DE GRUYTER

# Original Experimental

Eleonora Meister, Claudia Horn-Hofmann, Miriam Kunz, Eva G. Krumhuber and Stefan Lautenbacher\*

# Decoding of facial expressions of pain in avatars: does sex matter?

https://doi.org/10.1515/sjpain-2020-0078 Received May 22, 2020; accepted September 15, 2020; published online December 2, 2020

#### **Abstract**

**Objectives:** The decoding of facial expressions of pain plays a crucial role in pain diagnostic and clinical decision making. For decoding studies, it is necessary to present facial expressions of pain in a flexible and controllable fashion. Computer models (avatars) of human facial expressions of pain allow for systematically manipulating specific facial features. The aim of the present study was to investigate whether avatars can show realistic facial expressions of pain and how the sex of the avatars influence the decoding of pain by human observers.

**Methods:** For that purpose, 40 female (mean age: 23.9 years) and 40 male (mean age: 24.6 years) observers watched 80 short videos showing computer-generated avatars, who presented the five clusters of facial expressions of pain (four active and one stoic cluster) identified by Kunz and Lautenbacher (2014). After each clip, observers were asked to provide ratings for the intensity of pain the avatars seem to experience and the certainty of judgement, i.e. if the shown expression truly represents pain.

**Results:** Results show that three of the four active facial clusters were similarly accepted as valid expressions of pain by the observers whereas only one cluster ("raised eyebrows") was disregarded. The sex of the observed avatars influenced the decoding of pain as indicated by increased intensity and elevated certainty ratings for female avatars.

\*Corresponding author: Prof. Dr. Stefan, Lautenbacher Physiological Psychology, University of Bamberg, Markusplatz 3, 96047 Bamberg, Germany. Phone: +49 951 863 1851, Fax: +49 951 863 1976, E-mail: stefan.lautenbacher@uni-bamberg.de
Eleonora Meister and Claudia Horn-Hofmann, Physiological Psychology, University of Bamberg, Bamberg, Germany
Miriam Kunz, Department of Medical Psychology and Sociology, University of Augsburg, Augsburg, Germany
Eva G. Krumhuber, University College London, London, UK

**Conclusions:** The assumption of different valid facial expressions of pain could be corroborated in avatars, which contradicts the idea of only one uniform pain face. The observers' rating of the avatars' pain was influenced by the avatars' sex, which resembles known observer biases for humans. The use of avatars appeared to be a suitable method in research on the decoding of the facial expression of pain, mirroring closely the known forms of human facial expressions.

**Keywords:** avatar; decoding; pain; pain face; sex.

# Introduction

Facial expressions of pain can have great clinical relevance in pain diagnostics and pain management [1, 2]; however, they can only serve these functions when decoding by observers is largely correct. There are many influences on decoding, which make its study essential. In decoding studies, flexible and controllable stimulus material for the presentation of faces is necessary. The required systematic manipulation of specific features of facial expressions is limited even in trained actors, who have repeatedly been presented in decoding studies, whereas avatars may allow such experimental manipulations [3].

Under this perspective, the question arises how avatars have to be programmed to show convincing pain faces. The basic elements of facial activity (movements of individual facial muscles) can be sufficiently well described by the Facial Action Coding Systems (FACS [4]) as so-called Action Units (AUs). FACSGen is a computer program, using these AUs to create realistic avatars, so far mainly for studying the facial expression of emotions. The next question is which AUs have to be combined to realistically express pain. There seem to be certain Action Units that reliably occur during pain: 'furrowed brows' (AU4), 'narrowed eyes' (AU6, AU7), 'wrinkled nose' (AU9, AU10), 'opened mouth' (AU25, AU26, AU27) and sometimes 'eye closure' (AU43) [5, 6].

However, there is no universal pain face that includes all of these AUs, but these AUs are displayed in different combinations. As we could show with experimental pain [7], there are four distinct facial activity patterns of pain resulting from AU combinations: Cluster I ('narrowed eyes with furrowed brows and wrinkled nose'; AU6 7, AU4, AU9\_10), Cluster II ('opened mouth with narrowed eyes'; AU25 26 27, AU6 7), Cluster III ('raised eyebrows'; AU1 2) and Cluster IV ('furrowed brows with narrowed eyes'; AU4, AU6\_7). There also was a stoic cluster (V) with no visible facial expression.

Therefore, it is misleading to still present only one prototypical facial expression of pain when asking individuals in decoding studies to observe pain in others. However, it has not yet been investigated whether the five different clusters obtained by Kunz and Lautenbacher [7] are indeed all interpreted as pain by decoders. The present study should give first answers to this question.

Research about recognition of pain revealed that the sender's sex seems to modulate the observer's judgement [8-10], with varying patterns of sex differences across studies. To add answers to this question, the sex of the avatars was the individual feature varied in this study, with the exploratory hypothesis that female avatars showing the same intensity and combination of Action Units lead to higher ratings of pain by observers compared to male avatars.

In summary, we aimed at using dynamic avatars' faces to imitate the human expression of pain. The main motivation was the search for flexible and controllable stimulus material to study decoding of facial pain expression by human observers. In addition, we pursued two further objectives. (i) Since there is no universal expression of pain as shown in one of our earlier studies [7], several pain faces apparently exist and have to be imitated by avatars. The question is whether all of these pain faces help observers to judge pain in others. (ii) Sex of the observed other has a major influence on the perception of pain by the observer. Artificially created avatars promise strict experimental control of this feature and may promote the study of sex differences.

#### Methods

#### **Participants**

Participants were recruited using wall posters hung up in the Otto-Friedrich University of Bamberg, through email lists of students' councils and student groups in Facebook; they were excluded if they reported any psychological disease, chronic or current pain conditions, insufficient eyesight, or current prescription of analgesics,

tranquilizers, antidepressants or other centrally acting agents. Participants were instructed not to drink any alcohol on the day of testing.

The experimental procedure used in this study was approved by the ethics committee of the Otto-Friedrich University of Bamberg and all participants gave written informed consent. Participants either received monetary compensations (15€) or course credit1.

#### Experimental design and protocol

Our experiment was based on a within-subject design with two factors as we wanted to determine by repeated measurements (i) the effects of the various clusters of facial expression of pain and (ii) the effects of the sex of the avatar on the observers' judgements of pain. Dependent variables in this experiment were the ratings of pain intensity and certainty of pain in the observed avatar.

Sessions always took place in the same laboratory with the same experimenter under standardized conditions (light condition, seat position and distance to the computer screen). After oral and written instruction about the purpose and course of the experiment, the participants signed informed consent and filled in a demographic questionnaire.

The main experiment consisted of two computer tests, in which video clips of avatars were shown and the participants had to rate certain features of their facial expression. The avatars were presented with a height of 5.5 in. on a 14.0 in. color laptop monitor placed at a distance of 70 cm in front of the subjects.

In the first test, 80 videos of avatars displaying facial expressions of pain (16 individual avatars showing each the five different clusters of pain faces) were presented. The videos were presented in pseudorandom order in two blocks with 40 videos each (limitation to randomness guaranteed that sex and age of the avatars were equally distributed over the two blocks). Between these two blocks there was a break of 3 min to avoid a decline in vigilance. Participants rated the facial expression of pain by the avatars for intensity of pain and for certainty of pain being displayed on 11-point Likert-type scales either with verbal anchors from 0 - "no pain at all" to 10 - "extreme strong pain" or with verbal anchors from 0 - "uncertain" to 10 - "certain". Ratings were obtained after each video by mouse-click.

After the first test, the subjects had to fill in four questionnaires: Pain Anxiety Symptom Scale (PASS [11]), Pain Catastrophizing Scale (PCS [12]), Pain Vigilance and Awareness Ouestionnaire (PVAO [13]) and Saarbrueck Personality Questionnaire (SPF [14]) for the multidimensional assessment of empathy. The results from these questionnaires are not reported here.

In the second test, the subjects watched 16 videos (one video of the face of each individual avatar). However, this time they only saw videos with no emotional expression. Participants rated the faces for age (using a scale from 20 to 70 years, in 5-year steps), sex (using the categories "male"-"unsure"-"female"), similarity to oneself (using a 11-point Likert-type scale from 0 "not similar" to 10 "very similar") and attractiveness (using a 11-point Likert-type scale from 0 "not attractive" to 10 "very attractive"). The latter two ratings were not considered for the present analyses.

At the end, the participants got a short debriefing about the aims of the study.

<sup>1</sup> At the Otto-Friedrich University of Bamberg, students of the BSc Psychology have to collect a specific amount of course credits for having participated in psychological studies.

# **Materials**

#### Programs for producing the stimulus material

The faces of the avatars, that were used in this study, were modelled with the software FaceGen Modeller Core 3.14<sup>2</sup>, with which 2d presentations of 3d faces mirroring certain individual factors (e.g. sex, age) can be created. The sex of the avatar can be changed with a slider from "male" to "female". The avatars' age as well can be adjusted with a slider from 10 to 70 years. By default, the avatars are bald. They were left bald because hair style or hair colour might distract from the facial activity.

Short videos were used instead of static pictures because several studies have shown that dynamic expressions are perceived as more naturalistic and realistic compared to static faces [15, 16]. The videos were created with the software FACSGen<sup>3</sup>, which was developed on the basis of FaceGen and can be used to create dynamic facial expressions [16]. For that purpose, it orchestrates the Facial Action Coding System (FACS), which allows that every Action Unit can be animated separately, at levels of intensity from 0.0 to 1.0, and can be combined with every other Action Unit. Krumhuber and Tamarit [16] have demonstrated that this software produces emotionally valid and reliable facial expressions.

#### Stimulus material

For this study, 16 avatar faces were generated: four 20-year-old4 males, four 20-year-old females, four 60-year-old males, and four 60-year-old females. We chose the age of 20 and 60 years to guarantee a representative group of avatars with different ages, and not either only young or old avatars. The faces are shown in Figure 1.

With each avatar face, five videos were composed with a duration of 5 s. The aim of the design of the videos was to obtain a facial movement which is very close to the natural expressions of pain according to Kunz and Lautenbacher [7]. After a resting state of 1.5 s, unfolding of expression lasted for 1.0 s, full expression for 0.5 s, decline again for another 1.0 s, and another resting state for 1.0 s.

Figure 2 describes both the facial clusters identified Kunz and Lautenbacher [7] for comparison and how we tried to imitate these expressions by programming the corresponding AUs with intensities ranging from 0 to 1 in FACSGen. The animated five clusters were confirmed by Kunz and Lautenbacher as valid pain expressions, who are both FACS coders and the authors of the study of reference. Male and female avatars were animated equally, i.e. with identical facial activity patterns.

#### Data analyses

For data description, mean, standard deviation and frequency were computed.

As a manipulation check, the estimates of age of the avatars by the observers were subjected to a repeated-measurements ANOVA, with t-tests for post-hoc analyses. To verify the manipulation of the sex of the avatars, the distribution of the sex categorization by the observers was compared with a uniform distribution by use of a Pearson's chi-square test.

To assess whether observers' judgement of pain intensity and certainty was affected by (i) the five different clusters of facial activity and (ii) by sex of the avatar, a repeated measurement ANOVA with two within-subjects factors "cluster" (cluster I to cluster V) and "sex of avatar" (male vs female avatar) was computed. If sphericity could not be observed, we adjusted the degrees of freedom by a Greenhouse-Geisser correction. For F-tests, partial  $\eta^2$  is reported as an estimate of effect size. In case of significant F-tests, t-tests were conducted posthoc for closer examination.

It was examined whether intensity ratings and certainty ratings represent two different types of ratings or rather the same rating by use of a Pearson correlation analysis.

Significance was assumed at an alpha level ≤0.05. Data were analyzed using SPSS (version 20.0).

## Results

#### Participant characteristics

Eighty subjects (40 males and 40 females) aged from 19 to 35 years participated in this study (males' mean age: 24.63 years, range 19-35 years; females' mean age: 23.95 years, range 20-33 years). 57.5% of the participants had a university entrance qualification, 33.8% a university degree and 8.8% other qualifications. Nineteen females used hormonal contraceptives. Of the remaining 21 women, five were in the first, five in the second and 11 in the third third of their menstrual cycle. Nine of the 80 subjects were smokers.

#### Manipulation check

#### Sex of avatar

99.8% of the observers classified the male avatars as "male" and once as "unsure". For female avatars, 98.3% of the ratings were "female", while 10 times "unsure" (1.6%) and once "male". Thus, more than 99% of all ratings were congruent to the pre-set sex of the avatar. Not surprisingly, the distribution of the sex categorizations differed highly significantly from a uniform distribution, with  $\chi^2$ =2489.440, *df*=2, p<0.001.

#### Age of avatar

However, the avatars were not estimated to be 20 and 60 years old as they were pre-set in FaceGen Modeller Core

<sup>2</sup> FaceGen was founded 1998 by "Singular Inversions Inc.", a company which is specialized in creating realistic faces.

**<sup>3</sup>** FACSGen was developed by the Swiss Center for Affective Sciences at the University of Geneva [3].

<sup>4</sup> The age specifications refer to the slider settings.

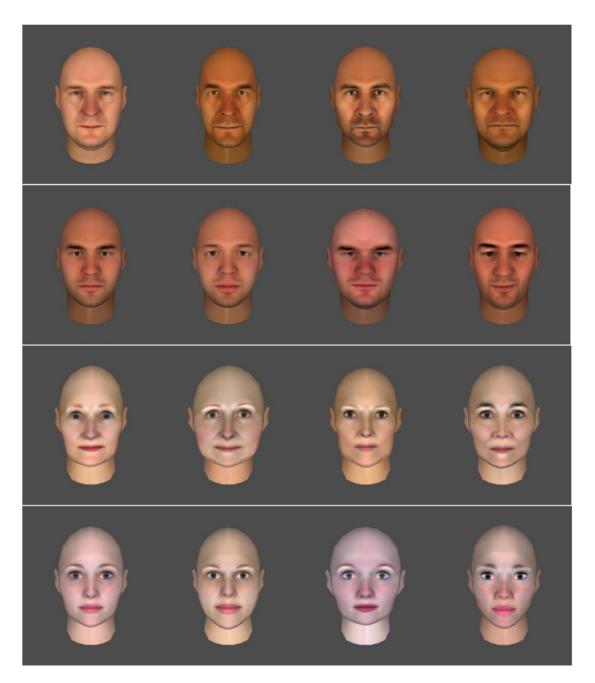


Figure 1: Avatar faces: 1st row: 60-year old men, 2nd row: 20-year old men, 3rd row: 60-year old women, 4th row: 20-year old women.

3.14. Their mean estimates were: M=42.75 years for old male avatars, M=33.98 years for young male avatars, M=47.81 years for old female avatars, and M=29.36 years for young female avatars. Although estimated and pre-set ages largely differed, analysis of variance showed a significant effect of pre-set age, F(1, 78)=0.456, p<0.001,  $\eta^2$ =0.916. Nevertheless, we refrained from further consideration of the age of the avatar in the present study as critical influence due to this distortion.

# Influence of the presentation of different facial clusters on observer judgements

#### **Intensity ratings**

Observers' pain intensity ratings significantly differed between clusters (F(2.21, 172.11)=236.794, p<0.001,  $\eta^2=0.752$ ). As can be seen in Figure 3 and Table 1, facial expressions of the clusters I, II and III were perceived to equally express moderate pain

#### Clusters based on Kunz & Lautenbacher 2014

# Cluster I: 'narrowed eyes with furrowed brows and wrinkled nose': Scoring highest on AU6\_7 with a slighter co-activation of AU4 and

slighter co-activation of AU4 and AU9\_10.

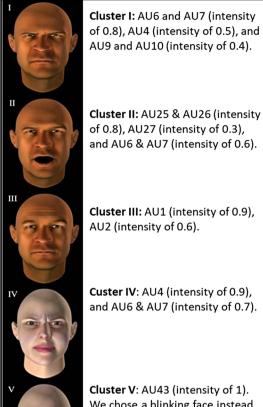
Cluster II: 'opened mouth with narrowed eyes':
Scoring highest on AU25\_26\_27 with a slighter co-activation of AU6\_7.

Cluster III: 'raised eyebrows': Scoring highest on AU1\_2.

Cluster IV: 'furrowed brows with narrowed eyes': Scoring highest on AU 4 with a coactivation of AU 6\_7.

Cluste V: `stoic:` No visible facial movements.

#### Clusters created by FACSGen



We chose a blinking face instead of a totally static one, so all faces are animated.

Figure 2: The five different clusters, shown by humans during experimental pain from Kunz & Lautenbacher (2014) and programmed for imitation in avatars (FACSGen).

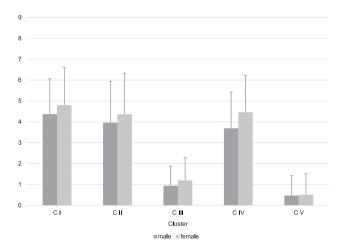


Figure 3: Mean intensity ratings (with SD) for male and female avatars in the five different clusters.

**Table 1:** Results of t-tests and Cohen's d, comparing the intensity ratings of the different clusters.

Comparison of clusters	T-test results	Cohen's d
Cluster I vs cluster II	T(79)=1.71, p=0.090	<i>d</i> =0.23
Cluster I vs cluster III	T(79)=22.60, p<0.001	d=2.34
Cluster I vs cluster IV	T(79)=5.59, p<0.001	d=0.30
Cluster I vs cluster V	T(79)=24.37, p<0.001	d=2.80
Cluster II vs cluster III	T(79)=15.31, p<0.001	d=1.91
Cluster II vs cluster IV	T(79)=0.37, p=0.712	d=0.05
Cluster II vs cluster V	T(79)=17.60, p<0.001	d=2.31
Cluster III vs cluster IV	T(79) = -19.92, p < 0.001	d=-1.98
Cluster III vs cluster V	T(79)=8.15, p<0.001	d=0.60
Cluster IV vs cluster V	T(79)=21.34, p<0.001	d=2.45

intensities, whereas clusters III and V yielded in contrast very low pain intensity scores.

Considering in addition the descriptive statistics and effect sizes (d) for the differences between clusters, a clear pattern of results emerged. In the avatars, pain – regarding its intensity - was substantially seen when the avatars expressed it by means of clusters I, II or IV, whereas clusters III and V were no credible expressions of pain.

## **Certainty ratings**

The results for the ANOVA conducted on the mean scores of the certainty ratings were quite similar, with a strong main effect of the different clusters: (F(2.47, 192.47)=179.016,p<0.001,  $\eta^2=0.697$ ).

Considering again also the descriptive statistics and effect sizes (d) for the differences between clusters (see Figure 4 and Table 2), the pattern of results resembled that obtained for pain intensity. Pain was certainly seen in the avatars by the observers when the avatars expressed it by means of clusters I, II or IV, whereas clusters III and V did not lead to much certainty that pain was expressed.

# **Correlation analyses**

The great similarity in the findings for pain intensity and pain certainty ratings casts doubt on the independency of these two judgements. Therefore, we conducted a correlation analysis to explore the relationship between these two dimensions. The 25 inter-dimensional correlations across clusters relating pain intensity and pain certainty ranged from 0.07 to 0.78, with only 64% of the correlations

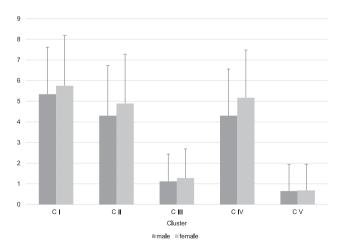


Figure 4: Mean certainty ratings (with SD) for male and female avatars in the five different clusters.

Table 2: Results of t-tests and Cohen's d, comparing the different certainty ratings of the different clusters.

Comparison of clusters	T-test results	Cohen's d
Cluster I vs cluster II	T(79)=3.21, p=0.002	d=0.41
Cluster I vs cluster III	T(79)=18.17, p<0.001	d=2.22
Cluster I vs cluster IV	T(79)=5.94, p<0.001	d=0.36
Cluster I vs cluster V	T(79)=18.75, p<0.001	d=2.58
Cluster II vs cluster III	T(79)=13.05, p<0.001	d=1.72
Cluster II vs cluster IV	T(79) = -0.48, p = 0.633	d=-0.06
Cluster II vs cluster V	T(79)=14.71, p<0.001	d=2.03
Cluster III vs cluster IV	T(79) = -16.21, p < 0.001	d=-1.84
Cluster III vs cluster V	T(79)=5.78, p<0.001	d=0.42
Cluster IV vs cluster V	T(79)=16.70, p<0.001	d=2.20

being significant. This indicates that pain intensity and pain certainty ratings are strongly related, but far from identical.

# Influence of sex of the avatars on observer judgement

The ANOVA conducted on the mean scores of the ratings of intensity of pain further yielded a significant main effect of the sex of the avatar, F(1, 78)=83.227, p<0.001,  $\eta^2$ =0.516, with higher intensity ratings for female avatars, T(0.465)=9.517, p<0.001, d=0.366 (see Figure 3). Similarly, for the ratings of certainty of judgement also a significant main effect of the sex of the avatar emerged, F(1, 78)=54.133, p<0.001,  $\eta^2=0.410$ . Just as for intensity ratings, the mean ratings of certainty of judgement were higher for female avatars, T(0.507)=7.609, p<0.001, d=0.309 (see Figure 4).

These effects of sex varied between clusters because we detected significant interaction between clusters and avatar's sex for the intensity rating (F(3.54, 276.05)=12.728,p<0.001,  $\eta^2$ =0.140) and the certainty ratings (F(3.56, 277.64)=11.086, p<0.001,  $\eta^2$ =0.124). Effect size comparisons showed that the biggest difference in the estimation of the intensity of pain as well as of the certainty of pain between the male and female avatars occurred in cluster IV and the smallest in cluster V (see Table 3).

# Discussion

In the present study, we started the usage of dynamic avatars' faces to imitate the human expression of pain over a few seconds. The background for this endeavor was the

Table 3: Results of t-tests and Cohen's d for the effects of the sex of the avatar on pain intensity and pain certainty ratings.

Cluster comparison	T-test results Pain intensity	Cohen's d	T-test results Pain certainty	Cohen's d
Cluster I: ♂ vs ♀	p<0.001	d=0.25	p<0.001	d=0.17
Cluster II: ♂ vs ♀	p<0.001	<i>d</i> =0.20	p<0.001	d=0.24
Cluster III: ♂ vs ♀	p<0.001	d=0.23	p=.049	d=0.12
Cluster IV: ♂ vs ♀	p<0.001	<i>d</i> =0.44	p<0.001	d=0.38
Cluster V: ♂ vs ♀	p=0.551	<i>d</i> =0.04	p=0.663	<i>d</i> =0.02

search for flexible and controllable stimulus material representing facial pain expression to study decoding by human observers. We mainly pursued two objectives. (i) Since there is no universal expression of pain as shown in one of our earlier studies [7], several pain faces apparently exist and have to be imitated by avatars. The question has arisen whether all of these pain faces help observers to estimate the intensity of pain in others and let the observers feel certain to see pain in others. (ii) Sex of the observed other has a major influence on the perception of pain by the observer. It is difficult to experimentally model the sex of the observed human subjects because samples differing in sex mainly differ also in other sex-independent personal characteristics. Artificially created avatars promise stricter control in this respect. Our findings regarding the two objectives will be discussed in this order.

# Usage of different facial expression patterns of pain

For the present study, we modeled five different types of facial expressions in our avatars, which should resemble as close as possible the five empirically determined clusters of facial expression of pain [7]. We tested how the different clusters of facial expressions were decoded. The estimated intensity of pain and certainty of judgement in the avatars were perceived differently for the five clusters. Largely, there were two levels of estimated pain intensity and certainty. Confronted with cluster I ('narrowed eyes with furrowed brows and wrinkled nose', AU6\_7, AU4, AU9\_10), II ('opened mouth with narrowed eyes', AU25\_26\_27, AU6\_7) and IV ('furrowed brows with narrowed eyes', AU4, AU6\_7), our observers perceived almost equally moderate levels of pain and appeared to be quite certain about the existence of pain in the observed avatars. In contrast, cluster III ('raised eyebrows', AU1\_2) and V (stoic cluster) did not let the observers perceive pain in the avatars because both the pain intensity and

certainty ratings were rather low. For cluster V as the stoic cluster with no visible signs of pain, it is easily understandable that it was interpreted as "no pain at all". Cluster III, however, was probably mainly decoded as expression of another internal state, which is associated to a greater degree with signs of startle than with pain [17]. The resulting interpretation errors may not be too serious in consequence as not many pain recipients in our earlier study have appeared to use cluster III for pain expression [7]. Coming back to the positive findings, our avatars could signal pain equally well with three different types of expression, which overlapped in AU 6 7 ('narrowed eyes') and varied in AU 4 ('furrowed brows'), AU 9\_10 ('wrinkled nose') and AU 25\_26\_27 ('opened mouth'). Thus, our decoding study corroborated the findings from the encoding studies, which suggested that AU 6\_7 is the key element of human facial pain language with other frequent facial responses combined leading to the variations in the pain face [5]. These findings produce important implications for future developments of trainings for medical staff [18], which are (a) that trainings should include different facial expressions of pain, (b) that avatars can help to exemplify these variations, and (c) that atypical or no visible facial signs of pain are not an absolute guarantee of facing a pain-free person.

#### Sex of the observed avatars

In this study, we could show that the avatars' sex has an influence on the observers' judgements of the intensity and certainty of pain in facial expressions as both ratings were higher for female avatars. This finding is in accord with that of Simon and colleagues [10] obtained in actors and agrees with the result of Hadjistavropoulos et al. [19] and Hirsh et al. [20], that male patients are perceived as experiencing less pain intensity than female patients. However, we used avatars, which allowed for very strict control of the sex-independent facial appearance, methodologically strengthening the validity of findings.

We consider two possible explanations of the higher rating scores of pain intensity and certainty for female senders. First, the tendency of women to encode emotions clearly visible [21] may lead to the perceptual habit to see more emotions in the face of women. Second, common knowledge or stereotypes about sex differences in pain perception, i.e. that women generally show an increased pain sensitivity compared to men, may lead to biased judgements [22]. Consequently, observers may preperceptually adopt the judgement algorithm to see more pain in the faces of women than in those of men. For note, not all decoding studies looking for sex differences in perception of pain in others have revealed this tendency to see more pain in the face of women (e.g. [9]).

# **Advantages of avatars**

The method of our study is innovative because of the firsttime usage of different dynamic facial expressions of pain by computer-generated avatars. The advantage of this man-made stimulus material is the high experimental controllability and flexibility due to the possibility to precisely model Action Units while simulating different individual factors. We know of only three other studies that conducted a pain experiment presenting computergenerated avatars. Hirsh et al. [20] and Wandner et al. [23] used virtual human technology to create virtual characters including short videos of computer-generated avatars and brief vignettes for each of them. Riva and his colleagues [9] used static pictures as well as dynamic sequences of avatar faces created with FaceGen that changed progressively from a neutral face to a maximum of the studied emotions (anger, disgust, and pain) to determine the identification threshold for the given states. Our aim was different because we assumed the idea of a universal pain face to be an unrealistic simplification and tried to compare different empirically derived pain faces [7] for their functional equivalence in observers. In many previous decoding studies [9, 10, 20, 23], mainly one pain face (assumed to be prototypical) was used. This prototypical pain face was very similar to our cluster I solution, which we found however to be the habitual pain expression in only one third of the people [7].

The analyses run by Riva et al. [9] and by our team in the present study were easily accomplished because face expressions in avatars can be flexibly manipulated and systematically varied in contrast to those of real people. Certain Action Units are not the strength of certain actors or confined to certain subjects or patients, but all our avatars were equally good in producing these key elements of facial activity during pain.

#### Limitations

There are also some limitations that should be mentioned. Participants had to rate the expression as pain and there was no choice to rate the shown expression as something else. We excluded these options in the present study because the focus was the question of equivalence of the different faces as pain expression. It is possible that some clusters would have been judged as alternative emotion because some participants reported after the experiment that they would have interpreted some avatars' expression e.g. as disgust or anger. Also, non-expressive behavior like vawning was mentioned as a possible interpretation of the perceived facial activity.

A certain degree of artificiality of the stimulus material may have limited the possibility to generalize our findings to real human faces. For example, the baldness of the avatars, which should avoid distraction by emotionally acting hair style, could be such an artificial feature because at least in Western cultures only a few people (and mostly men) are bald. Also, the participation of only young people as observers may have lowered the external validity and should lead to wider age range in replication studies. Finally, the avatars showed the racial characteristics of Caucasian people, which limited the possibility to generalize our findings to other ethnic groups. So far, Non-Caucasian people have largely been neglected in the studies on facial expressions of pain.

# Conclusion

The study of decoding – to better understand non-verbal pain communication – can benefit from the use of avatars as 'pain recipients' and senders of facial pain messages. They allow to systematically vary the facial pain signals by programming combinations of Action Units, which are known to form the common pain faces. The facial pain signals can thus be investigated by experimental variations without necessarily changing other facial features of the individual, which even goes beyond the alternatives human actors may offer. We could demonstrate that three of the four empirically derived active pain faces shown by our avatars were accepted by the observers in the present study as pain faces AU 6\_7 ('narrowed eyes') combined either with AU 4 ('furrowed brows'), AU 9\_10 ('wrinkled nose') or AU 25\_26\_27 ('opened mouth'). Only, AU1\_2 ('raised eyebrows') was not rated as resembling an expression of pain. Female avatars were consistently - with some minor variations between types of pain faces - rated as expressing

more pain than men. In future studies, the expression of other emotions like anxiety, anger and disgust should be added and compared with the expression of pain.

**Research funding:** None declared.

**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Competing interests:** Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study.

**Ethical approval:** The research complies with all the relevant national regulations and institutional policies, was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board (ethics committee of the Otto-Friedrich University of Bamberg).

#### References

- 1. Hadjistavropoulos T, Craig KD, Duck S, Cano A, Goubert L, Jackson PL, et al. A biopsychosocial formulation of pain communication. Psychol Bull 2011;137:910-39.
- 2. Kunz M, Scharmann S, Hemmeter U, Schepelmann K, Lautenbacher S. The facial expression of pain in patients with dementia. Pain 2007;133:221-8.
- 3. Roesch EB, Tamarit L, Reveret L, Grandjean D, Sander D, Scherer KR. FACSGen: a tool to synthesize emotional facial expressions through systematic manipulation of facial action units. J Nonverbal Behav 2011;35:1-16.
- 4. Ekman PA, Friesen WV. Facial action coding system. Palo Alto: Consulting Psychologists Press; 1978.
- 5. Kunz M, Meixner D, Lautenbacher S. Facial muscle movements encoding pain-a systematic review. Pain 2019;160:535-49.
- 6. Prkachin KM. The consistency of facial expressions of pain: a comparison across modalities. Pain 1992;51:297-306.
- 7. Kunz M, Lautenbacher S. The faces of pain: a cluster analysis of individual differences in facial activity patterns of pain. Eur J Pain 2014:18:813-23.
- 8. Kret ME, De Gelder B. A review on sex differences in processing emotional signals. Neuropsychologia 2012;50: 1211-21.
- 9. Riva P, Sacchi S, Montali L, Frigerio A. Gender effects in pain detection: speed and accuracy in decoding female and male pain expressions. Eur J Pain 2011;15:21435925.

- 10. Simon D, Craig KD, Gosselin F, Belin P, Rainville P. Recognition and discrimination of prototypical dynamic expressions of pain and emotions. Pain 2008;135:55-64.
- 11. Walter B, Hampe D, Wild J, Vaitl D. Die Erfassung der Angst vor Schmerzen: eine modifizierte deutsche Version der Pain Anxiety Symptoms Scale (PASS-D) [Assessment of anxiety of pain: a modified German version of the Pain and Anxiety Symptom Scale (PASS-D)]. Schmerz 2002;16:83.
- 12. Meyer K, Sprott H, Mannion AF. Cross-cultural adaptation, reliability and validity of the German version of the pain catastrophizing scale. J Psychosom Res 2008;64:469-78.
- 13. Kunz M, Capito ES, Horn-Hofmann C, Baum C, Scheel J, Karmann AJ, et al. Psychometric properties of the German version of the Pain Vigilance and Awareness Questionnaire (PVAQ) in pain-free samples and samples with acute and chronic pain. Int J Behav Med 2017:24:260-71.
- 14. Paulus C. Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie. Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index [The Saarbrueck Personality Questionnaire on Empathy: psychometric evaluation of the German version of the Interpersonal Reactivity Index]. PsyDoc 2009. Available from: http://hdl.handle.net/20. 500.11780/3343 [Accessed 24 Oct 2020].
- 15. Biele C, Grabowska A. Sex differences in perception of emotion intensity in dynamic and static facial expressions. Exp Brain Res 2006;171:1-6.
- 16. Krumhuber EG, Tamarit L. FACSGen 2.0 animation software: generating 3D FACS-valid facial expressions for emotion. Emotion 2012:12:351-63.
- 17. Ekman P, Friesen WV, Simons RC. Is the startle reaction an emotion? J Pers Soc Psychol 1985;49:1416-26.
- Elcigil A, Maltepe H, Esrefgil G, Mutafoglu K. Nurses' perceived barriers to assessment and management of pain in a university hospital. J Pediatr Hematol Oncol 2011;33:33-8.
- 19. Hadjistavropoulos T, McMurtry B, Craig KD. Beautiful faces in pain: biases and accuracy in the perception of pain. Psychol Health 1996;11:411-20.
- 20. Hirsh AT, Alqudah AF, Stutts LA, Robinson ME. Virtual human technology: capturing sex, race, and age influences in individual pain decision policies. Pain 2008;140:231-8.
- 21. Schwartz GE, Brown S-L, Ahern GL. Facial muscle patterning and subjective experience during affective imagery: sex differences. Psychophysiology 1980;17:75-82.
- 22. Plant EA, Kling KC, Smith GL. The influence of gender and social role on the interpretation of facial expressions. Sex Roles 2004;51: 187-96.
- 23. Wandner LD, Stutts LA, Alqudah AF, Craggs JG, Scipio CD, Hirsh AT, et al. Virtual human technology: patient demographics and healthcare training factors in pain observation and treatment recommendations. J Pain Res 2010;3:241-7.