Empirical validation of an agent-based model of emotion contagion

Erik S. van Haeringen, Emmeke A. Veltmeijer, and Charlotte Gerritsen

Abstract—In recent years, many agent-based models of human groups have implemented a mechanism of emotion contagion, yet empirical validation is lagging behind. The aim of the present paper is to validate an agent-based model of emotion contagion at the level of group emotion, by comparing simulations against the emotional development of real people in small groups. To study the effect of emotion contagion, the participants interacted via a video call, where they were virtually placed in different social environments while they played a quiz. This allowed the exchange of emotion among all, some or none of the participants. The patterns of emotional development in the empirical results supported our hypotheses based on literature of emotion contagion and social norms. Further, the simulations with the complete model resembled many of these patterns. When emotion contagion was disabled in the model, the resemblance decreased. These results give a first indication that emotion contagion occurs in groups that meet via video calls, and can in-part be predicted by the proposed model of emotion contagion. Yet, further study with a larger and more diverse empirical sample is needed, as well as comparisons across contagion mechanisms, to draw stronger conclusions and ultimately justify societal application.

Index Terms— Emotion contagion, Modeling human emotion, Multiagent systems, Validation

✦ ———————

1 INTRODUCTION

EMOTION contagion is a largely subconscious process where the emotions of people in groups become more similar as the result of the expression of those emotions themselves [1]. Emotion contagion encapsulates a number of processes that drive the formation of collective emotion in crowds that meet in-person as well as via media and online [2]. While the effects of emotion contagion in groups are often subtle, in some cases the effects can be extremely harmful to individuals and society. Every year there are outbreaks of mass panic and anger in crowds that cause injuries and deaths, of which the recent Astroworld stampede and riots at the US Capitol are examples that received much attention [3], [4]. Also hatred, anxiety, loneliness and depression have been suggested to be contagious [5]. Motivated by this, a number of computational models have been developed over recent years that are mostly aimed at the spread of negative emotions in large groups of people and their effect on behaviour, such as during evacuations [6].

In a literature review of agent-based mechanisms of emotion contagion, we concluded that empirical validation of these models of emotion contagion is lagging behind [6]. Moreover, most of the studies that did validate a crowd model against real people, compared the actions of people in videos to the actions of agents, like movement speed or direction [7], [8]. Since behaviour choices depend on numerous other factors besides emotion, this method provides indirect evidence for the validity of the contagion mechanism at best. Above all, establishing rigorous validation for models with emotionally interactive agents is important to eventually justify bridging the gap from scientific work to practical use cases, like for event planning, crowd management, warning systems and training purposes [9]–[12]. We argue that this should include more direct validation for the spread of emotions in groups, not merely action patterns that hint at an underlying emotional state.

What makes it challenging to validate a crowd model at the level of emotions, is the difficulty to collect detailed and reliable data about the emotional state of groups of people. Emotion, as well as entangled factors like personality, are generally seen as private, ethically limiting the data collection in the wild without informed consent. Also from a technical perspective, there is still an ongoing scientific challenge to reliably track the emotions of large groups of people in uncontrolled conditions [13], [14]. A notable exception can be found in the online crowd on public social media [2]. There, people share their expressions and react to others with the knowledge that this will be public, usually in the form of text, images or videos, which can relatively easily be collected. However, without direct face-to-face interaction and regulating feedback, it is not clear how representative these forms of contagion are for the spread of emotion in real crowds [15].

An environment that potentially bridges this gap, is that of video calls. There, participants are used to their expressions being recorded and shared with others whilst interacting face-to-face, albeit via a screen. We are not aware of any studies that have investigated the effectiveness of emotion contagion mediated by cameras and screens in groups. However, several studies have examined emotion

---

contagion in related settings. Rosenbusch et al. found emotion contagion in live streams on YouTube, where the expressions of one person are constantly recorded and shown, while those who were watching reacted via text in a group chat [16]. Mui et al. examined dyadic emotion contagion of an actor to a participant via virtual face-to-face contact mediated by webcams and screens [17]. While they found smile mimicry by the participants, they did not find an increase in the self-reported level of joviality. Hsu et al. compared the valence and arousal response measured via activity of facial muscles between facial expressions in pre-recorded video and via a live stream [18]. They found a stronger emotional reaction in the participants for live social interaction via a screen than for the pre-recorded video. Other work has also found evidence for mimicry and emotion contagion when participants watched emotional expressions in pre-recorded videos or images [19], [20].

To make a first step towards validation at the level of group emotions, the aim of the present paper is to compare agent-based simulations of emotion contagion against the emotional development of real participants in an experiment via a video call. The participants in this experiment play a competitive quiz in two teams via the video call, where the emotional state of each participant is annotated manually from the recorded video in small time steps. By modifying the composition of the virtual environment, different conditions are created with regard to the spread of emotions. These conditions include 1) virtually isolating the participants to disable emotion contagion, 2) virtually grouping the participant per team to allow contagion among agents with similar emotional stimuli, and 3) placing all participants in the same virtual space, allowing contagion among participants with conflicting emotional stimuli. Although the groups in this experiment are too small to be called a crowd, the context of the experiment could be argued to approach a slice of a crowd. This is because people with diverse traits, and who are mostly unfamiliar with each other, meet by chance in an unfamiliar environment to interact face-to-face, albeit via cameras and screens. Therefore, while individual traits, like gender, culture and group membership, have been shown to affect how emotions are expressed and read [21]–[23], in this initial study we chose not to control for individual traits to mimic the stochastic nature of the crowd.

Since emotion contagion is believed to drive emotional similarity [1], we hypothesised that the participants become more emotionally similar over time when they are in the same virtual space, forming a collective emotion. In contrast, we expected this does not happen when they are virtually isolated. Further, we expected that winning a quiz round results in a positive emotion, while losing a round triggers a negative emotion. Since one team wins and the other loses, we hypothesised that when the participants are virtually grouped per team, the emotion converges within a team, and the difference between the teams increases. On the other hand, when the teams are virtually placed in the same space, we expected the emotions of all participants to converge to some degree, decreasing the emotional difference between the teams. Finally, based on

literature that finds that there are larger constraints against the expression of negative emotions than most positive emotions in groups [24], we hypothesised a win is followed by relatively strong expressions of positive emotions, while a loss is followed by more diverse expressions that are weaker.

After performing the experiment, simulations with similar conditions to the experimental setting were performed using an agent-based model of emotion contagion. The model proposed in the present paper is an extension of the model DECADE (Dimensional Emotion Contagion via Agent-based Dyadic Exchanges) [25]. The emotional state of an agent in this model is defined as a location in a two-dimensional space of emotion, namely valence and arousal. How strongly valence and arousal spread, is determined by the social norms, personality and attention of the agents. Beside the process of emotion contagion, the proposed model includes processes for generation of emotions due to appraisal of game events (winning or losing) and the regulation of emotions over time due to natural decay.

2 METHODS

2.1 Empirical data

The empirical data was collected from an experiment with small groups of participants that played a quiz via a video-call on the platform Zoom [26]. The participants were divided over two teams, who competed in the quiz for an additional reward awarded to each member of the winning team. The aim of this setup was to elicit an emotional response from the participants that was to some degree predictable upon hearing the positive or negative result of the quiz question. Specifically, the facial emotional expression of the participant was recorded and annotated around them seeing the quiz results to deduce the emotional response. From these videos, the valence, arousal and categorical emotion of the participant were determined for 25 seconds around the result of each quiz question. Of this period, five seconds before the stimulus were examined to establish a base level of emotion to which the response can be related. The duration of twenty seconds following the stimulus was determined empirically to include most of the emotional responses to the stimuli, weighed against the total number of annotations that had to be made. While the participants were made aware pretrial that their video feed would be recorded, they were not aware of the specific focus of the study on emotions. Instead, the participants were told that the experiment was about group dynamics. To zoom in on the effect of emotion contagion, the group composition that the participant was exposed to on the screen was varied. Each participant was placed alone in a room behind a laptop. By digitally placing the participants in various virtual rooms, we created three conditions. In the first round there was one virtual room per team, allowing the team members to communicate and exchange emotions. In the second round all participants were isolated. In the third round all participants were in the same virtual room, allowing communication and emotion exchange between both teams as well as within the team.

Each quiz round consisted of two questions about quantitative facts that the participants were unlikely to know, like how many bridges there are in Amsterdam. The team with the closest answer got a point and this feedback was provided directly after each question. The researchers were not visible or audible to the participants during the experiment. Instead, the communication between participant and researcher took place via the chat function of the video call. In the first and third round a representative of the team answered the question in the chat. In the second round, each team member answered the question in the chat in isolation, after which the average was taken as the answer of the group.

Further, pretrial the participants were asked to fill in a digital consent form followed by a personality survey. The Big Five Inventory (BFI) was chosen to measure the personality of the participants [27], as it is widely used, freely available and its results can be translated to a profile for the OCEAN model of personality that is commonly used in agent-based simulations of crowds [28]. After the quiz, the participants were also asked to fill out a survey that aimed to assess their pre-existing relationship with the other participants. However, this form was not clear for everyone, causing some participants to misinterpret the question. Therefore, the effects of the social relationships among participants were not analysed in the present study. See Appendix A2 for a more detailed description of the experiment procedure. We obtained approval of the Ethics Committee for Information Science for the experiment.

2.2 Annotation

To establish a ground truth to validate the model against, we manually annotated the facial expressions of the participants in the recorded videos. Two researchers independently scored the videos of 25 seconds around the publication of the quiz answer in steps of one second. The dimensional emotions of valence and arousal were scored on a five-point Likert scale, ranging from very negative to very positive for valence, and from very passive to very active for arousal. Further, the assessor also assigned a categorical emotion label. These consisted of nine labels previously used in DECADE [25], plus the label 'unclear'.

The annotations were made using a tool that we developed in C++ using the Qt framework. This tool also provided the annotators with several reference clips for most of the valence-arousal combinations. These clips were extracted from the AffWild dataset that contains a collection of annotated videos from YouTube [29]. See appendix A3 for more details about how these examples were extracted from the dataset. As a pilot, the annotators rated 100 clips of one second randomly selected out of the total set of 3150 clips. They found that rating one-second slices of the videos in a random order was very difficult because of the short length and the missing context. The annotators expressed that they often felt unsure about their decision and the resulting agreement, measured as Krippendorff's alpha [30], was low. This was particularly the case for arousal (0.35) and the categorical label (0.39), while there was a higher agreement for valence (0.65). Social sciences commonly view alpha > 0.8 as a convincing agreement,

TABLE 1
MAPPING OF THE LIKERT SCALES OF VALENCE AND AROUSAL TO CONTINUOUS VALENCE/AROUSAL

Annotation	Continuous
0) Very negative/passive	-0.8
1) Slightly negative/passive	-0.4
2) Neutral	0
3) Slightly positive/active	0.4
4) Very positive/active	0.8

The Likert scales were used to annotate valence and arousal of facial expressions. The continuous valence-arousal space is used in the proposed model.

whereas 0.8 < alpha < 0.67 might be accepted to draw tentative conclusions [30].

To address this, the annotation tool was redesigned to allow the annotator to watch the full 25 second clip alongside the one second excerpt from the full clip they are rating to provide context for the expression and movement of the participant. The annotators expressed that their subjective confidence in the rating increased markedly when the context of the full clip was provided. The inter-rater agreement also increased significantly, with alpha-scores of 0.81, 0.85 and 0.63 for valence, arousal and the categorical labels respectively. It should be noted however that the actual agreement of the categorical labels is likely higher than indicated by this result. This is because some labels are closer to each other than others, like when one researcher scores happy and the other pleased, they agree more than when they score happy and sad. This can also be seen in Appendix A7 that shows the relation between valence, arousal and categorical emotion in the annotations. However, this cannot be taken into account when calculating Krippendorff's alpha as it is not an ordinal relationship, like the valence and arousal scale that were used. For this reason, we chose to show the empirical results for the emotion labels despite the low agreement score, but not to use them in validating the model. Further, when annotators disagree on the valence or arousal score, the average of the scores is used for the analysis.

2.3 The proposed model

To model the emotion dynamics of the participants, an agent-based model of emotion contagion was extended with personality, social norms and the appraisal of game events. The DECADE model formed the basis for the contagion mechanism among agents [25]. In DECADE, emotion spreads in the form of valence and arousal that together form the emotional state of the agent. The EEGS model (Ethical Emotion Generation System) served as inspiration to implement a stepwise structure to simulate the process from emotional stimulus to appraisal, integration, regulation and expression [31]. However, where the EEGS model is based on categorical emotions and focusses in higher detail on the appraisal process, for the proposed model these steps are simplified and performed in the context of dimensional emotion.
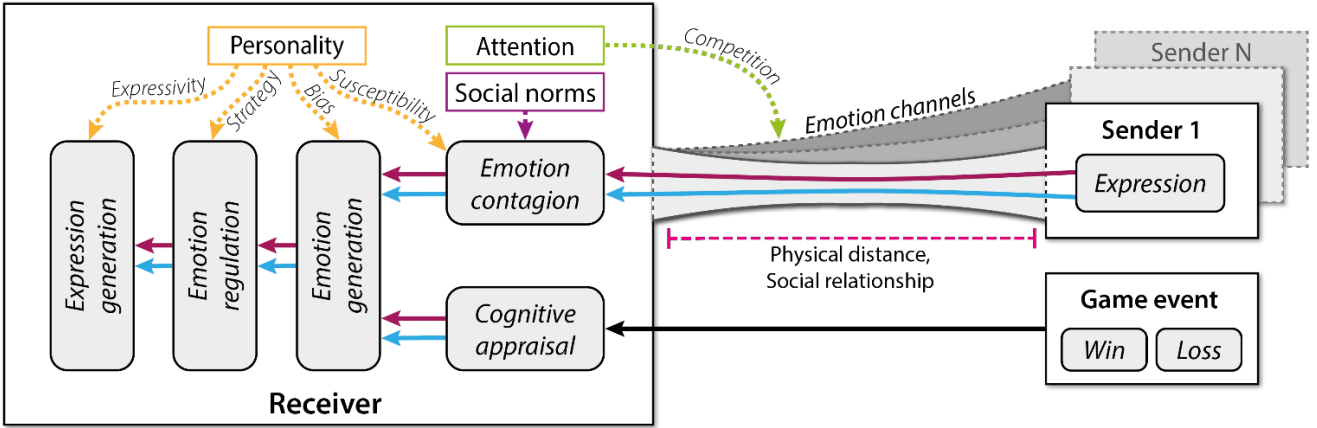
Fig. 1. Flow of the proposed model. The purple and blue arrows represent valence and arousal.

As shown in Fig. 1, emotional triggers are first acquired, either via inter-agent contagion or cognitive appraisal of external events. From these triggers an emotional response is generated, followed by emotion regulation considering the social norms of the agent, finally resulting in an emotional expression. These steps are influenced by the personality of the agent. For this, the personality profile of the agent was implemented using the OCEAN model of personality [32], also known as the Big Five. For brevity, the discussion below is limited to an explanation of the general process in each model step. We refer to Appendix A4 for the mathematical implementation of the model.

2.3.1 Emotion contagion

In DECADE emotion flows from a set of senders to a receiver through emotion channels [25]. The same mechanism is used in the proposed model, except that the receiver does not directly perceive the internal emotional state of the sender, but perceives its emotional expression. How well the expressed emotion flows from the senders to the receiver depends on the susceptibility of the receiver for the emotions of others, as well as the physical distance and social relationship between the sender and the receiver. The emotional susceptibility of the agent is derived from its personality, following the method used in several existing models of emotion contagion [33], [34], which is based on the empathy scale by Jolliffe and Farrington [35]. Since the experiment took place in a virtual room, the distance between the receiver and each visible sender on the screen was assumed to be equal. Further, we intended to set the social relationships among the agents based on the social relationships among the real participants. However, the survey that was used for this purpose was insufficiently clear to some of the participants. Since no reliable data was collected about the social relationships, the relations among the agents were set to be equal.

Next, the different emotions that reach the receiver through the emotion channels compete for the attention of the receiver, to establish a single weighted emotion. In DECADE, the receiver has an attention bias towards more emotional expressions, which is based on an empirical study that found that people overestimated the emotion of others in groups [36].

Finally, the weighted emotion is either dampened, absorbed or amplified by the receiver. In the proposed model, this is determined by a social norm for which emotions are acceptable to display in a group. Social norms for emotion display are known to vary depending on individual differences, culture and context, yet the rules against displaying joy were consistently found to be less restrictive than negative emotions like anger or fear [24]. However, a recent study found that this does not apply to all positive emotions, as for example less strict display rules were found for showing amusement and gratitude than for sensory pleasure or triumph [24]. Thus, although social display rules are more complex, for the purpose of the present study we implemented a simplified version of this concept via a parameter in DECADE that controls whether incoming emotions via contagion are dampened, absorbed or amplified by the agent [25]. Specifically, this is set such that agents have the tendency to amplify group emotion with a positive valence, absorb emotion with a neutral valence and dampen negative emotions of others.

2.3.2 Cognitive appraisal of game events

In appraisal theory of emotion, the process of stimulus perception to cognitive awareness is often divided in two steps [37], [38]. The first is a lower-level non-cognitive process where first-order phenomenological appraisal takes place. It results in a physical response to the event that is positively or negatively oriented and forms the basis for the following cognitive appraisal. In this second step, the desires, standards and attitudes of a person are applied to trigger one or more associated emotions. Similar to the EEGS model [31], the proposed model simplifies these two steps to a single step.

In the present study we did not test the individual attitudes, desires and standards of the participants to the performed experiment. Instead, we assume there is a universal goal to win from the other team, as this goal was instructed preceding the experiment and incentivised with an additional financial reward for the winners [39]. Specifically, the participants received a 15-euro gift card instead of a 10-euro card if their team won. Adjacent, we also assume a goal for at least some of the participants to enjoy themselves. This assumption is based on literature that links social competition in games to enjoyment [40], [41],

combined with the voluntary choice of the participants to take part in the experiment knowing it would comprise playing a quiz with others, while the financial reward of winning was relatively low. However, it is likely that there is variation in how strongly these desires were present in the participants. To represent this variation, the value for the cognitive appraisal of an event was drawn from a normal distribution. For the valence component the winning agents draw from distribution that has a positive mean, as the event is in agreement to both desires. In case of a loss, the goal to win and the goal to have a good time may conflict. Therefore, the agent draws from a distribution with a neutral mean, meaning that some agents react positively to the loss and other negatively. Since emotional arousal activates processes aimed to cope with both threatening and appealing situations [42], we expect arousal to increase either when perceiving a win or loss. Therefore, in the model the arousal component is always drawn from a distribution with a positive mean. We find that the directions of these assumptions are in line with the initial reactions of the real participants, as shown in section 3.1.1.

2.3.3 Emotion generation

In the next step the agent integrates the incoming emotions as the result of contagion and appraisal into its internal emotional state. Personality is believed to influence the tendency for a person to experience certain emotions, where for example the neuroticism trait has been linked to negative affect [43]. For the proposed model, we draw from a meta-analysis by Steel et al., who among others study the correlations between the OCEAN personality traits and the propensity for positive and negative affect [44]. The agents multiply the incoming emotions with a personality bias toward positive emotions if valence is above zero, or else with a personality bias toward negative emotions. These personality biases are calculated by multiplying the personality profile of the agent with the slopes found by Steel et al. between each trait and positive and negative affect. The result of this is that emotions that fit the personality of the agent are inflated, while those that do not are deflated to some degree. Then, since the emotional state of the agent is represented as a single point in valence-arousal space, the net emotion is calculated by adding the valence and arousal of each incoming emotion, that was adjusted by a personality bias, to the emotion location of the agent.

2.3.4 Emotion regulation

The last step in calculating the internal emotional state of the agent is in regulating its emotion. In previous work, the DECADE model included emotion regulation in the form of a decay over time towards a neutral valence-arousal state. This was based on the framework of Hudlicka who state that emotion decay is more likely to be exponential than linear [45]. The exponential decay implemented in DECADE results in an immediate and fast decline of emotion after a single stimulus that slows as it approaches a neutral state [25]. However, Ojha et al. argue it is more realistic for the agent to maintain an emotion for an amount of time, after which it decays exponentially [31]. In line
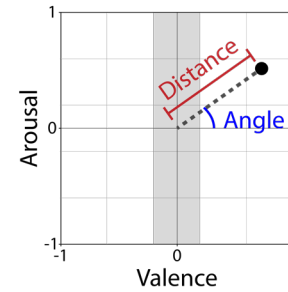

Fig. 2. The proposed metrics use the Euclidean distance of the agent's emotional state in the valence arousal space towards the origin [0, 0] and the direction angle towards the positive valence axis in degrees. When the emotional state equals the origin, no angle is calculated.

with this, a natural decay of emotion was implemented in the proposed model in the form of a hyperbolic tangential function. This results in the persistence of emotion for some time, after which valence and arousal decay with an inverted s-curve towards the neutral state, without overshooting the axis.

Further, people differ in how effectively they can regulate their emotional state [46]. Part of this variation is linked to differences in personality [47]. To simulate this, the maximum time that an emotion persists in an agent is drawn from a normal distribution, of which the mean depends on the personality of the agent. For the relation between regulation effectiveness and personality, we draw from a meta-analysis of personality traits and emotion regulation strategies by Baranczuk [48]. According to the authors three of the six strategies are effective in regulating emotion, namely reappraisal, problem solving and mindfulness. These strategies were found by Baranczuk to correlate significantly with each of the five traits of the OCEAN model. Therefore, to determine the regulation effectiveness of the agent in the proposed model, the sum is taken of each personality trait of the agent, modified respectively by the average slope that was found by Baranczuk for the three strategies.

2.3.5 Expression generation

Finally, once the internal emotional state of the agent has been updated, the agent will present its emotional state to a certain degree to the other agents as an emotional expression. How strongly the internal state is expressed depends on the personality of the agent. The expressivity trait of an agent is calculated by taking the sum of the personality traits of this agent modified by the correlations found empirically in study by Gross and John between each personality trait and emotional expressivity [49].

2.4 Quantitative metrics

To study the quantitative differences in emotional responses in groups, we propose four metrics. These are applied to both the empirical data and the model output from the time the result is published at second five, to the end of the video at second twenty-five. The metrics are inspired by the dimensions described by Gross and Jazaieri [46]. The authors discuss patterns in the intensity, duration, frequency and type of both healthy and psychopathological emotional reactions of individuals. We could not directly

apply these dimensions, as Gross and Jazaieri do not propose specific metrics and focus on categorical emotion on various timescales, while the present study examined dimensional emotion on a relatively short timescale. To adapt the described dimensions to our use case we propose the following metrics:

1. The duration of the emotional response is measured as the number of seconds valence is not within the neutral category of the Likert scale (grey area in Fig. 2). This purposefully does not require a consecutive series of emotion, as this would make it highly sensitive to noise in the expression or annotation process.
2. The emotion intensity is measured as the maximum of the distance in valence-arousal space toward the neutral origin [0, 0]. For the present study we assume there is a linear relationship between the five-point Likert scale for arousal and valence used to annotate the empirical data, and the valence-arousal space used in the model ranging from -1 to 1 (Table 1). Hence, each annotation category represents a span of 0.4.
3. Measuring the frequency of emotion is difficult for dimensional emotion and less relevant for the short timescale in this study. Instead, we measured the instability of the emotional response within an agent. Specifically, this is calculated as the variation in the emotion angle in the form of the standard deviation (Fig. 2).
4. The variation in the type of response among agents is measured as the absolute difference between the average emotion angle of an agent and the average emotion angle of all agents for a particular condition (e.g. isolated) and stimulus type (e.g. win).

2.5 Analysis

The proposed model was implemented in C++ using Microsoft Visual Studio. The statistical analysis and generation of figures were performed with RStudio. The annotations and personality profiles of the empirical data are included in the supplementary materials. The underlying video footage however is not shared for ethical reasons. Further, the model code and simulation data are available in the supplementary materials, as well as the R and Python scripts used in the analysis. See Appendix A1 for each of the above-mentioned materials. Appendix A5 contains the parameter settings for the simulations.

A qualitative analysis is performed via figures that show the average and standard error of the valence, arousal and categorical emotional response over time, split per social condition (isolated, together per team and both teams mixed) and per stimulus type (win, loss). A similar figure is made with the emotional distance between the teams for each of the grouping conditions.

A quantitative analysis is performed by statistically testing the effect of the social condition and stimulus type as independent variables on each of the metrics as dependent variables in the form of a Factorial ANOVA. To meet the normality assumption for the residuals, a square-root transformation was applied to the intensity and type deviation variables. For the instability variable a procedure called Tukey Ladder of Powers was used to find the transformation with the lowest deviation from normality [50],

which resulted in a power of 0.55. The assumptions for normality and homoscedasticity were checked visually, while the assumption for no multicollinearity was checked by testing whether the variance inflation factors were below the threshold value of three, as suggested by Zuur et al. [51]. We found no significant violations of these assumptions, except for the normality assumption of the instability variable, that despite the applied power transformation still showed a right skew. However, taking into account that the ANOVA test is reasonably resistant to violations of the normality assumption if the sample size and variance across groups are similar, and that the danger of such a violation is of a false positive [52], we conclude that the finding of no significant difference for the instability variable is justified. Subsequently, we performed Sidak post-hoc tests to establish which conditions differed significantly from each other within a stimulus type (win/loss).

3 RESULTS

3.1 Empirical results

In total 21 people participated in the experiments, divided over three sessions with six, nine and six participants per session respectively. 62% of the participants was male against 38% that was female, and all were either academic students or employees. The results of the personality survey are shown in Appendix A6. We extracted six videos of 25 seconds per participant (one per quiz question), resulting in 126 videos. These videos were annotated by two researchers once per second, resulting in 3150 data points in total.

3.1.1 Emotion development

Looking at Fig. 3, our general expectation is met, that the stimulus in the form of the quiz results, around second five, triggers an emotional response that subsequently fades over time. Winning the round results in an increase in valence and arousal, that is most pronounced when others in the virtual room win too, somewhat less pronounced when some of the other participants in the room lose, and lowest when the participant is isolated. Looking at the categorical emotions, the mixed and isolated conditions also contain a low percentage of negative emotions, despite winning the round. In the mixed round this may be the result of emotion contagion or other forms of interaction with the participants that lost. Since in the isolated round the team answer is the average of the individual guesses, the negative emotions may be a reflection of their own guess performance. Another explanation may be a lack of stimulation as suggested by the low arousal scores.

Next, losing the round results in a general increase in arousal, while the valence response is mixed within and between participants, as also illustrated by the categorical labels. In the isolated condition the response among losing participants is mixed, with some responding positive and others negative, balancing each other out to a net neutral response. The initial response of the team condition (from second five to eight) is characterised by a net decrease in valence. In contrast, during the same time window in the mixed condition there is a net increase of valence, where
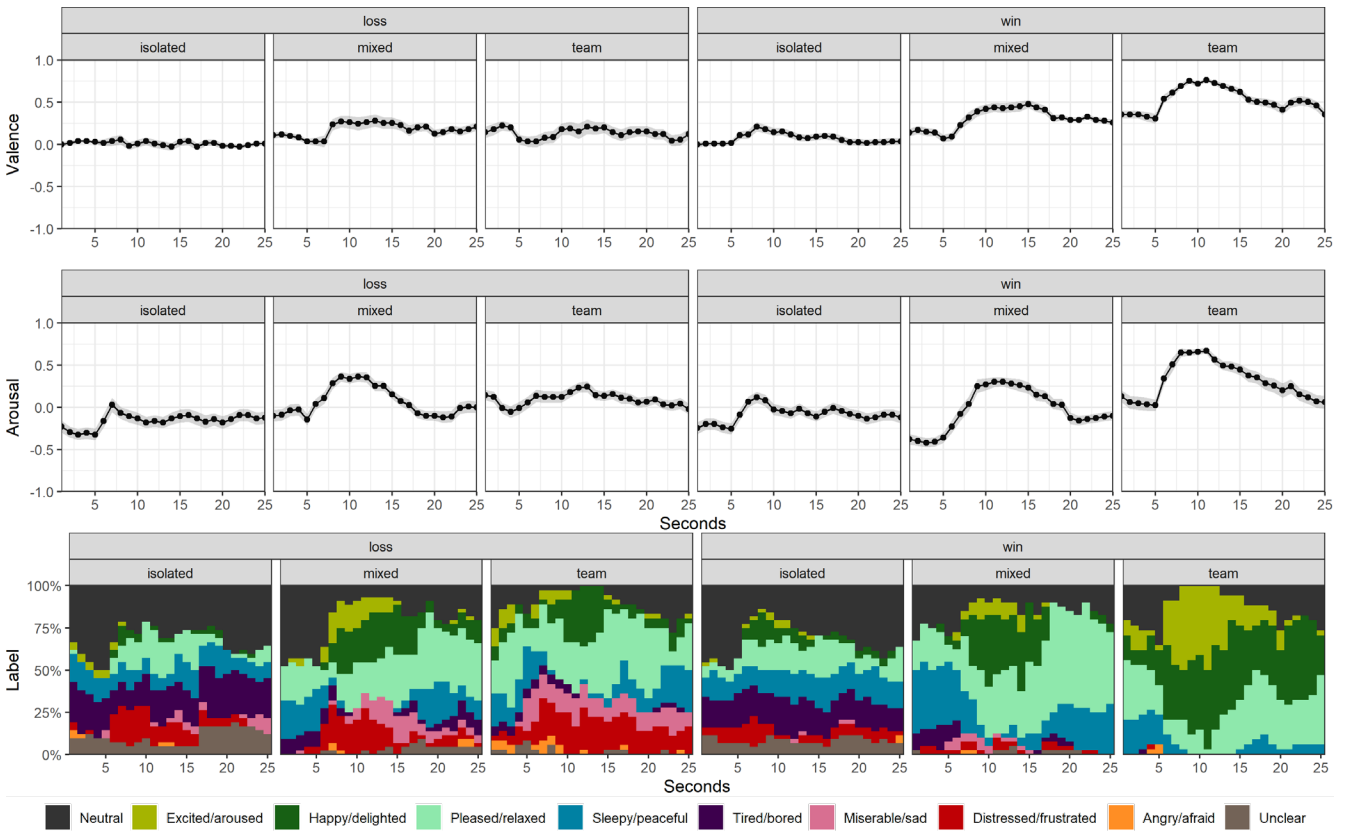
Fig. 3. Emotional development of the participants per condition and stimulus, shown as the mean ± SE for valence (top) and arousal (middle), and as the proportion of categorical emotion labels (bottom).
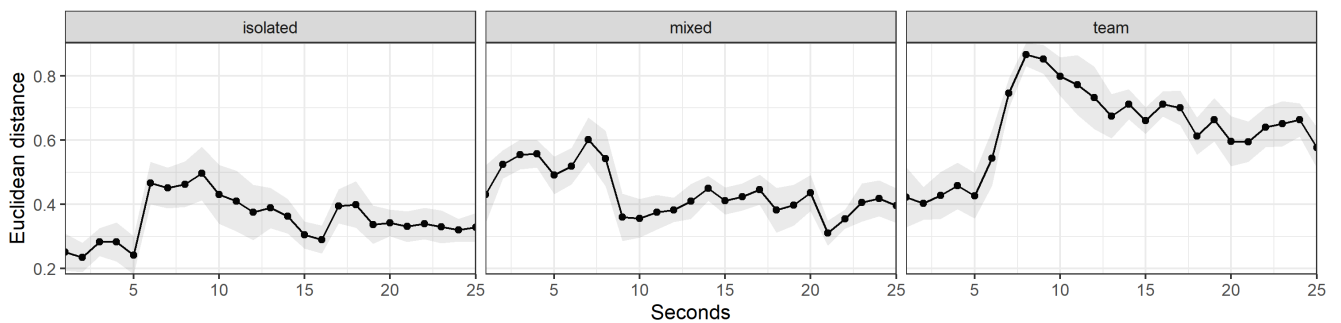

Fig. 4. Emotion deviation between the directly opposing teams, shown as the mean Euclidean distance in valence-arousal space ± SE.

the development looks similar to that of the winning participants in the mixed condition. This difference between the mixed and team condition, and similarity within the mixed condition for the two stimulus types are in line with our expectations for emotion contagion, where losing participants in the mixed condition are positively influenced by their winning counterparts. After this initial response however, follows in increase in valence in the team condition. An explanation for this could be the effect of social norms towards a positive attitude in social groups. Looking at the time before the stimulus at second five, valence is mildly positive in the conditions where others are visible, while valence is neutral in the isolated condition for this time period.

3.1.2 Emotional difference between opposing teams

The differences between the mixed and team condition in

Fig. 3. are in line with the expected effects of emotion contagion. However, these results are aggregated over multiple sessions and rounds. Therefore, it is not clear whether the opposing teams in the mixed condition indeed show more similar emotion responses, or whether the similarity comes as the result of averaging the rounds. Fig. 4 therefore explores the emotional difference between the two directly opposing teams as the distance in valence-arousal space of the mean emotions of the two teams. In the isolated and team condition the difference between the opposing teams increases when the stimuli are introduced around second five. In contrast, in the mixed condition the difference between the opposing teams first increases for two seconds before it sharply decreases. This pattern is in line with the idea that the participants first have to respond emotionally to the stimulus, before they can be affected via
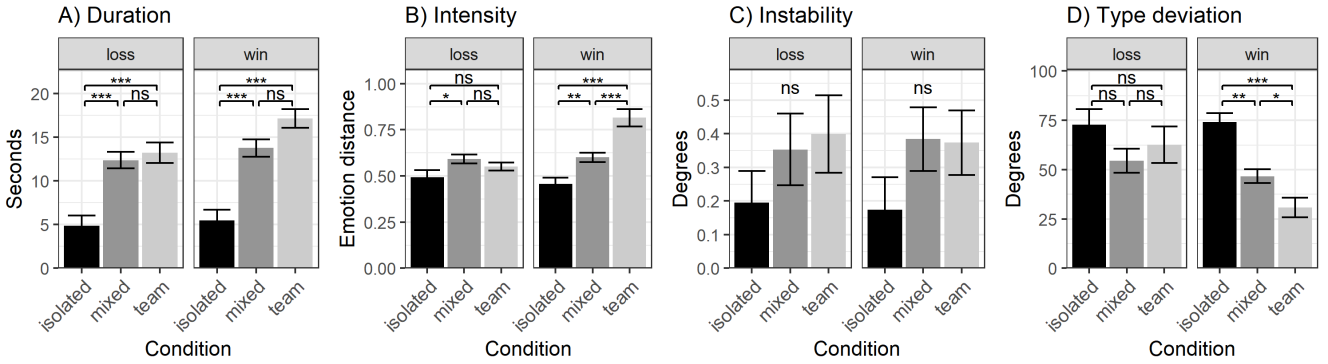
Fig. 5. Quantitative measures that describe the emotional response, given as the mean ± SE per condition and stimulus type. The result of the post-hoc tests are shown as stars, that indicate significant differences (p < 0.05 (*), 0.01 (**), 0.001 (***)) between conditions within a stimulus type (win/loss). For instability no significant effect was found for stimulus type or condition, thus no post hoc test was performed.

contagion by the emotions of others, which in turn causes the opposing emotions of the participants to become more similar.

3.1.3 Quantitative metrics of the emotional response

As shown in Fig. 5A, the conditions where participants were grouped with others resulted in a longer emotional response, compared to the isolated condition that did not depend on stimulus type. This suggests a form of social feedback for emotion expressions in groups, that maintains or reactivates the response. Important to note is that the duration may have suffered from a ceiling effect. For a share of the participants in the team condition (47% winners, 6% losers) the maximum duration for the chosen timeframe was measured. This potentially obscures a larger difference in duration between the team and mixed condition, as no participants in the mixed condition reached this ceiling.

The results for emotional intensity in Fig. 5B are similar to those discussed in the qualitative analysis of the dynamics, where the most intense response comes from winning participants grouped with their team. Being placed together with fellow losing participants results in an intermediate intensity, while the isolated participants show the least strong expressions.

Only a trend was found for the effect of the condition on the instability of the emotional responses. The trend as shown in Fig. 5C is in line with our expectations that isolated participants would show more stable emotional responses than participants in groups due to the emotional influence of the other participants. However, also in the isolated condition emotional instability was found. Since emotion contagion was not operational in the isolated condition, this indicates that there were other factors that caused the angle of emotion to vary.

Looking at Fig. 5D, the variation in emotion type among participants is higher when participants are isolated, compared to when they are placed with others, though this difference is not significant for the losing participants. This pattern matches our expectation that emotion contagion tends to converge emotions in groups, resulting in a collective emotional state. An explanation for the weaker effect of emotion contagion following the negative stimulus may be individual differences in the adherence and coping

strategies for the social norm against the display of negative emotions, hindering the formation of a collective emotion.

3.2 Model validation

The proposed model was configured such that the agents matched the participants in the experiments in how they were distributed over the virtual rooms and teams, who won and lost, and in their personality. Each configuration was simulated ten times as the proposed model contains stochastic factors, like in the appraisal of the game events by the agents, resulting in 180 simulations in total. Further the model was run with emotion contagion (fullModel), and without (noContagion) as a baseline comparison.

When comparing the dynamic response, the simulations with the full model reasonably match the patterns in the valence dimension of the real participants (Fig. 6). The most notable deviation is the lower valence of the winning agents in the mixed condition, compared to the real participants. Where the winning agents emotionally converged to high degree with the losing agents in this condition, the real participants seem to pertain some of the effect caused by the difference in stimuli. Conform our expectations, when emotion contagion is disabled in the model, the resemblance to the real participant significantly decreases in the grouped conditions.

When looking at the arousal dimension in Fig. 6, the general patterns that hint at emotion contagion in the empirical data are also found in the model output, like the stronger response in the winning team condition compared to the mixed condition. However, the simulated arousal response deviates markedly in the decay speed and the base level of arousal. For real participants, arousal (range -1 to 1) often started and sometimes ended well below zero in the examined timeframe, while the decay function in the model is aimed at a neutral state, which is defined as zero arousal [53]. An explanation for this may be that the model does not consider that the absence of stimuli leads to a lower arousal state, which translates to categorical emotions like boredom or sleepiness. Alternatively, it may be that the annotators rated the absence of expression and movement as a low arousal state, even though internally the agent was aroused. An indication along these lines may be that the lowest arousal across conditions is
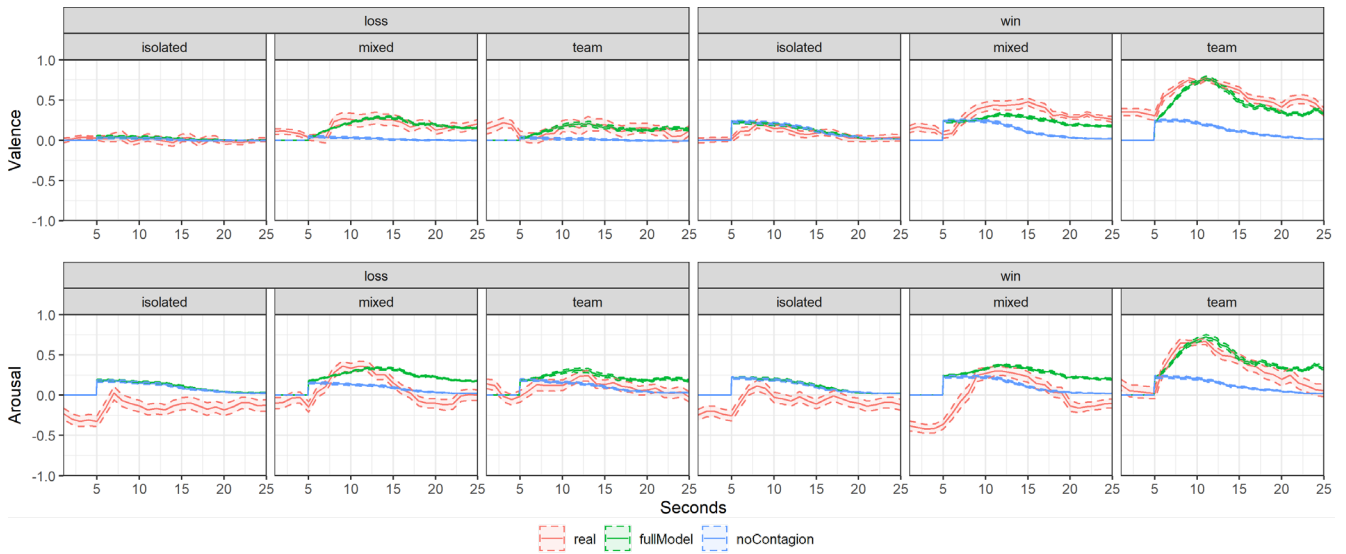
Fig. 7. Emotional development of the participants (red) and agents with emotion contagion (green) and without (blue) per condition and stimulus. This is shown as the mean ± SE for valence (top) and arousal (bottom).
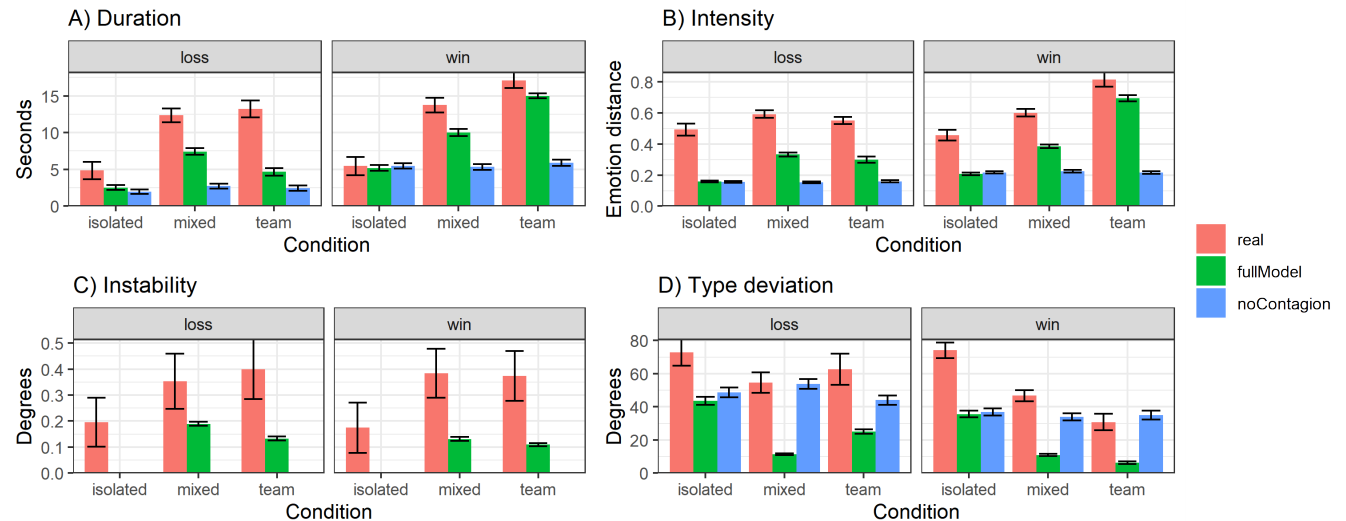

Fig. 6 Comparison of the four quantitative measures that describe the emotional response among the real participants, the full model and the model without emotion contagion, given as the mean ± SE per condition and stimulus type.

recorded upon the publication of the result around second four. Here, reading the message requires concentration and lack of movement that may have been interpreted by the annotator as a low state of arousal.

Even though the model reasonably resembles the average patterns in valence and arousal that indicate emotion contagion in the empirical data, when comparing the four quantitative metrics between the empirical data and model in Fig. 7, it becomes clear that the responses of the agents are less varied than those of the real participants. This is most obvious in the amount of variation within the individual responses (Fig. 7C). It shows there is more instability within participants than in the agents across conditions. Still, the expected pattern that the isolated condition is relatively more stable than the group conditions due to the emotional influence of the other participants, is found in both the agents and real participants. Similarly, the absence of additional factors from the model that result in

more individual variation in real participants may explain why the absolute duration, intensity and type deviation are all lower in the model compared to the empirical data, yet many of the relative differences among the conditions in the model follow a pattern similar to the real participants. This is not the case when emotion contagion is disabled in the model.

4 DISCUSSION

The aim of the present paper was to compare an agent-based model of emotion contagion, against the emotional development of groups of real people, who interacted face-to-face via a video connection in the context of a competitive quiz. With this objective, the first step was to test whether the emotional development in real groups matched the hypotheses based on the theory of emotion contagion. Congruent with our expectations, the combined

results show that the emotional responses in the experiment converge in groups. While emotion contagion has been found to be operational in a broad range of environments [2], [54], to our knowledge, these results show the first tentative evidence for emotion contagion via video calls in groups. This is important because the video call environment offers the possibility to record the face-to-face exchange of emotion in groups in a controlled setting, thereby providing a way forward to empirically validate models of emotion contagion. However, it should be noted that further study is required to draw more substantive conclusions, because of the relatively low number of trials and limitations with regard to the fixed order wherein the conditions were tested. Moreover, it is important that the context of the findings is taken into account. While this study did not explicitly control for individual traits, the participants were diverse in terms of gender and cultural background, yet similar in age range and level of education (primarily university students). The experiment was conducted physically in the lab, and as participants were placed in separate rooms, the unfamiliar environment and presence of strangers may have altered or dampened emotional expression, compared to other contexts where people are more at ease. Further, even though a high agreement was found for dimensional emotion between annotators, who differed in gender and age, commonalities like their cultural background may have coloured their observations. While it can be argued that it is the subjective experience of emotional expressions that drives emotion contagion, and therefore should not be ignored, it further highlights that the conclusions are preliminary.

Next, to empirically validate an agent-based model of emotion contagion, called DECADE, simulations were performed with conditions similar to the real experiment. Comparing the simulations to the real participants, we found that the agents reproduced many of the patterns that were found in the real participants. When emotion contagion was disabled in the model, the resemblance decreased substantially. However, while the relative differences among the conditions in the full model resembled the empirical study, in absolute sense there was a significant gap. The agents were emotionally more stable and similar to one another than the real participants. An explanation for this may be that the agents start each simulation emotionally neutral and only a single stimulus is considered per trial (winning or losing). In reality, the participants have diverse emotional histories and may have experienced additional internal or external stimuli, resulting a more intra- and interpersonal diversity in the emotional response.

Another aspect that may be worth exploring further is regulation due to social norms and ethical values. A form of emotion decay toward a neutral state is frequently considered in agent-based models of emotion contagion [6], including the proposed model. Additionally, we implemented a regulatory effect of social norms via the amplification-dampening tendency that was already part of the contagion mechanism of DECADE. In the proposed model, intense positive emotions are amplified more than weak positive emotions, while stronger negative emotions are dampened more than weaker negative emotions. Although we found that this simple mechanism to some degree captured the dynamics of the real participants, it is likely not suitable for many other scenarios. This mechanism for example ignores common wisdom that intense positive expressions are not always appropriate, nor does it distinguish between which types of positive emotion are appropriate. Instead, the applicable social norm with regard to emotions is known to depend strongly on individual traits and contextual features [55], [56]. To eventually develop agents that can function a broad range of social environments, future work may seek to combine emotion contagion with a more sophisticated regulation mechanism that includes social norms. A pioneering example is found in the work by Balint and Allbeck [57], who combine emotion contagion with emotion regulation via masking and substitution.

Further, there are several factors that were not examined in the present study, yet have been hypothesised in literature to influence the process of emotion contagion. These include the effects of cultural background and diversity [58], [59], group identity [60] and pre-existing social relationships [61]. We attempted to measure the latter factor for the present study using a digital form, that asked the participants how well they knew each of the other participants on four levels before the start of the experiment. However, the form design proved unclear, as some participants interpreted the question to ask for their estimate of the relationships between other participants, instead of their own relationship with others. Future work could use the proposed methodology to test assumptions of computational models of emotion contagion for the effect of emotional subgroups and crowd composition on spread of emotions [6].

In conclusion, validation of models of emotion contagion at the level of emotions remains a steep challenge, especially in large uncontrolled group settings. To make a start, this study demonstrated a novel methodology for empirical validation in a controlled setting in which small groups interacted via a video call. While the manual annotation process was manageable for the limited number of trials with small groups, it likely is too time-consuming for larger groups and data sets. Automated emotion recognition based on machine learning is an active field that in the future may provide a solution, yet to our knowledge there is currently no consensus on whether the state of the art in automated emotion recognition can accurately and robustly replace human annotation of emotions, especially in uncontrolled conditions [14]. The obtained data set and methodology in the present study provide an interesting opportunity for future study to test and compare emotion recognition models, as we hypothesise that while a real-world application, the semi-controlled conditions of the experiment are relatively favourable. This is discussed in more detail in appendix A8, where we also report a first exploration of the relation between automatically recognised facial motion using the OpenFace2 toolkit [62] and the human annotations of valence and arousal.

5 Acknowledgment

6 References

[1] G. A. van Kleef, M. W. Heerdink, and A. C. Homan, "Emotional influence in groups: the dynamic nexus of affect, cognition, and behavior," *Current Opinion in Psychology*, vol. 17. Elsevier B.V., pp. 156–161, Oct. 01, 2017. doi: 10.1016/j.copsyc.2017.07.017.

[2] A. Goldenberg and J. J. Gross, "Digital Emotion Contagion," *Trends Cogn Sci*, vol. 24, no. 4, pp. 316–328, Apr. 2020, doi: 10.1016/J.TICS.2020.01.009.

[3] The Guardian, "Astroworld: deaths of 10 people at Houston concert ruled accidental," Dec. 16, 2021. https://www.theguardian.com/music/2021/dec/16/astroworld-festival-deaths-ruled-accidental (accessed Aug. 24, 2022).

[4] Julian Borger, "American carnage: how Trump's mob ran riot in the Capitol," *The Guardian*, Jan. 07, 2021. https://www.theguardian.com/us-news/2021/jan/06/trump-capitol-american-carnage-washington (accessed Aug. 24, 2022).

[5] E. Hatfield, M. Carpenter, and R. L. Rapson, "Emotional contagion as a precursor to collective emotions," *Collective Emotions*, pp. 108–122, Apr. 2014, doi: 10.1093/ACPROF:OSO/9780199659180.003.0008.

[6] E. S. van Haeringen, C. Gerritsen, and K. V. Hindriks, "Emotion contagion in agent-based simulations of crowds: a systematic review," *Autonomous Agents and Multi-Agent Systems 2022 37:1*, vol. 37, no. 1, pp. 1–41, Nov. 2022, doi: 10.1007/S10458-022-09589-Z.

[7] Z. Tian, G. Zhang, C. Hu, D. Lu, and H. Liu, "Knowledge and emotion dual-driven method for crowd evacuation," *Knowl Based Syst*, vol. 208, 2020, doi: 10.1016/j.knosys.2020.106451.

[8] M. Xu *et al.*, "Emotion-Based Crowd Simulation Model Based on Physical Strength Consumption for Emergency Scenarios," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–15, 2020, doi: 10.1109/TITS.2020.3000607.

[9] K. Xie, Y. Mei, P. Gui, and Y. Liu, "Early-warning analysis of crowd stampede in metro station commercial area based on internet of things," *Multimed Tools Appl*, vol. 78, no. 21, pp. 30141–30157, Nov. 2019, doi: 10.1007/S11042-018-6982-5.

[10] H. Gayathri, P. M. Aparna, and A. Verma, "A review of studies on understanding crowd dynamics in the context of crowd safety in mass religious gatherings," *International Journal of Disaster Risk Reduction*, vol. 25, pp. 82–91, Oct. 2017, doi: 10.1016/J.IJDRR.2017.07.017.

[11] M. Volonte, Y.-C. Hsu, K.-Y. Liu, J. P. Mazer, S.-K. Wong, and S. V Babu, "Effects of Interacting with a Crowd of Emotional Virtual Humans on Users' Affective and Non-Verbal Behaviors," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020, pp. 293–302. doi: 10.1109/VR46266.2020.00049.

[12] Y. Mao, S. Yang, Z. Li, and Y. Li, "Personality trait and group emotion contagion based crowd simulation for emergency evacuation," *Multimed Tools Appl*, vol. 79, no. 5–6, pp. 3077–3104, May 2020, doi: 10.1007/s11042-018-6069-3.

[13] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, and L. S. Mattos, "Toward Emotion Recognition From Physiological Signals in the Wild: Approaching the Methodological Issues in Real-Life Data Collection," *Front Psychol*, vol. 11, p. 1111, Jul. 2020, doi: 10.3389/FPSYG.2020.01111/FULL.

[14] D. Kollias, "ABAW: Valence-Arousal Estimation, Expression Recognition, Action Unit Detection & Multi-Task Learning Challenges," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2022-June, pp. 2327–2335, 2022, doi: 10.1109/CVPRW56347.2022.00259.

[15] N. Aldunate and R. González-Ibáñez, "An integrated review of emoticons in computer-mediated communication," *Front Psychol*, vol. 7, no. JAN, p. 2061, Jan. 2017, doi: 10.3389/FPSYG.2016.02061/XML/NLM.

[16] H. Rosenbusch, A. M. Evans, and M. Zeelenberg, "Multilevel Emotion Transfer on YouTube: Disentangling the Effects of Emotional Contagion and Homophily on Video Audiences:," *https://doi.org/10.1177/1948550618820309*, vol. 10, no. 8, pp. 1028–1035, Jan. 2019, doi: 10.1177/1948550618820309.

[17] P. H. C. Mui, M. B. Goudbeek, C. Roex, W. Spierts, and M. G. J. Swerts, "Smile mimicry and emotional contagion in audio-visual computer-mediated communication," *Front Psychol*, vol. 9, no. NOV, p. 2077, Nov. 2018, doi: 10.3389/FPSYG.2018.02077/BIBTEX.

[18] C. T. Hsu, W. Sato, and S. Yoshikawa, "Enhanced emotional and motor responses to live versus videotaped dynamic facial expressions," *Scientific Reports 2020 10:1*, vol. 10, no. 1, pp. 1–11, Oct. 2020, doi: 10.1038/s41598-020-73826-2.

[19] U. Hess and S. Blairy, "Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy," *International Journal of Psychophysiology*, vol. 40, no. 2, pp. 129–141, Mar. 2001, doi: 10.1016/S0167-8760(00)00161-6.

[20] G. Isabella and V. A. Vieira, "The effect of facial expression on emotional contagion and product evaluation in print advertising," *RAUSP Management Journal*, vol. 55, no. 3, pp. 375–391, Sep. 2020, doi: 10.1108/RAUSP-03-2019-0038/FULL/PDF.

[21] U. Hess, M. Rychlowska, T. Storz, C. Blaison, A. Fischer, and E. G. Krumhuber, "Bridging the empathy gap: or not? Reactions to ingroup and outgroup facial expressions," *J Cult Cogn Sci*, vol. 6, no. 1, pp. 77–92, May 2022, doi: 10.1007/S41809-022-00097-W/FIGURES/2.

[22] J. Lange, M. W. Heerdink, and G. A. van Kleef, "Reading emotions, reading people: Emotion perception and inferences drawn from perceived emotions," *Curr Opin Psychol*, vol. 43, pp. 85–90, Feb. 2022, doi: 10.1016/J.COPSYC.2021.06.008.

[23] F. Prentice, H. Hobson, R. Spooner, and J. Murphy, "Gender differences in interoceptive accuracy and emotional ability: An explanation for incompatible findings," *Neurosci Biobehav Rev*, vol. 141, p. 104808, Oct. 2022, doi: 10.1016/J.NEUBIOREV.2022.104808.

[24] K. Manokara, A. Fischer, and D. Sauter, "Display rules differ between positive emotions: Not all that feels good looks good," *Emotion*, 2022, doi: 10.1037/EMO0001078.

[25] E. van Haeringen, C. Gerritsen, and K. Hindriks, "Integrating Valence and Arousal Within an Agent-Based Model of Emotion Contagion," in *Advances in Practical Applications of Agents, Multi-Agent Systems, and Social Good. The PAAMS Collection*, 2021, pp. 303–315. doi: 10.1007/978-3-030-85739-4_25.

[26] Zoom Video Communications, "Zoom," 2022. https://zoom.us/ (accessed Sep. 02, 2022).

[27] L. R. Goldberg, "The Structure of Phenotypic Personality Traits," *American Psychologist*, vol. 48, no. 1, pp. 26–34, 1993,

doi: 10.1037/0003-066X.48.1.26.

[28] E. S. van Haeringen, C. Gerritsen, and K. V. Hindriks, "Emotion contagion in agent-based simulations of crowds: a systematic review," Vrije Universiteit Amsterdam, technical report, 2022.

[29] D. Kollias *et al.*, "Deep Affect Prediction in-the-Wild: Aff-Wild Database and Challenge, Deep Architectures, and Beyond," *Int J Comput Vis*, vol. 127, no. 6–7, pp. 907–929, Jun. 2019, doi: 10.1007/S11263-019-01158-4/TABLES/16.

[30] A. F. Hayes and K. Krippendorff, "Answering the Call for a Standard Reliability Measure for Coding Data," *https://doi.org/10.1080/19312450709336664*, vol. 1, no. 1, pp. 77–89, Apr. 2007, doi: 10.1080/19312450709336664.

[31] S. Ojha, J. Vitale, and M.-A. Williams, "EEGS: A Transparent Model of Emotions," Nov. 2020, Accessed: Apr. 11, 2021. [Online]. Available: http://arxiv.org/abs/2011.02573

[32] R. McCrae and P. Costa, "The Structure of Interpersonal Traits: Wiggins's Circumplex and the Five-Factor Model," *J Pers Soc Psychol*, vol. 56, pp. 586–595, 1989, doi: 10.1037/0022-3514.56.4.586.

[33] F. Durupinar, U. Gudukbay, A. Aman, and N. I. Badler, "Psychological Parameters for Crowd Simulation: From Audiences to Mobs," *IEEE Trans Vis Comput Graph*, vol. 22, no. 9, pp. 2145–2159, Sep. 2016, doi: 10.1109/TVCG.2015.2501801.

[34] H. Jiang, Z. Deng, M. Xu, X. He, T. Mao, and Z. Wang, "An emotion evolution based model for collective behavior simulation," in *Proceedings - I3D 2018: ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, May 2018. doi: 10.1145/3190834.3190844.

[35] D. Jolliffe and D. P. Farrington, "Development and validation of the Basic Empathy Scale," *J Adolesc*, vol. 29, no. 4, pp. 589–611, Aug. 2006, doi: 10.1016/J.ADOLESCENCE.2005.08.010.

[36] A. Goldenberg, E. Weisz, T. Sweeny, M. Cikara, and J. Gross, "The Crowd Emotion Amplification Effect," *Psychol Sci*, vol. 31, no. 3, pp. 437–450, 2021, doi: 10.1177/0956797620970561.

[37] R. S. Lazarus, "Progress on a cognitive-motivational-relational theory of emotion," *Am Psychol*, vol. 46, no. 8, pp. 819–834, 1991, doi: 10.1037//0003-066X.46.8.819.

[38] J. Lambie and A. Marcel, "Consciousness and the Varieties of Emotion Experience: A Theoretical Framework," *Psychol Rev*, vol. 109, pp. 219–259, 2002, doi: 10.1037/0033-295X.109.2.219.

[39] C. G. McClintock and S. P. McNeel, "Prior dyadic experience and monetary reward as determinants of cooperative and competitive game behavior," *J Pers Soc Psychol*, vol. 5, no. 3, pp. 282–294, Mar. 1967, doi: 10.1037/H0024305.

[40] D. Liu, X. Li, and R. Santhanam, "Digital Games and Beyond: What Happens When Players Compete," *MIS Quarterly*, vol. 37, no. 1, pp. 111–124, Mar. 2013, Accessed: Mar. 15, 2023. [Online]. Available: https://misq.umn.edu/digital-games-and-beyond-what-happens-when-players-compete.html

[41] P. Vorderer, T. Hartmann, and C. Klimmt, "Explaining the Enjoyment of Playing Video Games: The Role of Competition," in *Proceedings of the Second International Conference on Entertainment Computing*, 2003, pp. 1–9.

[42] K. K. Imbir, J. Duda-Goławska, M. Pastwa, M. Jankowska, and J. Żygierewicz, "Event-Related Potential Correlates of Valence, Arousal, and Subjective Significance in Processing of an Emotional Stroop Task," *Front Hum Neurosci*, vol. 15, Feb. 2021, doi: 10.3389/FNHUM.2021.617861/FULL.

[43] D. Marengo, K. L. Davis, G. Ö. Gradwohl, and C. Montag, "A meta-analysis on individual differences in primary emotional systems and Big Five personality traits," *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–12, Apr. 2021, doi: 10.1038/s41598-021-84366-8.

[44] P. Steel, J. Schmidt, and J. Shultz, "Refining the relationship between personality and subjective well-being," *Psychol Bull*, vol. 134, no. 1, pp. 138–161, Jan. 2008, doi: 10.1037/0033-2909.134.1.138.

[45] E. Hudlicka, "Computational analytical framework for affective modeling: Towards guidelines for designing computational models of emotions," in *Handbook of Research on Synthesizing Human Emotion in Intelligent Systems and Robotics*, Hershey: IGI Global, 2014, pp. 1–62. doi: 10.4018/978-1-4666-7278-9.ch001.

[46] J. J. Gross and H. Jazaieri, "Emotion, emotion regulation, and psychopathology: an affective science perspective," *Clinical Psychological Science*, vol. 2, no. 4, pp. 387–401, Dec. 2014, doi: 10.1177/2167702614536164.

[47] D. J. Hughes, I. K. Kratsiotis, K. Niven, and D. Holman, "Personality traits and emotion regulation: A targeted review and recommendations," *Emotion*, vol. 20, no. 1, pp. 63–67, Feb. 2020, doi: 10.1037/EMO0000644.

[48] U. Barańczuk, "The five factor model of personality and emotion regulation: A meta-analysis," *Pers Individ Dif*, vol. 139, pp. 217–227, Mar. 2019, doi: 10.1016/J.PAID.2018.11.025.

[49] J. J. Gross and O. P. John, "Facets of emotional Expressivity: Three self-report factors and their correlates," *Pers Individ Dif*, vol. 19, no. 4, pp. 555–568, Oct. 1995, doi: 10.1016/0191-8869(95)00055-B.

[50] J. W. Tukey and others, *Exploratory data analysis*, vol. 2. Reading, MA, 1977.

[51] A. F. Zuur, E. N. Ieno, and C. S. Elphick, "A protocol for data exploration to avoid common statistical problems," *Methods Ecol Evol*, vol. 1, no. 1, pp. 3–14, Mar. 2010, doi: 10.1111/j.2041-210X.2009.00001.x.

[52] U. Knief and W. Forstmeier, "Violating the normality assumption may be the lesser of two evils," *Behav Res Methods*, vol. 53, no. 6, pp. 2576–2590, Dec. 2021, doi: 10.3758/S13428-021-01587-5/FIGURES/3.

[53] J. A. Russell, "A circumplex model of affect," *J Pers Soc Psychol*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980, doi: 10.1037/h0077714.

[54] C. Herrando and E. Constantinides, "Emotional Contagion: A Brief Overview and Future Directions," *Front Psychol*, vol. 12, p. 2881, Jul. 2021, doi: 10.3389/FPSYG.2021.712606/BIBTEX.

[55] D. R. Heise and C. Calhan, "Emotion Norms in Interpersonal Events," *Soc Psychol Q*, vol. 58, no. 4, p. 223, Dec. 1995, doi: 10.2307/2787125.

[56] C. M. Moran, J. M. Diefendorff, and G. J. Greguras, "Understanding emotional display rules at work and outside of work: The effects of country and gender," *Motivation and Emotion 2012 37:2*, vol. 37, no. 2, pp. 323–334, May 2012, doi: 10.1007/S11031-012-9301-X.

[57] T. Balint and J. M. Allbeck, "Is that how everyone really feels? Emotional contagion with masking for virtual crowds," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8637 LNAI, pp. 26–35. doi: 10.1007/978-3-319-09767-1_4.

[58] W. J. Chopik, E. O'Brien, and S. H. Konrath, "Differences in Empathic Concern and Perspective Taking Across 63 Countries:," *http://dx.doi.org/10.1177/0022022116673910*, vol. 48, no. 1, pp.

23–38, Oct. 2016, doi: 10.1177/0022022116673910.

[59]   B. Kuang, S. Peng, X. Xie, and P. Hu, "Universality vs. cultural specificity in the relations among emotional contagion, emotion regulation, and Mood State: An emotion process perspective," *Front Psychol*, vol. 10, no. FEB, p. 186, Feb. 2019, doi: 10.3389/FPSYG.2019.00186/XML/NLM.

[60]   J. van der Schalk *et al.*, "Convergent and divergent responses to emotional displays of ingroup and outgroup," *Emotion*, vol. 11, no. 2, pp. 286–298, Apr. 2011, doi: 10.1037/A0022582.

[61]   M. Kimura, I. Daibo, and M. Yogo, "The study of emotional contagion from the perspective of interpersonal relationships," *Soc Behav Pers*, vol. 36, no. 1, pp. 27–42, 2008, doi: 10.2224/SBP.2008.36.1.27.

[62]   T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. P. Morency, "OpenFace 2.0: Facial behavior analysis toolkit," *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, pp. 59–66, Jun. 2018, doi: 10.1109/FG.2018.00019.

Erik van Haeringen is a PhD student at the Computer Science department of the Vrije Universiteit Amsterdam. His current research focusses on modelling the spread of emotion and behaviour in human crowds. Previously, he completed a BSc and MSc degree in Ecology and Evolution, with a specific interest in social animal behaviour, biological modelling and complexity science.

Emmeke Veltmeijer obtained a BSc degree in Psychobiology, followed by a MSc degree in Artificial Intelligence at the University of Amsterdam in 2019. This sparked an interest in the interface between humans and technology. She is currently a PhD student in the Social AI Group at the Vrije Universiteit Amsterdam. In her research she focuses on affective computing in large groups.

Charlotte Gerritsen is associate professor within the Social AI group at Vrije Universiteit Amsterdam. She is recipient of a Vidi grant, a prestigious personal grant in the Netherlands. Her research is located at the intersection of Artificial Intelligence and the social sciences. She aims to enrich computational agent-based models with knowledge of human behaviour in order to make these models more realistic. These models are used to understand, predict and influence human behaviour in various domains.