in the classroom in a controlled way and to reflect on any stereotypical thinking patterns that may emerge, the objects were evaluated with the help of two online studies. The result of the evaluations is a database with 128 virtual objects named *InteractionSuitcase*. These objects were assigned to a cultural area (e.g., Anglo-American, European, or Latin American) and a contextual area (e.g., landmarks, food and drinks, or animals) by the participants. Furthermore, participants made a recognition task and rated their familiarity, the level of detail, and the stereotypical match with the assigned cultural area. The present work resulted in a standardized visual stimuli database that can be used in remote social VR encounters to trigger a critical review of stereotypical thinking and reflect on and avoid critical incidents. In addition, the virtual objects can serve as an intuitive and implicit measure to evaluate behavior depending on which objects are selected by the participants. Virtual objects can be used as facilitators of intercultural encounters. They can foster creativity, be used to learn vocabulary, or serve as icebreakers and communication aids [10]. In combination with Social VR, the *InteractionSuitcase* enables lively, tangible, and self-determined communication situations.

2. Related Work

Since Blascovich et al. [11] proposed to use VR as a tool for basic research in psychology and other fields, the number of such studies has increased [2,12–14]. However, Hamilton and colleagues [2] identified a lack of VR applications for affective and behavioral changes. As mentioned above, VR can significantly contribute to education by allowing a direct and self-determined experience of environments, a targeted replication of situations that are difficult to replicate using traditional teaching methods or virtual objects and artifacts [4,15,16]. Particularly in light of the current COVID-19 pandemic and climate crisis, where immersion programs are less viable, these potentials could be of great importance for language learning and support intercultural competence [3].

2.1. Inter- and Transcultural Encounters via VR

A literature review addressing foreign language learning and teaching with immersive media revealed that the potentials of VR are rarely utilized to gain inter- and transcultural competencies. In particular, the affective and conative levels of intercultural and transcultural competencies were rarely addressed [3]. To bridge this gap, models describing how and why VR can support behavioral change and learning might be applied to the context of inter- and transcultural learning. The BehaveFIT (Behavioral Model of Immersive Technologies) presented four VR factors influencing human behavior. One of the VR factors is virtual objects and artifacts and their manipulation capabilities, appearance, and interactivity [1]. However, the systematic literature review by Hein and colleagues revealed that virtual objects are less researched than the other factors of the BehaveFIT, beyond AR applications, and concerning inter- and transcultural learning [3]. Hein and colleagues concluded that manipulating object representation, spatial or temporal aspects, and physical laws can expand the inter- and transcultural learning space in VR. In this regard, the manipulation of virtual object representation is a promising feature to influence behavior change [1]. An obvious benefit of virtual objects in education is learning vocabulary or grammar, i.e., cognitive skills. This has already been recognized and used by many other researchers, especially in the field of AR research [3,6,17,18]. However, due to advancing globalization, not only language skills but also intercultural skills have become an important part of modern foreign language teaching [19]. Many researchers see intercultural competencies as a multidimensional construct consisting of three interrelated aspects: intercultural sensitivity (affective aspect), intercultural awareness (cognitive aspect), and intercultural fluency (behavioral aspect) [5,20,21]. Visual stimuli, such as virtual objects, could initiate and facilitate virtual encounters between people from different countries, bridge language barriers, foster creativity, and activate the different dimensions of inter- and transcultural learning [5]. To use virtual objects in social VR for inter- and transcultural encounters as stimuli they must be examined for certain factors. Thus, it is interesting to know how typical an object is in relation to a cultural area and which cultural area it is assigned to, but also other factors such as familiarity or level of detail play a role, especially in relation to the design aspect. To identify these factors, we examined previous research. There are already a few databases consisting of VR content that investigate associations and provide normative data. These will be examined in more detail in the next section and reviewed for feasibility for intercultural and transcultural purposes.

2.2. Visual Stimuli and Object Databases

Visual stimuli are often used in studies to improve memory or language learning or the perception and understanding of meta-cognitive processes. Within the field of psychological research and computer science, visual stimulus databases exist, albeit for different purposes. In the field of psychological research, databases exist with visual stimuli, (e.g., International Affective Picture System, IAPS; [22]). The stimuli with standardized norms vast design and impact space on human behavior, including predictable impact paths and manipulable variables (see Table 1) are to study various psychological aspects of language and visual cognition [23–25]. "Norms represent valuable information that can be used as experimental variables or systematically controlled to limit their potential influence on another experimental manipulation" ([23], p. 1). Hence, some papers offer databases with pictures and 2D normative data for different purposes. However, there are just a few comparable databases of normative data for the use of 3D models as visual stimuli via immersive applications [26–28].

Table 1. 3D and 360° normative databases, their dimension, number of visual stimuli, and participants' sample. N: number of visual stimuli; Dim: dimension; recog: recognition, recall, or name agreement; fam: familiarity; lod: level of detail; arp: arousal or presence; val: valence.

	Dim	*n*	Sample	recog	fam	lod	arp	val
Tromp et al. [26]	3D	123	*n* = 44 native German speakers	54.88	4.05 *	3.08 *	-	-
Popic et al. [28]	3D	121	*n* = 83 US-residents	74.00	4.37 *	2.42 *	-	-
Peeters [27]	3D	147	*n* = 168 native Dutch speakers	74.99	3.20 *	2.69 *	-	-
Schöne et al. [29] °	3D/360°	450	*n* = 41/*n*=68 students from the University of Osnabrück, Germany	55.56	-	-	1.59 **	-
Li et al. [12]	360°	73	*n* = 95 students from a West Coast university, USA	-	-	-	4.22 ***	5.61 ***
Singh et al. [30]	3D	125	-	-	-	-	-	-

* 5-point scale, ** 7-point scale, *** 9-point scale, ° VR condition

Tromp et al. [26] identified three, including their own, databases of three-dimensional virtual objects (see Table 1). The collections contain 121 to 147 objects related to everyday life. Furthermore, we identified three additional studies that examined 360° or 3D content as visual stimuli (see Table 1). We will review the six identified databases in more detail to understand where the objects originated and how they were investigated. Implications for the method of this investigation will be drawn from these findings. The Library for Universal Virtual Reality Experiments (luVRe) by Schöne and colleagues [29] includes 450 videos with 69 subjects recorded with an Insta 360° VR camera (Insta360, Shenzhen, China) and tripod. It includes everyday scenes and unusual encounters with varying arousal and valence (e.g., calming nature scenes, tourist attractions/cities, a visit to the dentist, male strippers, puppies playing, or a beer in a bar). Li and colleagues examined a collection of *n* = 73 360° clips of varying lengths. The collection was created after six months of web screening on websites such as YouTube, Vrideo, and Facebook, or through personal contacts. In total, more than 200 immersive VR clips were viewed and evaluated. Their goal is to provide a public database of virtual content to make experiments more comparable and reproducible in the future [12]. Both papers primarily examined 360° content rather than virtual, modeled objects and examined the affective associations of the content. Schöne et al. measured motivational tendencies using EEG (frontal alpha asymmetries (FAAs)) and compared 2D content with 3D/360° content. In comparison, Li et al. used self-assessment manikins and head movements as variables. In the first study, 15 videos were selected from the database. In the second study, 30 videos were selected to be examined for relative memory performance in free and cued recall. Thus, just a tenth of the library was analyzed. In Study 1, they categorized the 15 videos as positive, negative, and neutral. All negative videos in the 2D condition indexed negative affect, whereas 4 of 5 videos in the 3D condition went the other way. The researchers explain it by suggesting that the belief of being in the spatio-temporal frame of reference of the scene, VR experiences reduce both physical and mental shielding from the events. This raises the meta-awareness that the virtual environment can affect one and, in turn, one is affected [31,32]. Li et al. found in their videos that videos with high arousal indexing negative affect were the minority. Affective and conative experiences may be key to the development and implementation of VR in inquiry and behavior. However, they have rarely been studied in the educational domain [3]. The other four identified databases did not examine 360° content but 3D models. The focus was less on arousal, valence, or emotion but rather on familiarity, visual complexity, and corresponding lexical characteristics of the modal object names [27]. Singh et al. [30] rather pursue the goal of providing a collection with their database *bigBird*, in which certain technical data are specified for each object so that on this basis, machine learning is possible. The collection consists almost mainly of bottles and food packaging and is therefore not useful as a basis for educational purposes. Popic and colleagues' "virtual objects were gathered through discussions within the research group

and aligned with pre-existing 2D databases" ([28], p. 8). They focused on objects from everyday use [28]. Tromp et al. [26] extended the collection of Peeter's [27]. The Peeters et al. database [27] includes 147 color 3D objects standardized for name agreement, image agreement, familiarity, visual complexity, and corresponding lexical features of modal object names. An in-house graphic designer designed them for ongoing experimental VR studies. It is not described on which decision basis these objects were selected. These studies show that many objects affect recognition, recall, familiarity, level of detail, arousal, emotion, valence, or presence. Further differences between 2D and 3D objects have been described. Apart from Schöne and colleagues [29] who measured the emotional and motivational response of the videos using EEG and Li and colleagues [12] who recorded head movements, explicit and conscious measurement methods were most frequently used. Explicit measures provide information about what people want to disclose. Research areas that examine undesirable perceptions and attitudes rely on implicit measures [33]. Thus, implicit measures could complement findings of the connotation and effect of virtual objects. Especially in the case of stereotypes and prejudices towards cultural representations, it is highly likely that social desirability influences the participants' answers. Stereotypes can pave the way for prejudice and discrimination. On the other hand, they also reduce our complex social world to simple categories. They thus enable fast and efficient action in or between social groups and are thus important for inter- and transcultural learning.

2.3. Implicit Measuring

Attitudes towards stereotypes and prejudices towards other cultures are a very sensitive topic where the social desirability of responses has a major role to play [34]. There are a variety of different implicit measurement techniques that are used to examine unconscious perceptions and attitudes. One of these techniques is the Implicit Association Test (IAT). The IAT measures the extent to which non-associated terms make it difficult to answer questions. To this end, the IAT uses the principle of response competition. It measures the extent of unconscious stereotyping by measuring the reaction time required to associate certain terms. In response competition, two responses face each other, a habitual response and an opposing response. The stronger the habitual response, the longer it takes to give the opposite response. Probably the best-known IAT is the *race IAT*. Here, participants are shown pictures of black and white people. The pictures are then assigned either to pleasant or unpleasant words. White people prejudiced against black people usually associate pleasant words faster with white than with black faces and unpleasant words with black faster than with white faces. However, numerous other IATs relate to other concepts, such as gender-career, weight, sexuality, or countries [35,36]. For instance, Karpinski et al. (2004) [37] performed the Implicit Association Test (IAT) with the categories self and other to measure implicit self-esteem. If a person has many positive associations and few negative associations with self, then the association between self and pleasant comes very easily to them, and response times on these trials should be faster than on the other combinations [37]. We assume that a person who has a national bias will associate a virtual object that is stereotypical of his or her country of origin with the concept "self" faster than with the concept "others". Therefore, we asked participants to classify explicitly the objects into national categories (stereotypically German and stereotypically USA) in Study 1 (step 3, compare Figure 1). As an implicit measurement method, we implemented an IAT using the object database. The goal of the test was to find out whether participants have a national bias and whether this IAT can be implemented using our object database. To do this, we use Karpinski's approach [37] by applying the categories "self and others" (step 4, compare Figure 1). In sum, implicit measures can reveal cultural-related, affective, and stereotypical attitudes. Thus, we combined the research on object databases that applied mainly explicit associations and those implicit association methods to investigate virtual objects' cultural and affective association. In the context of English language teaching, learners acquire to develop a critical cultural (self-)reflection or the initiation of a change of perspective [5,38]. The binary concept of "self" and "other," which emphasizes

the experience of difference, multiperspective, and understanding of cultural "otherness," is part of intercultural learning [39].

2.4. Research Questions

Although the usefulness of virtual objects to enhance foreign language learning and teaching is evidenced [1,3,5], no research investigates those objects systematically concerning their cultural and affective associations. In general, object databases for VR are rarely investigated. The existing object databases show little information on cultural-related associations. Mostly, a linguistically and culturally homogeneous sample analyzes a set of objects for variables such as recognition, familiarity, daily use, or arousal [26,28,29]. The different linguistic and cultural contexts are rarely investigated [40]. Moreover, the connotations and associations of these objects were mainly investigated by explicit methods. Implicit measures help to gain a more accurate understanding of attitude dynamics. However, a holistic and systematic approach for such an investigation and provision of object databases for inter- and transcultural purposes is missing. The paper aims to present a proposal for such a database, *InteractionSuitcase*, and its empirical investigation with explicit and implicit methods. The *InteractionSuitcase* is a database of virtual objects designed to facilitate deeper insights into inter- and transcultural competencies under controlled conditions in either an educational or scientific context. Based on previous findings, we hypothesize that virtual objects initiate and facilitate inter- and transcultural encounters and learning processes, make learning tangible, and enrich EFL instruction with an unprecedented learning experience [5]. Therefore, we created a database of virtual objects specifically for inter- and transcultural encounters in a social VR learning setting. In the following, we will briefly present our procedure. In total, a collection of 150 semantic descriptions of objects and images remained (social VR applications: *n* = 100; textbook analyses: *n* = 50). In the next step, we searched for free or low-cost 3D models for each semantic description of the previously scanned objects in the commercial databases. "Natural scenes are more appealing [41], but also induce more complex responses that could contain confounding factors" [42]. Therefore, and to have more than one representative for each modal name, we searched for a more specific and a more generic object image for each identified item. For instance, Popic et al. [28] had at least two representatives for each modal name, too. Whenever possible, the choice of specific objects was based on a "German" and an "English" transliterated object. After expanding the collection, we proceeded to step 3 with a total of *n* = 293 objects (*n* = 141 generic, *n* = 152 specific) (compare Figure 1.
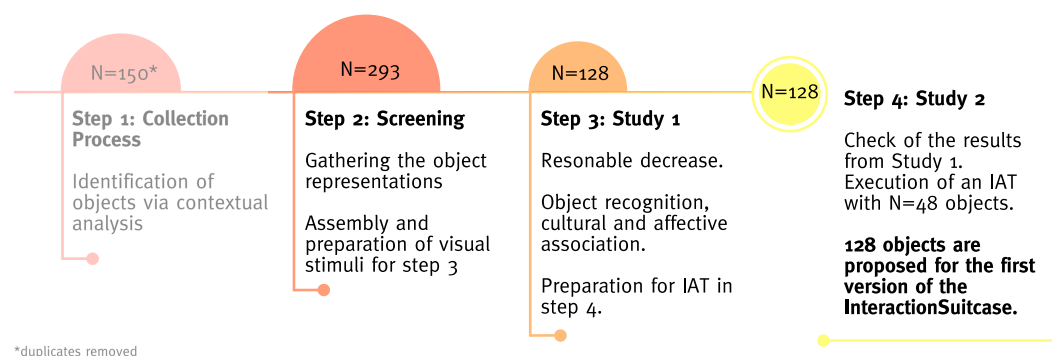

Figure 1. Step 1 describes the collection process. Step 2 screens the objects according to predefined criteria. Step 3 describes a study with a homogeneous sample. The objects were examined for recognition and cultural association i. a. in preparation for IAT in step 4. In step 4, a second study was conducted where the results from Study 1 were verified, and an implicit association test (IAT) was performed. Hereby, a heterogeneous sample was interviewed.

The present paper investigated the associations of these objects by the following research questions:

- **RQ1:** What are the explicit, normative properties of the identified visual stimuli (Study 1)?
- **RQ2:** Do English and German native speakers feel implicitly more or less connected with cultural (English or German)-associated objects (Study 2)?

To address these research gaps, the research followed four steps (Figure 1). Steps one and two contribute to the identification of visual stimuli, the gathering of representations, and the assembly and preparation of visual stimuli. These two steps have already been completed and published [5]. Steps three and four focus on this paper and address RQ 1 and RQ 2.

3. Study 1: Recognition, Cultural, and Affective Associations of the Virtual Objects

As earlier mentioned, intercultural competence is a multidimensional construct [5,20,21]. It has cognitive, social, affective, and conative correlations that make it possible for inter- and transcultural attributions to be formed due to unconscious and conscious processes. In the first online study, a total of 293 virtual representations identified through the context analysis of English textbooks and social VR applications were examined [5]. Below, we present the used method.

3.1. Method

3.1.1. Participants

Participants were asked about their age, biological gender, educational qualification, and occupation as part of the socio-demographic questionnaire. They were also requested to identify the country in which they currently live and the country where they have lived for the longest time. In Study 1, participated 230 participants ($n = 42$ males; $m = 21.47$ years ($sd = 3.14$)). Most of them ($n = 225$) were native German speakers (compare Figure 2). The other five were native speakers of English. Some respondents were recruited via the university sampling system, while others were recruited via a snowball system. The largest proportion was made up of students of media communication ($n = 191$). Five students are studying human–computer interaction in the Master's program and $n = 25$ in the Bachelor's program. Furthermore, one psychology student, one engineer, two research assistants, one flight attendant, one occupational therapist, one manager, and one business and modern dance student participated. One person indicated to be a student only.

3.1.2. Materials and Apparatus

The questionnaire was created with the tool *SoSci* Survey [43]. We used 2D images of the models as a representation in the questionnaire. All images were presented in the center of the screen and had an image height of 400 pixels. All links to the 3D models used ($n = 293$) can be found in the supplementary material section. We measured screen size and questionnaire width to ensure that subjects were presented with the visual stimuli in sufficient quality and under comparable conditions. The participants' screen sizes on which the study was conducted were on average $m = 1485 \times 878.2$ ($sd = 287.9 \times 150.7$) pixels in size and the questionnaire width averaged $m = 795.7$ pixels ($sd = 41.4$).

In this first study, explicit recognition and assignment to a cultural area and affective connotations were recorded for each object. We checked whether the objects were correctly recognized by asking the participants to name them. In addition, they were asked to name the country they most likely connect with the object and three other associations (affective connotation). In addition to the explicit method, a fast assignment task was executed as well. In an assignment task, the participants are shown the stimuli they must assign to one of two categories. Through this assignment, a range of indexes can be determined for each of the stimuli, which indicates which category the object was most frequently assigned to and the reaction time, and how quickly this assignment occurred.

3.1.3. Procedure

After completing the demographic questionnaire, participants performed a fast assignment task. The assignment task shows visual stimuli in the form of images of virtual objects.

An explanatory text is displayed before the first stimulus. Pressing the space key starts the assignment task, and the participant must classify stimuli into one of two categories. The categories or selection options are the same for each stimulus. The assignment is done by button press (arrow keys). The assignment task works similarly to a single test block of the IAT [43]. The reaction time is recorded in milliseconds with an accuracy of about 10 ms. The sequence of a stimulus always follows the same pattern:

- Display of a fixation cross (until the space key is pressed)
- Display of the visual stimulus (500 ms)—selection options

In the first run, all participants were shown pictures of generic and complex spheres in the center of the screen to practice this kind of task. Then, the participants sorted these spheres in time either to the categories generic (right arrow) or complex (left arrow); (Figure A1). We adapted this task from the first part of Experiment 2 by Wienrich and Latoschik [16]. In other words, the exercise did not include the target object or assignment categories. Then, the same assignment task was repeated with the 293 identified objects. Again, half of the participants ($n = 115$) categorized the objects as "positive" or "negative" (indicating the fast affective association), and the other half as "German" or "English" (indicating the fast cultural association; see Figure A2).
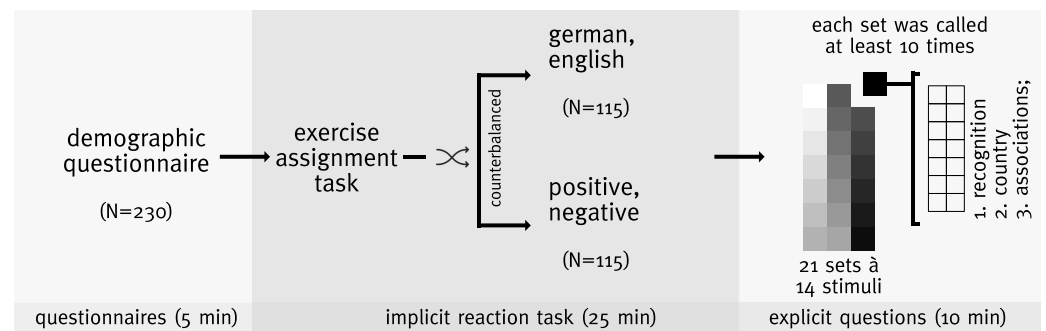

Figure 2. This figure shows the study design and timeline of the first study.

After the fast assignment task, all 293 visual stimuli should also be evaluated explicitly. To not cognitively overload the participants and be expected to obtain valid results, we divided the stimuli into 21 sets of equal size ($n = 14$), making sure that no two stimuli with the same or similar modal name were included in one set. One set contained a sphere from the exercise task as a filler to make all sets equal in size. The distribution can be found in the supplementary material section. A sequential analysis resulted in a systematic reduction of the virtual objects.

Recognition

First, explicit recognition was examined. Objects that were not recognized or were recognized incorrectly are deleted from further analysis.

Level of Detail

Only one representative of each item should be included within the first version of *InteractionSuitcase*. Therefore we examined which objects, the generic or the specific ones, were recognized faster (reaction time) or were assigned most frequently and unambiguously (number of assignments to positive, negative, "English", or "German").

Explicit Cultural Association

The visual stimuli were classified into priority groups based on the frequency of assignment to a cultural area. A total of ten cultural areas are identified in [44]. We are primarily interested in the European area (EU) (UK and Germany) and the Anglo-American (AngloAm) area.

Affective and Cultural Index and Descriptive Reaction Time

Both the affective ($aff.idx$) and the cultural index ($cult.idx$) have values between 1 and 2 and indicate which category the stimulus was assigned to more often. Therefore, a cultural index close to 2 would mean that the stimulus was more often assigned to the category "English" than to the category "German". Likewise, a value close to 1 indicates a more positive association for the affective index, while a value close to 2 suggests a negative one (compare Table 2).

Table 2. The table shows the affective index ($aff.idx$), the reaction time (rt in ms) in total, sorted by prioritization group (prio). Furthermore, it shows how many participants (part.) could make an assignment within the 500 ms.

prio	N (obj.)	aff.idx (sd)	aff. rt[ms] (sd)	N (part.)
1	41	1.095 (0.08)	412.44 (13.98)	40.63 (12.28)
2	21	1.108 (0.09)	412.92 (18.05)	39.95 (14.88)
3	41	1.104 (0.07)	404.10 (19.31)	37.71 (14.08)
4	25	1.086 (0.08)	417.19 (14.26)	39.80 (13.39)
Total	128	1.098 (0.08)	411.66 (16,40)	39.52 (13.66)

Reaction Time and Correlations

Moreover, we investigated whether there was a correlation between the fast and the explicit results. For this purpose, the cultural index of the assignment task was compared with the explicit assignment to a cultural area. In addition, we investigated the relationship between the affective and cultural indices.

Explicit Affective Association

Each participant named three associations for each stimulus in the assigned set (14 stimuli). Exploratory word clouds were created for each stimulus. These can be used for didactic purposes. The results of the explicit affective connotations are not presented in the results.

3.2. Results of Study 1

3.2.1. Recognition

The participants were explicitly asked if they could name the object. Objects that the participants did not recognize were discarded. A total of 13 objects were not correctly recognized and consequently excluded (8 specifics, 5 generics). Some objects (most notably from the context category) were not identified and therefore might not be well suited as objects for the suitcase. Here, another 19 objects were filtered, for example, "Berlin Wall" or "House of Commons". After this step, a total of $n = 261$ objects remain.

3.2.2. Level of Detail

Hence, the objects assigned more than 50 times, and thus by more than 25% of the participants, were rescanned. Objects that fewer participants could assign were excluded. The items whose counterparts were associated less frequently or more slowly were also rejected. Thus, other $n = 112$ specific objects were removed. There remain 22 specific objects, such as the Statue of Liberty, specific balls (soccer, basketball, etc.), specific cars (German ambulance, American police car), and different newspapers ("Die Zeit", "New York Times") or specific flags. Of the 144 objects identified, a further 16 were discarded because a similar representation was already present and elicited a higher affective response, i.e. some humanoid avatars or a watch. A total of 128 objects remain.

3.2.3. Explicit Cultural Association

The priority groups are defined as follows:

(1)　>67%: clearly assigned it to Eu_Germany, Eu_UK or AngloAm.

(2)　　>50%: assigned it to Eu_Germany, Eu_UK or AngloAm.

(3)　　>50%: assigned it to one or more other cultural areas.

(4)　　>33%: assigned it to no cultural area.

Based on this prioritization, 29 objects could be assigned to the AngloAm area, 31 to the Eu-Germany area, and 6 to the Eu-UK area. Thus, 44 objects were predominantly assigned to other cultures, and 18 were assigned to "no cultural area" by more than 40% (see Table 3).

Table 3. The table shows the cultural index, the reaction time, and the explicitly assigned cultural area (eca) inclusive prioritization (prio). Furthermore, it shows how many objects (obj.) are assigned to its cultural area and how many participants (part.) could make an assignment within the 500 ms. *n* = 18 visual stimuli were assigned to prioritization 4 and there associated with the none or neutral group. EU: European cultural area; EA: East Asian cultural area; LA: Latin-American cultural area; SSA: Subsahara-African cultural area; SA: South-Asian cultural area; O: Oriental cultural area; AUS: Australian cultural area.

N (obj.) *	eca (m)	prio	*cult.idx* (sd)	rt[ms] (sd)	N (part.)
20	AngloAM (87%)	1	1.58 (0.13)	411.45 (25.94)	18.83 (6.93)
9	AngloAM (57%)	2	1.53 (0.16)	397.82 (34.90)	
28	EU_Germany (71%)	1	1.47 (0.13)	415.22 (21.44)	17.90 (6.92)
3	EU_Germany (58%)	2	1.35 (0.15)	415.83 (12.90)	
5	EU_UK (61%)	1	1.72 (0.12)	413.55 (11.13)	21.67 (6.18)
1	EU_UK (54%)	2			
5	English ** (54%)	3	1.47 (0.11)	409.06 (22.22)	22.2 (5.56)
19	EU *** (70%)	3	1.52 (0.11)	406.25 (20.80)	19.25 (5.81)
2	EA (69%)	3	1.52 (0.02)	386.35 (55.45)	14.50 (1.50)
6	LA (54%)	3	1.49 (0.12)	404.23 (11.23)	15.83 (4.06)
3	SSA (55%)	3	1.42 (0.08)	405.73 (19.07)	23 (5.72)
3	SA (52%)	3	1.65 (0.13)	395.93 (41.81)	10.33 (3.30)
5	O (83%)	3	1.71 (0.15)	422.12 (12.39)	19.6 (6.28)
1	AUS (90%)	3	1,46 (-)	384.70 (-)	13 (-)
128	**In Total**	**-**	**1.51 (0.16)**	**409.25 (24.50)**	**18.38 (6.99)**

* according to eca, ** <50% associated with AngloAM and/or EU_UK, *** including <50% associated with EU_Germany.

3.2.4. Affective and Cultural Index and Reaction Time

Here we can confirm the results of Li et al. [12]. The presented objects rarely indicated negative affect. On average, participants in the affected group assigned objects to a category (positive, negative) within $m = 410.8$ ms$(sd = 16.9)$. None of the objects indicated a negative affect. The most negatively associated objects were, for example, a "mitra" ($aff.idx = 1.35$), a gun ($aff.idx = 1.36$), or even a dark-skinned avatar ($aff.idx = 1.29$). The most negative associated object was a soldier's helmet with an ($aff.idx = 1.38$).

The participants in the "culture group" assigned objects to a category (English, German) within $m = 409.2$ ms$(sd = 24.5)$ through the fast assignment task. Here the objects were distributed equally, i.e., $n = 61$ objects tended to indicate any association with "English", 58 with "German", and $n = 9$ were assigned to one category by exactly 50%. From the cultural index of all objects, it is also easy to see that the results are balanced around the middle (1.5) with a maximum of $cult.idx = 1.92$ (hijab) and a minimum of $cult.idx = 1.17$ (brain).

Reaction Time and Correlations

We investigated the correlation between affective and cultural indexes. The correlation, according to Pearson (r) and Spearman (ρ), shows a significant middle positive correlation between the affective and cultural index of the respondents (see Table 4, 1. row). However, since for the affective index 40 respondents and for the cultural index only 18, the condition of the bivariate normal distribution is not guaranteed for both variables. Therefore, a

correlation according to Pearson as well as Spearman was calculated. A higher cultural index represents an association with "English", while a high affective index represents a negative association. Thus, visual stimuli that were often associated as foreign ("English") by the mainly German sample were also often associated as negative.

Table 4. This table shows the strong significant positive correlation between the variables *cult.idx* and *aff.idx*. Furthermore, it shows the correlations between the explicit cultural association (AngloAM, EU_UK, and _Germany) and the cultural index. *cult.idx*: cultural index; *aff.idx*: affective index; eca: explicit cultural association.

			Pearson		Spearman	
			r	*p*	*ρ*	*p*
1. *cult.idx*	-	*aff.idx*	0.306 ***	<0.001	0.373 ***	<0.001
2. *cult.idx*	-	AngloAM (eca)	0.226 *	0.016	0.180 *	0.045
3. *cult.idx*	-	EU_UK (eca)	0.520 **	0.001	0.267	0.070
4. *cult.idx*	-	EU_Germany (eca)	−0.134	<0.904	−0.156	<0.936

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

We also investigated whether the explicit cultural association and fast assignment to a cultural space were correlated. Here, there was a significant, small correlation between the cultural index and assignment to the AngloAm area ($\rho = 0.18; p = 0.045; n \pm 18$). In parallel, there was no correlation between the cultural index and the explicit assignment to the European area, neither Germany (negative) nor the UK (positive). In the UK area, there was only a moderate tendency ($\rho = 0.267; p = 0.070; n \pm 18$).

In summary, the results from Study 1 show a correlation between the explicit cultural association and the fast assignment (*cult.idx*) to a cultural area (see Table 4). In addition, there was a correlation between visual stimuli that were predominantly positive and those that were classified by the German sample as predominantly German. However, the classification into the categories "German" and "English" was confusing. The subjects obviously had more difficulties classifying the visual stimuli into these categories than "positive" and "negative" ones.

3.3. Discussion of Study 1

The goal of the first study was to evaluate the virtual objects and gather initial information. Further, we want a systematic reduction of virtual objects ($n = 293$) to a manageable number. The result of Study 1 is an empirically based, reduced database of objects that are potentially supportive for inter- and transcultural exchange via social VR. Each virtual object was assigned both to a cultural area by a mainly german sample. The complete database, named *InteractionSuitcase* can be found in an Excel table in the supplementary material. An image of each included object is shown in Figure A3. Cultural and affective indexes describe each virtual object and the average reaction time needed to assign it to a cultural area and explicit information about recognition, cultural and affective association. However, the results of this database must be verified and extended for future studies and use cases. Schöne et al. [29] revealed that 3D/360° stimuli evoked a stronger sense of presence, altered motivational processing, and perceived valence of the stimuli than 2D content. Thus, the 2D images of our objects might elicit less arousal. However, we had to use 2D images due to economic reasons. To meet this limitation, videos of the 3D models will be presented in Study 2. Further, participants were mainly German, leading to the question of whether the results would be similar for English-speaking participants. Since the cultural associations are important for the intended use case of trans- and intercultural learning, Study 2 will investigate these associations more deeply. Moreover, the recognition task will be repeated to ensure the intuitive use of objects in a social VR application for a more heterogeneous sample.

4. Study 2: Implicit Cultural Associations of the Virtual Objects

Study 2 includes a more heterogeneous target group reflecting the desired target cultures, i.e., AngloAm, or European cultural area. This expanded target group repeated recognition and cultural attribution for all objects resulted from Study 1 ($n = 128$). However, while participants in Study 1 were explicitly asked to name the object name and cultural area, in Study 2, they were additionally asked to indicate how confident they were with the attribution and recognition. In addition to Study 1, an IAT is conducted with a selection of $n = 48$ objects of the priority groups 1 and 2 (compare Section 3.2, paragraph *explicit cultural association*). The IAT aims at the implicit re-investigation of the cultural associations. We used the "self"/"other" IAT in accordance with [37]. The binary concept of "self" and "other" emphasizes the experience of difference, multiperspectivity, and the understanding of cultural "otherness" [39] better than national categories. The main idea is that a person with German cultural heritage has stronger associations between stereotypic German visual stimuli and words referring to the category of "self". In contrast, non-stereotypical German stimuli should be stronger associated with words referring to the category of "other".

4.1. Method

4.1.1. Participants

Participants were recruited via the Prolific [45] and SurveyCircle [46] websites and via the university's internal volunteer system. They were credited with either EUR 3 or 0.25 participant hours. They were again asked about the same demographic information. A total of $n = 132$ participants took part in the second study ($n = 44$, male; $n = 3$, diverse). In total, $n = 64$ participants executed the questionnaire in English and $n = 68$ in German. All participants had a mean age of $m = 24.1 (sd = 6.4)$. Most of the respondents ($n = 61$) reported A-levels or higher education entrance qualification as their highest educational qualification. The second most common educational qualifications were Intermediate/General Certificate of Secondary Education ($n = 29$) or university degree ($n = 22$). The others reported still being in school ($n = 9$), having a secondary school-leaving certificate/Junior High Diploma ($n = 2$), or a completed apprenticeship ($n = 2$). The remaining seven reported having another degree. More than half of the participants reported being students ($n = 71$) or pupils ($n = 11$). The others are self-employed ($n = 7$), employed ($n = 30$), or unemployed ($n = 2$). Two respondents indicated something else (teacher and carer). Only participants with the longest period of residence in the USA, the UK, or GSA (German-speaking countries (Germany, Switzerland, or Austria)) countries were included in the survey. Hence, $n = 9$ of the participants were excluded. Furthermore, $n = 11$ participants did not complete the questionnaire. Thus, $n = 112$ respondents remain, of whom $n = 59$ completed the questionnaire in English.

A Mann–Whitney U-Test was calculated to determine if there were differences in age between German-, and English-speaking persons. There was a statistically significant difference in age between both groups ($W = 2787.00$, $p = 0.006$). English-speaking participants had a mean M = 26.41 years old (sd = 8.1), while German-speaking participants had a mean M = 21.96 (sd = 3.0) years old. There was also a difference in the level of education. While the English-speaking sample was well distributed, the majority of the German-speaking participants reported having completed A-levels/International Baccalaureate or higher education entrance qualification (79.4%). By gender, the groups English-speaking ($n = 25$, male; $n = 1$, diverse) and German-speaking ($n = 29$ male; $n = 2$, diverse) were balanced.

4.1.2. Materials and Apparatus

Due to the fact that the objects could now be purchased for the second study, this time, not only pictures but videos of the 3D models have been shown. For the experiment, all virtual models were integrated into *Unity* version 2020.1.10f1 [47], exported as a Windows standalone build. The objects were rotated around their central vertical axis in front of a white background at a speed of 60 degrees per second so that all sides of the object were

visible. In addition, a 25-second video of each object was created, which was presented via *SoSci Survey* [43]. All 128 objects can be seen in Figure A3.

4.1.3. Procedure

After the participants had selected the questionnaire, they filled out the same demographic questionnaire as in Study 1. They were then assigned randomly to one of eight stimuli sets. For example, one set contained 16 stimuli, to be answered with 10 questions each (see Table 5, and Figure 3). Finally, each participant was randomly assigned to one of three IAT sets. Via Sosci Survey, an IAT can be used with eight stimuli per category ("self" and "others"; see Table 6). Therefore, $n = 48$ visual stimuli with strong connotations to their cultural or stereotypical perception were selected (priority groups 1 and 2). The mean cultural index and explicit cultural association from Study 1 for each set can be seen in Table 7.

Table 5. This table shows the variables, instructions, their type, input, and specification.

Variable	Instruction	Type	Input	Specification
recognition	How well do you recognize this object?	metric	scale	0–100
name	What would you name this object?	text	open	
certainty	How sure are you about the naming of the object?	metric	scale	0–100
context category	Which category would you assign to this object?	nominal	selection	8 categories
context category certainty	How sure are you about the assignment?	metric	scale	0–100
familiarity	How familiar are you with this object?	metric	scale	0–100
level of detail	Rank the object on the scale according to its level of detail.	metric	scale	generic (0)–specific (100))
stereotypes	Rank the object on the scale according to the degree of stereotypical representation.	metric	scale	non-stereotypical (0)–stereotypical (100)
cultural area	To which part of culture would you most likely assign this object?	nominal	selection	10 cultural areas
cultural area certainty	How sure are you about the assignment?	metric	scale	0–100

Table 6. This table shows the targets words used for the binary concept "self and others" for both the German-speaking and the English-speaking target groups.

	English	German
self	me, myself, I, home, intimate, mine, we, known	mir, selbst, ich, Heimat, intim, meins, wir, bekannt
other	hers, his, them, they, others, people, yours, foreign	ihre, seine, die, sie, die Anderen, Menschen, deins, fremd

To obtain data for each stimulus, each set will be reviewed by an average of $n = 15$ participants (mixed group according to residence and language). The IAT sets will be conducted by at least $n = 30$ subjects each to ensure normal distribution.

Table 7. Descriptive statistics from Study 1 according to the IAT sets. N: number of participants for each set; cult.idx: cultural index; eca: explicit cultural association.

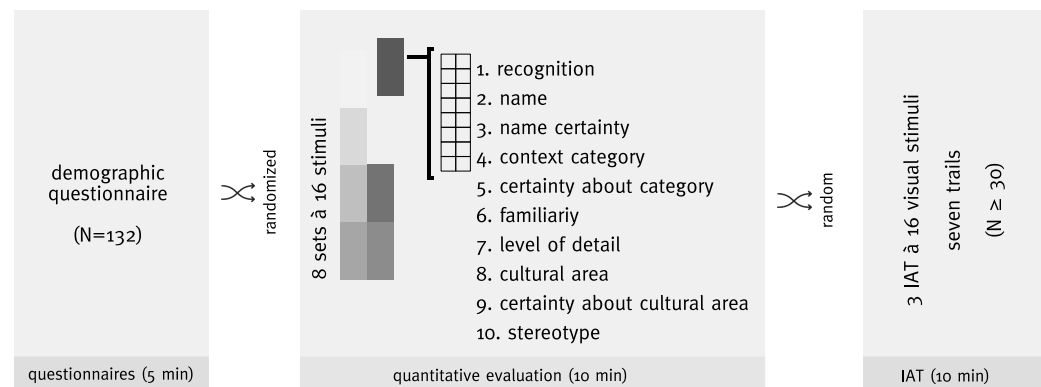|	IAT Set 1 ($n = 44$)		IAT Set 2 ($n = 34$)		IAT Set 3 ($n = 34$)	
	AngloAM	EU_ger	AngloAM	EU_ger	AngloAM	EU_ger
mean eca (Study 1)	98%	94%	83%	70%	67%	57%
mean cult.idx (Study 1)	1.6	1.4	1.6	1.5	1.6	1.5

Figure 3. This figure shows the study design and timeline of the second study.

4.1.4. Data Analysis

Quantitative Evaluation

To verify the results from Study 1, the explicit assessment of the recognition and cultural association was queried again. More variables could be asked without requiring too much cognitive effort from the participants by reducing the number of stimuli. To obtain comprehensive data for each stimulus in our database, familiarity and perceived level of detail were added to the variables in reference to Tromp et al. [26]. The following table shows the operationalization of the variables.

Implicit Association Test (IAT)

When performing the IAT on the computer, terms from different fields must be correctly classified by pressing a key. For example, the term "home" is to be classified as "self" or the term "abroad" is to be assigned to the category "others" (see Table 6). Due to the fact that, according to the stereotype, terms that do not belong together (such as "home" and a picture of the Statue of Liberty or "abroad" and "pretzel or beer") have to be operated by the same key (left (key E) or right (key I)), people with strong stereotypes need a relatively long time to make the correct classification. The IAT measures the reaction time needed to classify the different terms correctly and calculates for individuals from these reaction times a value for the strength of their unconscious stereotypes (D-score) [35]. An IAT consists of seven blocks. Blocks 3–4 and 6–7 are used to calculate the D-score. The others serve as test blocks. The D-score is usually between ±2. A value of ±0.15 is considered a slight association, a value of ±0.35 a moderate association, and a value of ±0.65 or more a strong association [48].

The resulting value is positive if the association between the categories "Self" and the stereotypical German objects is stronger (lower response times in blocks 3/4) than the association between "Others" and the stereotypical AngloAM objects (higher response times in blocks 6/7). Thus, the D-score is expected to be positive for German-speaking participants and negative for English speakers.

4.2. Results Study 2

4.2.1. Quantitative Evaluation

All objects were evaluated by an average of $n = 15$ persons and recognized by $m = 88.00\%(sd = 12.14)$. The names entered by the participants corresponded to the modal names with a few exceptions. Participants were very confident with the naming ($m = 83.80, sd = 18.92$). The participants classified each visual stimuli into one of eight predefined categories. According to frequency, $n = 7$ objects were assigned to the category persons, $n = 19$ to food and drinks, $n = 44$ to tools and miscellaneous, $n = 8$ to flags, $n = 2$ to landmarks, $n = 12$ to transportation, $n = 19$ to clothing, and $n = 17$ to animals. To provide comprehensive normative data for each stimulus in our database, familiarity ($m = 27.17, sd = 32.26$) and perceived level of detail ($m = 52.46, sd = 30.62$) were

also determined. The data can also be found for each stimulus in an excel table in the supplementary material part (Table 8).

Table 8. This table shows the mean for each metric variable and its standard deviation. It shows that there were no significant differences between the English- and German-speaking participants. The partially high values for standard deviation can be explained by the many different objects summarized here.

Variable	m(sd)		
	Total	**Eng**	**Ger**
recognition	88.0(12.14)	86.6(17.9)	89.6(12.6)
name certainty	83.8(18.9)	83.1(20.1)	85.0(16.7)
context category certainty	84.1(19.2)	82.8(20.7)	85.2(16.3)
familiarity	27.2(32.3)	32.2(33.4)	23.7(26.3)
level of detail	52.5(30.6)	51.4(33.0)	53.8(27.4)
stereotypes	59.7(35.5)	62.6(30.9)	58.4(29.8)
cultural area certainty	56.5(30.2)	63.0(30.1)	51.4(24.8)

4.2.2. Implicit Association Test (IAT)

In IAT set 1, for all objects that were sorted as stereotypical European-German, all objects could be sorted as such again ($n = 5$, prio 1; $n = 3$, prio 2). Likewise, for the stereotypical U.S.-American objects (IAT set1), the explicit classification from Study 1 was confirmed in Study 2 for all objects ($n = 6$, prio 1; $n = 2$, prio 2) (Table 9).

Table 9. Explicit assignment to the cultural areas for the object sets selected for the IAT.

IAT	obj. Set EU-ger			obj. Set AngloAM		
	Set 1	**Set 2**	**Set 3**	**Set 1**	**Set 2**	**Set 3**
no. obj.	8	8	8	8	8	8
mean(sd) *	71.4(19.4)	46.5(15.3)	52.7(16.3)	72.8(23.7)	75.9(14.8)	67.6(13.3)

* Mean of the percentage assignment of the given objects set to cultural area EU-ger or AngloAM by the participants of Study 2.

The assignment of the stereotypical European-German objects of IAT sets 2 and 3 could no further be confirmed unambiguously. In the following, we report the D-scores for all three IAT sets. Due to the low agreement, we will only interpret the results of the first set (Table 10).

Table 10. Descriptive statistics IAT. D-scores (D) according to the IAT set and the participants (English- or German-speaking).

	IAT Set 1		IAT Set 2		IAT Set 3	
	Eng ($n = 16$)	**Ger ($n = 28$)**	**Eng ($n = 18$)**	**Ger ($n = 16$)**	**Eng ($n = 19$)**	**Ger ($n = 15$)**
D (sd)	−0.174 (0.11)	0.391 (0.13)	−0.147 (0.13)	0.255 (0.31)	−0.042 (0.21)	0.334 (0.17)

As described above, a positive D-score is expected for German participants and a negative one for English speakers. A score of ± 0.15 is considered a slight association, a score of ± 0.35 a moderate association, and a score of ± 0.65 or more a strong association. In fact, German-speaking participants had a moderate bias, whereas English-speaking participants had a slight bias. Thus, both groups found it easier to associate the objects stereotyped for their cultural area with words of the category self and associate objects not stereotypical for their cultural area with the category "others".

4.2.3. Former, Qualitative Data

All $n = 112$ people also answered the multiple-choice question at the end of the questionnaire about whether these virtual objects could be useful for foreign language

teaching. The virtual objects were least attributed to having a creativity-enhancing effect (n = 33). More than half (n = 67) are convinced that these virtual objects can be used to bridge language barriers and that they can trigger critical thinking of stereotypical patterns (n = 70). A total of 78 participants indicated that the objects could be useful for vocabulary training. The assessment that virtual objects can be initiators and facilitators for conversations and inter- and transcultural encounters was held by the largest group (n = 80).

The qualitative results show that all participants in the second study recognize the potential of the *InteractionSuitcase* in terms of language education. In addition, the IAT results show that the virtual objects of the *InteractionSuitcase* can be promising parameters for manipulation in intercultural and transcultural encounters.

4.3. Discussion Study 2

Study 2 confirmed the results of Study 1 by a more heterogeneous sample and demonstrated that the virtual objects are recognized correctly. In addition, participants showed expected cultural associations. For each virtual object (n = 128), descriptive data are now available, which can be used in future studies. The results of the IAT showed that it was easier for the English-speaking AngloAm residents to associate the explicitly stereotyped AngloAm objects with the category "self". At the same time, the German-speaking participants did the same with the explicitly stereotyped German objects. All objects selected for the IAT were classified by participants in Study 2 as stereotypical of the EU and AngloAM cultural areas, too. There were only four exceptions. The Native American headdress was assigned to the culture area AngloAM by only 44%. Three other objects were also not assigned as in Study 1. However, the reason for this was that these were part of the eighth set and, in contrast to the others, were evaluated by only eight participants. These three objects can still be sorted out if necessary or have to be examined again in future studies. To give the database more weight and power, the results should at best be verified with an even larger and heterogeneous sample.

5. General Discussion

VR can significantly contribute to education by allowing a direct and self-determined experience of environments, a targeted replication of situations that are difficult to replicate using traditional teaching methods, or by allowing manipulation of embodiment or virtual objects and digital artifacts [2–4]. Visual stimuli are frequently used to improve learning. However, in the context of VR, there are few systematically and empirically derived databases. This paper proposes the first collection of virtual objects based on empirical evaluation for inter- and transcultural encounters between English- and German-speaking learners, named *InteractionSuitcase*. Two studies explored the stimuli implicitly and explicitly and analyzed the cultural and affective associations.

Study 1 examined the virtual objects explicitly and implicitly and gathered initial descriptive information. The results led up to a reduction of the initial collection of 293 objects to 128 objects. Study 2 confirmed the results of Study 1 by a more heterogeneous sample and demonstrated that the virtual objects are recognized correctly. In addition, participants showed the expected cultural associations in an explicit and implicit manner. Thus, for each virtual object (n = 128), descriptive data are now available, which can be used in future studies.

The qualitative feedback from the participants in Study 2 suggests that the rather young sample with an average age of 24 years is aware of the topic of cultural appropriation. Therefore no clear results came out, especially about the headdress. One respondent mentioned:

"I think the approach to learning in this way is excellent, but one should be careful not to stereotype countries and cultures, as I was asked to do in the sorting task. You could also focus on commonalities between cultures instead of classifying things as "foreign and familiar" and then defining what is right and

what is wrong. Perception is individual, what is foreign to me may be familiar to someone else."

This point is crucial, and we are aware of it. Therefore, it is indispensable that a reflection is performed before, after, or during the use of the objects. When using the *InteractionSuitcase*, it should not be the goal of the users to assign the objects to a country or a region. Rather, the *InteractionSuitcase* should serve as communication support and a basis for reflection. Further, scientists might use it as a behavioral measurement tool. "From a psychological point of view, it is not easy to make behavior quantitatively measurable" ([5], p. 95). Therefore, the *InteractionSuitcase* can be seen as an implicit, behavioral measurement tool. The use of the objects is intuitive and can be fully traced. In addition, learners using (non)stereotypical objects of the *InteractionSuitcase* can use them to negotiate the meaning of the concept of culture and identify shared views with exchange partners. Following the paradigm of action and task-based language learning, learners might use these objects as a basis for communicating with other learners in the VR environment [5]. Social desirability is highly likely to influence participants' responses, especially in the case of stereotypes and prejudices towards cultural representations. We used this advantage in these two studies and compared the explicit and implicit results. Therefore, the *InteractionSuitcase* is also valid in terms of social desirability.

6. Limitations and Future Work

It remains to be considered that the virtual objects were still presented and tested in an immersive context via VR. "*In VR, motivational tendencies and emotional reactions are related to objects or persons within the vicinity of the participant and not to the stimuli presented on a screen.*" ([29], p. 1). This and other considerations to be taken into account for further research with the *InteractionSuitcase* are presented below. Since the virtual stimuli have not yet been tested and evaluated in VR, a seminar with student teachers is planned in which the virtual stimuli will be exploratively examined. The prospective teachers, among others for the subject English at German high schools, will evaluate whether the objects are suitable for use in the teaching practice. Especially for this purpose, the *InteractionSuitcase* should become a freely accessible asset that educators and didactics can easily use via Social VR. Currently, we present only a database of virtual objects as visual stimuli. VR promotes learning as a situational process. The simulation of authentic, realistic situations is a strength of VR learning applications. The stimulative potential of the technology can be enhanced by multimodality [7]. So, in the future, we should think about using virtual objects as digital artifacts and as interactive, vibrant learning tools. When using the *InteractionSuitcase* via Social VR, whether as an educator or HCI scientist, the BehaveFIT [1] factors self- and other representation, but also situational context should always be considered. The *InteractionSuitcase* is a metaphor for the first proposal of a database of virtual objects that can be used either in an educational setting or in a scientific context under controlled conditions or for inter- and transcultural encounters via (social) VR. Scientists, teachers, and encounters can gain deeper conclusions about inter- and transcultural competencies and the choice of virtual objects in the suitcase. Who takes which objects? How often are they attributed with what? Is it possible to reduce culture to generic objects, or does this enable new ways of seeing and thinking? Such and other reflection tasks will be possible by the *InteractionSuitcase*. Not only students can reflect on stereotypical thinking, associations, and critical incidences, but also prospective teachers could expand their digital literacy and intercultural competence by using the suitcase. Not only students can reflect on stereotypical thinking, associations, and critical incidents, but also prospective teachers can expand their digital literacy and intercultural competence with the help of the suitcase. For instance, teachers can deliberately decide whether to confront learners with particularly stereotypical objects in order to provoke certain behaviors and then reflect on them. Depending on discretion, the *InteractionSuitcase* can also be extended with further objects. However, this should always be done systematically. Virtual objects can trigger or even encourage stereotypical thinking patterns. Therefore, a reflection, for example, in the context of a seminar concept, is mandatory.

7. Conclusions

Theoretically, digital artifacts or virtual objects can be used with all forms of modification and various interaction possibilities within VR simulations [1,7]. However, in the literature review, we and other researchers could not identify any analysis that specifically maps modes of impact on teaching and learning [7]. In addition, there is limited research that investigates virtual objects systematically concerning their cultural and affective associations. In modern English classrooms, teachers and learners need to acquire knowledge (recognize and define racism). They must develop a critical, sensitive awareness of racial thinking and act with respect and open-mindedness toward their own and others' perspectives. Cultural sensitivity and competence are not a one-time thing but require an ongoing commitment to regularly take time for ongoing dialogue to raise awareness and sensitivity. For this purpose, the virtual objects of the *InteractionSuitcase* were identified. With the help of this standardized visual stimuli database, the 128 virtual objects can be used in remote social VR encounters to trigger a critical examination of stereotypical ways of thinking, reflect, avoid critical incidents, and evaluate behavior during these encounters intuitively and implicitly. All 128 stimuli were rated by a heterogeneous sample of German- and English-speaking participants. Thus, the *InteractionSuitcase* can be used in mixed group studies as well as for inter- and transcultural encounters in English classes. Virtual objects enrich EFL instruction with an unprecedented learning experience. They make learning tangible and can initiate communication in ICT educational scenarios [5]. The research assumes that virtual objects can be useful for teaching and learning with all forms of alteration and various interaction possibilities in the context of VR interventions. A major advantage of VR is that emotional and cognitive processes can be studied under realistic conditions while maintaining stringent experimental control [16,29]. Hence, the *InteractionSuitcase* can serve as a quantifiable behavioral measure and a communication initiator in cross- and transcultural encounters in VR. Extensions of the *InteractionSuitcase* for other subjects or cultural domains (e.g., religion, ethics) are conceivable [5]. The *InteractionSuitcase*, in combination with social VR, will contribute to education, thereby enabling a vivid, tangible, and self-determined experience of inter- and transcultural encounters.
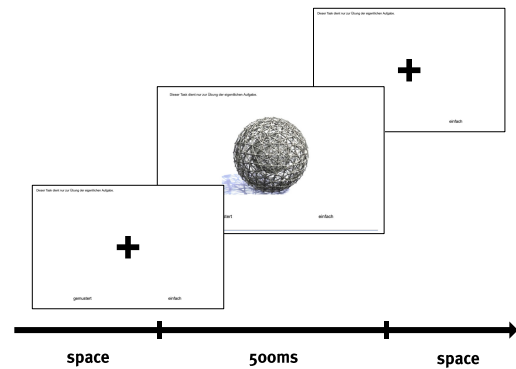
Appendix A

Appendix A.1

Figure A1. This figure shows the time sequence of an assignment task using the example of the exercise task.
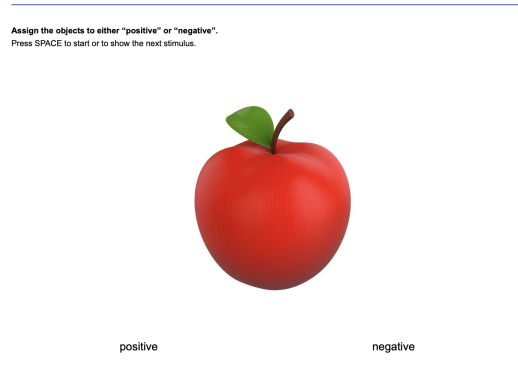

Figure A2. This figure shows an affective assignment task with positive and negative selection options using the visual stimulus "apple" as an example.

Figure A3. The figure shows the virtual objects identified and evaluated for the first proposal of the *InteractionSuitcase*.

References

1. Wienrich, C.; Döllinger, N.I.; Hein, R. Mind the Gap: A Framework (BehaveFIT) Guiding The Use of Immersive Technologies in Behavior Change Processes. *arXiv* **2020**, arXiv:2012.10912.
2. Hamilton, D.; McKechnie, J.; Edgerton, E.; Wilson, C. Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *J. Comput. Educ.* **2021**, *8*, 1–32. [CrossRef]
3. Hein, R.M.; Wienrich, C.; Latoschik, M.E. A systematic review of foreign language learning with immersive technologies (2001–2020). *AIMS Electron. Electr. Eng.* **2021**, *5*, 117–145. [CrossRef]
4. Wienrich, C.; Eisenmann, M.; Grafe, S.; Latoschik, M.E. CoTeach—Connected Teacher Education. *Vrinsight Greenpaper* **2020**, *124*, 53–55.
5. Hein, R.; Steinbock, J.; Eisenmann, M.; Latoschik, M.E.; Wienrich, C. Development of the InteractionSuitcase in virtual reality to support inter-and transcultural learning processes in English as Foreign Language education. *DELFI* **2021**, 91–96.
6. Yang, M.T.; Liao, W.C. Computer-assisted culture learning in an online augmented reality environment based on free-hand gesture interaction. *IEEE Trans. Learn. Technol.* **2014**, *7*, 107–117. [CrossRef]
7. Zender, R.; Weise, M.; von der Heyde, M.; Söbke, H. Lehren und Lernen mit VR und AR–Was wird erwartet? Was funktioniert. In Proceedings of the Daniel Schiffner (Hrsg.): Proceedings of DeLFI Workshops 2018 Co-Located with 16th e-Learning Conference of the German Computer Society, Frankfurt, Germany, 10 September 2018; Volume 16, pp. 1–12.
8. Latoschik, M.E.; Wienrich, C. Coherence and plausibility, not presence?! Pivotal conditions for XR experiences and effects, a novel model. *arXiv* **2021**, arXiv:2104.04846.
9. Vatter, C. *Interkulturelle Kompetenz und Fremdsprachenunterricht: Von der Konstruktion kultureller Differenz zur Verantwortung in der Migrationsgesellschaft?* Waxmann Verlag GmbH: Münster, Germany, 2020.
10. Steinbock, J.; Hein, R.; Eisenmann, M.; Latoschik, M.E.; Wienrich, C. Virtual Reality im modernen Englischunterricht und das Potenzial für Inter-und Transkulturelles Lernen. *MedienPädagogik Zeitschrift für Theor. Prax. Medien.* **2022**, *47*, 246–266. [CrossRef]
11. Blascovich, J.; Loomis, J.; Beall, A.C.; Swinth, K.R.; Hoyt, C.L.; Bailenson, J.N. Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* **2002**, *13*, 103–124. [CrossRef]
12. Li, B.J.; Bailenson, J.N.; Pines, A.; Greenleaf, W.J.; Williams, L.M. A public database of immersive VR videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures. *Front. Psychol.* **2017**, *8*, 2116. [CrossRef]
13. Schultheis, M.T.; Rizzo, A.A. The application of virtual reality technology in rehabilitation. *Rehabil. Psychol.* **2001**, *46*, 296. [CrossRef]
14. Fox, J.; Arena, D.; Bailenson, J.N. Virtual reality: A survival guide for the social scientist. *J. Media Psychol.* **2009**, *21*, 95–113. [CrossRef]
15. Ripka, G.; Grafe, S.; Latoschik, M.E. In Proceedings of the Preservice Teachers' Encounter with Social VR–Exploring Virtual Teaching and Learning Processes in Initial Teacher Education, online, 26 October 2020; pp. 549–562.
16. Wienrich, C.; Latoschik, M.E. eXtended Artificial Intelligence: New Prospects of Human-AI Interaction Research. *arXiv* **2021**, arXiv:2103.15004.
17. He, J.; Ren, J.; Zhu, G.; Cai, S.; Chen, G. Mobile-based AR application helps to promote EFL children's vocabulary study. In Proceedings of the 2014 IEEE 14th International Conference on Advanced Learning Technologies, Athens, Greece, 7–10 July 2014; pp. 431–433.
18. Liu, E.; Liu, C.; Yang, Y.; Guo, S.; Cai, S. Design and implementation of an augmented reality application with an English Learning Lesson. In Proceedings of the 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), Wollongong, Australia, 4–7 December 2018; pp. 494–499.
19. Fernández, S.S.; Pozzo, M.I. Intercultural competence in synchronous communication between native and non-native speakers of Spanish. *Lang. Learn. High. Educ.* **2017**, *7*, 109–135. [CrossRef]
20. Zhang, X.; Zhou, M. Interventions to promote learners' intercultural competence: A meta-analysis. *Int. J. Intercult. Relations* **2019**, *71*, 31–47. [CrossRef]
21. Chen, G.M.; Starosta, W. A review of the concept of intercultural awareness. *Hum. Commun.* **1998**, *2*, 27–54.
22. Lang, P.; Bradley, M.M. The International Affective Picture System (IAPS) in the study of emotion and attention. *Handb. Emot. Elicitation Assess.* **2007**, *29*, 70–73.
23. Brodeur, M.B.; Dionne-Dostie, E.; Montreuil, T.; Lepage, M. The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE* **2010**, *5*, e10773. [CrossRef]
24. Snodgrass, J.G.; Vanderwart, M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* **1980**, *6*, 174. [CrossRef]
25. Moreno-Martínez, F.J.; Montoro, P.R. An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS ONE* **2012**, *7*, e37527.
26. Tromp, J.; Klotzsche, F.; Krohn, S.; Akbal, M.; Pohl, L.; Quinque, E.M.; Belger, J.; Villringer, A.; Gaebler, M. OpenVirtualObjects: An Open Set of Standardized and Validated 3D Household Objects for Virtual Reality-Based Research, Assessment, and Therapy. *Front. Virtual Real.* **2020**, *1*, 37. [CrossRef]
27. Peeters, D. A standardized set of 3-D objects for virtual reality research and applications. *Behav. Res. Methods* **2018**, *50*, 1047–1054. [CrossRef] [PubMed]

28. Popic, D.; Pacozzi, S.G.; Martarelli, C.S. Database of virtual objects to be used in psychological research. *PLoS ONE* **2020**, *15*, e0238041. [CrossRef]

29. Schöne, B.; Kisker, J.; Sylvester, R.S.; Radtke, E.L.; Gruber, T. Library for universal virtual reality experiments (luVRe): A standardized immersive 3D/360° picture and video database for VR based research. *Curr. Psychol.* **2021**, 1–19. [CrossRef]

30. Singh, A.; Sha, J.; Narayan, K.S.; Achim, T.; Abbeel, P. Bigbird: A large-scale 3d database of object instances. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 509–516.

31. Kisker, J.; Gruber, T.; Schöne, B. Experiences in virtual reality entail different processes of retrieval as opposed to conventional laboratory settings: A study on human memory. *Curr. Psychol.* **2019**, *40*, 3190–3197. [CrossRef]

32. Pan, X.; Hamilton, A.F.d.C. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *Br. J. Psychol.* **2018**, *109*, 395–417. [CrossRef]

33. Ajzen, I.; Cote, N.G. Attitudes and the prediction of behavior. *Attitudes Attitude Chang.* **2008**, *13*, 289–305.

34. Fazio, R.H.; Olson, M.A. Implicit measures in social cognition research: Their meaning and use. *Annu. Rev. Psychol.* **2003**, *54*, 297–327. [CrossRef]

35. Greenwald, A.G.; McGhee, D.E.; Schwartz, J.L. Measuring individual differences in implicit cognition: The implicit association test. *J. Personal. Soc. Psychol.* **1998**, *74*, 1464. [CrossRef]

36. Project Implicit. 2011. Available online: https://implicit.harvard.edu/implicit/takeatest.html (accessed on 8 September 2021).

37. Karpinski, A. Measuring self-esteem using the Implicit Association Test: The role of the other. *Personal. Soc. Psychol. Bull.* **2004**, *30*, 22–34. [CrossRef]

38. Bennett, M.J.; Bennett, M.J. Intercultural Sensitivity. In *Principles of Training and Development*; Portland State University: Portland, OR, USA, 1993; Volume 25, pp. 185–206.

39. Eisenmann, M. Crossovers–Postcolonial Literature and Transcultural Learning. In *Learning with Literature in the EFL Classroom*; 2015; pp. 217–236.

40. Souza, C.; Garrido, M.V.; Carmo, J.C. A Systematic Review of Normative Studies Using Images of Common Objects. *Front. Psychol.* **2020**, *11*, 3326. [CrossRef] [PubMed]

41. De Cesarei, A.; Loftus, G.R.; Mastria, S.; Codispoti, M. Understanding natural scenes: Contributions of image statistics. *Neurosci. Biobehav. Rev.* **2017**, *74*, 44–57. [CrossRef] [PubMed]

42. Gall, D.; Latoschik, M.E. Visual angle modulates affective responses to audiovisual stimuli. *Comput. Hum. Behav.* **2020**, *109*, 106346. [CrossRef]

43. GmbH, S.S. SoSci Survey—Die Lösung für eine Professionelle Onlinebefragung. Available online: https://www.soscisurvey.de (accessed on 23 May 2022).

44. Reinke, C.; Bickel J. Infoblatt Kulturerdteile, Definition und Aufzählung der Kulturerdteile. Available online: https://www.klett.de/alias/1004296 (accessed on 21 June 2021).

45. Prolific. Quickly Find Research Participants You Can Trust. Available online: https://www.prolific.co (accessed on 23 May 2022).

46. SurveyCircle. Forschungswebseite SurveyCircle. 2016. Available online: https://www.surveycircle.com/de/ (accessed on 23 May 2022).

47. Unity. The World's Leading Platform for Real-Time Content Creation. Available online: https://unity.com (accessed on 23 May 2022).

48. Greenwald, A.G.; Nosek, B.A.; Banaji, M.R. Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Personal. Soc. Psychol.* **2003**, *85*, 197. [CrossRef]