Received: 10 July 2019 | Revised: 22 November 2019 | Accepted: 23 November 2019

DOI: 10.1002/ejp.1514

Check for updates

ORIGINAL ARTICLE

What does "moderate pain" mean? Subgroups holding different conceptions of rating scales evaluate experimental pain differently

Susanne Becker[1,2] | **Xaver Fuchs**[2,3] | **Karin Schakib-Ekbatan**[2,4] | **Marcel Schweiker**[2,4]

[1]Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Mannheim, Germany

[2]Heidelberg Academy of Sciences and Humanities, Heidelberg, Germany

[3]Biopsychology & Cognitive Neuroscience, Faculty of Psychology & Sports Science, Bielefeld University, Bielefeld, Germany

[4]Building Science Group, Karlsruhe Institute of Technology, Karlsruhe, Germany

Correspondence
Susanne Becker, University of Zurich, Balgrist University Hospital, Balgrist Campus, Lengghalde 5, 8008 Zurich, Switzerland.
Email: susanne.becker@balgrist.ch

Funding information
This work was funded by the Heidelberg Academy of Sciences and Humanities within the project "Thermal comfort and pain".

Abstract

Background: Pain ratings are almost ubiquitous in pain assessment, but their variability is high. Low correlations of continuous/numerical rating scales with categorical scales suggest that individuals associate different sensations with the same number on a scale, jeopardizing the interpretation of statistical results. We analysed individual conceptions of rating scales and whether these conceptions can be utilized in the analysis of ratings of experimental stimuli in pain-free healthy individuals and people with reoccurring/persistent pain.

Methods: Using a free positioning task, healthy participants ($N = 57$) and people with reoccurring/persistent pain ($N = 57$) ad libitum positioned pain descriptors on lines representing intensity and un-/pleasantness scales. Furthermore, participants rated experimental thermal stimuli on visual analogue scales with the same end anchors. A latent class regression approach was used to detect subgroups with different response patterns in the free positioning task, indicating different conceptions of pain labels, and tested whether these subgroups differed in their ratings of experimental stimuli.

Results: Subgroups representing different conceptions of pain labels could be described for the intensity and the un-/pleasantness scale with in part opposing response patterns in the free positioning task. Response patterns did not differ between people with and without pain, but in people with pain subgroups showed differential ratings of high intensity experimental stimuli.

Conclusions: Individuals' conceptions of pain labels differ. These conceptions can be quantified and utilized to improve the analysis of ratings of experimental stimuli. Identifying subgroups with different conceptions of pain descriptions could be used to improve predictions of responses to pain in clinical contexts.

Significance: The present results provide a novel approach to incorporate individual conceptualizations of pain descriptors, which can induce large distortions in the analysis of pain ratings, in pain assessment. The approach can be used to achieve better pain estimates, representing individual conceptions of pain and achieving a better comparability between individuals but also between pain-free persons and patients

with chronic pain. Particularly, in clinical settings this could improve quantification of perceived pain and the patient-clinician communication.

KEYWORDS

categorical rating scales, chronic pain, latent class regression, numeric rating scales, pain descriptors, visual analogue scales

1 | INTRODUCTION

Rating scales are almost ubiquitous in human pain assessment and indispensable in clinical settings because of the ease of application. Typically *visual analogue scales* (VAS), e.g. a 10-cm line with two endpoints such as "no pain" and "worst pain imaginable", *numerical rating scales* (NRS), e.g. 11, 21, or 101-point scales with similar endpoint anchors, and *verbal rating scales* (VRS), i.e. categorical scales with a list of descriptors such as "no pain", "mild pain", "moderate pain", "severe pain" and "most intense pain imaginable" are used (Williamson & Hoggart, 2005).

For VAS and NRS it is widely accepted that resulting data can be treated with parametric statistics (on a ratio level of measurement; e.g., Myles, Troedel, Boquest, & Reeves, 1999; Price, Bush, Long, & Harkins, 1994; Price, McGrath, Rafii, & Buckingham, 1983). However, some studies suggest that this assumption might be a statistical "illusion". When asking people what a certain number on an NRS means to them, variation between participants' answers is surprisingly high (Williams, Davies, & Chadury, 2000), also illustrated by low correlations between VAS/NRS and VRS ratings (e.g., Dijkers, 2010; Linton & Götestam, 1983; Lund et al., 2005). These low correlations have been explained by the fact that different verbal categories have different meanings to different people (Averbuch & Katzper, 2004; Dijkers, 2010; Linton & Götestam, 1983; Lund et al., 2005). Early studies intensely investigated the construction of reliable and valid VRS based on psychophysical cross-modal matching procedures (e.g., Gracely & Dubner, 1987; Gracely, McGrath, & Dubner, 1978; Heft, Gracely, & Dubner, 1980). However, these methods have been rarely applied (cf., Hall, 1981). Instead, ratings on VRS are typically converted directly to numbers and used as such in statistical analyses (cf., Gracely, 1990).

However, if individuals associate different sensations with the same number on a scale, this approach appears problematic and comparisons and interpretations of VAS/NRS ratings across individuals are difficult. Qualitative studies support these considerations. For example, some chronic pain patients redefine 0 on an NRS being as manageable pain to fit their personal experience (Williams et al., 2000). In addition, upper end anchors of rating scales such as "worst pain imaginable" can induce large variation, because these anchors appear unclear, dependent on individual experience, and are often perceived as indefinable (Dannecker, George, & Robinson, 2007; Yokobe, Kitahara, Matsushima, & Uezono, 2014). Thus, frequent/reoccurring pain likely alters the conception of pain descriptors. Nevertheless, no suggestion is available on how to handle such individual conceptions. Furthermore, since studies are lacking in pain-free individuals, it remains unclear whether pain indeed alters such conception.

The aim of this study was to assess individual conceptions of pain scales anchors and to utilize these conceptions to explain variance in the analysis of ratings of experimental stimuli. We hypothesized (a) that subgroups of individuals exist, which can be defined based on their conception of pain scale labels; (b) that these subgroups differ in their ratings of experimental stimuli; and (c) that these subgroups are different for people with reoccurring/persistent pain and pain-free healthy participants.

2 | MATERIALS AND METHODS

The Ethics committee and data protection officer of Karlsruhe Institute of Technology approved the study.

2.1 | Participants

In total, 82 participants took part in the study. Participants were assessed in two testing waves, one in January/February 2016 and one in January/February 2017. In the first testing wave, 30 healthy pain-free participants and in the second, a sample of 25 people with reoccurring/persistent pain and 27 sex- and age-matched healthy pain-free participants took part in the study.

Participants were recruited using announcements placed on a local online job market for students, a free newspaper, and on the institutional homepage. All participants were assured to be either native speakers or to have a comparable language level in German and aged above 18. Inclusion criterion for people with reoccurring/persistent pain was back pain for at least 6 months and at least 1 day every 2 weeks. Exclusion criteria for people with reoccurring/persistent pain and healthy participants were intake of opioid or psychotropic drugs, present of past mental disorders, sleep disorders. Healthy participants were excluded if they reported

TABLE 1 Number of participants, gender distribution, age, education level and occupation of all participants in the different sub-samples and scores of the sub-scales from part one of the West Haven-Yale Multidimensional Pain Inventory (MPI)

	1	2	
Testing wave	**Healthy participants**	**Healthy participants**	**People with reoccurring/ persistent pain**
N	30	27	25
Female/male (*N*)	17/13	16/11	13/12
Age [mean (*SD*)]	26.17 (0.50)	50.33 (2.55)	54.92 (2.68)
Education (*N*)[a]			
No general university entrance qualification for, no academic studies	1	9	6
General university entrance qualification, no academic studies	16	15	10
General university entrance qualification, academic studies	0	3	7
Academic studies without a general university entrance qualification	13	0	0
Occupation [*N*][a]			
Student	24	2	1
Employed	3	17	13
Unemployed	3	2	7
Retired	0	5	3
MPI [mean (*SD*)][b]			
Pain severity			2.31 (0.97)
Interference			2.43 (1.11)
Affective distress			2.45 (1.17)
Self-control			3.77 (1.30)
Support			1.65 (1.88)

Abbreviations: MPI, Multidimensional Pain Inventory; *SD*, standard deviation.

[a]Information from two participants on education level and occupation was missing.

[b]MPI was only assessed for patients. Information from three participants on MPI was missing.

any pain present for more than 6 months and more frequently than 1 day every 2 weeks. In- and exclusion criteria were assessed in a telephone interview before participation in the study. During this telephone interview, the interviewer also assessed whether the German language level was sufficient for participation.

Details on the sample are given in Table 1, including information on education level and occupation. Healthy participants of the first assessment wave were younger compared to healthy participants ($t(28) = -9.281$; $p < .001$) and people with reoccurring/persistent pain ($t(26) = -10.542$; $p < .001$) in the second testing wave. Healthy participants and people with reoccurring/persistent pain of the second testing wave did not differ in age ($t(50) = -1.239$; $p = .221$). Characteristics of the clinical pain of the people with reoccurring/persistent pain assessed with part one of the West Haven-Yale Multidimensional Pain Inventory (Kerns & Turk, 1985; German translation; Flor, Rudy, Birbaumer, Streit, & Schugens, 1990) are also reported in Table 1. These data show considerable pain severity ($M = 2.31$, $SD = 0.97$, on a scale from 0 to 6), interference ($M = 2.43$, $SD = 1.11$) and affective distress due to the pain ($M = 2.45$, $SD = 1.17$) in the people with reoccurring/persistent pain, because of which we assume that this pain could have lead to changes in the conceptions of pain scale labels.

2.2 | General procedure

After participants arrived at the testing facility, they received instructions about the procedures and methods upon which written informed consent was obtained. At the beginning of the testing session, participants performed a free positioning task assessing their conceptualization pain rating scale anchors (details below). Subsequently, participants' pain sensitivity was assessed and an experimental

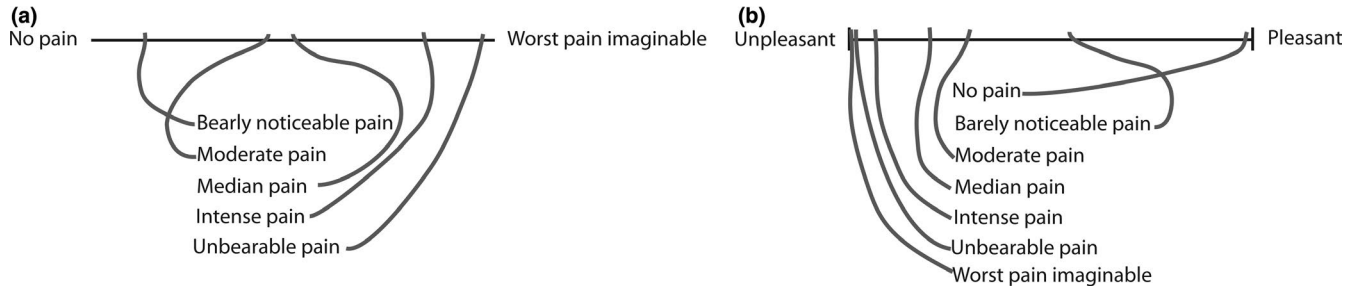

FIGURE 1 Free positioning task, showing an example for the intensity (a) and the un-/pleasantness (b) scale. Participants indicated the position of the pain labels shown below the horizontal line ad libitum on this line with the end anchors 'no pain' and 'worst pain imaginable' for the intensity scale (a) and 'unpleasant' and 'pleasant' for the un-/pleasantness scale

testing phase started during which participants rated experimental thermal painful and non-painful stimuli on VASs to allow a comparison of individual conceptions of pain labels and VAS ratings of experimental stimuli (details below). The free positioning task and the testing was part of a larger study incorporating in addition a qualitative interview, the performance of cognitive tests, the completion of questionnaires, and the assessment of physiological parameters of which the result are not presented here. All testing sessions were performed in the Laboratory for Occupant Behaviour, Thermal comfort, Satisfaction and Environmental Research (Wagner et al., 2018) belonging to the Building Science Group at the Karlsruhe Institute of Technology, Karlsruhe, Germany.

2.3 | Free positioning task and interview

The used free positioning task was established before (Schweiker et al., 2017) in studies of thermal comfort. The task assesses the individually perceived relative position of pain descriptors to each other, describing thereby the individual conceptualization of pain labels. Participants were presented with horizontal lines showing two verbal anchors at their endpoints (see Figure 1). Participants were instructed to position pain labels shown underneath the horizontal line (arranged in a vertical column; see Figure 1) ad libitum on these horizontal lines. The participants could choose the order in which they positioned the pain labels themselves. The positioning was combined with a "think-aloud" procedure (e.g., Boren & Ramey, 2000) allowing assessing participants' strategies, thoughts and reasons for choosing a particular position for a given pain label. Resulting qualitative data of this procedure was not included in the present analyses.

The first horizontal line presented to the participants represented an intensity scale with the verbal anchors "no pain" and "worst pain imaginable" at its endpoints. Participants were asked to position on the line the five pain labels from the German version of the McGill Pain Questionnaire (Stein & Mendl, 1988): barely noticeable ("gerade wahrnehmbar"), moderate ("mäßig"), median ("mittel"), intense ("stark") and unbearable ("unerträglich"; see Figure 1a). The endpoints "no pain" and "worst pain imaginable" are not part of the McGill Pain Questionnaire and were chosen to frame the dimension of pain intensity. Then, more lines with different end anchors were presented in a pseudo-randomized order. On all these lines, participants positioned all seven pain labels, i.e. the five labels from the McGill Pain Questionnaire and the labels "no pain" and "worst pain imaginable" on these lines (see Figure 1b). The dimensions and end anchors of these scales were: un-/pleasantness, "unpleasant"–"pleasant"; suffering, "no suffering"–"worst suffering imaginable"; acceptability, "not acceptable"–"acceptable", tolerability, "intolerable"–"tolerable". Here, only the dimensions intensity and unpleasantness are presented as the most widely used scales in pain research. Data on the other dimensions (suffering, acceptability, and tolerability) can be found in the supplementary material (see Results S6).

The positions of the labels in the free positioning task were quantified by measuring the distance of the positioned label to the left end of each horizontal line using a ruler. Each line was 10 cm long and values were coded in mm, resulting in values between 0 and 100.

2.4 | Thermal stimulation

During the experimental testing, participants received thermal stimuli applied with a contact thermode (SENSELab—MSA Thermotest, SOMEDIC Sales AB). The thermode size was 2.5 × 5 cm. Baseline skin temperature was kept constant at 30°C. For safety reasons, the maximal temperature was limited to 50°C. All thermal stimuli were applied to the volar forearm of participants' non-dominant hand after sensitization of the skin using 0.075% topical capsaicin cream. Capsaicin-induced sensitization of the skin was used to allow for painful stimulation without the risk of skin damage (Gandhi, Becker, & Schweinhardt, 2013). The cream was applied to the 2.5 × 5 cm stimulation area on the forearm. Capsaicin is the active ingredient of chilli pepper that induces heat sensitization

by activating temperature-dependent TRPV1 (vanilloid transient receptor potential 1) ion channels (Holzer, 1991). The cream was removed after 20 min (Dirks, Petersen, & Dahl, 2003; Gandhi et al., 2013) and the thermode was applied at the location on the forearm and fixed with a neoprene strap.

2.5 | Assessment of thermal sensitivity

To assess participants pain sensitivity and to determine stimulation intensities during the experimental pain testing, participants' warm detection threshold, heat pain threshold and heat pain tolerance were tested on the capsaicin-treated site of the non-dominant volar forearm and on a comparable untreated site on the dominant volar forearm as a control site before the experimental pain testing. Warm detection threshold, heat pain threshold and heat pain tolerance were assessed with the methods of limits, with three stimuli each increasing in temperature at a rate of 1°C/s until the participant felt the first warmth sensation (warm detection threshold), the slightest pain sensation (pain threshold) or could not tolerate the stimulus any longer (pain tolerance). Participants indicated their threshold/tolerance levels by pressing a mouse button after which the temperature of the thermode returned immediately to the baseline temperature.

2.6 | Experimental pain assessment

In order to assess participants VAS ratings of experimental thermal stimuli, participants received thermal stimuli of two intensities (high, low) at the capsaicin-treated stimulation site on their non-dominant forearm. Stimulation intensities were adjusted to participants' pain sensitivity by defining stimulation intensities ranging between the warm detection threshold and pain tolerance. Intensity of the high stimulation was set to the heat pain threshold plus 50% of the difference between the heat pain threshold and heat pain tolerance, aiming at a painful sensation (Herta Flor, Knost, & Birbaumer, 2002). The intensity of the low stimulation was set to just above the warm detection threshold with using the detection threshold plus 10% of the difference between the warm detection threshold and heat pain threshold, aiming at a non-painful sensation. To ensure that painful stimuli remained in the painful but bearable range and did not become too painful due to sustained heat stimulation leading to sensitization (Lautenbacher, Roscher, & Strian, 1995) and/or capsaicin-induced sensitization (Dirks et al., 2003; Sluka, 2002) and that non-painful stimuli remained non-painful, perceived painfulness of stimuli was assessed throughout the experiment using VAS and the intensity was adjusted when VAS indicated sensations outside the desired range. For the low stimuli, the stimulation intensity was lowered by steps of 0.5°C, if intensity ratings of these stimuli exceeded 25% on a 10 cm intensity VAS with the end anchors "no pain" (left) and "worst pain imaginable" (right). For the high stimuli, the stimulation intensity was lowered or increased by steps of 0.5°C, if intensity ratings of these stimuli were lower than 25% or exceeded 75% on the intensity VAS. Hence, ratings of <25 and between 25 and 75 were chosen to achieve two distinguishable sensations. Each stimulation intensity was applied three times. Order of the stimulation conditions was counterbalanced across participants. Participants rated how they perceived each experimental stimulus on different VAS. The end anchors of these VASs corresponded to the scales of the free positioning task in the interview.

2.7 | Statistical analysis

Analyses of the free positioning task and the ratings of the experimental stimuli were performed as before, for details see Fuchs, Becker, Schakib-Ekbatan, and Schweiker (2018).

2.7.1 | Polynomial regression models

A regression model for each rating scale was fitted to the positions of the pain labels on the scales in terms of their distance from the left end of the line. Endpoint anchors were not included in this analysis, because participants were not allowed to position them. The independent variable of the models was the pain labels participants placed on the line in the free positioning task, coded as an ordered factor from 1 to 5 for the intensity scale and from 1 to 7 for the remaining dimensions. The dependent variable was the position on the line from the left, coded as a value between 0 and 100. To allow also for non-linear relationships between the pain labels and the position on the line, polynomial regression models were used, including linear, quadratic, and cubic terms in the form of

$$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \varepsilon.$$

The models were first fit on to the level of the group mean to illustrate variation and divergence from this average curve on single subject level as displayed in Figure 3. Using polynomial contrasts, it was tested whether the average curves followed a linear, quadratic, or cubic trend.

2.7.2 | Latent class regression

To differentiate response patterns in the free positioning task expressing distinguishable conceptions of scales, the polynomial regression models described above were used within a latent class regression (LCR) model for each scale separately. LCR can be used to test whether a sample can be separated into a set of

unknown underlying (latent) clusters (i.e., subgroups) that differ with respect to certain parameters of statistical models or relationships between variables. For example, a sample could consist of subgroups in which the relationships between variables as described by regression models are different. LCR tests statistically whether a relationship between two (or more) variables can be explained significantly better by separating the sample in a number of different clusters/subgroups fitted by different regression models (Dayton & Macready, 1988; Grün & Leisch, 2008) compared to conventional models that apply the same parameter to all participants. For this purpose, LCR attempts to group participants that show similar relationships between variables and separate them from other subgroups of participants that are better described by different (regression) parameters. LCR can be used with repeated measurements as in the free positioning task where each participant positioned 5 or 7 labels (depending on the scale). LCR were applied using the package *flexmix* (Grün & Leisch, 2008) within the R statistical software (R Development Core Team, 2012). This package applies the expectation-maximization algorithms within a maximum likelihood framework for statistical comparisons of different models, e.g. assuming the abovementioned subgroups (see Grün & Leisch, 2008; for a detailed description of model testing within the LCR approach). In addition, the *flexmix* package allows determining the optimal number of subgroups statistically by implementing a stepwise procedure, in which models are fitted iteratively based on a user-defined range of numbers of subgroups. Subsequently, the best fitting model is selected. Here, 1 to 20 possible subgroups were allowed and for the estimation of each model, 500 iterations were used.

2.7.3 | Linear mixed models and model comparison

Linear mixed models (LMM) were used to compare participants' warm detection threshold, heat pain threshold, pain tolerance, VAS ratings of the experimental stimulation, and to test whether subgroups defined in the LCR procedure differed in this VAS ratings using IBM SPSS 21 (SPSS Inc.). In addition to the respective fixed factors, all LMM included all interactions of these fixed factors as well as participant ID as a random intercept factor, age as a covariate, and the interaction terms of age as a covariate with the fixed factors of each model controlling for the different age ranges in the sample. Age was added as a covariate in the LMMs because it was significant as a covariate in all (all p's $< .029$) but one ($p = .307$) analyses of the ratings of the experimental stimuli on the intensity as well as the un-/pleasantness. Sex and education level were not included in the analyses, because both variables were not significant as covariates in all analyses (all p's $> .16$).

In order to ensure that groups did not differ in the baseline pain sensitivity, possibly confounding the results, and to ensure the that sensitization procedure with the capsaicin was effective, separate LMM for the analysis of the warm detection threshold, heat pain threshold, and pain tolerance were calculated with 'stimulation site' as a within-subject fixed factor with two levels (capsaicin-treated, untreated), 'group' as a between-subject fixed factor with two levels (healthy participants, people with pain).

For the analysis of the VAS ratings of the experimental stimulation, 'stimulation condition' was included as a within-subjects fixed factor with two levels (low, high) for each scale separately (model 1). To test whether the LCR subgroup as a factor affected the VAS ratings in the experimental conditions, a second model (model 2) was computed for each scale. Model 2 included the fixed factor 'stimulation condition' (as in model 1, levels: low, high) and additionally a fixed factor for 'LCR subgroup' (with the number of levels depending on the LCR result for the respective scale). To test a significant improvement in model fits over model 1, both models were compared against each other using likelihood ratio tests. Post-hoc pairwise comparisons were used to test whether VAS ratings differed between stimulation conditions and LCR subgroups.

2.7.4 | Group distribution within and LCR subgroup distribution across scale dimensions

To test whether the distribution of healthy participants and people with reoccurring/persistent pain differed across LCR subgroups for each scale, χ^2-tests were used.

For all analysis, all 82 participants of the whole sample were pooled and analysed. The significance level was set to $\alpha = 0.05$. Where appropriate, correction for multiple testing was applied using false discovery rate to avoid alpha inflation. Exact p-values are reported with significances after correction for multiple testing indicated as follows where appropriate: $^{*}p < .05$; $^{**}p < .01$.

3 | RESULTS

3.1 | Warm detection threshold, heat pain threshold and tolerance

People with reoccurring/persistent pain and healthy controls did not differ in their basic pain sensitivity: Warm detection threshold, heat pain threshold, and heat pain tolerance did not differ between people with reoccurring/persistent pain and healthy participants (main effect 'group', all $p > .640$; for means and *SD* see Results S5; Table S3). However, heat pain threshold and heat pain tolerance were lower at the capsaicin-treated compared to the untreated stimulation site for both people with reoccurring/persistent pain and healthy participants (main effect 'stimulation site, heat pain threshold: $F(1, 78) = 15.411$; $p < .001$**; heat pain tolerance: $F(1, 79) = 34.949$; $p < .001$**),

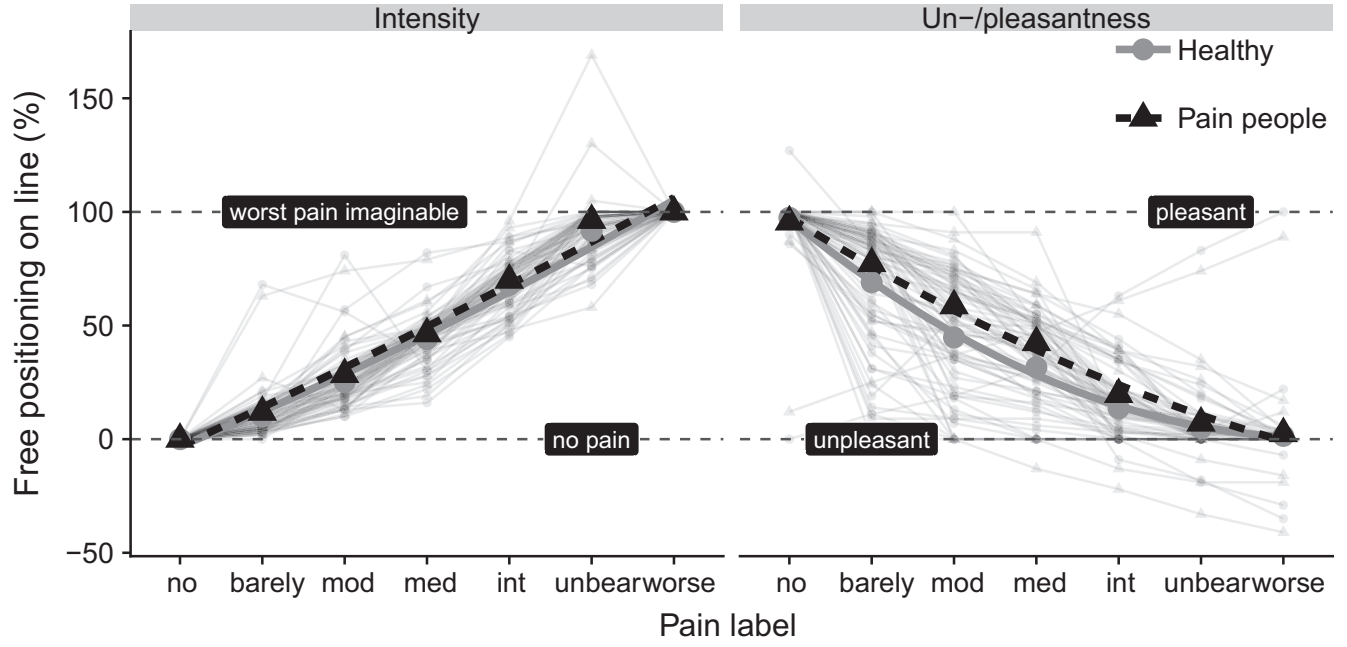

FIGURE 2 Positioning of the pain labels in the free positioning task for the intensity (left) and the un-/pleasantness scale (right). On the *x*-axis the pain labels are shown with the following abbreviation: barely, barely noticeable; int, intense pain; med, median pain; mod, moderate pain; no, no pain; unbear, unbearable pain; and worse, worst pain imaginable. Positioning of the participants of the pain labels are shown on the *y*-axis in percent relative to the left end of the scale. Endpoints of the scales with their respective anchors are depicted as horizontal dashed lines with the label attached. Average positioning of the labels by the healthy participants are depicted as grey dots (abbreviation in legend: Healthy) and as black triangles for the people with reoccurring/persistent pain (abbreviation in legend: Pain people). Positioning of single participants are represented as light grey points interconnected with thin grey lines

demonstrating the effectiveness of the sensitization with topical capsaicin. No difference between stimulation sites were found for the warm detection threshold (main effect 'stimulation site, $F(1, 76) = 0.003$; $p = .959$). No interactions between group and stimulation site were found for the warm detection threshold, heat pain threshold, and heat pain tolerance (all $p > .226$), showing that warm detection threshold heat pain threshold and heat pain tolerance did not differ between groups at the different stimulation sites. Thus, basic differences in sensitivity between people with reoccurring/persistent pain and controls did not confound the following results.

3.2 | Positioning of labels in the free positioning task

Participants' positioning of the pain labels on the intensity and un-/pleasantness scale are shown in Figure 2, demonstrating a high variability across participants. Mean positioning across all participants (healthy participants and people with reoccurring/persistent pain pooled) followed quadratic trends for both the intensity and un-/pleasantness scale (quadratic trend contrast, intensity: $t(320) = 5.784$; $p < .001$**; un-/pleasantness: $t(478) = -9.134$; $p < .001$**). Positioning of scale anchors in the free positioning task did not differ between healthy participants and people with reoccurring/persistent pain for both the intensity and the un-/pleasantness scale (main effect 'group', intensity: $F(1, 96) = 1.655$; $p = .201$; un-/pleasantness: $F(1, 77) = 0.262$; $p = .610$).

3.3 | Subgroups of ratings patterns based on free positioning task

The LCR revealed for the intensity and the un-/pleasantness scale four discriminable response patterns in the free positioning task each. Different intercepts and slopes in their regression models characterized these subgroups, as shown in Figure 3a,e. As displayed in Figure 3a,e, the curves indicating different response patterns from almost linear to cubic. Several curves are characterized, for example, by steep slopes toward one or both endpoints of the scale with shallow slopes in the middle or by an opposing pattern with a steep slope in the middle area of the scale and more shallow slopes toward the scale endpoints.

3.3.1 | Intensity scale

Most participants (subgroup 2, $N = 28$, and subgroup 4, $N = 28$) positioned the pain labels of the intensity scale in a S-shaped curve, although the degree of the curvature varied, with subgroup 2 showing a much stronger curvature than

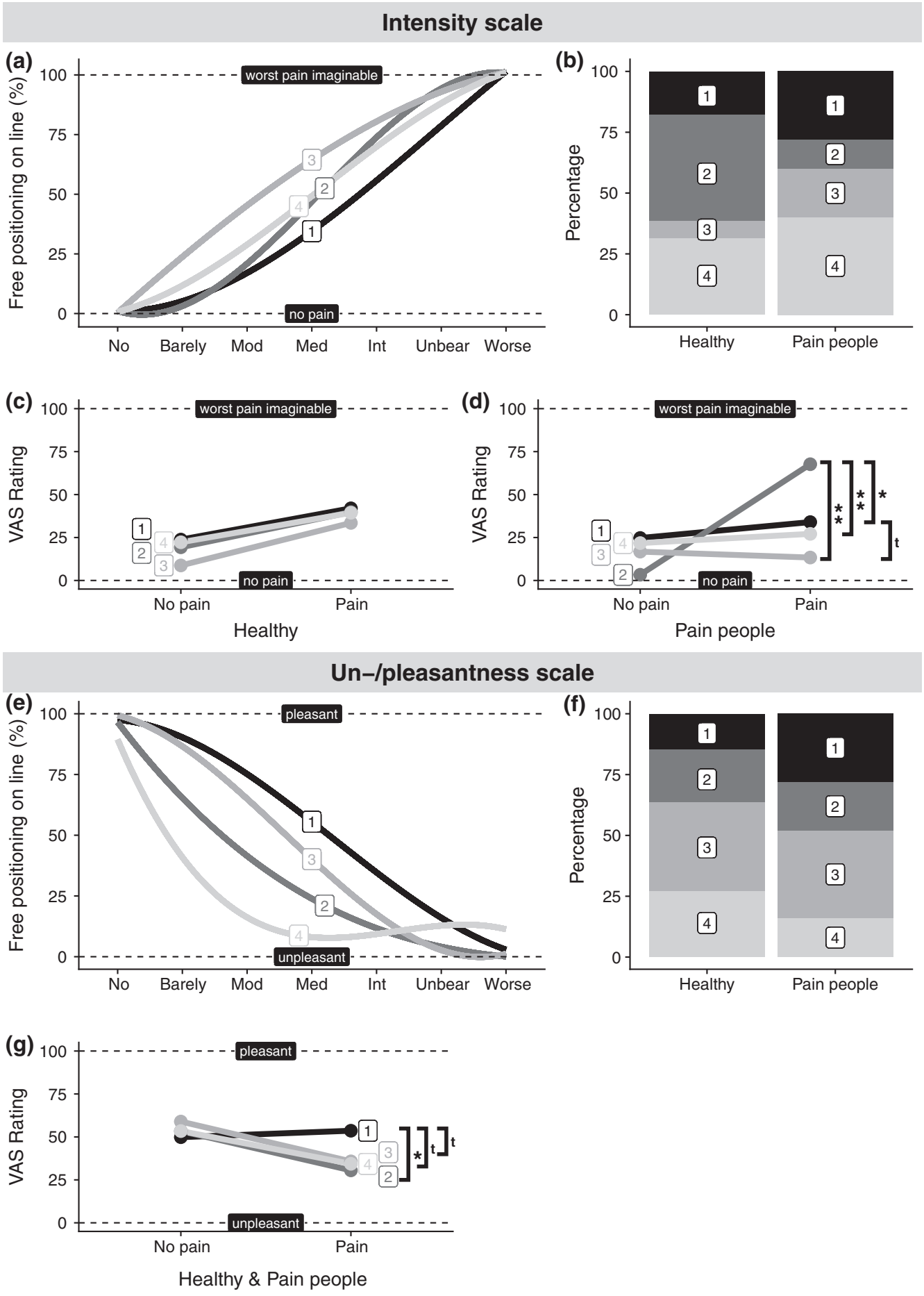
Intensity scale
(a)
Free positioning on line (%)
100
75
50
25
0
worst pain imaginable
no pain
No
Barely
Mod
Med
Int
Unbear
Worse
(b)
Percentage
Healthy
Pain people
(c)
VAS Rating
No pain
Pain
Healthy
(d)
Pain people
Un–/pleasantness scale
(e)
pleasant
unpleasant
(f)
(g)
Healthy & Pain people

FIGURE 3 Results from the latent class regression analysis (LCR) and ratings of the experimental stimuli of the subgroups of the LCR for the intensity (upper part) and un-/pleasantness scale (lower part). The different subgroups of participants with similar response patterns in the free-positioning task identified by the LCR are for each scale indicated by white labels and shades of gray. (a, e) The four different curves of the regression models of the subgroups determined by the LCR for the intensity (a) and the un-/pleasantness (e) scale are shown. On the *x*-axis the pain labels are shown with the following abbreviation: barely; barely noticeable; int, intense pain; med, median pain; mod, moderate pain; no, no pain, unbear, unbearable pain; and worse, worst pain imaginable. The *y*-axis represents positioning in the free-positioning task in percent relative to the left end of the scale. Endpoints of the scales with their respective anchors are depicted as horizontal dashed lines with the label attached. (b, f) Distribution of the percentage of healthy participants (abbreviation on the *x*-axis: Healthy) and people with reoccurring/persistent pain (abbreviation on the *x*-axis: Pain people) in each of the LCR subgroups for the intensity (b; healthy participants: $N_1 = 10$ (17.5%), $N_2 = 25$ (43.8%), $N_3 = 4$ (7%), $N_4 = 18$ (31.6%); people with reoccurring/persistent pain: $N_1 = 7$ (28%), $N_2 = 3$ (12%), $N_3 = 5$ (20%), $N_4 = 10$ (40%); index indicating LCR subgroup) and the un-/pleasantness scale (*F*; healthy participants: $N_1 = 8$ (14.3%), $N_2 = 12$ (21.4%), $N_3 = 21$ (37.5%), $N_4 = 15$ (26.8%); people with reoccurring/persistent pain: $N_1 = 7$ (28%), $N_2 = 5$ (20%), $N_3 = 9$ (36%), $N_4 = 4$ (16%); index indicating LCR subgroup). (c, d, g) Mean ratings on the visual analogue scale (VAS) of the low and high experimental heat stimuli (on the capsaicin-treated stimulation site) for the subgroups determined by the LCR are shown for the intensity scale for healthy participants (Healthy; c) and people with reoccurring/persistent pain (Pain people; d) separately and pooled for the un-/pleasantness scale (g; based on the statistical model). Significances of post-hoc comparisons are given as: $^{t}p < .10$, $^{*}p < .05$; $^{**}p < .01$

subgroup 4. These S-shaped curves indicate that participants positioned the middle anchors (moderate, medium, intense) with comparatively large distances between each other, while the anchors toward both ends (no pain, just noticeable and unbearable, most worst pain imaginable) were positioned close to each other. In contrast, subgroup 1 and 3 showed a more quadratic trend, although with opposing patterns. Subgroup 1 placed the lower scale anchors comparatively close to each other with the middle and upper anchors being more spread out but almost linearly arranged. Subgroup 3 showed an almost linear pattern for the lower and middle scale anchor, with the upper anchors being arranged closer to each other.

3.3.2 | Un-/pleasantness scale

For the un-/pleasantness scale, the largest subgroup (subgroup 3, $N = 30$) showed an S-shaped arrangement of the anchors. This subgroup 3 positioned the middle anchors (moderate, medium, intense) with comparatively large distances between each other, while the anchors towards both ends (no pain, just noticeable towards the pleasant end anchor and unbearable, worst pain imaginable towards the unpleasant end anchor) were positioned close to each other. In contrast, the other three subgroups showed more quadratic trends with subgroup 1 showing a pattern opposing the patterns of subgroup 2 and 4, with the pattern of subgroup 4 being much more extreme than the one of subgroup 2. Subgroup 1 placed the lower scale anchors (toward pleasant) comparatively close to each other with the middle and upper anchors (toward unpleasant) being more spread out but almost linearly arranged. Subgroup 2 and 4 showed such an almost linear pattern for the lower (toward pleasant) and middle scale anchor, with the upper anchors (toward unpleasant) being arranged closer to each other. Subgroup 4 placed the anchors medium, intense, unbearable and worst pain imaginable very close to each other or even at the same position, almost suggesting that medium and intense pain was positioned closer to the end anchor than unbearable and worst pain imaginable.

3.4 | Distribution LCR subgroup membership across people with reoccurring/persistent pain and healthy controls

The distribution of people with reoccurring/persistent pain and healthy participants across the different LCR subgroups (see Figure 3b) was different for the intensity scale (Log Likelihood = −1132; $\chi^2(5) = 9.112$; $p = .028$), while it did not differ for the un-/pleasantness scale (Log Likelihood = −2046; $\chi^2(5) = 2.640$; $p = .451$; Figure 3f). For the intensity scale a larger portion of healthy participants belonged to subgroup 2 compared to the people with reoccurring/persistent pain.

3.5 | Ratings of experimental stimulation dependent on response patterns in the free positioning task

Model fits of the analysis of the intensity and un-/pleasantness ratings of the experimental stimulation improved significantly by including LCR subgroup as a factor (intensity: $\chi^2(6) = 14.7$; $p = .023$; un-/pleasantness: $\chi^2(6) = 28.4$; $p < .001$; for details see Table S2). Results of the analyses of the ratings of the experimental stimuli including LCR subgroup as a factor are summarized above for the intensity scale for healthy participants in Figure 3c, for people with reoccurring/persistent pain in Figure 3d, and for the un-/pleasantness scale for healthy participants and people with reoccurring/persistent pain pooled in Figure 3g. Table 2 displays means and standard deviations of the ratings of the experimental stimuli on the intensity and the un-/pleasantness scale, independent of LCR subgroups as well as within subgroups.

TABLE 2 Means and standard deviations of ratings of low and high thermal stimuli on visual analogue scales for perceived intensity and un-/pleasantness independent and within subgroups resulting from the latent class regression (LCR) of the response patterns in the free positioning task

	Intensity scale		**Un-/pleasantness scale**	
independent of LCR subgroups				
Stimulation intensity	*Healthy participants*	*People with reoccurring/persistent pain*	*Healthy participants*	*People with reoccurring/persistent pain*
Low	20.36 (20.01)	19.62 (17.91)	56.46 (29.19)	52.74 (28.85)
High	39.66 (22.35)	31.32 (23.34)	37.02 (23.47)	41.22 (27.07)
within LCR subgroups[a]				
	Healthy participants	*People with reoccurring/persistent pain*	*Groups pooled*	
Low				
Subgroup 1	23.80 (21.02)	24.43 (17.38)	24.06 (19.03)	
Subgroup 2	19.46 (17.61)	3.33 (2.89)	17.67 (17.37)	
Subgroup 3	8.75 (7.85)	16.75 (24.05)	12.75 (17.10)	
Subgroup 4	22.22 (24.24)	22.30 (17.91)	22.25 (21.91)	
High				
Subgroup 1	42.00 (23.20)	33.71 18.46)	38.38 (21.01)	
Subgroup 2	40.00 (22.75)	67.67 (13.87)	42.96 (23.45)	
Subgroup 3	33.50 (6.86)	13.40 (21.38)	22.33 18.93)	
Subgroup 4	39.39 (24.73)	27.70 (18.49)	37.09 (22.84)	

[a]Healthy participants and people with reoccurring/persistent pain are presented separately for the intensity scale and pooled for the un-/pleasantness scale according to the applied statistical model.

Visual analogue scales ratings did not differ between people with reoccurring/persistent pain for the low and high stimulation and the different stimulation sites if LCR subgroup was not included in the model (see Results S1, Figure S1, and Table S2 for detailed results).

3.5.1 | Intensity scale

The model for the analysis of the ratings of the experimental stimulation on the intensity scale included the factor 'group', because the distribution across subgroups resulting from the LCR differed between people with reoccurring/persistent pain and healthy participants (see above). Table 3 summarizes the results of the LMM analysis of the ratings of experimental stimuli including LCR subgroup as a factor for the intensity scale.

Intensity ratings (for means and *SD* see Table 2) were different for low and the high stimulation as before in the analysis without LCR subgroup as a factor, as shown in Table 3 (main effect 'stimulation condition'). In line with the analysis without LCR subgroup as a factor, healthy participants and people with reoccurring/persistent pain did not rate the intensity of the stimuli differently (main effect 'group', Table 3). However, with respect to the LCR subgroups, the interaction of stimulation condition (low, high) with the LCR subgroup as well as the 3-way interaction of stimulation condition with LCR subgroup and group (healthy participants, people with pain) was significant (Table 3; Figure 3c,d). Thus, LCR subgroups rated perceived intensity of the high and low experimental heat stimuli differently dependent on group membership (healthy participants vs. people with pain). Post-hoc comparisons showed no differences in intensity ratings between subgroups for low stimuli across groups as well as within groups (all $p < .106$). Post-comparisons showed also no difference in intensity ratings of high stimuli between LCR subgroups for healthy participants (all $p < .519$), but did so in people with reoccurring/persistent pain a difference in the ratings of high stimuli between subgroup 1 versus 2 ($p = .020$), 2 versus 3 ($p = .007$), and 2 versus 4 ($p = .008$; Figure 3d).

3.6 | Un-/pleasantness scale

Since the distribution across subgroups was not different between people with reoccurring/persistent pain and healthy participants, 'group' (healthy participants, people with pain) was not included as a factor in the model for analyzing the ratings on the un-/pleasantness scale. Table 4 summarizes the results of the LMM analysis of the ratings of experimental

TABLE 3 Results of linear mixed model analysis of the intensity ratings with stimulation condition (high and low experimental stimulation), group (healthy participants and people with reoccurring/persistent pain) and subgroup ($N = 4$) from the latent class regression (LCR) of the response patterns in the free-positioning task as factors

	Intensity ratings	
Effect	***F* (*df* num, *df* den)**	***p*[a]**
Main effects		
Stimulation condition	10.769 (1, 71)	.002**
Group	1.283 (1, 68)	.261
LCR subgroup	1.338 (3, 70)	.269
Interactions		
Stimulation condition × LCR subgroup	6.235 (3, 72)	.001**
Group × LCR subgroup	0.347 (3, 70)	.850
Group × stimulation condition	0.126 (1, 72)	.723
Stimulation condition × group × LCR subgroup	4.048 (4, 72)	.002*

[a]Adjusted *F*-ratios, degrees of freedom for denominators (*df* den) and for numerators (*df* num) in brackets and exact probabilities for main effects and interactions. Significances given as per false discovery rate adjusted probabilities: *$p < .05$; **$p < .01$.

TABLE 4 Results of linear mixed model analysis of the un-/pleasantness ratings with stimulation condition (high and low experimental stimulation) and subgroup ($N = 4$) from the latent class regression (LCR) of the response patterns in the free-positioning task as factors. People with reoccurring/persistent pain and healthy participants are pooled in this analysis

	Un-/pleasantness ratings	
Effect	***F* (*df* num, *df* den)**	***p*[a]**
Main effects		
Stimulation condition	0.053 (1, 72)	.819
LCR subgroup	1.402 (3, 70)	.249
Interaction		
Stimulation condition × LCR subgroup	3.040 (3, 73)	.034[t]

[a]Adjusted F-ratios, degrees of freedom for denominators (*df* den) and for numerators (*df* num) in brackets and exact probabilities for main effects and interactions. Significances given as per false discovery rate adjusted probabilities: [t] < 0.10.

stimuli including LCR subgroup as a factor for the un-/pleasantness scale.

Overall un-/pleasantness ratings (for means and *SD* see Table 2) were not different for low and the high stimulation as before in the analysis without LCR subgroup as a factor, as shown in Table 4 (main effect 'stimulation condition'). Across stimulation conditions, LCR subgroups did not differ in their un-/pleasantness ratings of the experimental stimuli (main effect 'LCR subgroup'; Table 4). However, dependent on the stimulation condition LCR subgroups showed a trend for different un-/pleasantness ratings (interaction 'stimulation condition × LCR subgroup'; Table 4; Figure 3g). For low stimulation, post-hoc comparisons showed only a trend for difference in un-/pleasantness ratings between LCR subgroups 2 versus 3 ($p = .076$). Un-/pleasantness ratings of the high stimuli differed between subgroups 1 versus 2 ($p = .008$), 1 versus 3 ($p = .036$) and showed a trend for a difference between subgroups 1 versus 4 ($p = .054$; Figure 3g).

4 | DISCUSSION

The present results show that individuals have different conceptions of rating scale anchors, confirming previous results (Gracely & Dubner, 1987; Gracely et al., 1978; Heft et al., 1980; Morely, 1989; Tursky, Jamner, & Friedman, 1982). However, using a novel approach in utilizing individual conceptions of the relative position of verbal pain labels, revealed distinguishable subgroups of individuals with similar conceptions of pain scales and anchors. Confirming our hypothesis, including these subgroups when analyzing VAS ratings of experimental stimuli revealed differential ratings of these stimuli between the subgroups, hidden before in the overall average ratings.

Participants varied in their VAS ratings of experimental stimuli. Assessing individual conceptions of scales and their anchors might be a possibility to explain such variation. Interestingly, the subgroups revealed here showed conceptions on the relative position of scale anchors that were in part opposing. Possibly, this reflects previous observations that participants sometimes redefine a rating scale to fit their specific perception (Robinson-Papp, George, Dorfman, & Simpson, 2015; Williams et al., 2000). The present results suggest that some of the variability in pain ratings is due to different understandings of pain scale anchors and their relative position to each other.

Somewhat surprisingly and contradicting our hypothesis, we found no clearly separated clusters between people with reoccurring/persistent pain and healthy participants. People from both groups were present in all subgroups for the intensity and un-/pleasantness scale, albeit a small difference in the distribution of people from both groups was found for the intensity scale, suggesting that long-lasting/frequent pain experiences can result in some shift of pain conceptions, but not necessarily in a characteristic distinct shift or adjustment in conceptions of pain scale anchors. Previous qualitative studies that investigated variations in the understanding of rating scales focused on chronic pain patients (Dorfman et al., 2016; Robinson-Papp et al., 2015;

Williams et al., 2000), leaving it open whether pain-free individuals show similar variations in the understanding of pain scales. Although people with and without reoccurring/persistent pain appear comparable in several aspects in the present study, the analysis of the ratings of perceived intensity of high experimental stimuli in the different LCR subgroups suggests that the differences found in these ratings were driven by the people with reoccurring/persistent pain. People with reoccurring/persistent pain compared to the healthy participants belonging to the same LCR subgroup membership showed more extreme ratings of the experimental stimuli. It could be hypothesized that these differences in intensity ratings were, despite similar response patterns in the free positioning task, driven by different interpretations of the pain anchors, possibly resulting in more extreme answers on the VAS with only end anchors. This hypothesis has to be investigated in future studies.

Other factors than the presence of chronic pain might shape the understanding of pain rating scales and their anchors of individuals. For example, previous pain experience, unrelated to the presence of chronic pain, possibly influences how pain labels are personally anchored. Nevertheless, pain sensitivity, assessed here by thermal detection and pain threshold and tolerance, did not predict LCR subgroup membership in the present study (see Results S4).

Apart from aspects related to the experience of and sensitivity to pain, cognitive factors likely influence the understanding of pain labels and the ability to perform the free positioning task. Differentiating various pain labels and the free positioning task itself need good language mastery and an elaborated understanding of the pain descriptors. In line with this assumption, exploratory analyses showed that education level of participants predicted belonging to a specific LCR subgroup for the intensity, but not for the un-/pleasantness scale (see Results S4). Lower education level was associated with increased pain intensity for all pain labels. Whether this effect of education level was caused by a reduced understanding of the task and the pain labels or whether education level is generally associated with a specific conception of pain scales has to be investigated in future studies.

The present results show that depending on individual conceptions of rating scales, ratings of experimental stimuli on VAS can differ. In addition, results of the LCR illustrate that most resulting subgroups show deviations from a linear arrangement, with some patterns being directly opposing. While the pre-determined pain descriptions that had to be arranged in the free positioning task could be a reason, it nevertheless hints at different interpretations of rating scales that might affect NRS and VAS scales. For example, some people might perceive the lower part of a scale as compressed compared to the upper part or vice versa. Thus, the psychometric properties of NRS and VAS may be questioned, because resulting numbers possibly hide the fact that individuals interpret them in fundamentally different ways (cf., Dijkers, 2010; Lund et al., 2005; Ohnhaus & Adler, 1975; Williams et al., 2000).

The result that pain labels were not perceived in a linear, equidistant manner is particularly important with respect to the intensity scale. The pain descriptors positioned by the participants were retrieved from (Stein & Mendl, 1988), describing a validated German version of the McGill Pain Questionnaire (Melzack, 1975). Both the German and the original English versions of the Questionnaire were created with the goal to achieve a linear and equidistant scale. The current results contradict this assumption showing rather a quadratic trend of the distribution. Importantly, we only analyzed here the five pain labels from the McGill Questionnaire without including additional end scale anchors to avoid a bias in the analysis possibly inducing such a quadratic trend.

In qualitative studies, chronic pain patients expressed problems rating their multidimensional pain experience on unidimensional rating scales in terms of pain intensity/severity (Dorfman et al., 2016; Robinson-Papp et al., 2015; Williams et al., 2000). For example, patients mention that pain intensity/severity is not the dimension most prominent for them, but rather aversiveness and/or impaired functionality (Reading, 1980; Robinson-Papp et al., 2015; Serlin, Mendoza, Nakamura, Edwards, & Cleeland, 1995). Here, we focused on pain intensity and unpleasantness of pain. Results show that conceptions of scales can vary depending on the assessed scale dimension. Moreover, LCR subgroup belonging did not overlap across the intensity and un-/pleasantness scale, i.e. subgroup membership on one scale dimension was not associated with a specific subgroup membership on the other dimension (see Results S3). This finding emphasizes that interpretation of scales depend also on the assessed dimension and substantiates previous findings emphasizing the need to differentiate such dimensions. For example, distraction predominately affects intensity ratings, while emotional stimuli predominantly affect unpleasantness ratings (e.g. Villemure & Bushnell, 2002). Moreover, the described method might be of clinical utility possibly predicting differential effects on different dimensions of pain, which could be associated with different mechanisms underlying successful pain treatment. In line with this interpretation, it was recently shown that mindfulness-based compared to sham-mindfulness meditation engages different mechanisms resulting in pain relief with respect to perceived pain unpleasantness, while a similar differentiation between underlying mechanism could not be found in relation to perceived pain intensity (Adler-Neal et al., 2019).

The current study was a first of its kind in the context of pain showing that individual conceptions of pain labels can be assessed in a comparatively simple procedure. As such,

the study revealed promising results, but further work is needed to create an even easier to handle version that can be applied at the bedside. Most importantly, in its current form, individual predictions are not possible. Subgroups described in this study have to be replicated in a larger sample. Then, for each confirmed subgroup a non-linear polynomial function could be calculated that allows a "re-alignment" of individual ratings related to a specific scale conception to a linear ratio-scale VAS/NRS. In addition, the role of other factors likely modulating the response patterns in the free positioning task such as severity of chronic pain and pain duration, personality traits relevant in pain such as pain catastrophizing and optimism/pessimism should be investigated in future studies.

Nevertheless, the present study has limitations. Importantly, the unbalanced number of healthy participants and people with reoccurring/persistent pain could have biased the present results. Furthermore, participants of the first testing wave were significantly younger compared to those in the second testing wave. Age can affect pain perception, which was also observable in the present sample. Last, the criterion for pain presence used here, was comparatively low and possibly a stricter definition would have resulted in a more strongly impaired sample, possibly with different response patterns. Nevertheless, we like to point out that our sample reported considerable pain severity, interference and affective distress due to the pain (see Table 1) indicating that these participants were indeed affected by their pain.

In sum, the present study shows that conceptions of rating scales affect ratings of experimental stimuli. Including these conceptions in the analysis of these ratings can increase the model fit of statistical analyses and thus lead to an alignment of these assumptions with VAS ratings. These findings highlight a potential oversimplification of averaging ratings across participants in a sample, possibly leading to distortions of the resulting averages.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

AUTHOR CONTRIBUTION

SB, MS, and XF conceptualized and designed the experiment, SB, MS, XF, K S-E acquired the data, SB, MS and XF analysed the data, all authors discussed the data, were involved in the interpretation of the data, SB drafted the article, MS, XF and K S-E critically revised the article for important intellectual content. All authors approved the final version of this article.

ORCID

Susanne Becker https://orcid.org/0000-0002-5681-4084
Karin Schakib-Ekbatan https://orcid.org/0000-0002-3268-5935
Marcel Schweiker https://orcid.org/0000-0003-3906-4688

REFERENCES

Adler-Neal, A. L., Waugh, C. E., Garland, E. L., Shaltout, H. A., Diz, D. I., & Zeidan, F. (2019). The role of heart rate variability in mindfulness-based pain relief. *The Journal of Pain*. https://doi.org/10.1016/j.jpain.2019.07.003

Averbuch, M., & Katzper, M. (2004). Assessment of visual analog versus categorical scale for measurement of osteoarthritis pain. *The Journal of Clinical Pharmacology*, *44*(4), 368–372. https://doi.org/10.1177/0091270004263995

Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, *43*(3), 261–278. https://doi.org/10.1109/47.867942

Dannecker, E. A., George, S. Z., & Robinson, M. E. (2007). Influence and stability of pain scale anchors for an investigation of cold pressor pain tolerance. *Journal of Pain*, *8*(6), 476–482. https://doi.org/10.1016/j.jpain.2007.01.003

Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, *83*(401), 173–178. https://doi.org/10.1080/01621459.1988.10478584

Dijkers, M. (2010). Comparing quantification of pain severity by verbal rating and numeric rating scales. *Journal of Spinal Cord Medicine*, *33*(3), 232–242. https://doi.org/10.1080/10790268.2010.11689700

Dirks, J., Petersen, K. L., & Dahl, J. B. (2003). The heat/capsaicin sensitization model: A methodologic study. *The Journal of Pain*, *4*(3), 122–128. https://doi.org/10.1054/jpai.2003.10

Dorfman, D., George, M. C., Robinson-Papp, J., Rahman, T., Tamler, R., & Simpson, D. M. (2016). Patient reported outcome measures of pain intensity: Do they tell us what we need to know? *Scandinavian Journal of Pain*, *11*, 73–76. https://doi.org/10.1016/j.sjpain.2015.12.004

Flor, H., Knost, B., & Birbaumer, N. (2002). The role of operant conditioning in chronic pain: An experimental investigation. *Pain*, *95*(1), 111–118. https://doi.org/10.1016/S0304-3959(01)00385-2

Flor, H., Rudy, T. E., Birbaumer, N., Streit, B., & Schugens, M. M. (1990). Zur Anwendbarkeit des West Haven-Yale Multidimensional Pain Inventory im deutschen Sprachraum. *Der Schmerz*, *4*(2), 82–87. https://doi.org/10.1007/BF02527839

Fuchs, X., Becker, S., Schakib-Ekbatan, K., & Schweiker, M. (2018). Subgroups holding different conceptions of scales rate room temperatures differently. *Building and Environment*, *128*, 236–247. https://doi.org10.1016/j.buildenv.2017.11.034

Gandhi, W., Becker, S., & Schweinhardt, P. (2013). Pain increases motivational drive to obtain reward, but does not affect associated hedonic responses: A behavioural study in healthy volunteers. *European Journal of Pain*, *17*(7), 1093–1103. https://doi.org/10.1002/j.1532-2149.2012.00281.x

Gracely, R. H. (1990). Measuring pain in the clinic. *Anesthesia Progress*, *37*(2–3), 88–92.

Gracely, R. H., & Dubner, R. (1987). Reliability and validity of verbal descriptor scales of painfulness. *Pain*, *29*(2), 175–185. https://doi.org/10.1016/0304-3959(87)91034-7

Gracely, R. H., McGrath, P. A., & Dubner, R. (1978). Ratio scales of sensory and affective verbal pain descriptors. *Pain*, *5*(1), 5–18. https://doi.org/10.1016/0304-3959(78)90020-9

Grün, B., & Leisch, F. (2008). FlexMix version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal*

of Statistical Software, *28*(4), 1–35. https://doi.org/10.18637/jss.v028.i04

Hall, W. (1981). On "ratio scales of sensory and affective verbal pain descriptors". *Pain*, *11*(1), 101–107. https://doi.org/10.1016/0304-3959(81)90143-3

Heft, M. W., Gracely, R. H., & Dubner, R. (1980). A validation model for verbal descritor scaling of human clinical pain. *Pain*, *9*(3), 363–373. https://doi.org/10.1016/0304-3959(80)90050-0

Holzer, P. (1991). Capsaicin: Cellular targets, mechanisms of action, and selectivity for thin sensory neurons. *Pharmacological Review*, *43*(2), 143–201.

Kerns, D., & Turk, D. C. (1985). The West Haven-Yale Multidimensional Inventory (WHYMPI). *Pain*, *23*, 345–356. https://doi.org/10.1016/0304-3959(85)90004-1

Lautenbacher, S., Roscher, S., & Strian, F. (1995). Tonic pain evoked by pulsating heat: Temporal summation mechanisms and perceptual qualities. *Somatosensory & Motor Research*, *12*(1), 59–70. https://doi.org/10.3109/08990229509063142

Linton, S. J., & Götestam, G. K. (1983). A clinical comparison of two pain scales: Correlation, remembering chronic pain and a measure of compliance. *Pain*, *17*(1), 57–65. https://doi.org/10.1016/0304-3959(83)90127-6

Lund, I., Lundeberg, T., Sandberg, L., Budh, C. N., Kowalski, J., & Svensson, E. (2005). Lack of interchangeability between visual analogue and verbal rating pain scales: A cross sectional description of pain etiology groups. *BMC Medical Research Methodology*, *5*(1), 31. https://doi.org/10.1186/1471-2288-5-31

Melzack, R. (1975). The McGill Pain Questionnaire: Major properties and scoring methods. *Pain*, *1*(3), 277–299. https://doi.org/10.1016/0304-3959(75)90044-5

Morely, S. (1989). The dimensionality of verbal descriptors in Tursky's pain perception profile. *Pain*, *37*(1), 41–49. https://doi.org/10.1016/0304-3959(89)90151-6

Myles, P. S., Troedel, S., Boquest, M., & Reeves, M. (1999). The pain visual analog scale: Is it linear or nonlinear? *Anesthesia & Analgesia*, *89*(6), 1517–1520. https://doi.org/10.1097/00000539-199912000-00038

Ohnhaus, E. E., & Adler, R. (1975). Methodological problems in the measurement of pain: A comparison between the verbal rating scale and the visual analogue scale. *Pain*, *1*(4), 379–384. https://doi.org/10.1016/0304-3959(75)90075-5

Price, D. D., Bush, F. M., Long, S., & Harkins, S. W. (1994). A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. *Pain*, *56*(2), 217–226. https://doi.org/10.1016/0304-3959(94)90097-3

Price, D. D., McGrath, P. A., Rafii, A., & Buckingham, B. (1983). The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain*, *17*(1), 45–56. https://doi.org/10.1016/0304-3959(83)90126-4

R Development CoreTeam. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from Http://Www.r-Project.Org

Reading, A. E. (1980). A comparison of pain rating scales. *Journal of Psychosomatic Research*, *24*(3–4), 119–124. https://doi.org/10.1016/0022-3999(80)90032-X

Robinson-Papp, J., George, M. C., Dorfman, D., & Simpson, D. M. (2015). Barriers to chronic pain measurement: A qualitative study of patient perspectives. *Pain Medicine*, *16*(7), 1256–1264. https://doi.org/10.1111/pme.12717

Schweiker, M., Fuchs, X., Becker, S., Shukuya, M., Dovjak, M., Hawighorst, M., & Kolarik, J. (2017). Challenging the assumptions for thermal sensation scales. *Building Research & Information*, *45*(5), 572–589. https://doi.org/10.1080/09613218.2016.1183185

Serlin, R. C., Mendoza, T. R., Nakamura, Y., Edwards, K. R., & Cleeland, C. S. (1995). When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. *Pain*, *61*(2), 277–284. https://doi.org/10.1016/0304-3959(94)00178-H

Sluka, K. A. (2002). Stimulation of deep somatic tissue with capsaicin produces long-lasting mechanical allodynia and heat hypoalgesia that depends on early activation of the cAMP pathway. *Journal of Neuroscience*, *22*(13), 5687–5693. https://doi.org/10.1523/jneurosci.22-13-05687.2002

Stein, C., & Mendl, G. (1988). The German counterpart to McGill Pain Questionnaire. *Pain*, *32*(2), 251–255. https://doi.org/10.1016/0304-3959(88)90074-7

Tursky, B., Jamner, L. D., & Friedman, R. (1982). The pain perception profile: A psychophysical approach to the assessment of pain report. *Behavior Therapy*, *13*(4), 376–394. https://doi.org/10.1016/S0005-7894(82)80002-6

Villemure, C., & Bushnell, M. C. (2002). Cognitive modulation of pain: How do attention and emotion influence pain processing? *Pain*, *95*(3), 195–199. https://doi.org/10.1016/S0304-3959(02)00007-6

Wagner, A., Andersen, R. K., Zhang, H., de Dear, R., Schweiker, M., Goh, E., … Park, S. (2018). Laboratory approaches to studying occupants. In *Exploring occupant behavior in buildings* (pp. 169–212). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-61464-9_7

Williams, A. C. D. C., Davies, H. T. O., & Chadury, Y. (2000). Simple pain rating scales hide complex idiosyncratic meanings. *Pain*, *85*(3), 457–463. https://doi.org/10.1016/S0304-3959(99)00299-7

Williamson, A., & Hoggart, B. (2005). Pain: A review of three commonly used pain rating scales. *Journal of Clinical Nursing*, *14*(7), 798–804. https://doi.org/10.1111/j.1365-2702.2005.01121.x

Yokobe, J., Kitahara, M., Matsushima, M., & Uezono, S. (2014). Preference for different anchor descriptors on visual analogue scales among Japanese patients with chronic pain. *PLoS ONE*, *9*(6), 12–16. https://doi.org/10.1371/journal.pone.0099891

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Supinfo

How to cite this article: Becker S, Fuchs X, Schakib-Ektaban K, Schweiker M. What does "moderate pain" mean? Subgroups holding different conceptions of rating scales evaluate experimental pain differently. *Eur J Pain*. 2020;24:625–638. https://doi.org/10.1002/ejp.1514