THE EFFECTS OF ROBOTIC EMBODIMENT ON INTERGROUP BIAS: AN EXPERIMENT IN IMMERSIVE VIRTUAL REALITY

Katharina Jahn
University of Siegen, katharina.jahn@uni-siegen.de

Bastian Kordyaka
University of Siegen, bastian.kordyaka@uni-siegen.de

Follow this and additional works at: https://aisel.aisnet.org/ecis2019_rip

Recommended Citation

This material is brought to you by the ECIS 2019 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in Research-in-Progress Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
THE EFFECTS OF ROBOTIC EMBODIMENT ON INTERGROUP BIAS: AN EXPERIMENT IN IMMERSIVE VIRTUAL REALITY

Research in Progress

Katharina Jahn, University of Siegen, Siegen, Germany, katharina.jahn@uni-siegen.de
Bastian Kordyaka, University of Siegen, Siegen, Germany, bastian.kordyaka@uni-siegen.de

Abstract

The dissemination of immersive virtual reality (IVR) with body tracking offers new opportunities for communication across distant places by embodying virtual avatars or actual robots. Thus, when people meet via IVR, different forms of embodiment for humans and artificial intelligence become possible, which may influence the perception of the self, the encoding of information and the evaluation of communication partners. In this research-in-progress paper, we draw upon research on self-presence and social identity theory to investigate how robotic embodiment influences intergroup bias in the context of a user who sees a group discussion of individuals embodied in avatars with different degrees of visual and mental humanness. We hypothesise that less humanness leads to higher misattribution of debaters’ contributions and more negative evaluations of these debaters. Additionally, we assume that this effect is diminished when the users who watch the group discussion are embodied in an avatar with low humanness.

Keywords: Embodiment, Robots, Artificial Intelligence, Intergroup Bias, Laboratory Experiment, Self-presence.
1 Introduction

As immersive virtual reality (IVR) becomes increasingly affordable, new opportunities for computer mediated communication (CMC) arise. Immersion, as a technological characteristic, describes the degree to which a technology is inclusive, extensive, surrounding, vivid, matching, and provides a coherent plot (Slater and Wilbur, 1997). Through these means, the technology might strengthen the users’ illusion of being in a different place (tele-presence) (Slater and Wilbur, 1997; Witmer and Singer, 1998). With the use of full-body tracking, the matching aspect of immersion can be increased by recording individuals’ movements and transferring them to a virtual body. Recent research has shown that users can even adopt a robotic body in a distant place or a virtual body in IVR when their body is tracked (Kishore et al., 2016). Thus, when a camera is placed on the head of a robot, individuals can give a speech in front of an audience or talk with other people over a distance in the robotic body. Likewise, when the movements of all communication partners are recorded and head-mounted displays are used, individuals can meet in IVR, for example to instruct someone on how to enact movements.

When individuals communicate over IVR with body tracking, different combinations of interacting with artificial or actual bodies can occur. For a meeting with individuals from different places without the use of robots, communication partners may all have virtual avatars in IVR. When robots are used, communication partners may all be embodied as a robot (e.g. when robots are controlled over a distance in a place to which humans cannot go) or only some are embodied as robots while the others use their actual bodies (e.g. when a colleague working overseas takes part in a team meeting).

These application scenarios have two implications for CMC. First, from the perspective of the message receivers, they are confronted with the message senders who are embodied in a virtual avatar or robotic body - an experience that can vary fundamentally from real-world experiences as visual appearance can be changed to almost any degree in IVR. Importantly, this experience differs from usual CMC with Desktop computers in the higher immersion that is possible with IVR. Second, from the perspective of the senders, they are embodied in a body that can be different from their actual body to a varying degree, which is also an experience that can hardly be recreated without the use of IVR, except for dreaming (Metzinger, 2013).

On the one hand, individuals who are embodied in an avatar or robot with body tracking can experience so-called full-body ownership illusions. Research on body ownership has shown that the body influences the self and behaviour of individuals (Kilteni, Maselli, et al., 2015). On the other hand, individuals who interact with individuals who are embodied in an avatar or robot, may categorise the individuals according to their artificial representations. Research has already shown that group membership influences perception of robots, for example with regard to anthropomorphism (Eyssel and Kuchenbrandt, 2012). Robots who possess a human-like appearance are perceived as having mental states (Martini, Gonzalez, and Wiese, 2016). Additionally, mental schemas of robots that resemble a human in contrast to a box are associated with higher blood pressure and increased negative emotion (Broadbent et al., 2011) and robots resembling outgroup members are evaluated more negatively than robots representing ingroup members (Eyssel and Kuchenbrandt, 2012) and are more likely to be killed in a shooting game (Bartneck et al., 2018).

However, although these initial findings indicate that robotic bodies can be perceived differently than human bodies and intergroup bias for robots seems to exist, previous research mostly investigated reactions to pictures (Bartneck et al., 2018; Eyssel and Kuchenbrandt, 2012; Martini, Gonzalez, and Wiese, 2016), leaving it unclear how effects would be in real communication situations. This is especially the case when artificial intelligence and humans communicate in artificial bodies, as previous attempts did either manipulate permanent visual aspects of the robots’ design (Bartneck et al., 2018; Martini, Gonzalez, and Wiese, 2016) or information about the robot (Eyssel and Kuchenbrandt, 2012), but not both aspects in one experiment. Furthermore, research investigating the effects of embodying real humans in a robotic body (Kishore et al., 2016) did not investigate whether robot-related intergroup bias could change according to embodiment. Thus, designers of IVR communication systems cannot decide on a theoretical basis which design decisions are the most favourable in a given communication contexts, especially when multiple
individuals are present at the same time. We therefore want to address the following research question:

**Research Question.** How does robotic and artificial embodiment of senders and receivers influence intergroup bias?

To address our research question, we will conduct a laboratory experiment in IVR using a scenario in which the user watches a discussion group in IVR. For this, we vary both the embodiment of the user and the embodiment of the individuals in the discussion group on the dimensions of visual humanness (human vs. robot) and mental humanness (human vs. artificially intelligent). We assess how accurately the users can attribute statements from the discussion to the correct individuals and how the individuals are evaluated regarding affective and cognitive attributions.

The paper is structured as follows. First, we give an overview on the theoretical background regarding body ownership illusions and social identity theory. In the next section, we explain our research model and hypotheses. Subsequently, we describe the experimental design. To conclude, we discuss the implications our research can have for theory and design.

## 2 Theoretical Background

### 2.1 Self-Presence

Self-presence is defined as “the illusion that one’s virtual representation (e.g. avatar) is indeed oneself, that is, inhabiting the virtual body” (Schultze, 2010, p. 438). This construct is closely related to body ownership illusions, which are not constrained on virtual contexts. One typical way of experimentally inducing body ownership illusions is the rubber hand illusion. To trigger the illusory perception that a rubber hand is an individual’s own hand, the rubber hand is placed in front of the individual in the same position as their real hand (which is invisible to them). The experimenter subsequently strokes both the rubber hand and the real hand at the same time. After some time, participants start to perceive that the rubber hand is part of their own body (Botvinick and Cohen, 1998; Kilteni, Maselli, et al., 2015).

With the use of head-mounted displays and body tracking, individuals can embody a virtual avatar, which enables them to experience so-called full-body ownership illusions (Kilteni, Maselli, et al., 2015) that elicit a high sense of self-presence (Schultze, 2010) to the user and subsequently change a range of attitudinal and behavioural variables, even when visual appearance is not photorealistic (Jo et al., 2017). For example, light-skinned individuals who are embodied in a dark-skinned avatar drum in a larger radius (Kilteni, Bergstrom, and Slater, 2013) and show reduced intergroup bias (Hasler, Spanlang, and Slater, 2017). On the other hand, individuals who are embodied in an old avatar can show more favourable attitudes towards the elderly and feel closer to them (Oh et al., 2016). Thus, in the context of being embodied in a virtual body in a communication context, and subsequently experiencing high self-presence, it is likely that the design of the body changes how users perceive themselves and others. In the next section, we draw upon the social identity theory and the common ingroup identity model to understand this process.

### 2.2 Social Identity and Re-categorisation

According to the social identity theory (Tajfel and Turner, 1986), individuals strive to establish, improve, and maintain a positive self-image. To achieve this, they categorise themselves to groups they belong to (ingroup, e.g. human beings) and compare the value of this group to other relevant groups they do not belong to (outgroup, e.g. animals). As a consequence, to achieve a positive distinctiveness between the ingroup and the outgroup, individuals may engage in discriminatory cognition or behaviour towards outgroups and favour the own group over other groups (intergroup bias). However, the categorisations of individuals are not stable and rely on context-specific factors (Ellemers, Kortekaas, and Ouwerkerk, 1999). Therefore, the self-concept of an individual can be understood as a fluid entity.
The common ingroup identity model states that re-categorisation in a common, superordinate group can decrease intergroup bias for cognitive, affective, and behavioural outcomes. This model could be supported both in laboratory (Dovidio et al., 1997; Gaertner, Mann, et al., 1989) and field studies (Nier et al., 2001). An example of how re-categorisation can occur is displayed in Figure 1. If individuals encounter individuals of an outgroup (e.g. a light-skinned individual meets a virtual avatar with dark skin in IVR), they can perceive similarities based on physical appearance between themselves and the out-group member (e.g. the skin of the avatar the user embodies is as dark as the skin of the outgroup member) which helps re-categorising them both in a superordinate group, and can then result in more positive cognitive, affective, and behavioural outcomes and thus, less intergroup discrimination and bias.

In the context of communicating with virtual avatars, robotic bodies and artificially intelligence can both represent outgroups for human beings. On the basis of this assumption, we develop our specific hypothesis for the effects of robotic embodiment in communication contexts.

3 Research Model

Our research model is displayed in Figure 2. Hypotheses 1-4 are concerned with the effect of debaters’ embodiment in a group discussion on misattributions and evaluations of the individual who sees the group discussion. Hypotheses 5-8 are concerned with how the embodiment of the individual who sees the group discussion interacts with this misattribution and evaluation. Previous research has shown that individuals tend to misattribute contributions of outgroup members in a discussion more frequently than contributions of ingroup members (Sesko and Biernat, 2010; Vescio, Judd, and Kwan, 2004). As for human individuals, both humans or AI who are embodied in human bodies or AI in generel represent outgroup members, we hypothesise that the same effect will occur for these two groups.

Hypothesis 1. Users will misattribute the contribution of debaters with robotic appearance more frequently than contributions of debaters with human appearance.

Hypothesis 2. Users will misattribute contributions of debaters who possess artificial intelligence more frequently than contributions of debaters with human intelligence.
Besides increased errors in attribution of contribution, we assume that ingroup bias will also show in evaluation of robotic and AI individuals. This is in line with prior research that has shown that individuals tend to ascribe mental states to robotic agents only when they are designed with high humanness (Martini, Gonzalez, and Wiese, 2016).

**Hypothesis 3.** Users will evaluate individuals with robotic visual appearance more negatively than individuals with human visual appearance.

**Hypothesis 4.** Users will evaluate individuals who possess artificial intelligence more negatively than individuals who possess human intelligence.

When individuals are embodied in an outgroup avatar, the perception of self-other overlap can change towards the outgroup members (Oh et al., 2016). Additionally, they show increased outgroup mimicry behaviour (Hasler, Spanlang, and Slater, 2017). This research indicates that a process of re-categorisation takes place, in which members of the outgroup are perceived as more similar to the self than members of the ingroup, as predicted by the common ingroup identity model (Gaertner, Dovidio, et al., 1993). As it is unlikely to assume that individuals will completely change group membership based on their virtual avatar (e.g. from human to robotic), we assume that this re-categorisation process takes place by re-categorisation into a common superordinate group that is more inclusive than the original ingroup (e.g. from the human group towards a group that includes human and robotic individuals). Thus, we hypothesise that misattributions effect will decrease when individuals are embodied in a AI or robotic avatar.

**Hypothesis 5.** The effect of misattributing contributions of robotic debaters more frequently than human debaters will be diminished when users are embodied in a body with robotic visual appearance.

**Hypothesis 6.** The effect of misattributing contributions of artificially intelligent debaters more frequently than debaters with human intelligence will be diminished when users are embodied in an avatar with cues for categorising them as artificially intelligent.
Research on body ownership illusions in IVR has shown that embodying the body of an elderly avatar can lead to more favourable attitudes towards the elderly (Oh et al., 2016). Similarly, implicit bias towards dark-skinned people can be reduced when individuals embody an avatar of a dark-skinned person (Peck et al., 2013), and could last even one week after the last experience (Banakou, Hanumanthu, and Slater, 2016). Thus, in line with H5 and H6, we assume that this categorisation effect is also present in the context of robots.

**Hypothesis 7.** The negative effect of debaters’ visual humanness on evaluation is diminished when users are embodied in an avatar with robotic visual appearance.

**Hypothesis 8.** The negative effect of debaters’ mental humanness on evaluation is diminished when users are embodied in an avatar with cues for categorising them as artificially intelligent.

4 Method

4.1 Participants and Design

We will use a fully crossed 2 (users’ visual humanness: human vs. robot) x 2 (users mental humanness: human vs. AI) x 2 (debater’s visual humanness: human vs. robot) x 2 (debaters’ mental humanness: human vs. AI) mixed laboratory experiment to test our research model. Visual and mental humanness of the users’ avatar is a between-subjects factor whereas visual and mental humanness of the debaters’ avatars is a within-subjects factor. We plan to recruit 130 subjects.

4.2 Materials

*Immersive virtual reality:* We will use a HTC Vive head-mounted display with 5 Vive trackers (one for hip, two for feet and hands) and two controllers for interaction with the environment. Participants will be placed in a virtual room with a mirror at one side and a table at the front of the room, enabling participants to watch the panel discussion from the middle of the room.

![Figure 3. Design of visual humanness condition. A robotic discussion group member is displayed on the left and a human discussion group member is displayed on the right](image-url)
**Users’ Avatar, Visual Humanness:** We use Adobe Fuse CC (Beta) to create all avatars. The visual appearance of the users’ avatar is modelled in both conditions with a human body. However, in the robot condition, the body has a metallic, silvery skin, whereas the body in the human condition has a light-skinned skin tone. We decided to use a human-like appearance out of two reasons. First, previous research indicated that intergroup bias is lowest for human-looking robots compared to box-like avatars. Thus, our results will indicate a minimal difference for intergroup bias for robots. Second, from a methodological point of view, choosing a human-like appearance minimises visual differences between the conditions. As a result, differences between the conditions can more likely be attributed to the robotic nature of the avatar than to other possible confounders (e.g., other forms of facial expressions when robot has box-like appearance). The gender of the avatar is mapped to the actual gender of the participant to avoid reducing the sense of presence for female participants, which was indicated in previous research (Schwind et al., 2017).

**Users’ Avatar, Mental Humanness:** To implement the mental humanness condition for the users’ avatar, the avatar either wears a shirt that is labelled with “AI” (artificial intelligence condition) or “Human” (human condition). Apart from the shirt label, the design of the users’ avatar in both conditions is identical. We chose a shirt label because previous research showed that arbitrary, impermanent visual cues, such as cues in clothing, are sufficient to elicit categorisation processes (Kurzban, Tooby, and Cosmides, 2001). Thus, this operationalisation allows us to investigate whether these categorisation processes carry over to the user who is embodied in an avatar with this label.

**Group discussion Avatar, Visual Humanness:** The group membership of the group discussion avatars is created in congruence with the group membership of the users’ avatar. Whereas all avatars are modelled after human individuals and have different facial features, avatars in the robotic group possess a metallic, silver skin and avatars in the human condition are light-skinned.

**Group discussion Avatar, Mental Humanness:** The group membership of the group discussion avatars is created in line with the users’ avatar. Thus, avatars in the AI condition wear a shirt which is labelled with “AI” and avatars in the human condition wear a shirt that is labelled with “Human”.

**Group discussion:** To implement the group discussion, we draw upon the “Who said what?” paradigm (Taylor et al., 1978) and use the overall procedure from Sesko and Biernat (2010) but adapt it to IVR. We use a set of 8 virtual avatars with different facial appearance created in Mixamo Fuse (4 robotic, 4 human; 2 of each AI, 2 of each human intelligence) who take part in a group discussion. We develop 16 statements, with each avatar contributing 2 statements to the discussion. The order of the statements is equal for all participants, but the statements that are mapped to avatars vary randomly for each participant.

**Evaluation measurements:** We use measurements adapted to our context from Eyssel and Kuchenbrandt, 2012 to measure warmth and mind attribution with a Likert scale ranging from 1 (not at all) to 7 (very). An overview on the items is given in Table 1.

**Misattribution measurement:** The number of misattributions is calculated as means with 16 possible errors per participant, differentiating between four types of errors: a) within mental humanness/within visual humanness errors, b) between mental humanness/within visual humanness errors, c) within mental humanness/within visual humanness errors, d) between mental humanness/between visual humanness errors. For example, when a participant makes an error within mental humanness and within visual humanness, it could be that they have misattributed the contribution of an artificial intelligence with robotic appearance to the other artificial intelligence with robotic appearance. On the other hand, an individual that makes an error between mental humanness and between visual humanness could misattribute the contribution of a robot with robotic appearance to another human with human appearance. As for within/within errors, misattribution can only occur for one target, whereas for the other errors, two targets exist, within/within errors are multiplied by two (see, e.g., Sesko and Biernat, 2010; Taylor et al., 1978).
Table 1. Measurement items adapted from Eyssel and Kuchenbrandt (2012)

<table>
<thead>
<tr>
<th>Construct</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warmth</td>
<td>Please rate how you perceive the person regarding the following traits.</td>
</tr>
<tr>
<td></td>
<td>- helpful</td>
</tr>
<tr>
<td></td>
<td>- sensitive</td>
</tr>
<tr>
<td></td>
<td>- polite</td>
</tr>
<tr>
<td></td>
<td>- generous</td>
</tr>
<tr>
<td></td>
<td>- humble</td>
</tr>
<tr>
<td>Mind attribution</td>
<td>To what extent is this person capable of feeling hungry/joy/pain/fear?</td>
</tr>
<tr>
<td></td>
<td>To what extent is this person capable of hoping for things?</td>
</tr>
<tr>
<td></td>
<td>How likely is it that this person has a personality?</td>
</tr>
<tr>
<td></td>
<td>To what extent is the person capable of being aware of things?</td>
</tr>
<tr>
<td></td>
<td>How likely is it that this person has a soul?</td>
</tr>
</tbody>
</table>

4.3 Procedure

When participants enter the laboratory, they first answer questions for sociodemographic variables. To ensure that participants understand the meaning of the label that indicates mental humanness, we tell them that our research is the development of AI that can take part in group discussions. For this purpose, we would like them to watch the group discussion of AI and Human participants which use different avatars, but can be distinguished by their label on their shirt.

After the participants put on the trackers and the head-mounted display, they see a room with a mirror and are asked by the experimenter to look around and describe what they see. Next, they are asked to wait in the middle of the room to watch the group discussion. The eight virtual avatars enter the room and go to the table where the panel discussion takes place. Each of the eight avatars then present two statements regarding the topic of why they have lost as a team in an esports game. The statements are presented in random order and are randomised across debaters for every participant.

After the discussion is over, participants are consecutively presented with 32 statements (of which only half were actually present in the discussion) for which they have to indicate whether the statement was present in the discussion and if yes, which avatar had contributed the statement. Finally, they complete measures for warmth and mind attribution for each of the eight debaters and are thanked and debriefed.

5 Discussion

With the proposed experiment, we expect that we can draw conclusions on the effects that different forms of embodiment in IVR may have. On this basis, technology designers are able to estimate how design decisions regarding robotic or human AI design may influence the users’ perception in social interaction. This will be especially relevant when humans are embodied in a robotic body in another place, when people meet virtually or when they need to interact with artificial intelligence. On this basis, future research can investigate related questions, such as the effects of a box-like robotic appearance compared to human-like appearance or the effects of different levels of avatar realism. Furthermore, future research can investigate how intergroup bias is influenced when individuals do not merely watch a group discussion, but also interact within it. We are currently at the stage of programming the group discussion task, animating the characters and synchronising the lip movements to the spoken statements. When this part is completed, we will conduct the experiment and analyse the data.
References


