ARTICLE IN PRESS

USASP US ASSOCIATION FOR THE STUDY OF PAIN

PUBLISHED BY ELSEVIER

The Journal of Pain, Vol xxx, No. xx (Month),xxxx: pp xxx–xxx
Available online at www.jpain.org and www.sciencedirect.com

Original Reports

Experimental Pain Picture System (EPPS): Development and Validation

Maryna Alves,[*,†] Angelos-Miltiadis Krypotos,[*,‡] Geert Crombez,[†] and Johan W.S. Vlaeyen[*,§]

Research Group Health Psychology, KU Leuven, Leuven, Belgium, †Department of Experimental-Clinical and Health Psychology, Ghent University, Gent, Belgium, ‡Department of Clinical Psychology, Utrecht University, Utrecht, the Netherlands, §Experimental Health Psychology, Maastricht University, Maastricht, the Netherlands

Abstract: Pain-related pictures are useful for studying how individuals respond to pain-related stimulation. Such pictures can occasionally be found in databases for affective pictures. However, a validated database specifically for pain-related pictures is not available yet. In 2 experiments (N = 185 and 103, respectively), we developed and validated the Experimental Pain Pictures System (EPPS). In both experiments, negative valence, arousal, and painfulness ratings were compared between neutral-, sad-, and pain-related pictures. The pain-related pictures represented both deep and superficial somatic pain. Across the 2 experiments, pain-related pictures were judged as more negative, arousing, and painful than neutral pictures and more painful than sad pictures. The final EPPS contains 50 pictures of different painful events considered moderately to highly painful by participants. The EPPS is a valuable tool for studying pain-related responses, as it gives researchers a choice among many validated pictures depicting different types of pain, increasing the comparability between studies.

***Perspective:** This article presents the validation of the experimental pain pictures system, which consists of a set of pain-related pictures. The experimental pain pictures system is composed of pictures depicting different types of pain. Participants rated all the pictures as being negative, arousing, and painful.*

***Key words:** Pain-related pictures, Database validation, Negative valence, Arousal, Painfulness*

Pain is a communicative event; it helps to signal threats through verbal and non-verbal cues (eg, facial expressions, vocal expressions, and body posture).[1] Not surprisingly, there is a vast interest in the study of pain, whether at a fundamental or clinical level.

Multiple tools are used in research to study pain. Among them, are painful physical stimulation (eg, electrocutaneous stimulation or thermal stimulation),[2,3] questionnaires (eg, the chronic pain grade questionnaire[4] or fear of pain questionnaire[5]), and pictures.[6-9]

Pain-related pictures are often used in pain research. For example, the photograph series of daily activities is used to detect which activities are considered harmful by patients with low back pain.[10,11] Pain-related pictures are also useful to understand more basic pain mechanisms, like attentional deployment towards pain,[12-15] and empathy for pain.[6-9] Sets of pain-related pictures have also been developed in the context of pain communication. Walsh and colleagues,[1] for example, developed a set of pictures containing

Received December 13, 2022; Received in revised form June 13, 2023; Accepted June 18, 2023
The study was supported by an FWO grant (Reg. #G071118N).
The authors have no conflicts of interest to declare.
Supplementary data accompanying this article are available online at www.jpain.org and www.sciencedirect.com.
Address reprint requests to Maryna Alves, MSc Research Group Health Psychology, KU Leuven, Tiensestraat 102, 3000 Leuven, Belgium.

1526-5900/$36.00

https://doi.org/10.1016/j.jpain.2023.06.014

pain-related body postures, and the Delaware pain database is a set of pictures of facial expressions of pain.[16]

Research has shown that pain-related pictures can lead to self-reported discomfort/distress and physiological arousal (ie, heart rate and skin conductance).[17,18] Pain-related pictures can also have an impact on the experience of pain.[6,7] In their study, de Wied and Verbaten,[19] exposed participants to various pictural stimuli. Those who saw unpleasant pictures including pain cues (eg, injury) showed lower pain tolerance compared to participants who were exposed to unpleasant pictures without pain cues.

Several studies using pain-related pictures use ad hoc selected pictures, often lacking standardization, hampering comparison between studies. Other studies use pain-related pictures that originate from existing databases. To illustrate, various pictures of injuries and accidents that could be used as pain-related stimuli can be found in the DIsgust-RelaTed-Images (DIRTI),[20] the Nencki Affective Picture System,[21] and International Affective Picture System (IAPS)[22] databases. In none of these databases the painful aspects of the pictures have been evaluated making them not specific enough for the study of pain. Therefore, a more general database with pain-related pictures depicting different types of pain (eg, injuries or headaches) is still needed. Such general pain-related pictures database could be used as pain-related stimuli and would help in the study of various aspects of pain, such as empathy for pain, attentional processes of pain, and classical conditioning in pain, among others.

In the present study, we validate throughout 2 experiments a first set of pain-related pictures meant to be used in research. In Experiment 1, we aimed to develop and validate a set of pain-related pictures. Participants were instructed to classify pictures of 3 different classes (sad-, neutral-, and pain-related pictures) and evaluate these pictures in terms of negative valence, arousal, and painfulness. These ratings were then used to select the pictures with the highest scores in these classes, which were cross-validated in Experiment 2. We expected pain-related pictures to be judged 1) more negatively than neutral pictures, 2) more arousing than sad and neutral pictures, and 3) more painful than sad and neutral pictures.

Experiment 1

Methods: Experiment 1

Participants

The exclusion criteria based on self-report were: 1) out of the age range of 18 to 35 years old (ie, representing the population of young adults), 2) current diagnosis of a psychiatric disorder, 3) diagnosis of other severe medical conditions, 4) pregnancy (ie, many pictures included in this study can trigger strong disgust emotions), 5) recovering from severe psychological or physical trauma, 6) advice from a general practitioner to avoid stress, 7) hemophobia, and 8) highly sensitive to negative content. We recommended subjects who considered themselves highly sensitive to negative content to not participate as they may perceive some pictures as highly distressing.

We decided to collect data from a sample of healthy young adults as it is a population commonly targeted in experimental studies.

A sample of 207 participants participated in the online study via Prolific (https://www.prolific.co/). Prolific is an online platform useful for the rapid recruitment of participants as it counts more than 130 thousand persons coming from countries members of the Organisation for Economic Co-operation and Development (OECD) and has more than 250 pre-selection filters allowing the recruitment of representative samples. The data of 19 participants were excluded due to technical problems and/or missing data (ie, did not complete the experiment), and the data of 3 participants were not included in the analyses due to effortless participation (eg, rated all the pictures with the same scores or categorized many pictures within the same class). In the end, 185 participants were included in the final analyses (M = 25.18 years old, SD = 5.41; 128 men, 56 women, and 1 other).

Participants received 2.50£ as compensation. The Social and Societal Ethics Committee of the KU Leuven approved this experiment (reg# G-2020-2025).

Materials

Pictures Selection. We selected 300 pictures representing 3 different classes: pain, sad, and neutral (one-hundred pictures per class). The sad pictures were chosen among other categories of pictures having a negative connotation for 3 main reasons. First, the sad class offers a wide variety of pictures of human beings. Second, sad pictures are more distinguishable from pain than other negative classes (eg, some disgust-related pictures can be considered painful). Finally, sadness is a general emotion that can be elicited by pictures within the other classes facilitating the collection of pictures (eg, anger, regret, or fear).

We selected pictures depicting men or women across ages, focusing either on human body parts (ie, a hand, back, mouth) or a person executing activities (eg, cooking, having a car accident, or crying). Pictures that could be perceived as too distressing (eg, pictures of dead bodies or ripped body parts) were not included. Free copyright pictures were searched via databases or the worldwide web. Specifically, ninety-four photographs were taken from the Nencki Affective Picture System database,[21] 24 from the DIRTI database,[20] and 11 from the IAPS database.[22] Additionally, 94 photographs were chosen from Flickr under a creative commons license (https://www.flickr.com/creativecommons/), and 77 from Adobe Stock (https://stock.adobe.com/). To select the pictures, we created a list of different types of pain (eg, headache, back pain, or a broken bone), and sad moments (eg, loneliness, funeral, or breaking-up event) (See Table 1). Next, we selected pictures that best represented each pain and sad class element. For the neutral class, we selected some pictures with similar events to the other classes but with a neutral connotation

Table 1. List of Pain and Sad Events Depicted in the EPPS

Painful events	Sad events
Headache/migraine	Arrestment
Back pain	Loneliness
Earache	People crying
Animal bite	Beggar
Ankle pain	Bullying
Knee pain	Psychiatric institutionalization
Wrist pain	Child abandon
Medical pain/Visceral pain	Dismiss
Injury from a bike or a car accident	Child abuse
Shoulder pain	Drug addiction
Chest pain	People fleeing
Genital pain	Death
Wounds	Poverty
Sunburnt	Homeless
Burnt	Cancer
Broken nose	Breaking-up
Toothache	Fight
Throat pain	Unsanitary housing
Frozen bite	Defeat
Broken bone	House fire
Injection	Domestic abuse
Cramp	Famine
Firearm shoot	Funeral
Stab wound	Flood
Cut	War
Heart attack	Sad face
Necrosis	Landfill scavenger
Tongue bite	Capitivity
Childbirth	
Stitches	
Uterine contractions	
Bruises	
Black eyes	
Amputation	
Aphthous ulcer	

(ie, a healthy arm was selected in contrast to a broken arm) and pictures not depicting any negative connotation. Concerning the pain-related pictures, only pictures of adults were selected (ie, no children, adolescents, or elderly). Out of the 100 pain-related pictures, 22 represented deep somatic pain (eg, headache or earache), whereas 78 represented superficial somatic pain, including different types of bodily injury (eg, burnt, animal bite, or wound with blood). We decided to differentiate and explore the difference between the 2 types of pain because the pain in the deep somatic pain-related pictures is communicated through body posture and facial expression of pain. In contrast, the superficial somatic pain-related pictures often contain external signs of body injury (eg, blood) which could affect the reactivity to the pictures. In 60 of the pain-related pictures, the ethnicity of the individuals was not recognizable as the pictures depict isolated body parts. White individuals were present in 30 pictures, Asian individuals were in 3 pictures, Black individuals in 6 pictures, and 1 picture showed a Middle-Easter man. Regarding the sex of the persons photographed in the pain-related pictures, 42 were male, 24 were female, and 34 were not identifiable. The diversity in the pictures was restricted by the available copyright-free pictures depicting pain. Forty-eight sad pictures presented sad faces in various situations, 20 represented poverty, 7 depicted mistreatments, and finally, 13 represented physical (eg, funeral), and material (eg, burnt house) losses. In the neutral class, 85 pictures depicted neutral to happy faces in various daily life situations and 15 pictures represented isolated healthy body parts (eg, a hand, an eye, or a foot).

Rating Scales. We used 3 separate 10-point numeric scales to rate the pictures in terms of negative valence, arousal, and pain (see instructions in Appendix 1). The negative valence scale referred to the connotation of the picture; the scale ranged from 1 (very positive) to 10 (= very negative). In the experiment, we used a valence scale ranging from 1 (= very negative) to 10 (= very positive). However, the scores of negative valence were reversed for the analyses to align this scale in the same direction as the arousal and pain scale. The arousal scale referred to the intensity of the emotion provoked by the picture and ranged from 1 (= very weak) to 10 = (very strong). Such self-reported arousal measure has been used previously to validate pictural databases (eg, BAPS-Ado[23] or DIRTI.[20] The pain scale concerned the level of pain evoked by the picture and ranged from 1 (= not painful) to 10 (= very painful).

Procedure

The experiment was programmed in Psychopy (version 3.00, 2020)[24] and was run online via the Pavlovia server (https://pavlovia.org/). Participants first read the information brochure, gave their informed consent, and answered the demographical information questions (ie, age and gender). To reduce the duration of the experiment and avoid cognitive overload, we divided the 300 pictures into 4 sets of 75 pictures each (25 per class). The 4 sets of pictures were evaluated by respectively 45, 45, 48, and 47 participants. Participants were randomly allocated to one of the four sets of pictures.

The main experimental task was divided into two phases: classification and evaluation phase. These two phases were separated by a 60-sec break. During the classification phase, each of the 75 pictures was displayed at the center of the screen. At the bottom of the screen were 3 words representing the following classes:

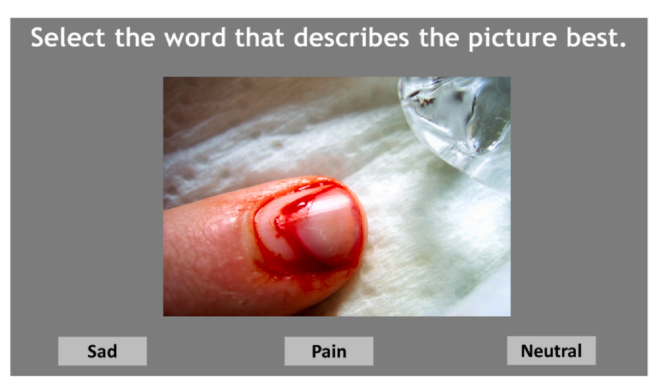

Figure 1. Example of a trial in the classification phase.

pain, sad, and neutral (see Fig 1 for an example of a trial of the classification phase). The position of each class was randomized across trials. On each trial, we instructed participants to carefully look at each picture and select one class that best matched the picture (see instructions in Appendix 1). The responses were collected using a mouse click. The pictures remained on the screen until a response was given. During the evaluation phase, each picture that had been previously presented was displayed again 3 times on the center of the screen, accompanied each time by one of the following scales: negative valence, arousal, or pain scale. The stimuli remained on the screen until a response was given.

Statistical Analyses

The data were analyzed with R.[25] The first step in our analyses was to calculate the classification score for each picture by attributing a score of 1 when the picture was classified according to the predefined class and 0 if not. Next, we calculated the mean score of the classification for each picture. Only those pictures that matched the predefined classes by at least 80% of the participants were for subsequent inferential analyses. This preselection was made to reduce the variability in ratings that pictures not representative of their classes could generate (eg, a sad picture identified as being pain-related by 40% of the participants would lead to higher scores of painfulness among these participants).

The second step in our analyses was to test our hypothesis following a multiverse analysis procedure that increased the transparency and validity of our results.[26] In other words, we conducted multiple analyses to show that our results are not test-depended.

First, differences in painfulness, negative valence, and arousal were analyzed with a factorial permutation test analysis using the ezPerm function from the ez package for R.[27] The ezPerm function is a non-parametric equivalent to the traditional analysis of variance (ANOVA), designed to be robust against violation of the normality and homoscedasticity assumptions. The permutation principle is a bootstrap process that repeatedly resamples the data (ie, 1,000 permutations), and calculates a statistic for each resampled dataset to form an empirical distribution for that statistic.[28] The *F*-statistic reported here is equivalent to the classical ANOVA, however, the *P*-value denoted as p_{perm} comes from the permutational test and is the probability of obtaining an *F*-statistic as extreme in our 1,000 permuted samples. The class (sad, pain, and neutral) was included in the ANOVA as a within-subjects factor, and gender (women and men) as a between-subjects factor. The gender other was not included in this analysis because of the low number of participants in this class ($n = 1$). Investigating the effect of gender on the rating of pictures was not the main objective of this study; however, this factor was included in the analyses because this is the usual analytical procedure in database validation (eg, IAPS).[22] Next, for the post hoc analyses, we conducted bootstrap pairwise contrasts (ie, 1,000 resamplings). As in the permutational test, we report for the post hoc the p_{boot} that is the *P*-value coming from the distribution of the 1,000 resampled t-test. We used non-parametric because the homogeneity of variances assumption required to run parametric tests was unmet (Appendix 2).

Second, we analyzed the differences between the different types of pain with a factorial permutation test with negative valence, arousal, and painfulness as dependent variables and type (deep somatic pain and superficial somatic pain) as the within-subjects factor. We conducted bootstrap pairwise contrasts to follow up on this test.

Third, we used the Spearman Rank Correlation test to explore associations among age, negative valence, arousal, and painfulness. The association was considered strong when $r \geq .60$, moderate when $.40 < r < .60$, and weak when $r \leq .30$.[29]

Supplementary tests were run to verify the reliability of our results. Specifically, we first analyzed the mean score of ratings of the pictures. To perform these analyses, we separately calculated a mean score of arousal, negative valence, and painfulness and the corresponding standard deviation for each picture. We compared the scores with a non-parametric Robust ANOVA with the classes (pain, sad, and neutral) and ratings (arousal, painfulness, and negative valence) as between-subject factors. The results of this test are in line with the results of the analyses mentioned above and can be found in Appendix 3. Next, we analyzed the data using the mixed effect models of the lmerTest package.[30] The negative valence, arousal, and painfulness scores were the dependent variables. The variables age, gender, and class were set as fixed effects and subjects as random effects. To select the best model to fit our data, we created 8 different models for each 1 of the dependent variables. Each model was composed of a specific combination of predictors. Finally, we ran a model comparison and selected the model with the lowest akaike information criterion (see Appendix 4 for more details). Finally, we conducted a cluster analysis using the k-means function from the stats package.[25] The cluster analysis allows us to verify if pictures from the same class are similar as it groups items sharing similar patterns. We used negative valence, arousal, and painfulness scores to cluster our initial pool of 300 pictures (Appendix 5).

Results Experiment 1

Classification Ratings

The classification analysis showed that 209 pictures (69.67%) were correctly classified by at least 80% of the participants. Among the final set 91, 82, and 36 pictures belonged to the neutral, pain, and sad class, respectively. Fig 2 presents a 3-D graph of the mean scores of negative valence, arousal, and painfulness for each 1 of the 209 pictures. The graph shows distinct separations among the means of the 3 variables, suggesting good discrimination between the different picture classes.

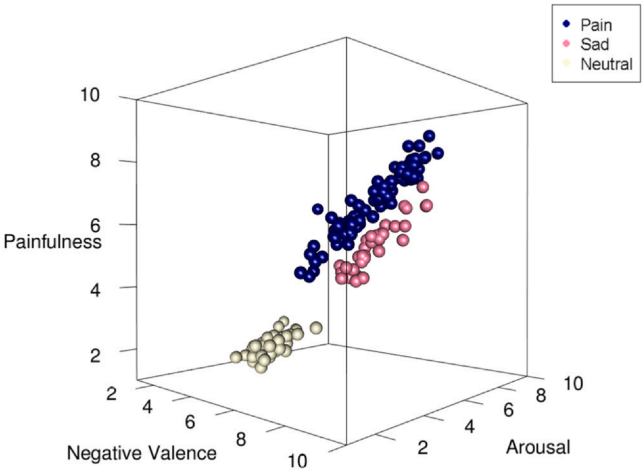

Figure 2. 3-D graph for the mean ratings per picture for negative valence, arousal, and painfulness in Experiment 1.

Negative Valence, Arousal, and Painfulness: Factorial Permutation Test

Negative Valence Ratings. The permutation factorial test revealed a main effect of class, $F(2, 180) = 1097.18$, $p_{perm} < .001$ (Fig 3, Table 2), no main effect of gender, $F(2, 179) = 2.22$, $p_{perm} = .108$, and a significant gender*class interaction, $F(4, 180) = 6.49$, $p_{perm} = .003$. Participants rated pain-related pictures as being more negative than neutral pictures, $t(180) = 40.50$, $p_{boot} = .002$ (estimate = 3.77, 95% bootstrap confidence interval (CI) [3.59, 3.95]). Sad pictures were also rated more negatively than neutral pictures, $t(180) = 39.42$, $p_{boot} = .002$ (estimate = 3.67, 95% bootstrap CI [3.49, 3.87]). No difference was found between sad and pain-related pictures for the negative valence ratings, $t(180) = 1.08$, $p_{boot} = .32$ (estimate = .10, 95% bootstrap CI [–.09, .29]). Negative valence scores among classes depended on gender: women judged sad pictures more negatively than men, $t(180) = 3.03$, $p_{boot} = .002$ (estimate = .45, 95% bootstrap CI [.17, .71]). No difference was found between gender for neutral pictures, $t(180) = -1.73$, $p_{boot} = .09$ (estimate = –.26, 95% bootstrap CI [–.56, .04]), or pain-related pictures, $t(180) = 1.49$, $p_{boot} = .10$ (estimate = .21, 95% bootstrap CI [–.03, .46]).

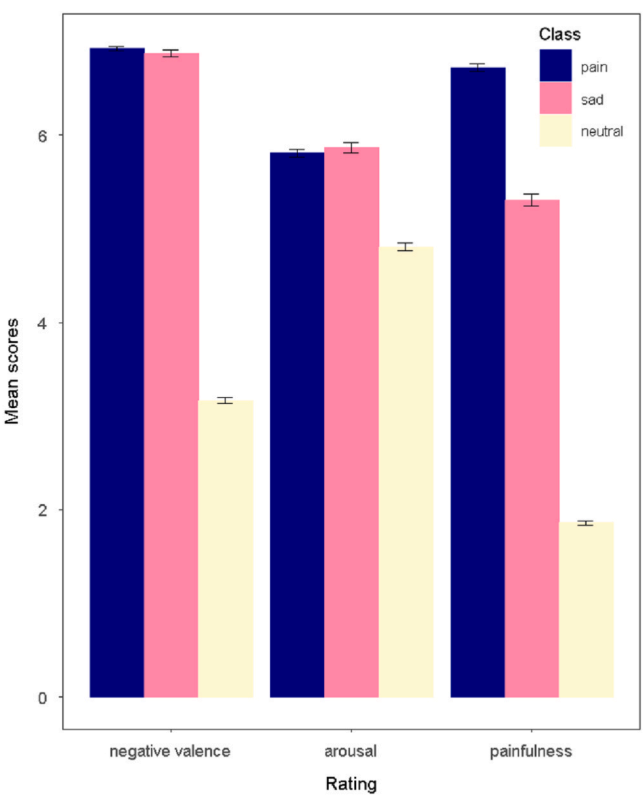

Figure 3. Interaction plot between the class (sad, pain, and neutral) and rating (negative valence, arousal, and painfulness) for Experiment 1. Error bars denote standard errors.

Table 2. Mean Scores and Standard Deviation Per Class for Experiment 1

	Negative valence		Arousal		Painfulness	
	Mean	SD	Mean	SD	Mean	SD
Pain	6.95	1.54	5.77	2.43	6.68	2.36
Sad	6.88	1.50	5.81	2.25	5.21	2.50
Neutral	3.13	1.99	5.81	2.57	1.179	1.50

Arousal Ratings. Arousal scores differed between classes, $F(2, 180) = 56.42$, $p_{perm} < .001$ (Fig 3, Table 2), but we found no significant evidence for a difference between gender, $F(2, 179) = 1.87$, $p_{perm} = .16$. The class*gender interaction was not significant, $F(4, 180) = 2.48$, $p_{perm} = .09$. Post hoc analyses showed that pain-related pictures are rated as more arousing than neutral pictures, $t(180) = 6.03$, $p_{boot} = .002$ (estimate = 1.00, 95% bootstrap CI [.64, 1.31]). Sad pictures were also rated as more arousing than neutral pictures, $t(180) = 6.32$, $p_{boot} = .002$ (estimate = 1.05, 95% bootstrap CI [.70, 1.39]). No difference was found between sad and pain-related pictures for arousal, $t(180) = -2.95$, $p_{boot} = .76$ (estimate = –.05, 95% bootstrap CI [–.39, .27]).

Painfulness Ratings. For painfulness, we found a significant main effect of the class, $F(2, 180) = 833.20$, $p_{perm} < .001$ (Fig 3, Table 2), a main effect of gender, $F(2, 179) = 5.01$, $p_{perm} = .02$, and a significant class*gender interaction, $F(4, 180) = 4.88$, $p_{perm} = .01$. The results showed that pain-related pictures are more painful than neutral pictures, $t(180) = 32.45$, $p_{boot} = .002$ (estimate = 4.85, 95% bootstrap CI [4.59 5.10]), and sad pictures, $t(180) = 9.86$, $p_{boot} = .002$ (estimate = 1.47, 95% bootstrap CI [1.14, 1.81]). Sad pictures were also rated as more painful than neutral ones, $t(180) = 22.59$, $p_{boot} = .002$ (estimate = 3.37, 95% bootstrap CI [3.07, 3.66]). The permutational test showed a significant effect of the gender. However, the post hoc analyses showed that this effect is small and leading no significant difference between women and men for the painfulness ratings, $t(179) = 1.64$, $p_{boot} = .07$ (estimate = .38, 95% bootstrap CI [–.03, .85]). The post hoc analyses for the class*gender interaction revealed that women rated sad pictures as being more painful than men, $t(180) = 3.03$, $p_{boot} = .001$ (estimate = .45, 95% bootstrap CI [.17, .70]). No difference between women and men was found for the painfulness rating of neutral, $t(180) = -1.73$, $p_{boot} = .09$ (estimate = −.26, 95% bootstrap CI [−.57, .03]), or pain-related pictures, $t(180) = 1.49$, $p_{boot} = .09$ (estimate = .21, 95% bootstrap CI [−.06, .44]).

Types of Pain: Factorial Permutation Test

Within the pain class, pictures can be divided into 2 subclasses: deep somatic pain (eg, headaches) and superficial somatic pain (eg, finger cut). The results show that superficial somatic pain-related pictures have been rated more negatively, $F(1, 182) = 280.9$, $p_{perm} < .001$, more arousing, $F(1, 182) = 109.2$, $p_{perm} < .001$, and more painful, $F(1, 182) = 176.3$, $p_{perm} < .001$ than deep somatic pain-related pictures (see Table 3 for mean scores).

Associations Among Negative Valence, Arousal, Painfulness, and Age

The results (Table 4) indicated that the more arousing or/and painful a picture is, the more negatively the picture is judged. We also found a positive correlation between painfulness and arousal indicating that pictures rated as highly painful are also more arousing. We found no any evidence that age was related to negative valence, arousal, or painfulness.

Discussion: Experiment 1

In Experiment 1, we aimed to develop a set of pain-related pictures. As expected, pain-related pictures were evaluated as more negative and arousing than neutral pictures and more painful than sad and neutral pictures. However, this study has some limitations. The first limitation concerns the instructions for the pain scale. We asked participants to rate the painfulness of the pictures without specifying the physical/bodily nature of the pain, which could explain the painfulness scores in the sad class. The second limitation is the selection of pictures. Several pictures did not reach a high classification score, indicating that they were not representative of their classes (ie, more specifically, 18 pain-related pictures, 9 neutral pictures, and 64 sad pictures were not representative of their classes).

Due to the above mentioned limitations, we conducted a second experiment. In Experiment 2, we modified the instruction for the painfulness scale and explicitly asked participants to rate the pictures 'physical/bodily' pain aspects. We specified the bodily/physical nature of the pain to avoid confusion between sad and pain-related pictures as there is a linguistic overlap between sadness and painfulness. This overlap is materialized by some 'sadness-pain concepts' such as heartache, broken heart, or hurt feelings that are associated with specific sad situations (eg, the death of a loved one or a breakup) and to specific body parts (eg, chest or heart).[31] To tackle the second limitation, we selected pictures of Experiment 1 with the highest classification scores.

Table 3. Mean Scores and Standard Deviation Per Type of Pain for Experiment 1

	Negative valence		Arousal		Painfulness	
	Mean	SD	Mean	SD	Mean	SD
Deep somatic pain	6.30	1.57	5.12	2.39	5.93	2.33
Superficial somatic pain	7.22	1.45	6.13	2.39	7.09	2.27

Table 4. Spearman Rank Correlation Between Age, Negative Valence, Arousal, and Painfulness for Experiment 1

	Valence	Arousal	Painfulness
Valence			
Arousal	.095		
Painfulness	.694	.406	
Age	-.0001	-.017	-.018

Experiment 2

Methods: Experiment 2

Participants

We recruited 115 participants via the Prolific online platform (https://www.prolific.co/). The exclusion criteria were the same as in the previous experiment, and none of the participants had participated in Experiment 1. Technical problems led to the exclusion of data from 6 participants, while 6 other participants' data were excluded due to missing data. One hundred and 3 participants were included in the analyses ($M = 24.25$ years old, $SD = 4.96$; 66 men, 35 women, and 2 others). Participants received compensation of 2.50£ for their participation in the experiment. The Social and Societal Ethics Committee of the KU Leuven approved this study (reg# G-2020-2025).

Materials

Pictures Selection. In the first experiment, several pictures did not reach a high classification score, meaning that some pictures did not represent their classes. To tackle this limitation, we selected the 150 pictures with the highest mean classification scores (50 pictures per class) from the set of Experiment 1. To have good variability among the pictures of the same class, if multiple similar pictures reached a high classification score, we selected only 1 of them (eg, if 2 pictures showing a toothache were among the 50 having the highest score of classification, we selected only 1 of them and replaced the other by the 51st picture). The set of pain-related pictures contains 9 pictures of deep somatic pain and 41 superficial somatic pain. The sad class includes 24 pictures of sad faces in different situations, 18 pictures of poverty, 2 of mistreatments, and finally, 6 pictures of physical and material losses (eg, funerals or burnt houses). The neutral class includes 46 pictures of neutral to happy faces (eg, working, fishing, camping), and 4 isolated healthy body parts (eg, torso, lips, hands).

Among the pain-related pictures, the ethnicity of individuals was not clearly recognizable in 35 pictures, Black individuals were presented in 2 pictures, Asian individuals in 2 pictures, White individuals in 10, and Middle Eastern man in 1 picture. The sex of individuals was represented as follows: 21 male, 10 female, and 19 unidentifiable (ie, body parts).

Ratings. We used the same negative valence and arousal scales as in Experiment 1. However, for the painfulness scale, we explicitly asked participants to report the bodily/physical pain associated with the picture ranging from 1 (= not painful) to 10 (= very painful).

Procedure

The procedure of this experiment was the same as the previous one. The experiment started with reading and signing the informed consent and was followed by the classification and evaluation phase. The pictures were divided into 2 sets of 75. Both sets were validated by 51 and 52 participants, respectively.

Statistical Analyses

We followed the same analytical procedure as in Experiment 1. Furthermore, to test any differences between the results of Experiments 1 and 2, we compared the mean ratings of both experiments using a factorial permutation test with negative valence, arousal, and painfulness, as dependent variables, class (pain, sad, and neutral) as within-subjects factor and experiment (Experiment 1 and Experiment 2) as between-subjects factor. For this analysis, we selected exclusively the pictures of Experiment 1 that were selected for Experiment 2. This analysis was conducted to quantify the difference in terms of negative valence, arousal, and painfulness ratings between experiments 1 and 2. As in Experiment 1, we also conducted supplementary analyses that can be found in Appendixes 3, 5, 6, and 7 (ie, mixed model analyses, cluster analysis, and ANOVA of the mean scores ratings).

Results: Experiment 2

Classification Ratings

From the 150 initial pictures, 133 (88.66%) were correctly classified by at least 80% of the participants, of which 50, 49, and 35 were pain-related, neutral, and sad pictures, respectively. Only pictures correctly classified were included in the remaining analyses.

Fig 4 shows a 3-D graph for the mean scores of the negative valence, arousal, and painfulness for the 133 pictures selected in this experiment.

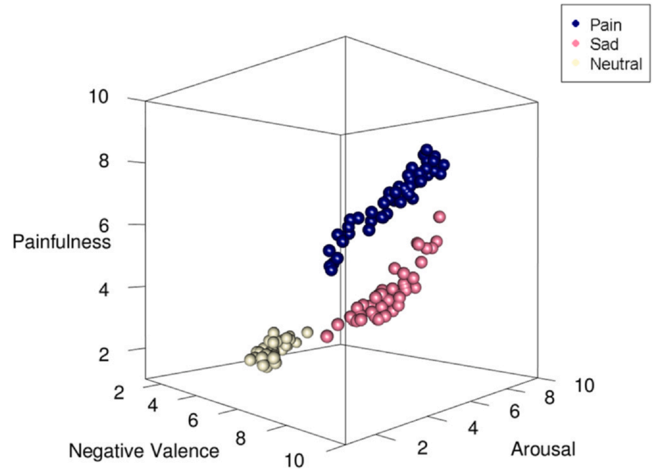

Figure 4. 3-D graph for the individuals' ratings of negative valence, arousal, and painfulness in Experiment 2.

Negative Valence, Arousal, and Painfulness: Factorial Permutation Test

Negative Valence Ratings. The permutation factorial test revealed a main effect of class, $F(2, 96) = 560.09$, $p_{perm} < .001$ (Fig 5, Table 5), and a main effect of gender, $F(1, 95) = 3.68$, $p_{perm} = .03$. The class*gender interaction was not significant, $F(4, 96) = .84$, $p_{perm} = .42$. The post hoc analyses show that neutral pictures are rated less negatively than pain-related pictures, $t(96) = 27.97$, $p_{boot} = .002$ (estimate = 4.30, 95% bootstrap CI

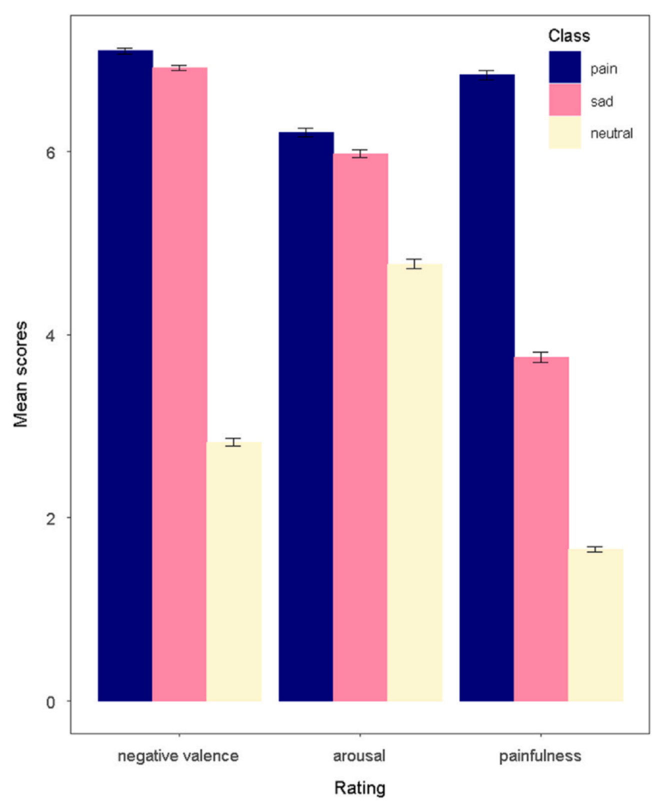

Figure 5. Interaction plot between the class (sad, pain, and neutral) and rating (negative valence, arousal, and painfulness) for Experiment 2. Error bars denote standard errors.

Table 5. Mean Scores and Standard Deviation Per Class for Experiment 2

	Negative valence		Arousal		Painfulness	
	Mean	SD	Mean	SD	Mean	SD
Pain	7.10	1.67	6.21	2.26	6.84	2.52
Sad	6.91	1.49	5.93	2.26	3.68	2.70
Neutral	2.80	1.99	4.76	2.57	1.64	1.46

Table 6. Mean Scores and Standard Deviation Per Type of Pain for Experiment 2

	Negative valence		Arousal		Painfulness	
	Mean	SD	Mean	SD	Mean	SD
Deep somatic pain	6.19	1.39	4.77	2.26	5.24	2.42
Superficial somatic pain	7.29	1.66	6.49	2.42	7.14	2.42

[3.99, 4.62]), and sad pictures, $t(96) = 26.79$, $p_{boot} = .002$ (estimate = 4.12, 95% bootstrap CI [3.83, 4.40]). No difference was found between sad and pain-related pictures for the negative valence, $t(96) = 1.18$, $p_{boot} = .19$ (estimate = .18, 95% bootstrap CI [−.10, .49]). The post hoc analyses did not show a difference between women and men for the negative valence, $t(95) = 1.18$, $p_{boot} = .30$ (estimate = .28, 95% bootstrap CI [−.27, .85]).

Arousal Ratings. For the arousal ratings, we found a main effect of the class, $F(1, 96) = 58.24$, $p_{perm} < .001$ (Fig 5, Table 5), no main effect of gender, $F(1, 95) = .41$, $p_{perm} = .60$, and a significant class*gender interaction, $F(4, 96) = 5.67$, $p_{perm} = .01$. Pain-related pictures were considered more arousing than neutral pictures, $t(96) = 6.22$, $p_{boot} = .002$ (estimate = 1.40, 95% bootstrap CI [.96,1.846]). Sad pictures were also rated more arousing than neutral ones, $t(96) = 5.20$, $p_{boot} = .002$ (estimate = 1.17, 95% bootstrap CI [.73, 1.59]). No significant difference was found between pain and sad pictures for arousal, $t(96) = 1.02$, $p_{boot} = .30$ (estimate = .23, 95% bootstrap CI [−.21, .633]). For the latter, men judged neutral pictures more arousing than women, $t(96) = -2.22$, $p_{boot} = .01$ (estimate = −.74, 95% bootstrap CI [−1.40, −.14]). No difference was found between women and men for the rating of pain-related pictures, $t(96) = -.05$, $p_{boot} = .99$ (estimate = −.02, 95% bootstrap CI [−.66, .70]), and sad pictures, $t(96) = .64$, $p_{boot} = .47$ (estimate = .20, 95% bootstrap CI [−.41, .83]).

Painfulness Ratings. The factorial permutation test shows a main effect of the class, $F(2, 96) = 443.96$, $p_{perm} < .001$ (Fig 5, Table 5), no main effect of gender, $F(2, 95) = 1.113$, $p_{perm} = .30$, and a significant class*gender interaction, $F(4, 96) = 4.50$, $p_{perm} = .015$. Pain-related pictures were considered as being more painful than sad, $t(96) = 13.40$, $p_{boot} = .001$ (estimate = 3.09, 95% bootstrap CI [1.63, 2.49]), and neutral pictures, $t(96) = 22.41$, $p_{boot} = .001$ (estimate = 5.17, 95% bootstrap CI [4.75, 5.55]). Sad pictures were also rated more painful than neutral ones, $t(96) = 9.01$, $p_{boot} = .001$ (estimate = 2.08, 95% bootstrap CI [1.63, 2.49]). Women rated sad pictures as being more painful than men, $t(96) = 2.20$, $p_{boot} = .021$ (estimate = .90, 95% bootstrap CI [.08, 1.78]). The rating for pain-related pictures, $t(96) = .28$, $p_{boot} = .79$ (estimate = .10, 95% bootstrap CI [−.59, .80]), or neutral pictures, $t(96) = -.69$, $p_{boot} = .42$ (estimate = −.16, 95% bootstrap CI [−.55, .26]), did not differ between gender.

Deep Types of Pain: Factorial Permutation Test

Deep somatic pain and superficial somatic pain were again rated differently. The results show that superficial somatic pain-related pictures have been rated more negatively, $F(1, 96) = 121.3$, $p_{perm} < .001$, more arousing, $F(1, 96) = 112$, $p_{perm} < .001$, and more painful, $F(1, 96) = 188.2$, $p_{perm} < .001$ than deep somatic pain-related pictures (see Table 6 for mean scores).

Associations Among Negative Valence, Arousal, Painfulness, and Age

The results (Table 7) indicate that the negative valence negatively correlates with arousal and pain. As the pain rating of a picture increased, so did its arousal and negative valence scores. The painfulness and arousal scores are positively correlated. No correlations were found between the age and the negative valence ratings, or between age and arousal, nor between age and painfulness.

Experiments Comparison

When looking into the negative valence, results revealed no main effect of the experiment, $F(2, 279) = .024$, $p_{perm} = .95$, and no significant class*experiment interaction, $F(2, 280) = .62$, $p_{perm} = .50$. As expected, results show a main effect of the classwhen analyzing the data of the 2 experiments together, $F(2, 280) = 1721.64$, $p_{perm} < .001$. Pain-related pictures were rated more negatively than neutral, $t(281) = -51.28$, $p_{boot} = .001$ (estimate = −4.38, 95% bootstrap CI [−4.56, −4.2]), and sad pictures, $t(281) = 3.58$, $p_{boot} = .001$ (estimate = .306, 95% bootstrap CI [.14, .48]). Sad pictures were also rated more negatively than neutral ones, $t(281) = -47.70$, $p_{boot} = .001$, (estimate = −4.07, 95% bootstrap CI [−4.2, −3.9]).

Our findings indicate a similar pattern of results for arousal scores. No main effect of the experiment, $F(2,$

Table 7. Spearman Rank Correlation Between Age, Negative Valence, Arousal, and Painfulness for Experiment 2

	Negative valence	Arousal	Painfulness
Negative valence			
Arousal	.181		
Painfulness	.545	.445	
Age	.019	.045	.037

280) = .018, p_{perm} = .95, or significant experiment*class interaction, $F(2, 279) = 2.09$, p_{perm} = .136, but a significant main effect of the class, $F(2, 280) = 89.41$, p_{perm} < .001. Neutral pictures were rated as less arousing than pain-related pictures, $t(281) = -8.92$, p_{boot} = .001, (estimate = −1.22, 95% bootstrap CI [−1.48, −.93]), and sad pictures, $t(281) = -6.66$, p_{boot} = .001, (estimate = −.91, 95% bootstrap CI [−1.17, −.64]). Pain-related pictures were seen as being more arousing than sad pictures, $t(281) = 2.25$, p_{boot} = .03, (estimate .30, 95% bootstrap CI [.05, .57]).

Regarding the painfulness ratings, we found a main effect of the experiment, $F(2, 279) = 2.09$, p_{perm} < .001, a main effect of the class, $F(2, 280) = 1298.58$, p_{perm} < .001, and a significant experiment*class interaction, $F(2, 220) = 23.53$, p_{perm} < .001. Pain-related pictures were rated as more painful than neutral, $t(281) = -40.32$, p_{boot} = .001, (estimate = −5.30, 95% bootstrap CI [−5.51, −5.08]), and sad pictures, $t(281) = 17.67$, p_{boot} = .001, (estimate = 2.32, 95% bootstrap CI [2.03, 2.58]). Sad pictures were also rated as more painful than neutral ones, $t(281) = -22.64$, p_{boot} = .001, (estimate = −2.97, 95% bootstrap CI [−3.23, −2.74]).

The general rating for painfulness in Experiment 1 was higher than in Experiment 2, $t(284) = 3.28$, p_{boot} = .003, (estimate = .62, 95% bootstrap CI [.24, 1.01]). The class*experiment interaction shows higher scores of painfulness for sad pictures in Experiment 1 compared to Experiment 2, $t(281) = 6.46$, p_{boot} = .001, (estimate = 1.48, 95% bootstrap CI [.99, 1.95]). We found no difference between the 2 experiments for neutral, $t(281) = .79$, p_{boot} = .44, (estimate = .10, 95% bootstrap CI [−.17, .35]), or pain-related pictures, $t(281) = 1.54$, p_{boot} = .15, (estimate = .29, 95% bootstrap CI [−.12, .66]).

Discussion: Experiment 2

Experiment 2 was conducted to confirm and cross-validate a set of pain-related pictures. We obtained the same pattern of results as in Experiment 1. Pain-related pictures were rated as more negative, arousing, and painful than neutral pictures. Pain-related pictures only differ from sad pictures in terms of painfulness showing that pain-related pictures were more painful than sad ones. Specifying the physical nature of the pain decreased the painfulness ratings for sad pictures, however, sad pictures were still judged as more painful than neutral ones.

General Discussion

This study aimed to develop and validate a set of pain-related pictures destinated to be used in future research. To achieve our aim, we compared the negative valence, arousal, and painfulness aspects of pain-related pictures to neutral and sad pictures. The statistical analyses revealed a similar pattern of results across the 2 experiments: neutral pictures were evaluated as less negative and less arousing than pain-related pictures; no differences were found between pain and sad pictures for arousal and negative valence ratings; pain-related pictures were rated as more painful than neutral or sad pictures. Collectively, all the items composing the final set of pain-related pictures were judged as moderately to highly painful by participants in both experiments.

Participants judged superficial somatic pain-related pictures as more negative, arousing, and painful than deep somatic pain ones. The difference between deep and superficial somatic pain-related pictures can be explained by former pictures containing depictions of external harm, such as blood or bruises. However, deep somatic pain-related pictures were still considered painful, negative, and arousing. The explanation for this finding is that pain can be communicated non-verbally via facial expressions or/and body posture, even in the absence of external injury.[1,32] Faces seemed to play an essential role in the painfulness ratings. For example, a picture of an isolated arm in a plaster cast was not recognized as painful, whereas a similar picture showing a man entirely with a leg in plaster was considered highly painful. The ability to recognize pain faces is already present in early childhood[15,33] and probably has an adaptive role as it signals the presence of a threat. Not surprisingly, individuals tend to pay more attention to facial expressions of pain than to neutrals ones.[34] Besides the facial expression, body postures relating to pain (ie, specific body postures that an individual in pain can take) are another good indicator of pain and pain intensity.[35] In all the deep somatic pain pictures, specific body postures of pain were depicted (eg, a man sited on the ground holding his knee or a woman holding her neck). In sum, the pain-related body postures and facial expressions could explain the negative valence and arousal ratings for the deep somatic pain-related pictures.

When considering the effect of gender on the rating of pain-related pictures, our results did not indicate a difference between women and men for the negative valence, arousal, or painfulness in either experiment. This finding suggests that women and men rated pain-related pictures similarly in both experiments. The literature on gender differences in visual pain recognition and evaluation presents contradictory evidence.[36,18,35] Preis and Kroener-Herwig[18] reported in their study that women rated pictures as more painful than men, whereas in the study of Walsh and colleagues[35] no difference between participants' sex was reported for the pain evaluation.

Unexpectedly, sad pictures were judged as somewhat painful in both experiments. In Experiment 1, the nature of pain was not mentioned in the instruction of the painfulness scale. In Experiment 2, we instructed participants to rate the physical/bodily pain depicted by the pictures, decreasing the painfulness ratings for the sad pictures. Despite this modification of instructions, sad pictures were still considered more painful than neutral ones. These sad pictures illustrated events such as a funeral or a woman crying in front of a burnt house, for example. Even when subjects were asked to evaluate the physical pain of pictures, sad pictures without any signs of bodily harm, injury, or sickness were considered painful. The painfulness scores of sad pictures could be due to the contrast between sad pictures (that induce a negative affect, like the pain-related pictures) and neutrals (inducing rather a positive affect). A similar contrast effect is reported by Herzog and collaborators,[37] who asked participants to rate the level of dyspnea in individuals depicted via positive, neutral, negative, and dyspnea-related pictures. Participants attributed higher scores of dyspnea to negative (not dyspnea-related) pictures compared to positive and neutral ones. Accordingly, the painfulness scores of sad pictures can be explained by sad pictures being perceived as more painful when being in compared to neutral pictures. Another explanation for the painfulness ratings of sad pictures could be the overlapping between facial expressions of pain and sadness. It has been shown that pain and sadness share some facial action units (eg, brow lower or jaw drop).[38] Furthermore, some sad pictures that we selected show complex scenarios (eg, poverty) which could trigger emotions other than sadness, such as anger, fear, or frustration. Sadness is a complex emotion that can be decomposed into more nuanced feelings like guilt, hopelessness, grief, shame, or fear.[31,39] Nevertheless, primary purpose of including the sad class was to serve as a negative control for the pain class, and this objective was met.

Our study has some limitations. First, we conducted both experiments online, which reduced the control over the experimental conditions in which the task was completed. To prevent possible biases, we took several precautions. We excluded participants who took an extended time to complete the experiment or showed effortless performance. Nevertheless, the clear concordance between the results of the 2 experiments increases our confidence that the data are reliable. Second, we collected a limited amount of demographic information about participants (ie, gender and age), and we have no information about other factors such as socio-economic background, race, or medical history that impact how individuals perceive the pictures. Third, the set of pictures contains mainly pictures of acute pain and was validated by a group of healthy individuals making the EPPS less relevant for clinical populations. Fourth, we did not match the pictures between the different classes in terms of the complexity of the pictures (eg, colors or luminosity) which could influence the perception of the pictures. Finally, we did not measure the inter-rate reliability when selecting the first pool of pictures, which could have impacted on the clustering of pictures (eg, some pictures might not be well representative of their respective classes). To address this limitation, we included only pictures with a classification score of at least 80% in the analyses.

Conclusions

In conclusion, pain-related pictures have been used in the literature, but no validated database containing various types of pain existed yet. The present study fills this gap by testing the face validity and reliability of a set of 50 pain-related pictures, which is available on request for research purposes. We hope that the wide use of this dataset will prove invaluable in understanding pain.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jpain.2023.06.014.

References

1. Walsh J, Eccleston C, Keogh E: Pain communication through body posture: the development and validation of a stimulus set. Pain 155(11):2282-2290, 2014. https://doi.org/10.1016/j.pain.2014.08.019

2. Tsuji T, Arikuni F, Sasaoka T, Suyama S, Akiyoshi T, Soh Z, Hirano H, Nakamura R, Saeki N, Kawamoto M, Yoshizumi M, Yoshino A, Yamawaki S: Peripheral arterial stiffness during electrocutaneous stimulation is positively correlated with pain-related brain activity and subjective pain intensity: an fMRI study. Sci Rep 11(1):4425, 2021. https://doi.org/10.1038/s41598-021-83833-6

3. Umeda M, Okifuji A: Comparable conditioned pain modulation and augmented blood pressure responses to cold pressor test among resistance exercisers compared to healthy controls. Biol Psychol 153:107889, 2020 https://doi.org/10.1016/j.biopsycho.2020.107889

4. Smith BH, Penny KI, Purves AM, Munro C, Wilson B, Grimshaw J, Chambers WA, Smith WC: The Chronic Pain Grade questionnaire: validation and reliability in postal research. Pain 71(2):141-147, 1997. https://doi.org/10.1016/s0304-3959(97)03347-2

5. McNeil DW, Rainwater 3rd AJ: Development of the fear of pain questionnaire III. J Behav Med 21(4):389-410, 1998. https://doi.org/10.1023/a:1018782831217

6. Botan V, Bowling NC, Banissy MJ, Critchley H, Ward J: Individual differences in vicarious pain perception linked to heightened socially elicited emotional states. Front Psychol 9:2355, 2018. https://doi.org/10.3389/fpsyg.2018.02355

7. Fitzgibbon BM, Giummarra MJ, Georgiou-Karistianis N, Enticott PG, Bradshaw JL: Shared pain: from empathy to synaesthesia. Neurosci Biobehav Rev 34(4):500-512, 2010. https://doi.org/10.1016/j.neubiorev.2009.10.007

8. Ren Q, Lu X, Zhao Q, Zhang H, Hu L: Can self-pain sensitivity quantify empathy for others' pain? Psychophysiology 57(10):e13637, 2020 https://doi.org/10.1111/psyp.13637

9. Timmers I, Park AL, Fischer MD, Kronman CA, Heathcote LC, Hernandez JM, Simons LE: Is empathy for pain unique in its neural correlates? A meta-analysis of neuroimaging studies of empathy. Front Behav Neurosci 12:289, 2018. https://doi.org/10.3389/fnbeh.2018.00289

10. Leeuw M, Goossens ME, van Breukelen GJ, Boersma K, Vlaeyen JW: Measuring perceived harmfulness of physical activities in patients with chronic low back pain: the photograph series of daily activities—short electronic version. J Pain 8(11):840-849, 2007. https://doi.org/10.1016/j.jpain.2007.05.013

11. Trost Z, France CR, Thomas JS: Examination of the photograph series of daily activities (PHODA) scale in chronic low back pain patients with high and low kinesiophobia. Pain 141(3):276-282, 2009. https://doi.org/10.1016/j.pain.2008.11.016

12. Chan FHF, Suen H, Jackson T, Vlaeyen JWS, Barry TJ: Pain-related attentional processes: a systematic review of eye-tracking research. Clin Psychol Rev 80:101884, 2020. https://doi.org/10.1016/j.cpr.2020.101884

13. Crombez G, Van Ryckeghem DML, Eccleston C, Van Damme S: Attentional bias to pain-related information: a meta-analysis. Pain. 154(4):497-510, 2013. https://doi.org/10.1016/j.pain.2012.11.013

14. Fashler SR, Katz J: Keeping an eye on pain: investigating visual attention biases in individuals with chronic pain using eye-tracking methodology. J Pain Res 9:551-561, 2016. https://doi.org/10.2147/JPR.S104268

15. Pilch M, O'Hora D, Jennings C, Caes L, McGuire BE, Kainz V, Vervoort T: Perspective-taking influences attentional deployment towards facial expressions of pain: an eye-tracking study. Pain 161(6):1286-1296, 2020. https://doi.org/10.1097/j.pain.0000000000001827

16. Mende-Siedlecki P, Qu-Lee J, Lin J, Drain A, Goharzad A: The Delaware pain database: a set of painful expressions and corresponding norming data. Pain Rep 5(6):e853, 2020 https://doi.org/10.1097/PR9.0000000000000853

17. Fusaro M, Tieri G, Aglioti SM: Seeing pain and pleasure on self and others: behavioral and psychophysiological reactivity in immersive virtual reality. J Neurophysiol 116(6):2656-2662, 2016. https://doi.org/10.1152/jn.00489.2016

18. Preis MA, Kroener-Herwig B: Empathy for pain: the effects of prior experience and sex. Eur J Pain 16(9):1311-1319, 2012. https://doi.org/10.1002/j.1532-2149.2012.00119.x

19. de Wied M, Verbaten MN: Affective pictures processing, attention, and pain tolerance. Pain 90(1-2):163-172, 2001. https://doi.org/10.1016/s0304-3959(00)00400-0

20. Haberkamp A, Glombiewski JA, Schmidt F, Barke A: The disgust-related-images (DIRTI) database: validation of a novel standardized set of disgust pictures. Behav Res Ther 89:86-94, 2017. https://doi.org/10.1016/j.brat.2016.11.010

21. Marchewka A, Zurawski Ł, Jednoróg K, Grabowska A: The Nencki Affective Picture System (NAPS): introduction to a novel, standardized, wide-range, high-quality, realistic picture database. Behav Res Methods 46(2):596-610, 2014. https://doi.org/10.3758/s13428-013-0379-1

22. Lang PJ, Bradley MM, Cuthbert BN: International affective picture system (IAPS): affective ratings of pictures and instruction manual. Technical Report A-8. Gainesville: The Center for Research in Psychophysiology. University of Florida; 2008

23. Szymanska M, Monnin J, Noiret N, Tio G, Galdon L, Laurent E, Nezelof S, Vulliez-Coady L: The Besançon affective picture set-adolescents (the BAPS-Ado): development and validation. Psychiatry Res 228(3):576-584, 2015. https://doi.org/10.1016/j.psychres.2015.04.055

24. Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, Kastman E, Lindeløv JK: PsychoPy2: experiments in behavior made easy. Behav Res Methods 51(1):195-203, 2019. https://doi.org/10.3758/s13428-018-01193-y

25. Core Team R: R Software: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing; 2020

26. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W: Increasing transparency through a multiverse analysis. Perspect Psychol Sci 11(5):702-712, 2016. https://doi.org/10.1177/1745691616658637

27. Lawrence, A. ez: easy analysis and visualization of factorial experiments. R package version 4.3; 2015. Available at: https://cran.rproject.org/web/packages/ez. Accessed March 9, 2021.

28. Feys J: Nonparametric tests for the interaction in two-way factorial designs using R. R J 8(1):367-378, 2016. https://doi.org/10.32614/RJ-2016-027

29. Akoglu H: User's guide to correlation coefficients. Turk J Emerg Med 18(3):91-93, 2018. https://doi.org/10.1016/j.tjem.2018.08.001

30. Kuznetsova A, Brockhoff PB, Christensen RHB: lmerTest package: tests in linear mixed effects models. J Stat Softw 82(13):1548-7660, 2017. https://doi.org/10.18637/jss.v082.i13

31. Shirai M, Soshi T: Why is heartache associated with sadness? Sadness is represented by specific physical pain through verbal knowledge. PLoS One 14(5):e0216331, 2019 https://doi.org/10.1371/journal.pone.0216331

32. Necka EA, Amir C, Dildine TC, Atlas LY: Expectations about pain and analgesic treatment are shaped by medical providers' facial appearances: evidence from five online clinical simulation experiments. Soc Sci Med 281:114091, 2021 https://doi.org/10.1016/j.socscimed.2021.114091

33. Deyo KS, Prkachin KM, Mercer SR: Development of sensitivity to facial expression of pain. Pain 107(1-2):16-21, 2004. https://doi.org/10.1016/s0304-3959(03)00263-x

34. Yan Z, Pei M, Su Y: Children's empathy and their perception and evaluation of facial pain expression: an eye tracking study. Front Psychol 8:2284, 2017. https://doi.org/10.3389/fpsyg.2017.02284

35. Walsh J, Eccleston C, Keogh E: Sex differences in the decoding of pain-related body postures. Eur J Pain 21(10):1668-1677, 2017. https://doi.org/10.1002/ejp.1072

36. Keogh E: Gender differences in the nonverbal communication of pain: a new direction for sex, gender, and pain research. Pain 155(10):1927-1931, 2014. https://doi.org/10.1016/j.pain.2014.06.024

37. Herzog M, Sucec J, Van Diest I, Van den Bergh O, Chenivesse C, Davenport P, Similowski T, von Leupoldt A: Observing dyspnoea in others elicits dyspnoea, negative affect and brain responses. Eur Respir J 51(4):1702682, 2018 https://doi.org/10.1183/13993003.02682-2017

38. Kappesser J, Williams AC: Pain and negative emotions in the face: judgements by health care professionals. Pain 99(1-2):197-206, 2002. https://doi.org/10.1016/s0304-3959(02)00101-x

39. Arias JA, Williams C, Raghvani R, Aghajani M, Baez S, Belzung C, Booij L, Busatto G, Chiarella J, Fu CH, Ibanez A, Liddell BJ, Lowe L, Penninx BWJH, Rosa P, Kemp AH: The neuroscience of sadness: a multidisciplinary synthesis and collaborative review. Neurosci Biobehav Rev 111:199-228, 2020. https://doi.org/10.1016/j.neubiorev.2020.01.006