Abstract

Objective. Patients with dementia, whose ability to provide self-report of pain is often impaired, are in crucial need of observers who can detect and judge the patients’ pain-indicative behaviors appropriately, in order to initiate treatment. The facial display of pain promises to be especially informative for that purpose. The major aim of the study was to investigate, whether facial pain displays of patients with dementia can be as easily interpreted as facial displays of individuals without cognitive impairment and whether nurses have learned—through their professional training and work experience—to better read the facial display of pain compared with a control group.

Design, Subjects and Outcome Measures. For that purpose, 21 nurses and 21 age-matched controls watched 120 video clips, showing facial expressions of young and old individuals with and without dementia during slight and moderate noxious stimulation. After each clip, subjects were asked to rate how much pain the observed individual might have experienced.

Results. The observer ratings did not differ between nurses and controls as regards pain level and concordance. On the other hand, level and concordance of the pain judgments were strongly influenced by the cognitive status, age, and gender of the observed individuals, with higher and more accurate pain judgments for older and for female individuals.

Conclusion. Without further contextual information, “professional” observers do not show a superior competence in inferring pain in others by reading their facial display. Therefore, additional training seems needed to reliably prevent that pain goes unnoticed in patients with dementia.

Key Words. Pain; Dementia; Decoding; Facial Expressions; Observer

Introduction

There is ample evidence that patients with cognitive impairment, mainly dementia, do not report pain spontaneously as often as unimpaired individuals do [1,2]. In accordance, authors of several studies have repeatedly reported of less frequent prescription of analgesics in patients with dementia [3,4]. As recent studies indicate, this reduced pain report and less frequent prescription of analgesics in patients with dementia is not likely due to a decrease in painful conditions but instead is due to failure of verbal communication [5]. As the verbal report of pain is the gold standard, which guides also the management of pain in patients with dementia, there is an urgent need for alternative ways of pain assessment when the verbal report starts to fail. The facial display of pain promises to be the most informative nonverbal form of communication, which might help an observer to infer whether an individual suffers from pain [6].

The question arises how successful observers are in identifying and quantifying pain in others when relying solely on the facial display. Findings have been a bit controversial. There is evidence that observers are well capable of distinguishing pain from other affective states and can even differentiate between different pain intensities [7]. Other findings are less optimistic as they suggest that untrained observers do not perform well. However, even a short training in recognizing the prototypical facial display of pain improved the decoding...
ability of these observers [8–10]. Surprisingly, findings also show that professionals with long-standing experiences with patients in pain nevertheless showed a limited decoding ability, given that they tend to underestimate pain in others more than nonprofessionals [11–14]. An explanation might be that the continuous confrontation with intense pain in others leads to an adaptation-level effect, with the tendency to reduce succeeding judgments [15,16].

All these findings have been obtained when judging the facial expression in individuals without cognitive impairment, and even though studying the ability of observers to decode facial expressions of pain in individuals without cognitive impairment is of great interest, it might be of even greater clinical relevance to investigate the ability of observers to decode facial pain displays in individuals who are limited in their ability to communicate verbally, like patients with dementia. Can pain be as easily inferred from the facial expression in patients with dementia as in individuals without cognitive impairment? Studies from our laboratory suggest that the facial display of pain is preserved in its specificity in patients with mild to moderate degrees of dementia but amplified in its intensity compared with age-matched control subjects without cognitive impairment [17,18]. This conclusion was based on fine-grained analyses (using the Facial Action Coding System [FACS] [19]), which allow for determining single facial actions and their combinations but do not interpret the facial “gestalt” [19]. Nevertheless, this fine-grained analysis suggests that an observer should receive all the necessary facial signals needed to decode the pain-related messages in patients with dementia. Evidence in line with this suggestion also stems from a recent study which found that caregivers (who spend 10 or more hours with a patient with dementia) are able to correctly infer pain based on nonverbal pain signals (composite score of different pain behaviors) in patients with dementia [20].

With the present study, we now want to investigate whether the competence of observers to decode pain from the facial display of patients with dementia is indeed comparable to the decoding competence for young and older individuals without cognitive impairment. As stimulus material, we used videotapes of patients and control subjects from our former facial encoding studies [17,18]. Based on previous findings, we expected that the correlations between self-report ratings and the observer ratings of pain are above chance level, but that the strength of the correlation coefficients will vary depending on the characteristics of the individual being observed (young vs old; older adults with cognitive impairment vs old adults without cognitive impairment; women vs men). For example, if the observed individual is female, the concordance between observer ratings and self-report of pain is likely higher than in males, given that women tend to show better fine grade in facial pain expressions [21] and given that observers have been found to rate pain higher in females compared with males [22]. Moreover, concordance between observer ratings and self-report of pain might also vary depending on the age of the observed individual, as it has been shown that observers see more pain in facial expressions of older compared with younger patients [23]. With regard to our main question of whether correlation coefficients will also vary depending on the cognitive status of the individual in pain, this has yet to be answered.

Next to the characteristics of the observed individual, we also raised the question of whether working as a nurse or a caregiver for years and thus, facing daily the challenge of validly detecting pain and initiate treatment affects the competence of decoding facial expressions of pain. Based on the literature [11–14], we assumed that professional experts might underrate the level of pain in others, and that they do not show higher concordance.

## Methods

### Subjects

The subjects were 21 “professionals” (P) (hospital nurses and nurses for older adults) and 21 “nonprofessionals” (NP) (no paramedical professions, mainly secretaries and administrative officers) of the same sex (16 women each) and comparable age (P: 39.5 years ± 8.4; NP: 39.2 years ± 9.9; P > 0.05). The professionals had to have work experiences either for a minimum of 3 years as a nurse for older adults or of a minimum of 5 years as a hospital nurse including positions at geriatric wards (for further details regarding the work experience of the nurses, please see Table 1). In order to select a representative sample of nurses who work with older individuals with and without pain, nurses were recruited from a large local hospital (Bezirkskrankenhaus Obermain) (N = 16) with different specialized units as well as from local nursing

### Table 1 Descriptive data on study nurses regarding their work experiences

<table>
<thead>
<tr>
<th>Work Experience as a Nurse</th>
<th>Mean</th>
<th>SD</th>
<th>Area of Medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of work</td>
<td>15.5</td>
<td>8.0</td>
<td>Gerontopsychiatry (19%)</td>
</tr>
<tr>
<td>Percentage of patients suffering from pain</td>
<td>57.8%</td>
<td>24.4</td>
<td>Geriatrics (24%)</td>
</tr>
<tr>
<td>Percentage of patients with dementia</td>
<td>37.3%</td>
<td>29.2</td>
<td>Orthopedics (57%)</td>
</tr>
</tbody>
</table>

SD = standard deviation.
Decoding Facial Pain Expressions of Patients with Dementia

homes (N = 5). Exclusion criteria for both groups were acute or chronic pain, mental disorders in the last 10 years, somatic diseases with likely affection of the pain system, self-reported impaired vision, disorder of attention, and prosopagnosia. These criteria were recorded by the use of an anamnesis questionnaire. The study protocol was approved by the ethics committee of the Otto-Friedrich University of Bamberg. All participants gave written informed consent. All individuals were paid for participation (20 €).

Material

The video segments, which were presented to the observing subjects of the present study, were recorded in earlier studies on the facial expressions of pain (for more detail, see description in Kunz et al. [17,24]). Three groups of individuals were shown in these videos, namely young individuals (20–38 years), older individuals without cognitive impairment (above 65 years), and older individuals with mild to moderate forms of dementia (above 65 years [mini mental state examination [MMSE] score of 17.0, standard deviation: 4.9]). Patients with dementia were tested with the MMSE [25], and in addition, they were examined by a neurologist or psychiatrist and diagnosed according to the criteria specified by the International Classification of Disease, tenth revision, NINCDS-ADRDA [26], and NINDS-AIREN [27]. According to these diagnostic guidelines, analyses of cerebrospinal fluid, blood chemistry analysis, electroencephalogram, computed tomography or magnetic resonance imaging, and electrocardiogram analyses were conducted besides clinical assessment.

The face of the individuals was videotaped while they received phasic pressure stimulation induced by a Fischer pressure algometer. All subjects from the earlier studies had agreed on a later presentation of these videotapes. The videos showed two stimulation conditions: one in which subjects received a slightly painful stimuli (2 kg) and one in which they received a moderately to strongly painful stimuli (5 kg). The pressure increased steadily at an application rate of 1 kg per second until the target intensity was reached and then continued at that level for another 5 seconds. Subjects were shown the facial expression that occurred during these 5 seconds at the target intensity level (either 2 kg or 5 kg). For each of the three groups (young, old without dementia, old with dementia), the video material of 20 individuals (20 young; 10 old) was randomly selected out of the much greater samples studied (see Kunz et al. [17,24]). Given that each individual received two stimulation intensities (2 kg and 5 kg), we presented 120 video segments (2 intensities × (20 young + 20 old + 20 patients with dementia)) to each observer in the present study in a randomized order.

Only frontal views without other facial displays, e.g., due to vocalization or yawning were used. Moreover, all 5-second video clips started off with a “neutral expression.” Another inclusion criterion of a video segment was the availability of a concurrent rating on a 6-point verbal rating scale (VRS) (“no pain,” “slight pain,” “moderate pain,” “strong pain,” “very strong pain,” “unbearably strong pain”) of the observed subject for later comparison with the rating of the observers. Due to the cognitive impairment in some of the videotaped individuals, we used this relatively simple VRS to assess the self-report of pain (whereas a numerical rating scale [NRS] was used by the observers to rate the pain of the videotaped individuals [see more details later]). Moreover, only videos of those patients with dementia were selected who were still able to provide self-report of pain. All video clips were also analyzed using the FACS [19], and these findings have been published elsewhere [17].

Experimental Protocol

The presentation of the video segments and the assessment of the observer ratings were made possible by the use of a laptop (screen width of 15.4 inches). Testing took place in a quiet room either in our laboratory at the University of Bamberg (N = 10) or in a clinical center (Klinikum Kutzenberg; N = 32). Light conditions could be adjusted to guarantee good visibility.

The sessions lasted for approximately 90 minutes and were divided into three blocks. In each block subjects watched and rated 40 video segments (120 video segments in total). There were 5-minute breaks between blocks to allow for a short recreation period.

Subjects were told that the individuals in the videos were recorded while they were experiencing different levels of pain intensities as well as during nonpainful sensations. After each video, subjects were asked to complete three 10-point NRS. First, a pain scale was presented with the verbal anchors “no pain”—“extremely strong pain” and the subjects were instructed to rate the intensity of the pain, which the person in the video appears to experience, on the basis of the observable facial display. Following the pain scale, subjects also had to rate the perceived similarity of the observed individual with themselves (“not similar”—“very similar”) and the perceived attractiveness of the person in the video (“not attractive”—“very attractive”); however, these two additional scales were not considered any further in the present report. Each scale was presented for 7.5 seconds, and the subjects rated by mouse-click.

Statistics

To investigate the competence of the two observer groups to decode the facial expressions of pain, we used two parameters for further computations:

1. **Level**: On the one hand, we were interested on how much pain the observers saw in the facial expressions presented to them. Therefore, we simply used the NRS pain ratings as provided by the observers.

2. **Concordance**: On the other hand, we were interested in the agreement between observers’ pain judgments and the self-report ratings as provided by the
videotaped individuals. Therefore, we calculated for each observer the Pearson correlation between his/her own pain ratings (NRS) and the self-report pain ratings of those individuals having been observed (VRS). These individual correlations, which were Fisher’s z-transformed for better comparisons between the subject groups, are based on up to 120 pairs of ratings (given the 120 video segments) and thus, can be considered to be robust parameters.

These two parameters were subjected to the same type of variance analysis with one between-subject factor group (professional vs nonprofessionals) and three within-subject factors, namely age and cognitive status (young vs old; adults without cognitive impairment vs old patients with dementia), gender (women vs men), and stimulus intensity (2 kg vs 5 kg pressure). The within-subject factors all refer to variations in the subjects and experimental conditions shown in the video segments; only the between-subject factor reflects variations in the observer. Only in case of significance, single comparisons (t-tests) were computed for post hoc testing.

As the correlations between observer ratings (NRS) and self-report ratings (VRS) were supposed to reflect the competence of the observer of matching the self-report of the observed subject (concordance), these correlations should be positive and significantly greater than zero if such a competence can be assumed to exist at all. For that purpose, one-sample t-tests were computed to test whether the correlations were significantly greater than zero.

The α-level was set to 0.05 throughout, and analyses were conducted using SPSS 20 (IBM, SPSS, Chicago, IL, USA).

**Results**

**Level**

Descriptive statistics of the NRS ratings for the observed intensity of pain are given in Table 2 separately for the nurses (professionals) and the control subjects (nonprofessionals). Analysis of variance showed a nonsignificant result for the between-subject factor “group” (F<sub>1,40</sub> = 0.294, P = 0.591); thus, professionals did not differ from nonprofessionals as regards the level of perceived pain in others. There was, however, a significant interaction between the between-subject factor “group” and the within-subject factor “gender” (F<sub>1,40</sub> = 5.760, P = 0.021). As can be seen in Table 2, professionals tended to rate the pain intensity in women lower than the nonprofessionals did. However, the post hoc t-test comparing the two groups of observers specifically for their judgments in women failed to reach significance (t = -0.935, P = 0.178).

Not surprisingly, all observers saw more pain in the facial display of the observed subjects when the stimulus intensity was higher. Accordingly, the main effect of the within-subject factor stimulus intensity was significant (F<sub>1,40</sub> = 197.588, P < 0.001). However, there was also a significant interaction with the within-subject factor age.

**Table 2**  Mean and SD of observer ratings (NRS; 1–10) in “professionals” (hospital nurses and nurses for the elderly) and “nonprofessionals” (secretaries and administrative officers) for observed young and old subjects with and without dementia of both sexes (male vs female) under two stimulation conditions (2 kg vs 5 kg)

<table>
<thead>
<tr>
<th>Video Attributes</th>
<th>Professionals</th>
<th></th>
<th></th>
<th>Nonprofessionals</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Young Male 2 kg</td>
<td>2.55</td>
<td>1.18</td>
<td></td>
<td>2.48</td>
<td>1.38</td>
<td></td>
</tr>
<tr>
<td>5 kg</td>
<td>2.66</td>
<td>1.23</td>
<td></td>
<td>2.53</td>
<td>1.28</td>
<td></td>
</tr>
<tr>
<td>Female 2 kg</td>
<td>2.94</td>
<td>1.08</td>
<td></td>
<td>2.94</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>5 kg</td>
<td>2.95</td>
<td>0.80</td>
<td></td>
<td>3.08</td>
<td>1.56</td>
<td></td>
</tr>
<tr>
<td>Old Male 2 kg</td>
<td>3.51</td>
<td>1.15</td>
<td></td>
<td>3.59</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td>5 kg</td>
<td>3.52</td>
<td>1.17</td>
<td></td>
<td>3.53</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td>Female 2 kg</td>
<td>3.62</td>
<td>1.24</td>
<td></td>
<td>4.01</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td>5 kg</td>
<td>4.28</td>
<td>1.18</td>
<td></td>
<td>5.14</td>
<td>1.86</td>
<td></td>
</tr>
<tr>
<td>Demented Male 2 kg</td>
<td>3.72</td>
<td>1.41</td>
<td></td>
<td>3.82</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td>5 kg</td>
<td>4.55</td>
<td>1.30</td>
<td></td>
<td>4.89</td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td>Female 2 kg</td>
<td>4.33</td>
<td>1.23</td>
<td></td>
<td>4.73</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td>5 kg</td>
<td>6.05</td>
<td>1.41</td>
<td></td>
<td>6.46</td>
<td>1.40</td>
<td></td>
</tr>
</tbody>
</table>

NRS = numerical rating scale; SD = standard deviation.
and cognitive status. As can be seen in Table 2, observers only saw more pain in facial displays at higher stimulus intensities in old individuals with (t = 17.468, P < 0.001) and without (t = 6.945, P < 0.001) dementia, whereas observers saw no differences between intensities in young individuals (t = 1.429, P = 0.161).

Moreover, the level of signaled pain appeared to differ between the three groups of observed subjects (F1.287/51.463 = 211.819, P < 0.001). All observers saw least pain in the facial display of young subjects (young vs old adults without cognitive impairment: t = 10.909, P < 0.001; young vs old adults with cognitive impairment: t = 15.974, P < 0.001). Highest levels of pain were perceived in patients with dementia (post hoc comparison with old adults without cognitive impairment: t = 15.964, P < 0.001).

Also noteworthy, there was also significant main effect of the factor “gender”. As can be seen in Table 2, observers saw more pain in the facial display of women compared with men (F1/40 = 141.421, P < 0.001).

Concordance

Our concordance parameter was derived from the pain intensity estimates of each single observer compared with the self-report of the observed subjects, by computing individual correlations between the two rating parameters. Accordingly, such concordance parameters were available for each observer under each observational condition and could be subjected to the calculation of descriptive statistics, which are displayed in Table 3.

Analysis of variance showed a nonsignificant result for the between-subject factor “group” (F1/40 = 1.155, P = 0.289); thus, professionals did not differ from nonprofessionals as regards the concordance of their estimates of pain in the observed subjects.

Whereas “stimulus intensity” did not change the concordance of the observer ratings (F1.420 = 1.639, P = 0.222), “age and cognitive status” of the observed subject did (F2/80 = 47.724, P < 0.001). On average, the concordance scores for the three groups of observed individuals were: young = 0.10; older adults without cognitive impairment = 0.25; older adults with cognitive impairment = 0.08 (see also Table 3 for more detailed scores). The concordance of the observer ratings were clearly highest when older cognitive unimpaired individuals were judged, which can also be seen in significant group differences between this group and the other two groups (young vs old adults without cognitive impairment: t = 8.423, P < 0.001; old adults without cognitive impairment vs old adults with cognitive impairment: t = 5.185, P < 0.001). Surprisingly, the concordance was lowest when judging facial pain displays in young individuals, given that the correlation coefficients were significantly reduced compared with judging old adults with cognitive impairment (t = 5.739, P < 0.001).

Table 3 Mean and SD of Pearson correlations between observers’ judgment and self-report ratings in “professionals” (hospital nurses and nurses for the elderly) and “nonprofessionals” (secretaries and administrative officers) for observed young and old subjects with and without dementia of both sexes (male vs female) under two stimulation conditions (2 kg vs 5 kg)

<table>
<thead>
<tr>
<th>Video Attributes</th>
<th>Professionals</th>
<th></th>
<th></th>
<th></th>
<th>Nonprofessionals</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>P</td>
<td>Mean</td>
<td>SD</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>2 kg</td>
<td>0.115</td>
<td>0.313</td>
<td>0.107</td>
<td>0.044</td>
<td>0.335</td>
<td>0.553</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 kg</td>
<td>-0.169</td>
<td>0.304</td>
<td>0.019</td>
<td>-0.072</td>
<td>0.342</td>
<td>0.347</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2 kg</td>
<td>0.164</td>
<td>0.360</td>
<td>0.049</td>
<td>0.075</td>
<td>0.375</td>
<td>0.368</td>
</tr>
<tr>
<td></td>
<td>5 kg</td>
<td>-0.437</td>
<td>0.212</td>
<td>&lt;0.001</td>
<td>-0.549</td>
<td>0.293</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Old</td>
<td>2 kg</td>
<td>-0.045</td>
<td>0.310</td>
<td>0.513</td>
<td>-0.129</td>
<td>0.313</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 kg</td>
<td>0.306</td>
<td>0.231</td>
<td>&lt;0.001</td>
<td>0.286</td>
<td>0.355</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2 kg</td>
<td>0.028</td>
<td>0.540</td>
<td>0.014</td>
<td>0.158</td>
<td>0.261</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>5 kg</td>
<td>0.691</td>
<td>0.401</td>
<td>&lt;0.001</td>
<td>0.684</td>
<td>0.430</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Demented</td>
<td>2 kg</td>
<td>-0.212</td>
<td>0.181</td>
<td>&lt;0.001</td>
<td>-0.325</td>
<td>0.251</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 kg</td>
<td>0.251</td>
<td>0.249</td>
<td>&lt;0.001</td>
<td>0.380</td>
<td>0.200</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2 kg</td>
<td>0.449</td>
<td>0.286</td>
<td>&lt;0.001</td>
<td>0.386</td>
<td>0.234</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>5 kg</td>
<td>-0.070</td>
<td>0.379</td>
<td>0.406</td>
<td>-0.213</td>
<td>0.131</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

P values indicate whether the correlations significantly differed from zero (one-sample t-test). Significant concordance values are printed in bold (significant concordance is only given if correlation coefficient are positive and significant from zero). SD = standard deviation.
The facial display of women could be better read by all observers than that of men (see Table 3), given that the main effect “gender” also reached significance \( F_{1/40} = 6.535, P = 0.014 \). This effect was accompanied by two significant interaction effects, also including the factor “gender” \( \times \) “stimulus intensity” \( F_{1/40} = 82.509, P < 0.001 \) and “gender” \( \times \) “stimulus intensity” \( \times \) “group” \( F_{1/40} = 5.816, P = 0.021 \). Post hoc testing revealed that these interactions were mainly due to the fact that the facial display of women was most correctly judged when a 2 kg pressure was applied (mean concordance scores: \( Q = 0.21; \sigma = -0.09; t = 7.072, P < 0.001 \) and that this difference reversed just to the opposite at the higher stimulation intensity of 5 kg (mean concordance scores: \( Q = 0.02; \sigma = 0.16; t = -3.936, P < 0.001 \)).

The findings earlier inform about whether concordance parameters differed between groups and between pain intensities (2 vs 5 kg), but they do not allow to conclude whether the concordance values were better than chance or in other words, whether they significantly differed from zero. To test this, one-sample t-tests were computed. This was done for each combination of “group,” “stimulus intensity,” “age and cognitive status,” and “gender,” (e.g., professionals [“group”] rating those videos of male [“gender”], young individuals [“age and cognitive status”] receiving a 2 kg pressure stimulus [“stimulus intensity”], resulting into a total of 24 tests. Only 10 out of these 24 tests showed that concordance values (positive correlations) were significant from zero (see Table 3). Interestingly, we found nearly no significant positive correlations when the observed individuals were young (only one of the one-sample t-tests was significant). In contrast, the agreement between observers’ ratings and self-ratings of the observed individuals proved to be significant in 56% of the tests when old individuals were observed (see Table 3). Furthermore, concordance values also differed from zero in 50% of the tests when women were observed but only in 33% of the tests on men. In old women, even 63% of the tests proved that observers’ pain judgments and self-report ratings correlated above chance level.

**Discussion**

To answer the question of how capable “professionals” (hospital nurses and geriatric nurses) and the “nonprofessionals” (secretaries and administrative officers) as regards the competence in inferring pain from the facial expression might even be considered to be good news, considering previous reports of an underestimation of pain by “professionals” [11–14]. However, it turned out to be also true that long-standing experiences with patients or inhabitants of nursing homes, who likely suffer from acute and chronic pain quite frequently, are not sufficient per se to accurately judge the pain intensity experienced using facial cues. Given the amount of rather weak correlations that we found between self-report and observers’ pain ratings (that were apparent in nonprofessionals as well as in health care professionals), there is much room for improvement. This means that the professional training of nurses and caregivers probably does not include sufficient guidance in how to read and interpret the facial display of pain in others and that such a competence does not develop out of itself simply due to experience. This conclusion applies at least to the health care situation in Germany.

The lack of differences between the “professionals” (hospital nurses and geriatric nurses) and the “nonprofessionals” (secretaries and administrative officers) as regards the competence in inferring pain from the facial expression might even be considered to be good news, considering previous reports of an underestimation of pain by “professionals” [11–14]. However, it turned out to be also true that long-standing experiences with patients or inhabitants of nursing homes, who likely suffer from acute and chronic pain quite frequently, are not sufficient per se to accurately judge the pain intensity experienced using facial cues. Given the amount of rather weak correlations that we found between self-report and observers’ pain ratings (that were apparent in nonprofessionals as well as in health care professionals), there is much room for improvement. This means that the professional training of nurses and caregivers probably does not include sufficient guidance in how to read and interpret the facial display of pain in others and that such a competence does not develop out of itself simply due to experience. This conclusion applies at least to the health care situation in Germany.

The question arises why we could not replicate the phenomenon that health care professionals attribute less pain to others compared with nonhealth care professionals, which other authors reported to be common among professional observers [11–14]. There have been many differences between studies such as differences in samples and methods of assessing judgment quality. We like to stress one difference in particular, namely the availability of contextual information. We deprived the observers strongly of any contextual information showing exclusively the facial display (no other body parts or other persons) and informed the participants only that the observed subjects might have experienced various levels of pain as well as nonpainful episodes. Therefore, our observer situation was quite apart from in vivo situations in hospitals and nursing homes, which may have made our study different from those suggesting underestimation in professional observers. It is well known that contextual information is
critical for the judgment of pain in others [11,28]. We excluded this information deliberately because our primary intent was to learn how informative the facial display is solely by itself. Therefore, we used a typical experimental approach by reducing the number of influences and controlling for the remaining ones. It is the classical dilemma that the real life is too ambiguous, and the experimental life is too limited. By creating this “artificial” pain judgment situation and solely isolating the facial expression, we were able to show that the face alone is able to convey the necessary signals needed to notice pain in different groups of individuals to different degrees. The reduction of contextual information may—however—have been the reason why adaptation-level effects [13], which are thought to be responsible for the underestimation by “professionals,” were not obvious in the present study. Nurses and caregivers become used to pain in particular professional contexts. Outside these contexts, this adaptation-level effect may disappear, and no substantial judgment generalization over contexts may occur.

Judging Facial Pain Displays in Male, Female, Young, Old and Cognitively Impaired Individuals

Although there were no differences between observer groups, there were differences due to the type of the observed individuals. All observers saw more pain in older individuals with dementia than in those without and least pain in young individuals. This matches partially the order of objectively assessed facial pain expressiveness. According to FACS analyses conducted on the videos presented, the patients with dementia showed the strongest facial display of pain, however, with no differences between young and old individuals without cognitive impairment [17,29]. Therefore, seeing more pain in facial expressions of patients with dementia is supported by fine-grained FACS analyses of facial expressiveness in this patient group. The tendency to also see more pain in old compared with young individuals (although FACS analyses did not reveal different degrees of facial expressiveness between age groups [29]) was also reported by other authors [23,30]. It has been suggested that seeing more pain in older individuals is based on a stereotype that older individuals are frailer and thus more likely to suffer under painful experiences [23]. Interestingly, however, the concordance of the observer ratings was also best for older individuals, considering the match with the self-report ratings. Thus, in spite of potential judgment biases due to such stereotypes of older individuals, the ability to accurately infer pain from facial expressions was even increased in elders.

The poor correlations found between observer ratings and self-report in patients with dementia might seem surprising. We made sure to only select patients with dementia who were able to give a self-report rating of pain in order to be able to correlate self-report ratings with observer ratings. However, given their cognitive decline (mean MMSE score of 17), it is possible that they were less capable of providing as fine-grained evaluations of their subjective experience compared with the individuals without cognitive impairment. Self-reporting of pain is a very challenging cognitive maneuver. Trained, educated, and cognitive unimpared subjects may be apt to match an external scale to the internal metrics of their sensations and feelings with perfect linear slope and definite intercept that their ratings on external scales gain ratio or—at least—interval properties. In many older adults, we cannot expect this kind of cognitive fitness, and ratings may gain not more than ordinal quality. In the population under discussion, we think that the scale quality often declines even to nominal level. Patients with dementia might not be saying much more than “yes” and “no” to pain by giving numbers or words, which lost their original metric meanings. In sum, although we selected patients who were able to give a self-report rating, we think that the ability of self-report is not either existent or nonexistent in this patient group but fades out in parallel to the cognitive decline. Therefore, potentially more coarse-grained pain ratings in patients with dementia might be responsible for the lack of concurrence between observer ratings and patients’ pain ratings. However, given that observers were able to differentiate correctly between facial responses to 2 kg and 5 kg stimuli in patients with dementia, it can be concluded that observers are able to correctly identify different levels of pain based on the facial expressions in patients with dementia (despite the lack of concurrence between self-report and observer ratings), allowing for satisfactory pain judgment.

In tendency, women were observed as suffering more from pain and were judged with a higher concordance than men. Seeing more pain in women is also not a new finding [22,31,32]. Using FACS analyses, we found that women did not differ in their levels of facial pain expressiveness but are better able than men to fine-grade their facial display to express different levels of pain [21]. This suggests that observers could have made partially use of objectively existing gender differences in facial display as well as of gender role-related stereotypes.

Taken together, the competence of inferring the pain in others on the basis of the facial display appeared to be strongly dependent on features of the observed subjects. In other words, observer competence for one group of observed individuals does not necessarily hold for another group. In consequence, when trying to develop a training for health care professionals to improve decoding competence of facial pain displays, the training material should be manifold (including different prototypes of individuals) to allow for generalization.

Limitations of the study are 1) that the sample of professionals (hospital nurses and geriatric nurses) were mainly recruited in one institution, which was, however, not atypical for this kind of health services in Germany and 2) that the observed individuals were videotaped subjects from earlier studies, which did not allow full control over all relevant features but guaranteed similar noxious stimulation and objective FACS coding of the facial display.
Conclusions

In summary, hospital nurses and geriatric nurses did not differ from professions without paramedical backgrounds in their competence of judging pain in others by observing their facial display. Both observer groups only showed poor to moderate concordance in inferring pain from facial expressions. Therefore, our findings suggest that nurses do not seem to be well prepared to make adequate use of the facial expression to detect and judge pain in patients with dementia, at least without additional contextual information. Future studies have to show, whether a specific training might improve the decoding ability in health care professionals. Furthermore, given that we only included patients with mild to moderate forms of dementia who were still able to provide a self-report rating, further studies are needed to investigate, whether observers are also able to decode pain from facial expressions in more severe forms of dementia, where self-report is lacking.

Acknowledgment

This study was supported by grants of the Bamberg University (FNK) and of the European program COST (TD 1005). We thank Wiebke Hofer, MPsy for the support in video extractions.

References


