Laughing Together: The Role of Virtual Agents in Emotional Contagion, Conformity, and Opinion Shaping in a Virtual Stand-Up Comedy Club

Similde Mair*
University of Hamburg

Fariba Mostajeran†
University of Hamburg

Joan Llobera‡
Artanim Foundation

Mel Slater§
University of Barcelona

Frank Steinicke¶
University of Hamburg

Figure 1: Virtual Environment, seen from the participant's point of view.

Abstract

Interaction between humans and virtual agents is becoming increasingly prevalent. This research investigates the influence of virtual agents on users' opinions within a virtual stand-up comedy club setting, focusing on the socio-psychological mechanisms of emotional contagion and group conformity. We aimed to combine these two areas in a novel study approach. To this end, we designed and conducted an experiment in which participants rated the funniness of jokes in VR, under varying conditions of virtual agent laughter (absence, medium intensity, strong intensity). Our analysis shows a significant positive linear effect on the log odds of higher funniness ratings when virtual agents exhibit laughter. Moreover, our study shows that this effect is mainly due to emotional contagion processes, individual differences in susceptibility to group conformity, and the interaction of both concepts, showing that emotional contagion happens only when the individual susceptibility to group conformity is high. The results offer valuable insights into the social impact of VR environments and the potential of virtual agents to shape user perception and opinion.

Index Terms: Virtual agents, emotional contagion, conformity, opinion-shaping, virtual reality.

1 Introduction

Interaction between humans and Virtual Agents (VAs) is becoming increasingly important as immersive technologies, such as Virtual Reality (VR), open up new possibilities for social experiences. VAs are widely used in applications ranging from healthcare and psychotherapy to education and entertainment. For instance, they assist in healthcare settings [28], facilitate VR-based therapies for psychiatric disorders [11], and serve as lecturers in digital learning environments [30, 38]. Additionally, VAs play an important role in gaming [20] or act as virtual coaches [35].

As the usage of VR technology increases, a particularly important question is the extent to which a VA is able to influence a user in those environments. In the settings mentioned above, VAs have the potential to influence not only users' emotions but also their opinions, particularly when they engage in social or emotionally charged interactions.

To explore this influence, examining the socio-psychological processes of emotional contagion and group conformity is essential, as both can shape user behaviour in virtual environments. Emotional contagion and group conformity are well-researched phenomena since 1950s. A well-known example of emotional contagion is "canned laughter"—the use of pre-recorded laughter in comedy shows to encourage the audience to laugh. Studies indicate that such external stimuli can shape the perception and evaluation of humour [8]. Group conformity, on the other hand, refers to how individuals adjust their behaviour and opinions to align with the group to gain social acceptance [1]. For example, a person might laugh at a joke in a group setting to avoid standing out, even if they do not find it particularly funny. These two research areas are closely linked, as emotional expressions within groups not only act as triggers for emotional contagion but also contribute to forming social norms to which group members conform. [49].

Previous studies have tested either emotional contagion (e.g., [37]) or group conformity (e.g., [29]) with a VA, but not together. In a novel approach, our study combines these two areas. It examines how VAs influence participants through emotional contagion and group conformity in virtual environments, specifically within a

*e-mail: similde.mair@studium.uni-hamburg.de
†e-mail: fariba.mostajeran.gourtani@uni-hamburg.de
‡e-mail: joan.llobera@artanim.ch
§e-mail: melslater@ub.edu
¶e-mail: frank.steinicke@uni-hamburg.de

virtual stand-up comedy club. To this end, we conducted an experiment in which participants rated the funniness of jokes in a VR environment, influenced by the presence or absence of VAs' laughter. The goal of the study is to assess whether VAs' laughter affects these ratings, providing insights into the extent to which social influence mechanisms operate in virtual spaces when no other humans are present.

This research integrates insights from the fields of Human-Computer Interaction (HCI), particularly Human-Agent Interaction, and social psychology. It contributes to ongoing studies on the social impact of VR environments and VAs, highlighting the importance of emotional contagion and group conformity in the design and application of VAs.

2 Related Work

This section builds on related work investigating the influence of humans in real life and virtual humans in a digital context and extended realities (XRs). The focus here is on the possibilities of influence through emotional contagion and group conformity.

2.1 Emotional Contagion

Before exploring the topic of digital emotional contagion in more depth, it is first necessary to explain the general mechanisms involved.

2.1.1 Definition

Emotional contagion is the transfer of mood and emotions between people in a group. The mechanisms behind this can be divided into three parts: Mimicry, category activation, and social appraisal [16]. Mimicry refers to behaviour in which people mimic the other person's facial expressions and other characteristics [3]. The concept of category activation refers to the process by which the brain perceives an emotional expression by the body, thus activating the corresponding emotion [36]. Social appraisal is the process by which individuals use the emotions of others as a guide for their own emotional appraisals, resulting in similar emotional experiences [31]. Due to these three mechanisms, the contagion of emotions can occur through many different types of contact with other people's emotions, including through direct face-to-face interactions and digital contagion without meeting in person.

Emotional contagion was confirmed in several studies. For instance, Barsade [3] showed that the mood of a confederate in group discussions transferred to the members. She concluded that emotional contagion changes people's moods and makes them continuously influence the moods and then the judgments and behaviors of others. Also, Schachter and Singer [40] proposed that emotional experience is a result of two components: physiological arousal and cognitive labeling of that arousal based on the context. In their experiment, they found that people interpreted their emotional state not only from their physical state but also by taking into account the environment around them.

2.1.2 Digital Emotional Contagion

A famous phenomenon of digital emotional contagion is the so-called *canned laughter*. Canned laughter is pre-recorded laughter used in television programmes, especially sitcoms or comedy shows. This artificial laughter gives viewers the impression of humour and results in an increased rating of mirth, as shown by Chapman [8].

Digital emotional contagion is also possible through other modalities than audio. An experiment by Kramer et al. [27] showed that emotions can be transferred even through text posts on Facebook alone.

Emotional contagion has also been studied in the context of virtual humans, as interactions with VAs trigger similar reactions to those between humans and create affective responses [39]. Studies show that animations [52], appearance [50], and verbal/non-verbal behaviours [51] promote emotional contagion in such interactions. It was shown by Qu et al. [37] that a virtual woman showing positive or negative facial expressions influenced participants' emotions. Tsai et al. [47] have shown that emotional contagion between a person and a VA is possible only through a picture of the agent, where the results showed increased happiness when a smile is added to an image of a virtual character.

In the studies mentioned, either emotional contagion and the associated change in opinion were explored, or emotional contagion in XR, without the consequences that this has on the participant's opinion. This paper aims to combine the two areas and measure the impact of emotional contagion on the rating in a VR setting through a human-agent contagion.

2.2 Group Conformity

In the context of the influence of VAs on humans, group conformity is an important topic for discussion, as studies have shown that group conformative behaviour towards VAs is possible [29]. In this section, we will first discuss in detail the fundamental mechanisms of group conformity, before exploring group conformity within the paradigm of digital environments and XR.

2.2.1 Definition

Group conformity describes an individual behavioural phenomenon in which one's own behaviour is adapted to the behaviour of a group, i.e. one's own behaviour is aligned with the norms of others and personal ideas move into the background [43]. In the context of group conformity, we can distinguish between two processes: Facilitation and Inhibition. Both processes describe the way in which group norms influence the behaviour of individuals. Facilitation describes the process of group members being more inclined to express their opinions or emotions. In contrast, inhibition describes the tendency to suppress certain emotional expressions or behaviours.

One of the best-known studies on this topic is Asch's group conformity experiment [1]. They investigated how group pressure influences judgment behaviour. In a group of apparent participants, who were in fact confederates, one participant had to solve a perceptual task (e.g. comparing the length of lines). The confederates deliberately gave incorrect answers to see if the participants would contradict them or join the group. The results showed that many participants followed the majority despite clear evidence for their perceptions, illustrating the strong influence of social conformity.

Asch's research stimulated a wealth of follow-up studies. Gutiérrez et al. [17] have shown that group norms influence the perception of humour: Jokes are rated as funnier when the group norm is positive towards humour. This illustrates that group conformity changes humour evaluations. Johnson [24] has shown that participants rate an advert as funnier if confederates in the room also rate the advert as funnier.

2.2.2 Influential Factors for Group Conformity

The following factors have been identified in a group that influence the conformity of the members:

- **Group Size**: Asch also tested the influence of the group size on the measured influence - they concluded that a group size of three people was sufficient to achieve the full effect [1]. However, more recent theories attribute greater importance to different group sizes [5]. In sum, the influence of group pressure increases strongly up to three people. For bigger groups, the group size still is a relevant factor but the impact is weaker.

- **Salience of social group**: The more a person identifies with the group (in-group), the greater the influence on their behaviour [48].

- **Unanimity of group responses**: Conformity is strongest when the group has an unanimous opinion. Even a single dissenter can weaken the influence of the group [48].

- **Public or private response**: If the response is publicly visible, conformity increases as the person seeks social acceptance and is afraid of rejection [48].

2.2.3 Conformity in Digital World and XR

Several studies have already investigated whether virtual humans are also capable of producing similar social conformity effects; Asch's line experiment in particular has often been replicated [26, 29, 19]. Kraemer [26] conducted a VR study where participants informed that the avatars in the VR experience were other participating students, while they were in reality controlled by confederates. The results demonstrated that the participants were influenced by a group of three avatars. Kyrlitsias [29] replicated Asch's experiment, but with only VAs. The results show no influence, therefore no group confirming behaviour. In a following experiment, Kyrlitsias tested the influence dependent on how realistic the agents behave [29]. Realistic behaviour was implemented here by different animations, e.g., by blinking and directing the gaze toward the speaker. The results of the experiment showed that contrary to the first experiment, conformity can be caused by VAs in virtual environments. This aligns with the threshold model of social influence by Blascovich [4], which posits that an agent cannot elicit social reactions from humans unless its behaviour seems highly realistic. In addition, a study by Hertz and Wiese indicates that conformity is significantly influenced by task ambiguity, with higher ambiguity leading to greater influence [19].

Most previous studies on digital conformity have focused on tasks with objectively correct answers, such as perceptual judgment tasks based on Asch's line experiment. Mostajeran et al. [33] examined how the size of a group of virtual humans impacts the induction of social anxiety. They measured the physiological reactions of the participants along with other objective and subjective measures. Their results indicated that the group size of three virtual humans induced the highest psycho-physiological anxiety compared to the group sizes of six and fifteen virtual humans. This study explores group conformity in a completely subjective context: humour perception. This shift in focus allows us to examine whether group norms can shape personal opinions when no objectively correct response or psycho-physiological stressors exist.

3 User Study

We conducted a user study to investigate the influence of VAs on participants' behaviour and opinions, and whether this influence occurs due to emotional contagion, social conformity, or a combination of both. To investigate this we designed a VR environment where a virtual comedian performed a series of jokes, accompanied by VAs displaying different intensities of laughter. Participants were then asked to rate the funniness of the joke. The aim of the study was to assess whether the agents' laughter influenced these ratings.

3.1 Methods and Materials

The study's experimental design comprised a 3 (laughter intensity: no laughter, medium laughter, strong laughter) $\times$ 6 (group size: 0-5 agents at the table) within-subject design. With this structure, we made sure that the participant saw each group size of agents thrice with each intensity of laughter. The exception is the condition without agents, where the three laughter intensities could not be represented; this condition was presented only once. Thus, the total number of trials per participant was 16 ($3 \times 5 + 1$).

Participants were exposed to all conditions in a randomised order, determined by the permutation of the 16 conditions. This setup required the inclusion of 16 jokes, which were presented in the same fixed order for all participants, pairing each joke with a unique condition. To mitigate potential gender biases, eight jokes were performed by a male virtual comedian and eight by a female virtual comedian. The order of the comedians was counterbalanced based on participant IDs, ensuring that in 50% of trials, the male comedian performed first, while in the other 50%, the female comedian performed first.

3.1.1 Virtual Environment

We designed a virtual environment using the Unity game engine. We used the *Music Bar* asset from the Unity Asset Store to create the background. This environment consists of a stage and a large area in front of it. Two VAs, who alternated the role of the comedian, were placed standing on the stage. We placed multiple round tables in front of the stage to facilitate the participants and the VAs. The participant sat at one table, together with 0-5 VAs (Figure 1).

In front of the participant and on the virtual table, there was a virtual sheet of paper on which was written in German "Rate the humour of the joke on a scale of 1-5" and five virtual buttons from 1 to 5, which could be used to input the rating. To increase the realism of a stand-up comedy club, we made the virtual club full of virtual audience. To do so, we placed a total of 6 other tables and 15 other virtual humans in the room (Figure 2). Depending on the number of agents at the table, the number of agents in the audience also changed: At the smallest group size, there were 10 agents in the audience, with the number increasing by 1 in parallel with the group size at the table, up to a maximum of 15 agents in the audience. The agents in the audience had different animations, ranging from blinking and looking around the room to laughing and shaking their shoulders when a joke was over.

3.1.2 Virtual Agents

The models for the agents came from different providers, namely *Adobe Fuse* [21] in combination with *Adobe Mixamo*, or *Character Creator*, the models for the comedians were received from *Didimo* [12]. The animations were created by the experimenters or came from *Adobe Mixamo*. As declared in 2.2.3, to generate conformative behaviour, the attending agents need to seem realistic and show

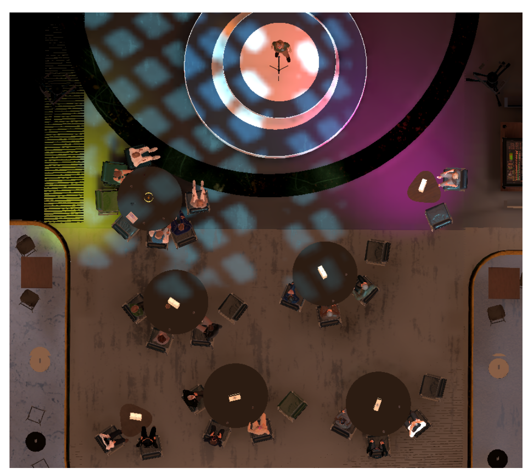

Figure 2: Virtual Environment, seen from a bird's eye view. The participant sits at the table in the upper left, in the free space where the sheet of paper is in front of them.

realistic behaviour. For this purpose, we developed several animations that we assigned to each agent. We developed three levels of laughing behavior, each matched with a corresponding sound. In the non-laughter condition and while listening to the joke, the agents sat quietly, looking around or at the participant while blinking. In the medium laughter condition, the agents shook their shoulders slightly, making a giggling noise and opening their mouths to smile slightly. In the strong laughter condition, a laughter animation is playing, which consists of strong shoulder movements, leaning forwards and backwards, and/or patting the thighs, with a sound corresponding to an open laugh. In addition, each agent plays their own laughter sound, in a male or female voice depending on their gender, which was provided by *Pixabay* and *artlist*. The laughter sound was positioned to come from the each VA, therefore creating the illusion that the VAs are creating the sound

For this study, we applied a maximum group size of five agents sitting at the same virtual table as the participant. With five agents the table was then fully occupied. The position of the VAs was consistent across the different conditions. VAs sitting furthest away from the participant were presented the most often (condition from 1 VA to all VA). This was done because they were the most visible to the participant, thus ensuring that the participant was aware of a single VA. To control for influences on group conformity, as seen in Section 2.2.2, we took the following considerations into account for the study: To avoid in-group bias, care was taken when selecting the agents to ensure that gender was always represented 50/50. If an odd number of agents were present, the last gender was selected at random. To avoid unanimity of group responses (see Section 2.2.2) and mixing experimental conditions, all the agents present at the table where the participant sat always laughed with the same intensity.

In order to ensure that the effects measured were specifically due to the agents' laughter and not just the presence of an ambient laughter track, which would in turn trigger canned laughter effects (see 2.1.2), an audio track of laughter was consistently played in the background, independently of the current condition. The sound was positioned with different sources around the tables in the background. This gave the impression that the laughter was from the existing virtual audience at the other tables in the club. As a result, any observed differences in ratings could be attributed solely to the laughter generated directly by the VAs at the virtual table where the participant sat.

3.1.3 Setup

We used the Meta Quest Pro Headset for the experiment. Controllers were unnecessary because the participants could use their hands to input in the VR environment, i.e. press the buttons on the virtual sheet of paper. During the trial, the Head-Mounted Display (HMD) tracked the lip movement (left and right corner of the mouth) of the participant at a rate of 30 samples per second using the *Face Tracking for Movement SDK* for Unity and the *OVR Face Expressions* script. In addition, we tracked the noises of the participants via the integrated microphone.

3.1.4 Jokes

To reduce the influence of the individual jokes on the rating, we needed to make sure that they were equally funny. To this end, we initially looked for papers in which jokes had already been evaluated using scientific methods. However, we could not use these jokes, as they were always English papers and therefore English jokes, which could not be translated into German. ChatGPT could also not be used for the evaluation and/or generation of the jokes, as various papers show that this does not correspond to human evaluations if it has not been trained accordingly beforehand [15, 22]. The *LaughLab* project offered a solution. In this large scientific project, a vast number of jokes (40.000) were rated for funniness by a large number of people (N=350.000) [7]. In the corresponding book, the authors showed that a certain type of jokes, so-called shallow jokes (in German *"Flachwitze"*), were very robustly always rated as equally funny. With this background, we collected a set of 16 shallow jokes. Using shallow jokes had the additional advantage, that they are not very complex and are therefore easier to understand.

For the audio generation of the jokes, the pro-version of the AI audio platform *elvenlabs* was used. For the male voice, jokes were first recorded by a male acquaintance and then transformed using the voice changer tool from the platform. For the female voice, the text-to-speech generator was used. The reason for this was that there was a realistic-sounding AI female voice for the German language, while the male voices sounded too synthesized when only text-to-speech was used.

3.2 Measures and Hypotheses

The participants completed an online questionnaire consisting of the German version of the following questionnaires:
Demographic Data: age, gender, vision corrections, and impairments, as well as their previous experience with VR which was administered before starting with the experiment.
Simulator Sickness Questionnaire (SSQ): measures the subjective feeling of simulator- and cybersickness. There are three subscales, Nausea, Oculomotor, and Disorientation, and 16 symptoms asked on a scale from 0 to 3. It was administered two times, once before starting with the VR experiment and once after that[25].

The rest of the questionnaires were administered once after the experiment:
Igroup Presence Questionnaire (IPQ): has 14 items for measuring the sense of presence experienced in a virtual environment, divided into 4 subgroups: Spatial Presence, Involvement, Experienced Realism, and Sense of Being There. The rating uses a 7-point Likert scale (0–6) with varying anchors: some range from *Fully disagree* to *Fully agree*, while others use specific terms like *Did not feel present* to *Felt present* [41].
Conformity Scale: assesses tendencies toward conformity, reliance on others' opinions, and willingness to go along with others versus independence, with 11 items on a 9-point Likert scale (-4 = *Very Strong Disagreement*, +4 = *Very Strong agreement*) [32].
Agreeableness Subscale (BFI-AGR): is a subscale of the Big Five Inventory (BFI) and assesses traits such as compassion, cooperation, and trust in social interactions, using a 1–5 Likert scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*) [23].
Personal perception of the experiment: collects feedback about the experiment with open-ended questions to understand whether the participant was able to guess the purpose of the study, was a joke not understood/already known, what differences were noticed between the conditions, whether they thought that the mood of the other virtual agents at the table influenced their own mood, and was the effect of emotional contagion known.

Additionally, after each condition, the participants were asked to rate the funniness of the joke on a 5-point Likert scale (1 = *Not Funny at All*, 5 = *Very Funny)*. To make sure that the participants did not have to leave the virtual environment, the rating was implemented in VR. In addition, the participants' lip movements were recorded to measure whether and when the participants laughed during the experiment with a scale between 0 (no lip activity) and 100 (high lip activity, i.e. smiling). These values were taken using the Meta Movement SDK face tracking for Unity, whereby the extent to which the corners of the participants' mouths moved upwards was tracked (high value = corners of mouth high up). An audio recording was made via the microphone to record sounds like giggles and laughter with the same scale from 0 (no sound) to 100 (loud sound). These data are used to measure whether the agents' laughter can influence the participant's rating. In order to test this assumption, a number of hypotheses were put forward.

Based on emotional contagion and facilitation processes, the first hypothesis is: *H1a: The perceived funniness of the joke is positively influenced by the agents' laughter.* Based on inhibition processes, the following hypothesis is put forward: *H1b: The perceived funniness is inhibited when there is no laughter from the agents.* The second hypothesis is supported by the work on group conformity about group size, where it is assumed that the strength of conformity behaviour increases with the size of the group [5]: *H2: The agents' influence increases with an increase in their group size.*

As mentioned in Section 2, the line between emotional contagion and group conformity is blurred. For example, it is not fully clear if the participants rate a joke funnier, whether it might be because other agents in the group are laughing, or because they have been emotionally "affected" or because they are conforming to the group. Or is it an interaction of both? To explore the mechanisms underlying this influence, we stated some additional hypotheses.

We assumed that the observed effects are partly due to processes of emotional contagion, especially category activation (see Section 2.1.1), in combination with facial feedback theory, as it has been shown that facial feedback, in this case, a smiling mouth, influences participants' reactions to humour [44]. We, therefore, assumed that the influence on the rating took place through a combined reaction of emotional contagion and facial feedback theory. To challenge this hypothesis, we measured and recorded the participants' smiles and laughter: *H3: The participants measured laughter mediates the effect on rating.*

Furthermore, we hypothesized that the observed effects are partly due to individual differences in susceptibility to group conformity. To examine this hypothesis we used Mehrabian's Conformity Scale [32] and the sub-part of the Big 5 Inventory Agreeableness [23], in which high scores correlate strongly with normative conformity [18]: *H4: Participants with higher scores on Social Conformity and Agreeableness are more strongly influenced by the agents' laughter.*

3.3 Procedure

The experiment took place in the [anonymized]. Participants first filled out the consent form for the experiment and started the online questionnaire with initial demographic and SSQ (Baseline) questionnaires. After that, the experimenter introduced the participant to the task, which was to visit the virtual Stand-Up Comedy Club and rate the jokes told by the comedian on the stage from 1 (not funny at all) to 5 (very funny). The real purpose of the experiment (measuring group conformity and emotional contagion) was not revealed at this moment so as not to influence the participants. Then the participants put on the Meta Quest Pro HMD and started with the experiment. In the virtual environment, a moderator performed the introduction for the participants (only audio), before the first comedian came to stage. After rating 16 different jokes, the virtual part of the experiment was finished and the participants took off the HMD. Thereafter, the participant filled out the post questionnaires (SSQ-post, IPQ, Conformity, BFI, concluding questions). After all the questionnaires were answered by the participants, the experiment was concluded, lasting a total of about 20 minutes.

3.4 Participants

As the experiment was a pilot study conducted at the [anonymized], no ethics application was required. Twenty participants (11 female, 8 male, 1 non-binary) with an average age of 25.1 ($SD = 3.38$) participated in the study. The participants came from different backgrounds, including students from the Department of Computer Science and students from the Department of Psychology. To avoid language barriers during the trial, we ensured, at the recruiting stage, that all the participants possessed adequate German listening comprehension. All of them had normal or corrected to normal vision and none of them had reported other vision disorders. Four of them had no previous experience with VR before the experiment, six rarely, five occasionally, three often, and two very often.

3.5 Results

3.5.1 Data Analysis

First, we transformed the lip tracking data to minimise any noise effects. For this purpose, we calculated a moving average with a window size of 5 samples. The moving average was calculated for each sample, shifting the window one sample at a time, creating an overlap between the windows. This means that we calculated a new average for each sample, with each window overlapping with the previous and subsequent ones. For the boundary values, we filled the missing values with the original data after the moving average calculation.

Although we intended to use the recorded sound for our analysis, we finally excluded it. The main reason was that some participants spoke during the recording, resulting in non-usable audio.

We identified multiple outliers for the mean lip value with Grubbs' test and by visualizing the data on a box plot. Several values were identified for the mean lip value. The possible reason for these outliers could lie in the natural variability of the measured characteristics, as the reaction to humorous stimuli and the decision to laugh or not to laugh strongly depend on individual factors. Therefore, no measures were taken to remove outliers. No outliers were found for any other variables.

The normality of the data was checked using the Shapiro-Wilk test. The results showed that some variables were normally distributed (e.g., Big-5), while others were not (e.g., participants' measured smiles).

To check the mean lip values for log-normality, the log-normal distribution was applied (with a small shift of .1 to avoid zero values). The Kolmogorov-Smirnov test for log-normality showed $p < .001$, indicating that the data did not fit well with a log-normal distribution. Consequently, no data transformations were performed.

As suggested by Taylor et al. [45], a Cumulative Link Mixed Model (CLMM) using the Laplace approximation was used to analyse the influence of laughter intensity on the rating. We decided to choose this model because it is well suited for ordinal scaled dependent variables, as the rating was interpreted in this study, and takes into account random effects that may arise from the variation between the participants and the specific jokes. A Likelihood Ratio Test (LRT) was carried out to check whether adding independent variables as predictors significantly improves the quality of the model, always in comparison to a null model without any independent variables.

To ensure the validity of the model, assumption tests were conducted, including a test for the proportional odds assumption as suggested by Brant [6] and a Variance Inflation Factor (VIF) analysis to check for multicollinearity [14].

In order to recognise possible random effects, we carried out a CLMM analysis with the jokes as independent variables. The results showed that three out of 16 jokes were rated significantly funnier. Therefore, we included jokes in the analysis as a random factor in order to exclude the effects that the joke has on the rating.

Additionally, a Repeated Measures Analysis of Variances (RM ANOVA) was carried out. Two subsets were created for this purpose. The first subset had the number of agents from 1 - 5, meaning that agents were always present. The second subset had every possible number of agents, i.e. 0 - 5, but only the laughter intensity 0, i.e. no laughter on the part of the agents. The two subsets were used to analyse the two different directions of group conformity influence (facilitation and inhibition). As an RM ANOVA was

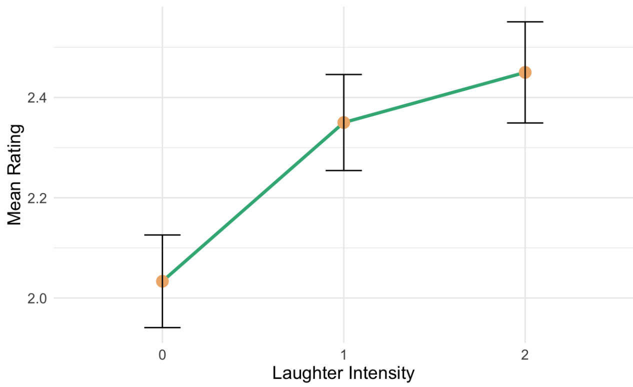

Figure 3: Mean rating for funniness across different laughter intensities, with error bars representing standard errors.

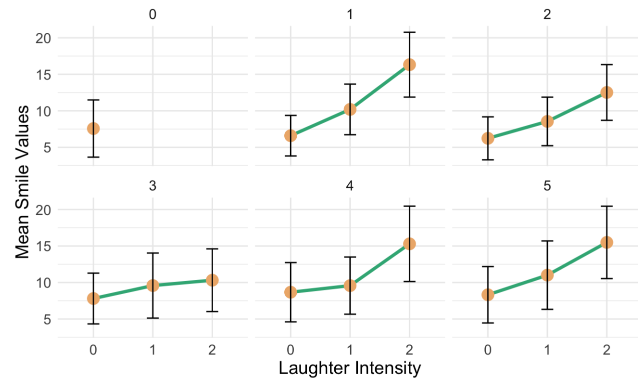

Figure 4: Mean smile across different laughter intensities, with error bars representing standard errors. The plot is faceted by VAs' group size and shows how the relationship between laughter intensity and smile activity varies.

performed in the analysis, Mauchly's test for sphericity was first performed to check the assumption of sphericity for the data. As effect size, the partial eta-square (η_p^2) is reported, whereby a value of .01 is considered a small effect, .06 a medium effect, and .14 a large effect [9].

In addition, we performed a series of mediation and moderation analyses using CLMMs. The mediation model was estimated using linear mixed-effects models, as suggested by Baron & Kenny [2]. One participant had to be removed from the meditation analysis, as the face-tracking was not working in this trial.

For the SSQ analysis, we used a paired t-test to compare the pre- and post-measurements. We evaluated the results of the IPQ using the method developed by Tran et al. [46]. They analysed 243 studies that used the IPQ and provided a reference value against which researchers can compare and interpret their study results. Using these reference values, we were able to compare our mean results with the percentile thresholds and categorise them into ranking classes: Low [-3.00, 0.28], Moderate [0.28, 0.73], High [0.73, 1.07], Very high [1.07, 1.30], Exceptional [1.30, 3.00] (Table 4, [46]. For this purpose, our results were transformed to [-3, 3].

3.5.2 Influence on Participants Opinion

The proportional odds assumption test was not significant ($chi^2 = 11.08, df = 21, p = .96$), indicating that the assumption was met and the coefficients are therefore constant among all conditions. The VIF analysis indicated no problematic multicollinearity among the predictors laughter intensity and agents quantity ($GVIF = 1.106, df = 2, GVIF^{1/(2Df)} = 1.025$ and $GVIF = 1.106, df = 5, GVIF^{1/(2Df)} = 1.010$). Therefore, we could proceed with the analysis.

The results of the LRT showed that the model including laughter intensity fitted significantly better ($LR = 15.286, df = 2, p < .001$).

The results of the CLMM showed that laughter intensity had a significant effect on the probability for higher rating ($b = .651$, $z = 3.529$, $p < .001$), indicating that as laughter intensity increases, the rating tends to increase. Figure 3 shows the differences in mean rating depending on the laughter intensity. We conducted the same analysis for inhibition effects using a subset with no laughter from the VAs, testing if the rating is lower, when a bigger group size is not laughing, in comparison when being alone. The LRT showed that adding group size to the model did not significantly improve the model fit ($LR = 8.60, df = 5, p = .126$).

Turning to the RM ANOVA, Mauchly's test for sphericity for the first subset was not violated ($W = .87, p = .299$), indicating that the assumption of sphericity holds. The results of the RM ANOVA for this subset showed, however, that the effect of laughter intensity is not significant ($F(2,38) = 2.40, p = .105, \eta_p^2 = .112$), suggesting that laughter intensity does not significantly impact rating in this analysis.

For inhibition effects, again a subset was created. Mauchly's test for sphericity was not violated for this subset ($W = .32, p = .146$), indicating that the assumption of sphericity holds. The results for this subset show no significant inhibition effects ($F(5,95) = 1.48, p = .202, \eta_p^2 = .072$).

3.5.3 Group size

The LRT showed that the model with the interaction between group size and laughter intensity did not explain the data significantly better than the model without group size ($LR = 11.835, df = 13, p = .541$).

The results of the ANOVA also show that the interaction between group size and laughter intensity is not significant ($F(8,152) = .93, p = .494, \eta_p^2 = .047$). Mauchly's test for sphericity shows no violations for all effects ($p > .05$)

3.5.4 Mediators for the effect

The influence of laughter intensity with lip movement (i.e. smiling) as a mediator on ratings was analysed using linear models to conduct a mediation analysis (see Figure 4).

In one model, we analysed the influence of laughter intensity on lip activity. Laughter intensity showed a significant positive linear effect on lip activity ($b = 4.805$, $SE = 1.065$, $t = 4.511$, $p < .001$). The second model included both laughter intensity and lip activity as predictors of ratings. Results show that lip activity had a significant positive effect on ratings ($b = .022$, $SE = .004$, $t = 5.657$, $p < .001$), whereas the effect of laughter intensity became non-significant when controlling for lip movement ($b = .1$, $SE = .088$, $t = 1.136$, $p = .257$).

The mediation analysis revealed a significant mediation effect of laughter intensity on the rating via lip activity. Table 1 shows the results.

3.5.5 Moderators for the effect

We tested the hypothesis that individual differences in the tendency to conform or agree moderate the influence of laughter intensity on the evaluation of the joke using two separate CLMMs.

For the analysis of the moderation by conformity, the results of the LRT showed that the model with the interaction between laugh-

Measure	Estimate	95% CI	p-value
ACME	**.1089**	.0567 – .17	**< .001**
ADE	.0992	-.0592 – .28	.292
Total Effect	**.2081**	.0360 – .39	**.016**
Prop. Mediated	**.5184**	.2046 – 2.23	**.016**

Table 1: Causal Mediation Analysis Results. Significant p-values are shown in bold.

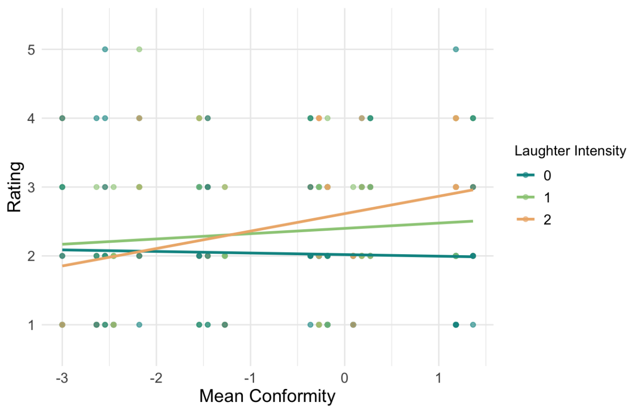

Figure 5: Mean conformity in interaction with different laughter intensities. It can be seen that laughter intensity 2 only influences the rating if the participant's conformity values are high. Intensity 1 has a small influence combined with a high conformity value.

ter and conformity explained the data significantly better than the model without conformity ($LR = 11.582, df = 3, p = .009$).

The results of the CLMM showed that the interaction between laughter intensity and conformity was significant ($b = .411$, $SE = .139$, $z = 2.952$, $p = .003$).

This emphasises that virtual groups are especially influential when users are already inclined to conform. This is illustrated in Figure 5: Especially in the case of strong laughter, it can be seen that rating is positively influenced when conformity is high, while there is no influence when conformity values are low.

Regarding agreeableness, the LRT showed that the extended model, including the interaction between laughter intensity and agreeableness, did not fit significantly better ($LR = 3.909$, $df = 35$, $p = .27$).

3.5.6 Exploratory Analysis

As both hypothesized influential factors for opinion shaping (emotional contagion and group conformity) have shown significant results, we investigated with an exploratory analysis the interaction of both. To this end, we conducted a linear mixed model analysis.

The results of the LRT showed that the model including the interaction between laughter intensity and conformity explained the data significantly better than the model without this interaction ($LR = 23.01, df = 3, p < .001$).

The results of the LMM showed that the interaction between laughter intensity and conformity was significant ($b = 3.569$, $SE = .77$, $t = 4.637$, $p < .001$), indicating that the effect of laughter intensity on lip movement was stronger for individuals with higher conformity scores.

3.5.7 Questionnaires

SSQ A paired t-test analysis of the SSQ results between pre-measurements ($M = .176, SD = .158$) and post-measurements ($M = .161, SD = .172$) showed no significant difference ($t(19) = .88$, $p = .391$). The observed mean difference of .015 (95% CI [-.021, .051]) indicates no relevant changes or symptoms of simulator sickness.

IPQ The analysis showed that spatial presence was very high ($M = 1.08$, $SD = 1.09$), involvement was high ($M = .79$, $SD = 1.11$), and the general feeling of "being there" was moderate ($M = .6$, $SD = 1.50$). However, experienced realism was low ($M = -.64$, $SD = 1.07$). The overall presence score was moderate ($M = .47$, $SD = .87$).

Demographic Data The results of the CLMM showed that laughter intensity had no significant linear effect on the rating in interaction with either gender ($p > .26$) or age ($p > .8$).

3.6 Discussion

Results show mixed findings regarding the proposed hypotheses. While discussing the findings, it is important to keep in mind that the sample size of 20 participants may limit the generalisability of the results.

3.6.1 Influence on Participants' Opinion

Even though the RM ANOVA did not show any significant effects, the LRT did indicate a good model fit. Therefore, we found evidence that laughter intensity can influence rating. The analysis showed that laughter from the VAs significantly increased the probability for higher humour ratings ($p < .01$). These findings suggest that the VA's laughter influences the participants' ratings, supporting Hypothesis 1a.

Regarding potential inhibition effects, neither the RM ANOVA nor the LRT provided statistical evidence for a significantly better model fit when including these effects. Therefore, hypothesis 1b is not supported by the data.

3.6.2 Group size

The LRT and the RM ANOVA showed no significant effects. Therefore, H2 is not supported. Reasons for this could be, apart from the small sample size, the VR environment, and the salience of a particular VA. Firstly, a head movement to the left and right was necessary to see all VAs in full group size (see Figure 1). Secondly, one particular male VA who was always present when a male VA was needed was very prominent. In the debriefing, several participants said that this VA had a particular impact on them, as his laughter and movement were generally very contagious.

3.6.3 Mediators for the effect

The linear mixed models analysis showed that laughter affected lip activity (i.e. smiling), which indicates that the participants were physically reacting to the laughter, therefore experiencing emotional contagion. These findings align with prior research on emotional contagion, such as Chapman's study on canned laughter [8]. It can also be seen in Figure 4 that big group size does not play a vital role, even a single smiling VA is enough to trigger smiling. This supports the conclusion that one VA is enough to elicit emotional contagion, as seen in 2.1.2 with the work of Qu et al. [37] and Tsai et al. [47].

In the second model, the results show that smiling had a significant positive effect on funniness ratings, which implies that participants' smiles influenced their overall ratings of humour. This aligns with previous research (e.g., Strack et al., [44]) suggested that facial feedback enhances humour perception.

When we turned to the mediation analysis, it revealed that lip activity (i.e. smiling) plays a crucial role in how perceived laughter influences humor ratings. While the direct effect of laughter on humor ratings became nonsignificant when lip activity is taken into account, the indirect effect (mediated by lip activity) is significant.

This suggests that smiling can moderate the relationship between perceived laughter and humor ratings. Overall, more than half of the effect of laughter on humor ratings is mediated by lip movements, highlighting their importance in the perception of humor.

In conclusion, the hypothesised mediation effect through a chain reaction of emotional contagion and facial feedback was supported. Therefore, H3 is supported by the data.

3.6.4 Moderators for the effect

While agreeableness does not affect the influence of laughter, we found a significant interaction for the values of group conformity, which shows that the influence on the rating takes place especially when the conformity values of the participant are high: Participants with high conformity scores adapted their ratings more strongly to the laughter of the VAs. This finding aligns with Kyrlitsias' replication of Asch's experiment [29], which demonstrated that VAs can induce conformity — consistent with our results. H4 is partially supported by the data.

3.6.5 Exploratory Analysis

The results of the exploratory analysis have shown that both hypothesised factors influence the rating in joint interaction. Subjective conformity behaviour can act as a moderator that influences whether emotional contagion occurs or not. This joint interaction then influences the opinion of the participant. If the conformity values are low, i.e. a person does not act group conform, the perceived laughter will not influence their opinion, not even through emotional contagion. However, if a person has high group conformative behaviour, emotional contagion can be triggered and the rating will be adjusted.

3.6.6 Questionnaires

The results of the SSQ questionnaire show that the study did not induce cybersickness. The results of the IPQ are moderate and equivalent to the values from the comparative analysis [46].

4 Limitations and Future Work

The small sample size (N = 20) limits the statistical power and generalisability of the results. A larger sample size would avoid this limitation in future work. Another limitation of our work is that we had to exclude the participants' audio data. In future studies, this could be avoided by telling the participants not to speak during the experiment or by recording the audio and extracting laughter afterward.

Another limitation relates to the very subjective aspect of the variable that was tested, i.e., the funniness assessment of jokes. Humour is subjective, and although precautions were taken in the study (different conditions per joke, approximately equally funny jokes), individual differences could play a role. Future work could eradicate this with a pre-study in which participants evaluate jokes beforehand, and equally funny jokes are then selected. Another approach could be to conduct a comparative study outside of VR. This would help to better quantify and compare the agents' influence.

Furthermore, to avoid salience effects and recognise what influences the participant, including a gaze detector that records where participants look before rating the funniness would be helpful. This could help, for example, assess whether the male character reported to be particularly salient played a dominant role in the results.

A large proportion of the participants were also psychology or HCI students. They are already familiar with the concepts of emotional contagion and group conformity, which could lead to an increased awareness of the purpose of the study, possibly influencing their responses. In addition, the use of within-groups can lead to fatigue or more likely demand characteristics, since it may become apparent what the experiment is about. Thus, future works shall include more diverse participants and consider conducting between-group studies.

Future work could investigate whether a stronger ingroup feeling, e.g. through equal gender or age groups, would strengthen the effect. Intelligent Virtual Agent (IVA) could be used here, who not only talks to the participants during the study to build rapport, but can also reinforce the ingroup feeling with certain statements (for example the IVA claims to be a fan of the same football team, or similar).

Another interesting aspect for future work would be to require participants to state their opinions out loud instead of submitting them anonymously. Research suggests that public responses increase social conformity, as individuals tend to align their answers with the group due to social pressure and fear of standing out (see Section 2.2.2, [5]). By making the rating process public, future studies could explore whether the conformity effect becomes even stronger in a virtual setting.

The role of immersion could be also studied in future work. Some studies suggest that although the sense of presence is higher in VR, it does not always contribute to a stronger effect of the stimuli on the users. For example, Mostajeran et al. [34] discovered that exposure to a virtual forest environment has a positive effect on cognition regardless of the mode of presentation (immersive videos versus photo slideshows, all presented in a VR HMD). Similarly, since people laugh together in non-immersive media such as TV, a question arises whether we can replicate these results if we play only a video of the scene to the user or would they need to be immersed in the scene to see the effects that we observed in this study.

The findings of this study highlight the importance of considering group conformity and emotional contagion in the design and implementation of VAs across various domains.

In healthcare, where VAs are increasingly used to support patients and therapeutic interventions, VAs could use emotional contagion to create a more engaging and supportive environment. However, future research should investigate whether group conformity effects could unintentionally influence patients' responses, for example by adjusting their behaviour to VAs to avoid conflict. Understanding the underlying dynamics of group conformative behaviour towards VAs could help to develop agents that promote emotional engagement without social pressure.

In educational contexts, VAs serve as lecturers or tutors in digital learning environments. Here, the potential of emotional contagion could be used to boost student motivation and engagement, for example, an agent who behaves enthusiastically regarding a topic could have a positive influence on student motivation. In training contexts where VAs act as trainers, emotional contagion could be used to improve endurance and enthusiasm. For example, a virtual fitness trainer expressing enthusiasm and encouragement during a workout could have a positive impact on motivation and commitment to a training programme.

Beyond current applications with VAs, these findings could support the development of new training methods to counteract social pressure and group conformity, i.e. new applications could be developed to train people to avoid peer pressure and the influence of others. For example, VR-based simulations could be used to teach students how to form and defend their opinions independently, even when confronted with opposing group perspectives. By exposing them to controlled group pressure (e.g. through AI-driven group members (i.e. IVAs)), participants could learn how to manage group dynamics and keep their independent judgement. In high-stakes professions, such as law enforcement, emergency response and military operations, resisting groupthink and making independent decisions is critical. VR-based training could help professionals recognise and counteract these tendencies by putting them through simulated group decision-making scenarios in which they have to critically evaluate information and challenge group consensus. For example, police officers could train in VR scenarios in which they have to make split-second decisions under peer pressure, reducing the risk of conformity-based errors.

The findings can also help understand different aspects of the bystander effect through VR. The bystander effect describes the phenomenon that an individual is less willing to help a victim if passive bystanders are present in the situation [10]. This effect has been also observed in VR [42] where participants intervened more in an argument between two VAs if they shared a common social identity (i.e., supporters of the same sports club) with the victim VA. Another aspect of this effect is the group size of the bystanders. It has been suggested that the greater the size of the group of bystanders, the less the chance that any of them will intervene [13]. This can be systematically studied in VR by simulating critical situations and VAs as bystanders in different group sizes.

By answering these questions, future research can ensure that VAs are designed to take advantage of the positive aspects of emotional contagion and group conformity while minimising the potential risks, increasing their effectiveness in a wide range of applications.

5 Conclusions

This study investigated the influence of VAs on participants' humour ratings, focusing on the interaction between emotional contagion and group conformity. The results showed that the intensity of VAs' laughter significantly influenced participants' ratings, supporting the idea that emotional cues like laughter from VAs can influence personal opinions about humour. However, the analysis also yielded mixed results about the proposed hypotheses.

Notably, the results suggest that there is a significant relationship between laughter intensity and ratings of funniness, with increased laughter intensity leading to a higher probability of a higher rating. This suggests that laughter by VAs can indeed influence how funny jokes are rated, aligning with previous research on emotional contagion and group conformity in VR. However, there was no evidence for influence depending on group size, which was inconsistent with the original hypotheses.

We identified lip activity (i.e. smiling) as a significant mediator in the relationship between laughter and ratings.

We also observed that individual differences, particularly in conformity, significantly moderated the impact of laughter intensity on humour ratings. Participants with higher levels of conformity showed higher ratings in response to stronger laughter, highlighting the influence of social factors on humour perception. This aligns with established theories on group conformity and suggests that virtual interactions can evoke certain social behaviours.

These findings contribute to the growing body of research on Human-Computer Interaction, particularly within the context of Human-Agent Interaction, and offer valuable insights into the dynamics of virtual group conformity and emotional contagion in VR settings. We emphasise the importance of further exploration into these dynamics, calling for larger sample sizes and diverse methodologies.

Acknowledgments

This work has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101135025, PRESENCE project.

References

[1] S. E. Asch. Opinions and Social Pressure. *Scientific American*, 193(5):31–35, 1955. 1, 2

[2] R. M. Baron and D. A. Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6):1173–1182, 1986. Place: US Publisher: American Psychological Association. doi: 10.1037/0022-3514.51.6.1173 6

[3] S. G. Barsade. The Ripple Effect: Emotional Contagion and its Influence on Group Behavior. *Administrative Science Quarterly*, 47(4):644–675, Dec. 2002. Publisher: SAGE Publications Inc. doi: 10.2307/3094912 2

[4] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson. TARGET ARTICLE: Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, 13(2):103–124, Apr. 2002. Publisher: Routledge. doi: 10.1207/S15327965PLI1302_01 3

[5] R. Bond. Group size and conformity. *Group Processes & Intergroup Relations*, 8(4):331–354, Oct. 2005. doi: 10.1177/1368430205056464 2, 5, 8

[6] R. Brant. Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics*, 46(4):1171–1178, 1990. Publisher: [Wiley, International Biometric Society]. doi: 10.2307/2532457 5

[7] British Association for the Advancement of Science. *Laughlab: The Scientific Quest for the World's Funniest Joke*. Random House, Jan. 2011. 4

[8] A. J. Chapman. Funniness of Jokes, Canned Laughter and Recall Performance. *Sociometry*, 36(4):569–578, 1973. doi: 10.2307/2786252 1, 2, 7

[9] J. Cohen. A power primer. *Psychological Bulletin*, 112(1):155–159, 1992. Place: US Publisher: American Psychological Association. doi: 10.1037/0033-2909.112.1.155 6

[10] J. M. Darley and B. Latané. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8(4, Pt.1):377–383, 1968. doi: 10.1037/h0025589 9

[11] L. Dellazizzo, S. Potvin, M. Luigi, and A. Dumais. Evidence on Virtual Reality–Based Therapies for Psychiatric Disorders: Meta-Review of Meta-Analyses. *Journal of Medical Internet Research*, 22(8):e20889, Aug. 2020. doi: 10.2196/20889 1

[12] I. Didimo. Didimo: Automated character creation, infinite variations, and asset customization. `https://www.didimo.co`. Accessed on March 18, 2025. 3

[13] P. Fischer, J. I. Krueger, T. Greitemeyer, C. Vogrincic, A. Kastenmüller, D. Frey, M. Heene, M. Wicher, and M. Kainbacher. The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, 137(4):517–537, 2011. doi: 10.1037/a0023304 9

[14] J. Fox and G. Monette. Generalized Collinearity Diagnostics. *Journal of the American Statistical Association*, 87(417):178–183, Mar. 1992. doi: 10.1080/01621459.1992.10475190 5

[15] L. F. Góes, P. Sawicki, M. Grzes, D. Brown, and M. Volpe. Is gpt-4 good enough to evaluate jokes? 2023. 4

[16] A. Goldenberg and J. J. Gross. Digital Emotion Contagion. *Trends in Cognitive Sciences*, 24(4):316–328, Apr. 2020. Publisher: Elsevier. doi: 10.1016/j.tics.2020.01.009 2

[17] C. A. Gutiérrez, H. Carretero-Dios, G. B. Willis, and M. M. Morales. "It's funny if the group says so": Group norms moderate disparaging humor appreciation. *HUMOR*, 31(3):473–490, Aug. 2018. Publisher: De Gruyter Mouton. doi: 10.1515/humor-2017-0055 2

[18] R. R. Henares. *A New Look at Social Influence: Exploring the Personality Predictors of Conformity*. Phd dissertation, Purdue University, West Lafayette, IN, USA, 2024. 5

[19] N. Hertz and E. Wiese. Influence of Agent Type and Task Ambiguity on Conformity in Social Decision Making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1):313–317, Sept. 2016. Publisher: SAGE Publications Inc. doi: 10.1177/1541931213601071 3

[20] E. Hudlicka. Chapter 4 - Virtual Affective Agents and Therapeutic Games. In D. D. Luxton, ed., *Artificial Intelligence in Behavioral and Mental Health Care*, pp. 81–115. Academic Press, San Diego, Jan. 2016. doi: 10.1016/B978-0-12-420248-1.00004-0 1

[21] A. Inc. Adobe fuse. `https://helpx.adobe.com/de/support/fuse.html`. Accessed on March 18, 2025. 3

[22] S. Jentzsch and K. Kersting. Chatgpt is fun, but it is not funny!

humor is still challenging large language models. *arXiv preprint arXiv:2306.04563*, 2023. 4

[23] O. P. John, L. P. Naumann, and C. J. Soto. Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. In *Handbook of Personality: Theory and Research*, pp. 114–158. Guilford Press, New York, NY, 3rd ed., 2008. 4, 5

[24] J. F. Johnson. Conformity and humor: Group effects on laughter. 2

[25] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, 3(3):203–220, July 1993. doi: 10.1207/s15327108ijap0303_3 4

[26] P. Kraemer. Do As We Do, Not As You Think: The Effect of Group Influence on Individual Choices in a Virtual Environment. *Journal For Virtual Worlds Research*, 6(1), Apr. 2013. doi: 10.4101/jvwr.v6i1.7002 3

[27] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, June 2014. doi: 10.1073/pnas.1320040111 2

[28] L. Kruse, J. Hertel, F. Mostajeran, S. Schmidt, and F. Steinicke. Would You Go to a Virtual Doctor? A Systematic Literature Review on User Preferences for Embodied Virtual Agents in Healthcare. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 672–682, Oct. 2023. doi: 10.1109/ISMAR59233.2023.00082 1

[29] C. Kyrlitsias and D. Michael-Grigoriou. Asch conformity experiment using immersive virtual reality. *Computer Animation and Virtual Worlds*, 29(5):e1804, 2018. doi: 10.1002/cav.1804 1, 2, 3, 8

[30] J. Li, R. Kizilcec, J. Bailenson, and W. Ju. Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior*, 55:1222–1230, Feb. 2016. doi: 10.1016/j.chb.2015.04.005 1

[31] A. Manstead and A. Fischer. Social appraisal: The social world as object of and influence on appraisal processes. *Appraisal processes in emotion: Theory, methods, research*, Jan. 2001. 2

[32] A. Mehrabian and C. A. Stefl. Basic temperament components of loneliness, shyness, and conformity. *Social Behavior and Personality: An International Journal*, 23(3):253–263, Jan. 1995. doi: 10.2224/sbp.1995.23.3.253 4, 5

[33] F. Mostajeran, M. B. Balci, F. Steinicke, S. Kühn, and J. Gallinat. The effects of virtual audience size on social anxiety during public speaking. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 303–312. IEEE, 2020. doi: 10.1109/VR46266.2020.00050 3

[34] F. Mostajeran, J. Krzikawski, F. Steinicke, and S. Kühn. Effects of exposure to immersive videos and photo slideshows of forest and urban environments. *Scientific Reports*, 11(1):3994, 2021. doi: 10.1038/s41598-021-83277-y 8

[35] F. Mostajeran, F. Steinicke, O. J. Ariza Nunez, D. Gatsios, and D. Fotiadis. Augmented Reality for Older Adults: Exploring Acceptability of Virtual Coaches for Home-based Balance Training in an Aging Population. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, Honolulu HI USA, Apr. 2020. doi: 10.1145/3313831.3376565 1

[36] P. M. Niedenthal, P. Winkielman, L. Mondillon, and N. Vermeulen. Embodiment of emotion concepts. *Journal of Personality and Social Psychology*, 96(6):1120–1136, 2009. Place: US Publisher: American Psychological Association. doi: 10.1037/a0015574 2

[37] C. Qu, W.-P. Brinkman, Y. Ling, P. Wiggers, and I. Heynderickx. Conversations with a virtual human: Synthetic emotions and human responses. *Computers in Human Behavior*, 34:58–68, May 2014. doi: 10.1016/j.chb.2014.01.033 1, 2, 7

[38] J. Rickel. Intelligent Virtual Agents for Education and Training: Opportunities and Challenges. In A. de Antonio, R. Aylett, and D. Ballin, eds., *Intelligent Virtual Agents*, pp. 15–22. Springer, Berlin, Heidelberg, 2001. doi: 10.1007/3-540-44812-8_2 1

[39] A. Robb, R. Kopper, R. Ambani, F. Qayyum, D. Lind, L.-M. Su, and B. Lok. Leveraging Virtual Humans to Effectively Prepare Learners for Stressful Interpersonal Experiences. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):662–670, Apr. 2013. doi: 10.1109/TVCG.2013.35 2

[40] S. Schachter and J. Singer. Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69(5):379–399, 1962. Publisher: American Psychological Association. doi: 10.1037/h0046234 2

[41] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001. doi: 10.1162/105474601300343603 4

[42] M. Slater, A. Rovira, R. Southern, D. Swapp, J. J. Zhang, C. Campbell, and M. Levine. Bystander responses to a violent incident in an immersive virtual environment. *PLoS ONE*, 8(1):e52766, 2013. doi: 10.1371/journal.pone.0052766 9

[43] Stangl, W. Gruppenzwang. Online Lexikon für Psychologie & Pädagogik, Feb. 2022. 2

[44] F. Strack, L. L. Martin, and S. Stepper. Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54(5):768–777, 1988. doi: 10.1037/0022-3514.54.5.768 5, 7

[45] J. E. Taylor, G. A. Rousselet, C. Scheepers, and S. C. Sereno. Rating norms should be calculated from cumulative link mixed effects models. *Behavior Research Methods*, 55(5):2175–2196, Aug. 2023. doi: 10.3758/s13428-022-01814-7 5

[46] T. Q. Tran, T. Langlotz, J. Young, T. W. Schubert, and H. Regenbrecht. Classifying Presence Scores: Insights and Analysis from Two Decades of the Igroup Presence Questionnaire (IPQ). *ACM Trans. Comput.-Hum. Interact.*, 31(5):61:1–61:26, Nov. 2024. doi: 10.1145/3689046 6, 8

[47] J. Tsai, E. Bowring, S. Marsella, W. Wood, and M. Tambe. A Study of Emotional Contagion with Virtual Characters. In Y. Nakano, M. Neff, A. Paiva, and M. Walker, eds., *Intelligent Virtual Agents*, pp. 81–88. Springer, Berlin, Heidelberg, 2012. doi: 10.1007/978-3-642-33197-8_8 2, 7

[48] J. Ullrich, W. Stroebe, and M. Hewstone, eds. *Sozialpsychologie*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2023. doi: 10.1007/978-3-662-65297-8 3

[49] G. A. van Kleef and A. H. Fischer. Emotional collectives: How groups shape emotions and emotions shape groups. *Cognition and Emotion*, 30(1):3–19, Jan. 2016. doi: 10.1080/02699931.2015.1081349 1

[50] M. Volante, S. V. Babu, H. Chaturvedi, N. Newsome, E. Ebrahimi, T. Roy, S. B. Daily, and T. Fasolino. Effects of Virtual Human Appearance Fidelity on Emotion Contagion in Affective Inter-Personal Simulations. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1326–1335, Apr. 2016. doi: 10.1109/TVCG.2016.2518158 2

[51] M. Volonte, A. Robb, A. T. Duchowski, and S. V. Babu. Empirical Evaluation of Virtual Human Conversational and Affective Animations on Visual Attention in Inter-Personal Simulations. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 25–32, Mar. 2018. doi: 10.1109/VR.2018.8446364 2

[52] Y. Wu, S. V. Babu, R. Armstrong, J. W. Bertrand, J. Luo, T. Roy, S. B. Daily, L. C. Dukes, L. F. Hodges, and T. Fasolino. Effects of Virtual Human Animation on Emotion Contagion in Simulated Inter-Personal Experiences. *IEEE Transactions on Visualization and Computer Graphics*, 20(4):626–635, Apr. 2014. doi: 10.1109/TVCG.2014.19 2