

Article

Do Spanish Affective Picture Norms Transfer to Russian Speakers? Valence and Arousal Ratings of a Low-Level Matched Subset of the EmoMadrid Database

Alexander Kirsanov ^{1,*}, Anton Karapuzov ², Elena Levchenko ³, Taisia Loginova ², Daria Marochko ³ and Evgeny Blagoveshchensky ¹

¹ Laboratory of Behavioural Neurodynamics, Saint Petersburg State University, Saint Petersburg 199034, Russia; e.blagoveshchensky@spbu.ru

² Laboratory for Cognitive Psychology of Digital Interfaces Users, HSE University National Research University Higher School of Economics (HSE), Moscow 101000, Russia; antonkarapuzov@gmail.com (A.K.); tloginova@hse.ru (T.L.)

³ Department of Psychology, Saint Petersburg State University, Saint Petersburg 199034, Russia; cemezku@mail.ru (E.L.); marochkodaria2004@gmail.com (D.M.)

* Correspondence: a.kirsanov@spbu.ru

Featured Application

The EMselec subset of the EmoMadrid database can be used with Russian-speaking participants in protocols that contrast negative, neutral and positive pictures. When pictures are selected by their absolute valence or arousal values, the Spanish norms should be recalibrated with the category-specific differences reported here. Negative and positive pictures should not be assumed to be matched in arousal, and neutral and positive pictures should not be ranked by their Spanish valence norms.

Abstract

Norms of standardised affective picture sets obtained in one sample may not transfer to another. We examined whether the EMselec subset of the EmoMadrid database (40 negative, 40 neutral and 40 positive pictures matched for luminance, spatial frequency and chromatic complexity) retains its normative properties in a Russian-speaking sample. Sixty-three participants ($M = 25$ years, $SD = 6$; 43 women) rated the pictures online on five-point valence and arousal scales under the exposure parameters of the Spanish norms. Rank-order agreement was high across all pictures (valence $\rho = 0.91$; arousal $\rho = 0.86$) and lower within categories (arousal $\rho = 0.40$ to 0.73 ; neutral and positive valence $\rho = 0.27$ and 0.07). Absolute agreement was good for valence ($ICC = 0.89$) and imprecisely estimated for arousal ($ICC = 0.70$, 95% CI $[-0.02, 0.89]$). Arousal ratings were lower in all categories, by 0.24 to 0.59 points. Valence ratings were compressed towards the middle of the scale (calibration slope 0.68), and the overall valence difference was not significant. Unlike in the Spanish norms, negative pictures were rated as more arousing than positive ones (difference 0.35), so their arousal matching did not hold. EMselec keeps its category structure, but its absolute norms require local recalibration.

Keywords: affective picture database; EMselec; EmoMadrid; cross-cultural validation; valence; arousal; normative ratings



Academic Editor: Alexander N. Pisarchik

Received: 8 September 2026

Revised: 10 October 2026

Accepted: 20 October 2026

Published: 23 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

Emotions organise much of human behaviour, and the induction and measurement of affective states are standard methods in affective science. Stimuli used for this purpose have to meet the technical requirements of the recording method and should produce comparable effects in different populations. Emotions are commonly induced with pictures of objects, scenes and faces [1–3], video clips [4–6], sounds and music [7–9], emotional words and phrases [10,11], and virtual reality and video games [12–14]. Static emotional scenes are widely used in laboratory research because they are relatively easy to standardise and can be used in many classical paradigms, such as oddball and affective priming tasks. Which stimuli are suitable depends on the method and the paradigm. Studies that record brain activity with functional magnetic resonance imaging (fMRI), electroencephalography (EEG), magnetoencephalography (MEG) or functional near-infrared spectroscopy (fNIRS) also have to control stimulus properties that affect the recorded signal [15].

Visual affective stimuli range from symbolic (words, sentences, pictograms) to non-symbolic (faces and scenes), and non-symbolic stimuli can be static or dynamic. Static scenes are the most widely used, because they are easy to standardise, present and replicate, and their low-level properties (brightness, contrast, spatial frequency) can be controlled, which makes valence and arousal ratings more comparable across studies [1,2,16–18]. In neurophysiological experiments this control is essential, because uncontrolled differences in visual parameters can affect the amplitudes and latencies of event-related potentials (ERPs) [19]. Faces are processed by specialised neural networks, so responses to them are difficult to compare with responses to other stimuli [20,21]. Video clips have higher ecological validity, but their perceptual characteristics are much harder to standardise and control than those of static scenes [22,23]. Static scenes are also more effective in inducing specifically negative affect, whereas videos evoke a broader and less specific range of experiences [24]. Static pictures are therefore often chosen as a compromise between ecological and affective validity, provided that they are carefully selected and standardised.

Most visual stimulus databases, however, do not control the low-level visual parameters of their images (brightness, spatial frequency and colour complexity), although these parameters matter in neurophysiological experiments such as ERP studies [25,26]. Uncontrolled visual differences between stimuli can confound measures of neural activity, especially at early processing stages that depend mainly on sensory information. To address this problem, the Brain, Affect and Cognition research group (Cerebro, Afecto y Cognición, CEACO) of the Universidad Autónoma de Madrid created the EmoMadrid database of emotional scenes [15]. The database contains 813 pictures with normative valence and arousal ratings. For each picture it also reports brightness, spatial frequency and colour complexity, so that stimuli can be selected for methods that are sensitive to these parameters, including EEG. The database is freely available for research, which makes results easier to reproduce and compare between laboratories.

Emotional experience is commonly described within a dimensional framework defined by valence (unpleasant–pleasant) and arousal (calming–activating) [27]. Some models add further dimensions such as dominance [28], but most affective picture sets are normed on valence and arousal. Cross-cultural studies show that the relative evaluations of stimuli often agree between samples, whereas the absolute levels of valence and arousal can differ [29]. This combination of similar ordering and different absolute values is the main reason why affective picture sets need to be validated when they are used in a new population.

The EmoMadrid norms were also collected with short exposures, which are typical of neurophysiological studies [30]. Each picture was shown for 1 s and rated during the following 4 s [15]. In the normative procedures of the International Affective Picture System

(IAPS) and the Nencki Affective Picture System (NAPS), pictures were shown for 6 s [31] and 3 s, respectively, and in NAPS a reduced copy of the picture remained on the screen during the ratings [18]. The EmoMadrid presentation conditions are therefore closer to those of typical event-related potential studies. The database also codes the social content of the scenes, for example the presence of people or animals, so that the content of the stimuli can be varied systematically, and its pictures were chosen to elicit emotion without relying on clichéd or outdated scenes. These properties make EmoMadrid suitable for studies that need to separate the sensory and emotional contributions to stimulus processing [15,19,32].

Standardised picture databases such as the IAPS, the Geneva Affective Picture Database (GAPED), NAPS and the Open Affective Standardized Image Set (OASIS) have been widely used in research on affective processing. However, when their pictures were rated in other countries, the ratings often differed from the original norms. Adaptations of the IAPS outside the United States, reviewed by Branco et al. [33], show that the valence and arousal values of the original norms shift when the pictures are rated in other countries, for example in Brazil, Chile and China. These studies are directly relevant here, because they show that the norms of a widely used picture set do not transfer automatically to a new national sample, which is the question addressed for EMselec in the present study. Emotional expressiveness, the tendency to give extreme ratings and norms of emotional expression also vary across cultures [34,35]. In a 23-country study of emotion regulation, Matsumoto et al. [36] found that suppression of emotional expression was more common in cultures that emphasise social order and hierarchy and less common in cultures that value affective autonomy and egalitarianism. Russian participants, in particular, have been found to inhibit the expression of positive emotions towards strangers [37]. Differences in response style between the Russian-speaking and the Spanish samples can therefore be expected a priori.

Along with behavioural differences, the linguistic representation of emotions also varies: the same emotional states can be described with different words or may not have exact equivalents [38,39]. For example, Vietnamese speakers distinguish between ‘shame’ and ‘suffering’ as separate categories, whereas in English, they often overlap [40]. These factors can also influence the neurophysiological processing of stimuli. Perception of consonance and dissonance, for example, depends on enculturation as well as on the physical properties of sound [41–43]. Even formally standardised stimulus sets such as EmoMadrid therefore need to be validated when they are used in a new linguistic and cultural environment.

The present study used EMselec, a subset of EmoMadrid pictures selected by the database authors so that the negative, neutral and positive categories do not differ in spatial frequency, brightness (luminance) or chromatic complexity [15]. Differences in these parameters could otherwise affect both the affective ratings and the neuroimaging data. The pictures were downloaded from the EmoMadrid website [44] on 4 April 2025.

Three considerations frame the present comparison. The first concerns the balance between universal and culture-specific components of affective responses. The valence and arousal dimensions describe affective experience across cultures [34], whereas emotional expressiveness, response styles and norms of emotional expression differ between them [34–36]. The second concerns the level at which normative ratings are compared. A stimulus set may keep the relative ordering of its pictures in a new sample while its absolute values shift or are compressed [29], and the two properties matter for different designs: contrasts between affective categories depend on the ordering, whereas the selection or matching of pictures by their normative values depends on the absolute ratings. The third concerns recalibration. When a normative set is used in a new linguistic and cultural setting,

its absolute values, and the matching criteria on which subsets such as EMselec are built, have to be checked in the new population before they can be relied upon [33].

This study had two aims: (1) to compare valence and arousal ratings between a Spanish reference sample and a Russian-speaking sample, and (2) to determine which properties of the Spanish EMselec norms can be used with Russian-speaking participants without recalibration [15,33]. The study concerns the applicability of EMselec in a new linguistic and cultural setting, since the capacity of the pictures to elicit emotions was established in the original normative study [15].

2. Materials and Methods

Figure 1 summarises the study. Of the 70 volunteers who took part, 63 were included in the analyses (Section 2.1). Each participant rated the EMselec pictures (Section 2.2) online in two blocks, one for valence and one for arousal, presented in counterbalanced order (Section 2.3). The ratings were compared with the Spanish norms at the picture level and analysed at the participant level within the Russian sample (Section 2.4).

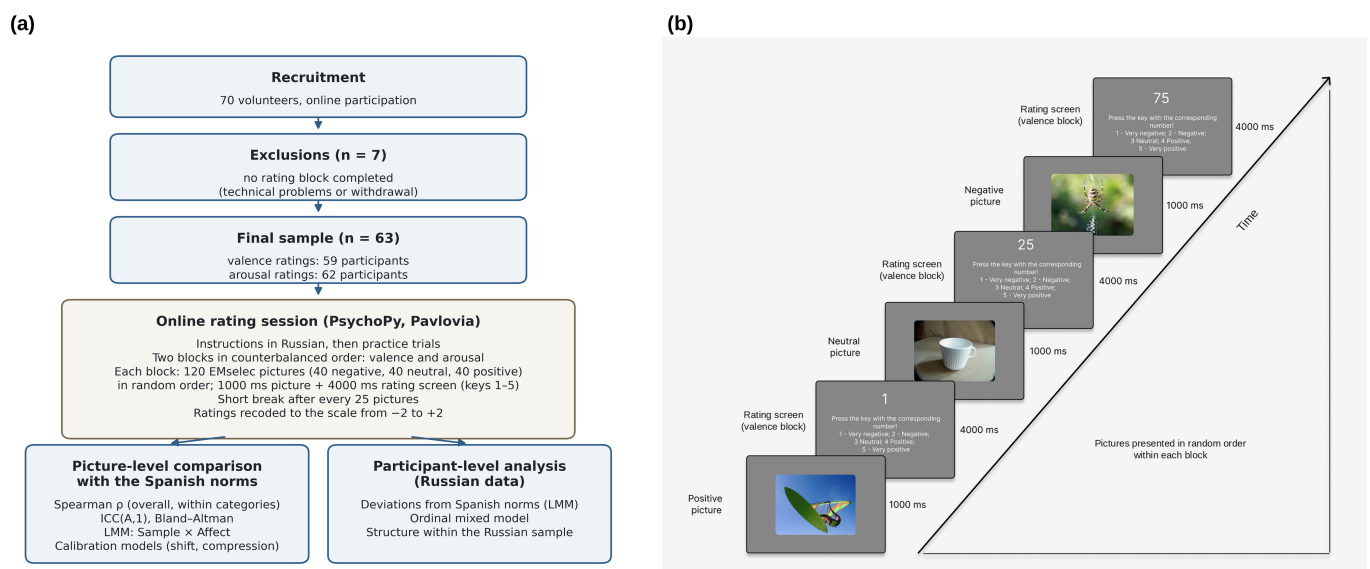


Figure 1. Overview of the study. (a) Workflow: recruitment and exclusions, the online rating session and the picture-level and participant-level analyses. (b) Trial structure. Each picture was shown for 1000 ms and followed by a 4000 ms response screen showing the picture number and the rating scale of the current block (valence or arousal; the valence block is shown). Participants rated each picture by pressing keys 1 to 5. The pictures shown are examples of positive, neutral and negative EMselec pictures; the pictures were presented in random order within each block.

2.1. Participants

The final sample consisted of 63 participants aged 18 to 45 years ($M = 25$, $SD = 6$), of whom 43 were women. Of the 70 volunteers who took part, 7 were excluded because they did not complete either block, owing to technical problems or because they stopped participating. Valence ratings were available from 59 participants and arousal ratings from 62, because 4 participants completed only the arousal block and 1 only the valence block. Participation was voluntary. The sample size follows the recommendation of the database authors for contributed norms [45].

The Russian ratings were compared with the normative values of the EmoMadrid database [15]. These norms were collected between 2006 and 2018 from 768 undergraduate psychology students at three Spanish universities, aged 17 to 50 years ($M = 20.1$, $SD = 3.4$), of whom 593 were women. Each participant rated only part of the database (on average

155 pictures), so the number of ratings varies across pictures [15]. For the EMselec pictures, the Spanish norms are based on 104 to 319 ratings per picture for valence (median 178) and 103 to 322 for arousal (median 179). The demographic characteristics above refer to the normative sample of the database as a whole and are not reported separately for the raters of individual pictures. Participants rated the pictures in group sessions of 17 to 44 people in a quiet, dimly lit room, with the pictures projected onto a screen, using paper forms until 2008 and individual computers afterwards [15]. The rating scales, exposure parameters and block structure of the present study followed this procedure. The two samples differed in testing environment, display equipment and period of data collection and, judging by the database sample as a whole, probably also in age, sex composition and student status (Table 1).

Table 1. Characteristics of the EmoMadrid normative sample and the Russian-speaking sample.

Characteristic	EmoMadrid Normative Sample, Whole Database [15]	Russian-Speaking Sample (Present Study)
Participants	768 in total; each rated a part of the database (on average 155 pictures)	63 (valence 59, arousal 62); each rated all 120 pictures
Ratings per EMselec picture	104–319 (valence), 103–322 (arousal)	53–59 (valence), 56–62 (arousal)
Age, years	17–50 (M = 20.1, SD = 3.4) ^a	18–45 (M = 25, SD = 6)
Sex	593 women, 175 men ^a	43 women, 20 men
Population	Undergraduate psychology students, three universities ^a	Adult volunteers recruited online
Period of data collection	2006–2018	2025
Setting	Group sessions of 17–44 people, quiet dimly lit room	Individual, online (Pavlovía)
Display	Projection screen, computer screen	Participants' own computers, not controlled
Response recording	Paper forms until 2008, then individual computers	Keyboard keys 1–5
Exposure/response window	1 s/4 s	1 s/4 s
Scales	Valence and arousal, 5 points, –2 to +2	Same, labels translated into Russian
Dimensions	Separate runs, order counterbalanced	Separate blocks, order counterbalanced
Picture order	Randomized per participant and block	Randomized per participant and block

Note. Characteristics of the Spanish sample are taken from the original database publication [15]. ^a Values refer to the normative sample of the whole database; the composition of the participants who rated the EMselec pictures is not reported.

2.2. Materials

The study used the EMselec subset of the EmoMadrid database [15], which comprises 120 pictures, 40 in each of the negative, neutral and positive categories. The subset was designed to minimise differences in physical image properties between the affective categories. According to the validation data reported by Carretié et al. [15], the categories do not differ significantly in spatial frequency, luminance, chromatic complexity or contrast (one-way ANOVA with Tukey HSD post hoc comparisons). In the Spanish norms, valence ratings differ significantly between all three categories, whereas arousal ratings do not differ between negative and positive pictures and are higher for both than for neutral pictures.

2.3. Procedure

The experiment was programmed in PsychoPy 2024.2.4 (Open Science Tools Ltd., Nottingham, UK) and ran online on the Pavlovia platform (Open Science Tools Ltd., Nottingham, UK; <https://pavlovia.org>). Participants took part from their own desktop or laptop computers. A physical keyboard was required because responses were given with the number keys 1 to 5, so the task could not be completed on smartphones or tablets. Screen size, resolution, brightness and colour rendering were not controlled. The physical matching of EMselec refers to the image files, and each participant saw all pictures on the same display. Because the categories are matched in luminance and colour, uncontrolled display properties are unlikely to have produced systematic differences between the categories within the Russian sample. Display conditions did, however, differ between the Russian and Spanish samples (Table 1), so part of the differences in absolute ratings between the samples may reflect display conditions rather than language or culture.

The experiment consisted of two blocks. In one block participants rated all pictures for valence and in the other for arousal. The order of the blocks was counterbalanced, so that half of the participants started with the valence block and half with the arousal block. Separate blocks were used to reduce a halo effect between the two dimensions, which can arise when they are rated together [46].

Each picture (1024×768 pixels) was presented for 1 s, followed by a 4 s response interval (Figure 1b). During the interval, a blank background was presented with the number of the previous image in the centre and the rating scale. Participants pressed a key (1, 2, 3, 4 or 5) corresponding to their rating within the 4 s interval. The valence scale was labelled as 1 = very negative, 2 = negative, 3 = neutral, 4 = positive and 5 = very positive, and the arousal scale as 1 = very calming, 2 = calming, 3 = neutral, 4 = arousing and 5 = very arousing. The Russian instructions and labels are given in Supplementary Table S2. Key presses were recoded as the key value minus 3, so that both scales ranged from -2 to $+2$ with 0 as the neutral point, as in the EmoMadrid norms. The Spanish normative values are reported on the same scale [15] and were used without transformation. The order of pictures was randomized independently for each participant and each block. A short break of about 15 s was provided after every 25 images. Before the main procedure, participants completed a practice block with several trial images. The instructions were translated into Russian by the authors and checked by a second author. They were presented as text in the PsychoPy interface before the experiment began.

2.4. Statistical Analysis

The main analyses were conducted at the picture level. Each picture contributed one mean rating per sample and dimension, which gives 240 observations per dimension. Because the Spanish norms are available only as picture-level means and standard deviations, these analyses treat pictures, not participants, as the unit of replication. Their p values and effect sizes therefore describe differences between the two sets of ratings of these pictures, not differences between the populations from which the samples were drawn. The Russian data were additionally analysed at the level of individual ratings (see the participant-level analysis below). Analyses were performed in Positron 1.103.0 (Posit Software, PBC, Boston, MA, USA) with R 4.3.0 (R Foundation for Statistical Computing, Vienna, Austria), using the packages lme4 1.1-34, lmerTest 3.1-3, emmeans 1.8.7, effectsize 1.0.1, performance 0.14.0, irr 0.84.1, car 3.1-2, nlme 3.1-163, ordinal 2026.7-26 and tidyverse 2.0.0.

Rank-order correspondence between the Russian-speaking sample and the Spanish normative values was assessed with Spearman's rank correlation coefficient (ρ); because of tied ranks, p -values are based on the normal approximation. Correlations were computed across all pictures and separately within each affective category. To remove the contribution

of the separation between categories, we also computed a partial rank correlation, the Pearson correlation between the ranks of the Russian and Spanish values after the category means of the ranks had been removed. Confidence intervals (95%) were obtained with the Fisher z transformation, using the standard error of Bonett and Wright for Spearman coefficients [47]. Absolute agreement was quantified with intraclass correlation coefficients estimated in a two-way random-effects model for single measurements with absolute agreement, ICC(A,1) in the notation of McGraw and Wong [48], reported with 95% confidence intervals and interpreted as poor below 0.50, moderate from 0.50 to 0.75, good from 0.75 to 0.90 and excellent above 0.90 [49].

Systematic shift was examined with the method of Bland and Altman. For each picture, the Russian minus Spanish difference was computed, and the mean difference was tested against zero with a one-sample t test. Limits of agreement are reported as the mean difference ± 1.96 SD, together with 95% confidence intervals for the mean difference and for each limit [50]. Category-dependent differences were modelled with linear mixed-effects models fitted separately for valence and arousal, with random intercepts for stimuli (EM_CODE), accounting for the non-independence of observations within images. The following model was applied:

$$Y_{sk} = \beta_0 + \beta_1 \text{Sample}_s + \beta_2 \text{Affect}_k + \beta_3 (\text{Sample} \times \text{Affect})_{sk} + u_k + \varepsilon_{sk}, \quad (1)$$

where Y_{sk} is the mean rating of picture k in sample s , Affect_k is the affective category of picture k , u_k is the random intercept for picture k (EM_CODE) and ε_{sk} is the residual.

Sample (Russian-speaking vs. Spanish normative) and Affect (Positive, Neutral, Negative) were specified as fixed effects. Models were estimated by restricted maximum likelihood in the lme4 package, with tests of fixed effects performed using the lmerTest package. Type III sums of squares were used, with the Kenward–Roger approximation for degrees of freedom.

Effect sizes (partial η^2) were computed from F-statistics using the effectsize package, are reported with 90% confidence intervals, and were interpreted according to conventional benchmarks (0.01 small, 0.06 medium, ≥ 0.14 large).

To assess differences between samples within each affect category, pairwise contrasts were computed from estimated marginal means and are expressed throughout as Russian minus Spanish values. The three between-sample contrasts of each dimension were treated as one family, and their p values and confidence intervals were Bonferroni-adjusted for three comparisons. The effect size for each contrast was expressed as a correlation coefficient,

$$r = \sqrt{\frac{t^2}{t^2 + df}}, \quad (2)$$

which provides direct interpretability on a scale from 0 to 1 [51]. Degrees of freedom for the contrasts and marginal means were obtained with the Kenward–Roger method. The same model structure was additionally applied to the within-picture standard deviation of ratings as an index of rater dispersion, with the same adjustment of the contrasts.

Model fit was evaluated using marginal (R_m^2) and conditional (R_c^2) coefficients of determination following Nakagawa and Schielzeth [52], reflecting variance explained by fixed effects alone and by the full model, respectively.

Residual normality was assessed with the Shapiro–Wilk test. Constancy of error variance was assessed with the Breusch–Pagan score test [53], which relates the residual variance to the fitted values (check_heteroscedasticity in the performance package), and with Levene’s test across affective categories. Because the residuals of mixed models are not independent, the Breusch–Pagan test is only approximate in this setting. Both models

were therefore also refitted with a separate residual variance for each Sample $\times$ Affect cell (varIdent variance structure, nlme package) and compared with the common-variance models by likelihood ratio tests.

To distinguish a shift in the rating scale from a compression of its range, the Russian picture means were regressed on the Spanish norms separately for each dimension:

$$RU_k = a + bES_k + e_k, \quad (3)$$

where k indexes pictures, RU_k and ES_k are the Russian and Spanish means of picture k , and e_k is the residual. Three models were compared: a shift-only model with b fixed at 1, a compression-only model with a fixed at 0, the neutral point of the scale, and a model with both parameters free. A slope below 1 indicates a compression of the range, and a non-zero intercept a displacement of the scale. The models were compared with AIC and BIC and with F tests of the nested models, and an interaction between the Spanish norm and Affect was added to test whether the slope differed between categories.

For the Russian sample, each individual rating was also expressed as its deviation from the Spanish norm of the same picture and analysed with a linear mixed model,

$$D_{pk} = R_{pk} - ES_k = \gamma_0 + \gamma_1 Affect_k + v_p + u_k + \varepsilon_{pk}, \quad (4)$$

where R_{pk} is the rating of picture k by participant p , ES_k is the Spanish normative mean of picture k , and v_p and u_k are random intercepts for participants and pictures. In this model the Spanish norms act as a fixed reference, and the category means of D estimate the between-sample differences while taking into account the variability among Russian participants and among pictures. Because of the number of observations, confidence intervals for these means are asymptotic (Wald) intervals. As a sensitivity analysis for the ordinal nature of the ratings, a cumulative link (proportional odds) mixed model with Affect as the predictor and random intercepts for participants and pictures was fitted to the raw Russian ratings (ordinal package). The Spanish sample could not be included in these models, because its data are available only as picture-level summaries.

3. Results

3.1. Agreement Between Valence and Arousal Ratings

Figure 2 shows the distributions of the picture means in the two samples. Each picture mean is an average of 53 to 62 discrete ratings in the Russian sample and 103 to 322 in the Spanish norms and can therefore take intermediate values on the -2 to $+2$ scale. The distributions have one peak for each affective category and deviate from normality, so rank-order agreement between the Russian (RU) and Spanish (ES) values was assessed with Spearman's correlation coefficient and absolute agreement with intraclass correlation coefficients. The relation between the two sets of values is shown in Figure 3.

Rank-order agreement between samples was high for valence ($\rho = 0.91$, 95% CI [0.86, 0.94], $S = 26,082$, $p < 0.001$) and arousal ($\rho = 0.86$, 95% CI [0.80, 0.91], $S = 39,009$, $p < 0.001$). Within categories, the correlations were lower (Table 2). For arousal they ranged from $\rho = 0.40$ for neutral to $\rho = 0.73$ for negative pictures, and the partial correlation controlling for category was $\rho = 0.62$ (95% CI [0.49, 0.72]). For valence, the within-category correlation was $\rho = 0.67$ for negative pictures but 0.27 for neutral and 0.07 for positive pictures, with confidence intervals that include zero, and the partial correlation was $\rho = 0.34$ (95% CI [0.17, 0.49]). The high overall valence correlation therefore mainly reflects the separation between categories. Part of this attenuation is expected, because the Spanish valence norms vary little within the neutral and positive categories (SD of picture means 0.16 and 0.17; Supplementary Table S4). Absolute agreement, estimated as ICC(A,1), was good for valence

(ICC = 0.89, $F(119, 115) = 16.7$, $p < 0.001$, 95% CI [0.84, 0.92]) and imprecisely estimated for arousal (ICC = 0.70, 95% CI [−0.02, 0.89], $F(119, 2.96) = 13.2$, $p = 0.028$) (Figure 3). The confidence interval for arousal is wide and extends below zero. This instability follows from the systematic shift reported below. Absolute-agreement coefficients penalise a constant difference between raters, and with a shift of this size relative to the variance between pictures, the approximate denominator degrees of freedom of the F test become small. The arousal coefficient should therefore be read as an approximate value.

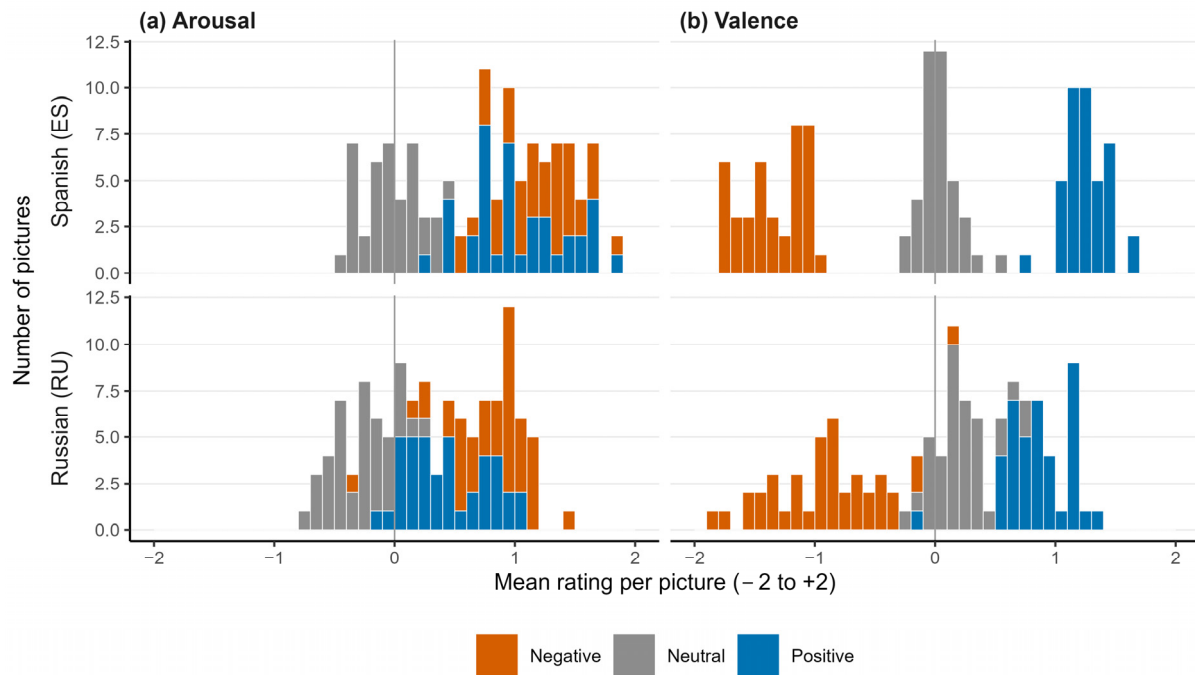


Figure 2. Distributions of picture mean ratings for (a) arousal and (b) valence in the Spanish normative sample (upper row) and the Russian-speaking sample (lower row). Each bar shows the number of pictures whose mean rating falls in an interval of 0.1 scale points; colours indicate the affective category. Russian arousal means lie below the Spanish norms in all categories, and Russian valence means cover a narrower range of the scale.

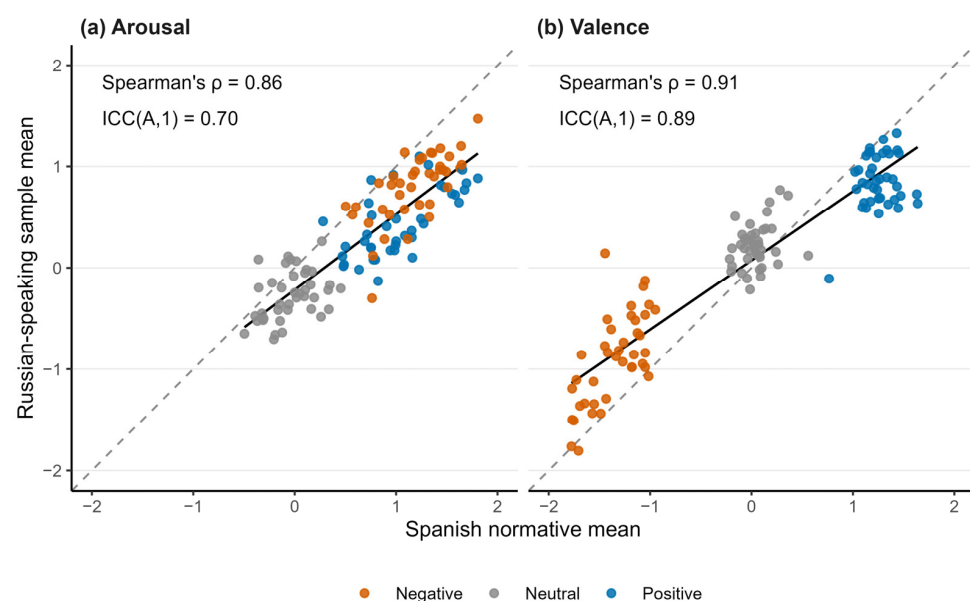


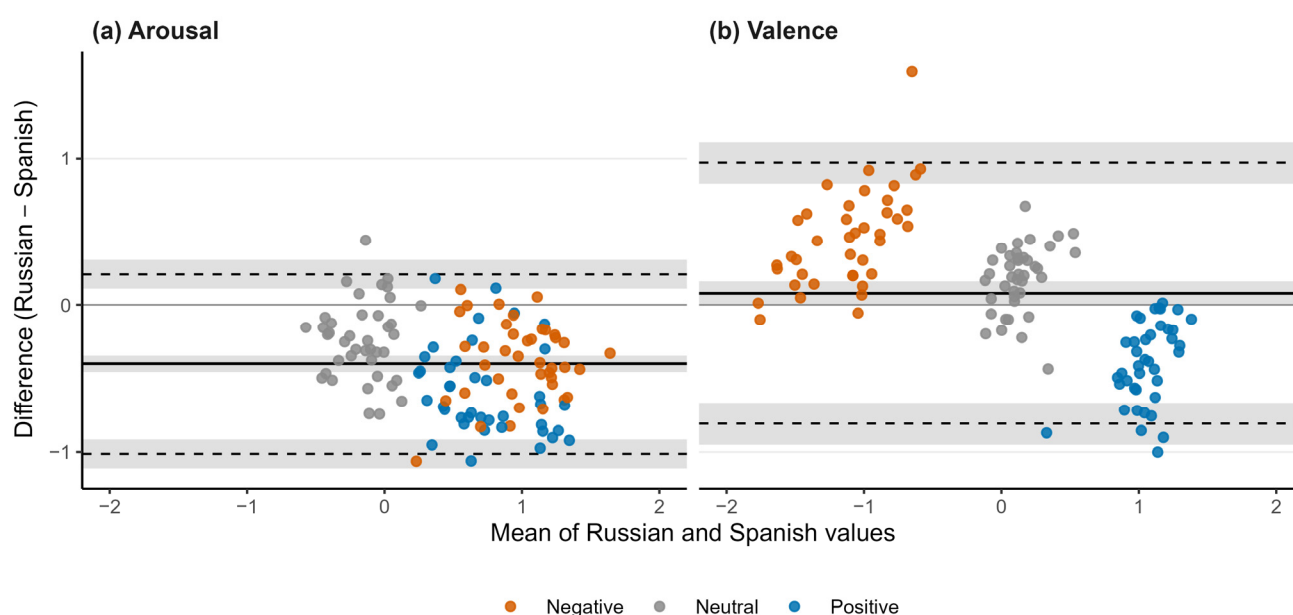
Figure 3. Russian picture means plotted against the Spanish normative values for (a) arousal and (b) valence. Each point is one picture, coloured by affective category (negative, neutral, positive). Each panel reports Spearman's ρ and ICC(A,1). The diagonal line indicates equal ratings in both samples.

Table 2. Spearman correlations between the Russian picture means and the Spanish normative values, across all pictures and within affective categories.

Pictures	<i>n</i>	Valence ρ [95% CI]	Arousal ρ [95% CI]
All	120	0.91 [0.86, 0.94]	0.86 [0.80, 0.91]
Negative	40	0.67 [0.42, 0.82]	0.73 [0.51, 0.86]
Neutral	40	0.27 [−0.05, 0.54]	0.40 [0.09, 0.64]
Positive	40	0.07 [−0.25, 0.37]	0.64 [0.38, 0.80]
Partial, category removed	120	0.34 [0.17, 0.49]	0.62 [0.49, 0.72]

Note. Confidence intervals were obtained with the Fisher *z* transformation (standard error of Bonett and Wright [47] for Spearman coefficients). The partial coefficient is the correlation between the ranks after removal of the category means of the ranks. $p < 0.001$ for all coefficients except neutral valence ($p = 0.090$), positive valence ($p = 0.67$) and neutral arousal ($p = 0.011$).

Agreement between the arousal ratings of the two samples was examined with a Bland–Altman plot (Figure 4a). The mean difference was -0.40 (95% CI $[-0.46, -0.35]$), $SD = 0.31$, and the 95% limits of agreement ranged from -1.01 (95% CI $[-1.11, -0.92]$) to 0.21 (95% CI $[0.11, 0.31]$). Negative and positive pictures were more widely scattered than neutral pictures.

**Figure 4.** Bland–Altman plots for (a) arousal and (b) valence. Each point is one picture, coloured by affective category. The x-axis shows the mean of the Russian and Spanish values, and the y-axis shows their difference (Russian–Spanish). The solid line indicates the mean difference, the dashed lines indicate the 95% limits of agreement (mean difference ± 1.96 SD), and the grey line marks zero.

For valence (Figure 4b), the mean difference was $+0.08$ (95% CI $[-0.004, 0.16]$; $t(119) = 1.88$, $p = 0.063$), $SD = 0.45$, and the 95% limits of agreement ranged from -0.81 (95% CI $[-0.95, -0.67]$) to 0.97 (95% CI $[0.82, 1.11]$). The differences formed three clusters corresponding to the affective categories.

3.2. Linear Mixed-Effects Regression Analysis

The effects of Sample (Russian-speaking vs. Spanish normative), affective category (positive, neutral, negative) and their interaction on the mean arousal ratings of the pictures

were examined with a linear mixed-effects model with a random intercept for each picture (Equation (1); Table 3).

Table 3. Results of the two-way analysis of variance (Type III) for arousal: effects of Sample, Affect, and their interaction.

Effect	F(df1, df2)	<i>p</i>	η^2_p
Sample	250.96 (1, 117)	<0.001	0.68 [0.63, 1.00]
Affect	179.70 (2, 117)	<0.001	0.75 [0.71, 1.00]
Sample $\times$ Affect	15.98 (2, 117)	<0.001	0.21 [0.13, 1.00]

Note. η^2_p = partial eta squared, with 90% confidence intervals in brackets; df = degrees of freedom. Upper bounds are fixed at 1.00 because the intervals are one-sided. No correction for multiple comparisons applies to omnibus effects.

The main effect of Sample was significant— $F(1, 117) = 250.96, p < 0.001, \eta^2_p = 0.68$, 90% CI [0.63, 1.00]—with lower arousal ratings in the Russian-speaking sample. The effect of Affect was also significant— $F(2, 117) = 179.70, p < 0.001, \eta^2_p = 0.75$, 90% CI [0.71, 1.00]. A significant Sample $\times$ Affect interaction ($F(2, 117) = 15.98, p < 0.001, \eta^2_p = 0.21$, 90% CI [0.13, 1.00]) indicated that the magnitude of between-sample differences varied depending on the affect category.

Bonferroni-adjusted contrasts (Table 4; Figure 5a) showed that arousal ratings were lower in the Russian-speaking sample in all three categories, although the size of the difference depended on the category. The difference was largest for positive pictures ($M_{diff} = -0.59, t(117) = -13.37, p_{adj} < 0.001, r = 0.78$), intermediate for negative pictures ($M_{diff} = -0.38, t(117) = -8.66, p_{adj} < 0.001, r = 0.62$) and smallest for neutral pictures ($M_{diff} = -0.24, t(117) = -5.42, p_{adj} < 0.001, r = 0.45$).

Table 4. Estimated marginal means and between-sample contrasts (RU–ES) of arousal within each affective category (Bonferroni adjustment for three contrasts).

Affect	M RU [95% CI]	M ES [95% CI]	Mdiff	95% CI adj.	t(117)	p _{adj}	r
Positive	0.44 [0.35, 0.54]	1.03 [0.93, 1.13]	−0.59	[−0.70, −0.48]	−13.37	<0.001	0.78
Neutral	−0.27 [−0.37, −0.17]	−0.03 [−0.13, 0.06]	−0.24	[−0.35, −0.13]	−5.42	<0.001	0.45
Negative	0.79 [0.70, 0.89]	1.18 [1.08, 1.27]	−0.38	[−0.49, −0.27]	−8.66	<0.001	0.62

Note. M RU and M ES = estimated marginal means with unadjusted 95% confidence intervals. Mdiff = contrast RU–ES; positive values indicate higher arousal ratings in the Russian sample. Degrees of freedom were obtained with the Kenward–Roger method (df = 117 for all contrasts). The confidence intervals of Mdiff and p_{adj} are Bonferroni-adjusted for the three between-sample contrasts of this dimension. Effect size *r* was computed with Equation (2).

The marginal and conditional coefficients of determination were $R^2_m = 0.742$ and $R^2_c = 0.900$. Residuals were normally distributed (Shapiro–Wilk $W = 0.995, p = 0.54$), and the Breusch–Pagan test did not indicate non-constant error variance ($p = 0.94$). A model with separate residual variances for the six Sample $\times$ Affect cells did not improve the fit (likelihood ratio = 4.83, df = 5, $p = 0.44$), and all effects and contrasts were unchanged (Supplementary Table S4).

The same model was fitted to the valence ratings (Table 5).

Bonferroni-adjusted contrasts (Table 6; Figure 5b) showed that the direction of the between-sample difference reversed across categories. Russian valence ratings were lower than the Spanish norms for positive pictures ($M_{diff} = -0.40, t(117) = -8.90, p_{adj} < 0.001, r = 0.64$) and higher, that is, less negative, for negative pictures ($M_{diff} = +0.45, t(117) = 10.11, p_{adj} < 0.001, r = 0.68$), with a small upward difference for neutral pictures ($M_{diff} = +0.18$,

$t(117) = 4.01$, $p_{adj} < 0.001$, $r = 0.35$). The Russian sample therefore used a narrower portion of the valence scale at both poles, which is tested directly in Section 3.3. Averaged over all pictures, the difference was small and not significant (+0.08; Section 3.1), because effects of opposite sign cancel at the aggregate level.

Table 5. Results of the two-way analysis of variance (Type III) for valence: effects of Sample, Affect, and their interaction.

Effect	F(df1, df2)	<i>p</i>	η^2_p
Sample	9.07 (1, 117)	0.003	0.07 [0.02, 1.00]
Affect	888.03 (2, 117)	<0.001	0.94 [0.93, 1.00]
Sample $\times$ Affect	94.16 (2, 117)	<0.001	0.62 [0.55, 1.00]

Note. η^2_p = partial eta squared, with 90% confidence intervals in brackets; df = degrees of freedom. Upper bounds are fixed at 1.00 because the intervals are one-sided. No correction for multiple comparisons applies to omnibus effects.

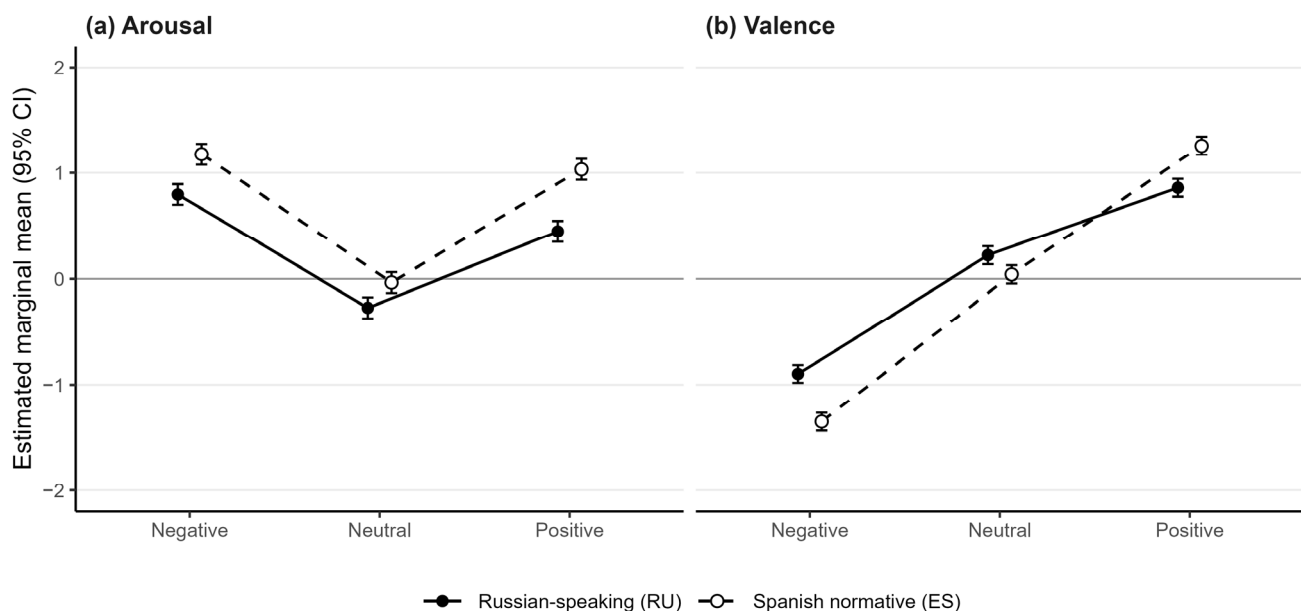


Figure 5. Mean ratings with 95% confidence intervals for each affective category in the Russian-speaking (RU, solid line) and Spanish normative (ES, dashed line) samples for (a) arousal and (b) valence.

Table 6. Estimated marginal means and between-sample contrasts (RU–ES) of valence within each affective category (Bonferroni adjustment for three contrasts).

Affect	M RU [95% CI]	M ES [95% CI]	Mdiff	95% CI adj.	t(117)	<i>p</i> _{adj}	<i>r</i>
Positive	0.86 [0.77, 0.94]	1.26 [1.17, 1.34]	−0.40	[−0.51, −0.29]	−8.90	<0.001	0.64
Neutral	0.22 [0.14, 0.31]	0.04 [−0.04, 0.13]	+0.18	[0.07, 0.29]	4.01	<0.001	0.35
Negative	−0.90 [−0.98, −0.81]	−1.35 [−1.43, −1.26]	+0.45	[0.34, 0.56]	10.11	<0.001	0.68

Note. M RU and M ES = estimated marginal means with unadjusted 95% confidence intervals. Mdiff = contrast RU–ES; positive values indicate higher valence ratings in the Russian sample. Degrees of freedom were obtained with the Kenward–Roger method ($df = 117$ for all contrasts). The confidence intervals of Mdiff and *p*_{adj} are Bonferroni-adjusted for the three between-sample contrasts of this dimension. Effect size *r* was computed with Equation (2).

The marginal and conditional coefficients of determination were $R_m^2 = 0.919$ and $R_c^2 = 0.956$. Residual normality was violated (Shapiro–Wilk $W = 0.982$, $p = 0.004$), whereas the Breusch–Pagan test did not indicate non-constant error variance ($p = 0.46$). Levene’s

test showed unequal variances across categories in both samples (Russian $F(2, 117) = 9.28$, $p < 0.001$; Spanish $F(2, 117) = 11.28$, $p < 0.001$). The standard deviations of picture means in each cell (Supplementary Table S4) show that this reflects the larger spread of negative pictures in both samples (Russian $SD = 0.45$, compared with 0.22 for neutral and 0.27 for positive pictures; Spanish 0.26, 0.16 and 0.17). A model with a separate residual variance for each Sample $\times$ Affect cell fitted better than the common-variance model (likelihood ratio = 45.88, $df = 5$, $p < 0.001$), and all effects kept their direction and significance (Sample $\times$ Affect $F(2, 117) = 79.16$, $p < 0.001$; all Bonferroni-adjusted between-sample contrasts $p < 0.001$; Supplementary Table S4).

3.3. Calibration of the Russian Ratings Against the Spanish Norms

Table 7 compares the three calibration models. For valence, the model with both parameters free fitted clearly better than the shift-only model ($F(1, 118) = 164.87$, $p < 0.001$) and somewhat better than the compression-only model ($F(1, 118) = 7.27$, $p = 0.008$; $\Delta AIC = 5.2$). Its slope was 0.68 (95% CI [0.63, 0.73]) and its intercept 0.07 (95% CI [0.02, 0.13]). Russian valence ratings were thus compressed to about two thirds of the Spanish range, with a small upward displacement. The slope differed between categories ($F(4, 114) = 5.33$, $p < 0.001$), so the compression was not uniform across the scale. For arousal, the model with both parameters free also fitted best (shift-only model $F(1, 118) = 43.22$, $p < 0.001$; compression-only model $F(1, 118) = 33.91$, $p < 0.001$), with a slope of 0.74 (95% CI [0.67, 0.82]) and an intercept of -0.22 (95% CI [-0.29 , -0.14]), and the slope again differed between categories ($F(4, 114) = 7.45$, $p < 0.001$). The arousal difference therefore combined a downward shift, a compression and a category-dependent component.

Table 7. Calibration models relating the Russian picture means to the Spanish norms (Equation (3)).

Dimension	Model	a [95% CI]	b [95% CI]	AIC	BIC
Valence	Shift only (b = 1)	0.08	1 (fixed)	153.53	159.11
	Compression only (a = 0)	0 (fixed)	0.68	55.79	61.36
	Shift and compression	0.07 [0.02, 0.13]	0.68 [0.63, 0.73]	50.62	58.98
Arousal	Shift only (b = 1)	-0.40	1 (fixed)	63.57	69.15
	Compression only (a = 0)	0 (fixed)	0.57	56.44	62.02
	Shift and compression	-0.22 [-0.29 , -0.14]	0.74 [0.67, 0.82]	28.13	36.49

Note. a = intercept, b = slope; n = 120 pictures per dimension. Lower AIC and BIC indicate better fit. Test of b = 1: $F(1, 118) = 164.87$ (valence) and 43.22 (arousal), both $p < 0.001$; test of a = 0: $F(1, 118) = 7.27$, $p = 0.008$ (valence) and 33.91, $p < 0.001$ (arousal). Slope $\times$ Affect: $F(4, 114) = 5.33$, $p < 0.001$ (valence); $F(4, 114) = 7.45$, $p < 0.001$ (arousal).

A parallel analysis of the within-picture standard deviation of ratings showed that Russian ratings were, on average, more dispersed than the Spanish ratings, for valence (main effect of Sample $F(1, 117) = 19.95$, $p < 0.001$, $\eta^2_p = 0.15$) and for arousal ($F(1, 117) = 27.40$, $p < 0.001$, $\eta^2_p = 0.19$). For valence, the Sample $\times$ Affect interaction was significant ($F(2, 117) = 4.81$, $p = 0.010$, $\eta^2_p = 0.08$), and the difference was reliable only for positive pictures (Mdiff = +0.15, 95% CI [0.08, 0.22], $p_{adj} < 0.001$, $r = 0.43$; negative +0.04, $p_{adj} = 0.46$; neutral +0.03, $p_{adj} = 0.71$). For arousal, the interaction was not significant ($F(2, 117) = 0.72$, $p = 0.49$), and the contrasts were +0.07 ($p_{adj} = 0.073$), +0.08 ($p_{adj} = 0.016$) and +0.12 ($p_{adj} < 0.001$) for negative, neutral and positive pictures. The Breusch–Pagan test indicated non-constant error variance in both dispersion models ($p < 0.001$), so these

tests are approximate. Each picture received 53 to 59 valid valence ratings (median 56) and 56 to 62 valid arousal ratings (median 61) in the Russian sample (Supplementary Table S1).

3.4. Participant-Level Analysis of the Russian Ratings

The participant-level models included 6749 valence ratings from 59 participants and 7244 arousal ratings from 62 participants. The deviation of the Russian ratings from the Spanish norms depended on the affective category for both valence ($F(2, 117) = 93.75$, $p < 0.001$) and arousal ($F(2, 117) = 16.02$, $p < 0.001$). The estimated deviations were +0.45 (95% CI [0.35, 0.55]) for negative, +0.18 (95% CI [0.07, 0.28]) for neutral and -0.40 (95% CI $[-0.50, -0.30]$) for positive pictures for valence, and -0.38 (95% CI $[-0.49, -0.28]$), -0.24 (95% CI $[-0.35, -0.14]$) and -0.59 (95% CI $[-0.70, -0.49]$) for arousal (Supplementary Table S5). These estimates agree with the picture-level contrasts in Tables 4 and 6, with somewhat wider confidence intervals. The cumulative link mixed models fitted to the raw Russian ratings gave the same pattern of category differences as the linear models. Valence ratings increased from negative to neutral to positive pictures, and arousal ratings were highest for negative, intermediate for positive and lowest for neutral pictures. All three pairwise differences had the same direction as in the linear models and were significant on both dimensions (all Tukey-adjusted $p < 0.001$), including the higher arousal of negative than positive pictures (Supplementary Table S5).

3.5. Affective Structure Within the Russian-Speaking Sample

Within the Russian sample, valence ratings followed the expected order (Supplementary Table S3). Negative pictures were rated lower than neutral pictures (difference -1.12 , 95% CI $[-1.29, -0.95]$) and neutral pictures lower than positive pictures (-0.64 , 95% CI $[-0.81, -0.46]$; negative vs. positive -1.76 , 95% CI $[-1.93, -1.58]$; all Tukey-adjusted $p < 0.001$). The corresponding differences in the Spanish norms were -1.39 , -1.21 and -2.61 . For arousal, negative ($+1.07$, 95% CI $[0.91, 1.23]$) and positive ($+0.72$, 95% CI $[0.56, 0.88]$) pictures were rated as more arousing than neutral pictures, which reproduces the V-shaped profile of the Spanish norms. Unlike in the Spanish norms, where negative and positive pictures did not differ in arousal (0.14 , 95% CI $[-0.03, 0.31]$, $p = 0.12$), negative pictures were rated as more arousing than positive pictures in the Russian sample (0.35 , 95% CI $[0.19, 0.51]$, $t(117) = 5.12$, $p < 0.001$). The same difference was obtained in the participant-level model (0.35 , 95% CI $[0.19, 0.51]$). The arousal matching of negative and positive pictures, one of the selection criteria of EMselec, therefore did not hold in the Russian-speaking sample.

3.6. Summary of Findings

Three results summarise the comparison. First, the category structure of EMselec was preserved in the Russian-speaking sample. The three categories were ordered in the same way as in the Spanish norms on both dimensions, and the ordering of individual pictures within categories was preserved for arousal (partial $\rho = 0.62$) but only weakly for valence (partial $\rho = 0.34$). Second, absolute ratings differed systematically. Arousal ratings were lower in all categories, by 0.24 to 0.59 points, and valence ratings were compressed towards the middle of the scale (calibration slope 0.68) with a small upward displacement. Third, negative pictures were rated as more arousing than positive pictures in the Russian sample (difference 0.35), so the arousal matching of these two categories did not hold. For the practical use of EMselec, these results mean that designs contrasting the three categories can use the set with Russian-speaking participants, whereas designs that select or match pictures by their absolute valence or arousal values, or that assume equal arousal of negative and positive pictures, should rely on the Russian values reported in Supplementary Table S1.

4. Discussion

The present study compared valence and arousal ratings of the EMselec subset of the EmoMadrid database in a Russian-speaking sample with the Spanish normative values. Across all pictures, rank-order correspondence was high on both dimensions ($\rho = 0.91$ and 0.86). Within categories it remained moderate to high for arousal, whereas for valence it was clear only for negative pictures, so the high overall valence correlation mainly reflects the separation of the three categories. Absolute ratings differed in dimension-specific ways. Arousal ratings were lower in the Russian sample in all three categories, and the size of the difference depended on the category. Valence ratings were less positive for positive pictures and less negative for negative ones, and the calibration analysis described this pattern as a compression of the range with a small upward displacement. High rank-order agreement together with lower absolute agreement is compatible with a systematic offset, a compression of the scale, a category-dependent bias or a combination of these, and the calibration models indicate a combination for both dimensions. The norms therefore need local recalibration, but the stimulus set itself remains usable.

The combination of a largely stable ordering of pictures and different absolute rating levels is compatible with a distinction between invariant and context-dependent components of affective evaluation. The high correlations between the samples are consistent with dimensional models of emotion [2], whereas the systematic differences in mean ratings may reflect how the two samples used the rating scales. Culturally embedded evaluative frameworks are one possible source of such differences, but the present design cannot separate them from the other differences between the samples listed in Table 1.

This pattern is also compatible with hierarchical models of affective processing, according to which the emotional significance of a stimulus is encoded early in relatively universal terms, whereas later evaluative judgements are shaped by higher-order cognitive and sociocultural factors [30,54]. From this perspective, the preserved ordering would reflect shared mechanisms of rapid affective evaluation, and the differences in absolute values would arise at later stages. Because the present study measured ratings only, it cannot identify the processing stage at which the differences arise. This question could be addressed with event-related potentials recorded to the same pictures in both samples.

From a constructivist perspective, these findings may also be understood in terms of the diversity of conceptualisations of affective experience, based on the arbitrary assignment of meanings. Whereas basic affective signals provide a common basis for experience (as discussed earlier), their classification and quantification depend on culture-specific semantic and interpretative systems [55,56]. Identical stimuli could then be mapped onto partially different rating scales in different communities, which would produce systematic differences in ratings.

The between-sample differences were more dispersed for valence than for arousal (SD of differences 0.45 vs. 0.31). For valence, the difference changed in size and sign across categories, whereas for arousal it had the same sign in all categories and differed only in size. Arousal is less clearly defined than valence, which is closely tied to approach and avoidance tendencies, and may therefore be construed differently across linguistic and cultural settings [57,58], which makes it sensitive to expressive norms and response calibration strategies [59]. Differences in the interpretation of the arousal scale have also been reported between normative samples of affective pictures [18]. The translation of the scale is a concrete instance of this problem. In the Russian instructions the arousal dimension was introduced as «возбуждение» (*vozbuzhdenie*), a word that denotes excitement and agitation and can carry connotations that the Spanish pole labels, from very calming to very activating, do not have. Part of the lower Russian arousal ratings may therefore reflect how

this term was understood, and future Russian norming studies should use labels closer to the activation dimension.

The lower arousal ratings are also compatible with the theory of ideal affect, which holds that cultures differ in the levels of activation they consider desirable [60,61], and the reduced polarisation of valence ratings is compatible with the finding that Russians are less expressive towards strangers than towards friends and family [37]. Expression towards strangers and anonymous ratings of pictures are, however, different phenomena, and rating behaviour, subjective experience and overt expression are related but distinct constructs. These accounts are therefore possible interpretations that the present data cannot test. They could be examined in studies that measure ideal affect, display rules and picture ratings in the same participants.

For studies with Russian-speaking participants, three practical consequences follow. First, EMselec can be used without change in designs that contrast the three affective categories. Second, when pictures are selected or thresholded by their absolute values, the Spanish norms should be corrected with the category-specific differences in Tables 4 and 6. Because the calibration slope differed between categories on both dimensions, these category-specific corrections are likely to be more accurate than the single calibration line in Table 7. Individual neutral or positive pictures should not be ranked, or selected within their category, on the basis of the Spanish valence norms, because the within-category correlations for these categories were low ($\rho = 0.27$ and 0.07). Third, negative and positive pictures should not be assumed to be matched in arousal, so studies comparing these two conditions should control arousal statistically or select pictures on the basis of local ratings.

The study has several limitations. First, the Russian-speaking sample and the Spanish normative sample differed in language and culture, in testing environment, display equipment and period of data collection, and probably in age, sex composition and student status (Table 1). The demographic characteristics of the Spanish raters are known only for the database as a whole, not for the raters of the EMselec pictures. The present design cannot separate these factors, so the observed differences describe the two samples and should not be read as a general effect of Russian versus Spanish culture. Second, women predominated in both samples (68% in the Russian sample and 77% in the normative sample of the whole database), and the Russian participants were older on average ($M = 25$ vs. 20.1 years). Sex-disaggregated analyses were not conducted because the Russian sample was not powered for subgroup comparisons. Third, the ratings were collected online, and display parameters such as screen size, brightness and colour calibration were not controlled. Because the categories are matched in luminance and colour and each participant saw all pictures on the same display, this is unlikely to have affected the differences between categories, but part of the differences in absolute ratings between the samples may reflect these display conditions, because the Russian participants used their own screens and the Spanish participants viewed projected pictures. Fourth, the Spanish normative values were available only as picture-level means and standard deviations. The between-sample comparisons therefore treat pictures, not participants, as the unit of replication, and the variability of the Spanish sample is not represented in any model; the participant-level analysis addresses this problem only for the Russian data. Fifth, the translation of the arousal scale may have contributed to the arousal difference.

Future studies could extend the validation of EMselec to samples from other countries. Such data would show which properties of the norms are stable across populations and which depend on the population in which they were obtained.

5. Conclusions

Across all pictures, the Russian-speaking sample and the Spanish norms showed high rank-order correspondence ($\rho = 0.91$ for valence and 0.86 for arousal). Within categories, the ordering of pictures was moderately preserved for arousal and for negative valence but not for neutral and positive valence. Absolute agreement was good for valence ($ICC = 0.89$) and imprecisely estimated for arousal ($ICC = 0.70$, 95% CI $[-0.02, 0.89]$). Arousal ratings were lower in the Russian sample in all three categories, by an amount that depended on the category, and valence ratings were compressed towards the middle of the scale with a small upward displacement. Negative pictures were rated as more arousing than positive pictures, so the arousal matching of the EMselec categories did not hold in the Russian-speaking sample. Ideal affect and display rules offer possible explanations of these differences that the present data cannot test.

The results also indicate that the EMselec subset is a potentially useful standardised stimulus set for studies with Russian-speaking participants. Its matched physical characteristics support comparisons between groups, provided that absolute norms are recalibrated and the arousal difference between negative and positive pictures is taken into account.

More generally, the results show that the normative values of an affective picture set should be checked when the pictures are used in a new population, even when their ordering is preserved.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/app162110490/s1>, Table S1: Picture-level ratings of the 120 EMselec pictures in the Russian-speaking sample; Table S2: Russian instructions and rating scales with English translation; Table S3: Pairwise comparisons between affective categories within each sample; Table S4: Dispersion of picture means by cell, models with separate residual variances and analysis of within-picture standard deviations; Table S5: Participant-level models of the Russian ratings.

Author Contributions: Conceptualization, A.K. (Alexander Kirsanov); methodology, A.K. (Alexander Kirsanov); software, A.K. (Anton Karapuzov) and T.L.; formal analysis, A.K. (Alexander Kirsanov); investigation, D.M. and E.L.; data curation, D.M. and E.L.; writing—original draft preparation, A.K. (Alexander Kirsanov); writing—review and editing, all authors; supervision, E.B.; funding acquisition, E.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a grant from Saint Petersburg State University (No. 125022002751-7).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee for Research Involving Human Participants of the Saint Petersburg Psychological Society (protocol No. 22, approved on 25 May 2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Picture-level data supporting the findings of this study (Russian means, standard deviations and numbers of valid ratings for each picture) are provided in Supplementary Table S1. Participant-level data and the analysis code are available from the corresponding author upon reasonable request. The EmoMadrid normative values are available from the database authors at <http://www.psicologiauam.es/CEACO/EmoMadrid.htm>, accessed on 19 October 2026.

Acknowledgments: We thank all participants who took part in the study and the authors of the EmoMadrid database for making the pictures and normative data available for research. We are also grateful to the three anonymous reviewers for their careful reading of the manuscript and constructive comments, which helped to improve the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Carretié, L. Exogenous (automatic) attention to emotional stimuli: A review. *Cogn. Affect. Behav. Neurosci.* **2014**, *14*, 1228–1258. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lang, P.J.; Bradley, M.M. Emotion and the motivational brain. *Biol. Psychol.* **2010**, *84*, 437–450. [\[CrossRef\]](#) [\[PubMed\]](#)
- Siedlecka, E.; Denson, T.F. Experimental methods for inducing basic emotions: A qualitative review. *Emot. Rev.* **2019**, *11*, 87–97. [\[CrossRef\]](#)
- Hu, W.; Huang, G.; Li, L.; Zhang, L.; Zhang, Z.; Liang, Z. Video-triggered EEG-emotion public databases and current methods: A survey. *Brain Sci. Adv.* **2020**, *6*, 255–287. [\[CrossRef\]](#)
- Romeo, Z.; Fusina, F.; Semenzato, L.; Bonato, M.; Angrilli, A.; Spironelli, C. Comparison of slides and video clips as different methods for inducing emotions: An electroencephalographic alpha modulation study. *Front. Hum. Neurosci.* **2022**, *16*, 901422. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zempelin, S.; Sejunaite, K.; Lanza, C.; Riepe, M.W. Emotion induction in young and old persons on watching movie segments: Facial expressions reflect subjective ratings. *PLoS ONE* **2021**, *16*, e0253378. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gerdes, A.B.M.; Wieser, M.J.; Alpers, G.W. Emotional pictures and sounds: A review of multimodal interactions of emotion cues in multiple domains. *Front. Psychol.* **2014**, *5*, 1351. [\[CrossRef\]](#) [\[PubMed\]](#)
- Västfjäll, D. Emotion induction through music: A review of the musical mood induction procedure. *Music. Sci.* **2001**, *5*, 173–211. [\[CrossRef\]](#)
- Weninger, F.; Eyben, F.; Schuller, B.W.; Mortillaro, M.; Scherer, K.R. On the acoustics of emotion in audio: What speech, music, and sound have in common. *Front. Psychol.* **2013**, *4*, 292. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fischler, I.; Bradley, M. Event-related potential studies of language and emotion: Words, phrases, and task effects. *Prog. Brain Res.* **2006**, *156*, 185–203. [\[CrossRef\]](#) [\[PubMed\]](#)
- Landis, T. Emotional words: What's so different from just words? *Cortex* **2006**, *42*, 823–830. [\[CrossRef\]](#) [\[PubMed\]](#)
- Felnhofer, A.; Kothgassner, O.D.; Schmidt, M.; Heinzle, A.-K.; Beutl, L.; Hlavacs, H.; Kryspin-Exner, I. Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *Int. J. Hum.-Comput. Stud.* **2015**, *82*, 48–56. [\[CrossRef\]](#)
- Freeman, D. Creating emotion in games: The craft and art of Emotioneering™. *Comput. Entertain.* **2004**, *2*, 15. [\[CrossRef\]](#)
- Hemenover, S.H.; Bowman, N.D. Video games, emotion, and emotion regulation: Expanding the scope. *Ann. Int. Commun. Assoc.* **2018**, *42*, 125–143. [\[CrossRef\]](#)
- Carretié, L.; Tapia, M.; López-Martín, S.; Albert, J. EmoMadrid: An emotional pictures database for affect research. *Motiv. Emot.* **2019**, *43*, 929–939. [\[CrossRef\]](#)
- Dan-Glauser, E.S.; Scherer, K.R. The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behav. Res. Methods* **2011**, *43*, 468–477. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kurdi, B.; Lozano, S.; Banaji, M.R. Introducing the Open Affective Standardized Image Set (OASIS). *Behav. Res. Methods* **2017**, *49*, 457–470. [\[CrossRef\]](#) [\[PubMed\]](#)
- Marchewka, A.; Żurawski, 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* **2014**, *46*, 596–610. [\[CrossRef\]](#) [\[PubMed\]](#)
- Delplanque, S.; N'diaye, K.; Scherer, K.; Grandjean, D. Spatial frequencies or emotional effects? A systematic measure of spatial frequencies for IAPS pictures by a discrete wavelet analysis. *J. Neurosci. Methods* **2007**, *165*, 144–150. [\[CrossRef\]](#) [\[PubMed\]](#)
- Haxby, J.V.; Hoffman, E.A.; Gobbini, M.I. The distributed human neural system for face perception. *Trends Cogn. Sci.* **2000**, *4*, 223–233. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sabatinelli, D.; Fortune, E.E.; Li, Q.; Siddiqui, A.; Krafft, C.; Oliver, W.T.; Beck, S.; Jeffries, J. Emotional perception: Meta-analyses of face and natural scene processing. *NeuroImage* **2011**, *54*, 2524–2533. [\[CrossRef\]](#) [\[PubMed\]](#)
- Maffei, A.; Angrilli, A. E-MOVIE—Experimental MOVies for Induction of Emotions in neuroscience: An innovative film database with normative data and sex differences. *PLoS ONE* **2019**, *14*, e0223124. [\[CrossRef\]](#) [\[PubMed\]](#)
- Samson, A.C.; Kreibig, S.D.; Soderstrom, B.; Wade, A.A.; Gross, J.J. Eliciting positive, negative and mixed emotional states: A film library for affective scientists. *Cogn. Emot.* **2016**, *30*, 827–856. [\[CrossRef\]](#) [\[PubMed\]](#)
- Boğa, M.; Koyuncu, M.; Kaça, G.; Bayazit, T.O. Comparison of emotion elicitation methods: 3 methods, 3 emotions, 3 measures. *Curr. Psychol.* **2023**, *42*, 18670–18685. [\[CrossRef\]](#)
- Bradley, M.M.; Lang, P.J. Emotion and motivation. In *Handbook of Psychophysiology*, 3rd ed.; Cacioppo, J.T., Tassinary, L.G., Berntson, G.G., Eds.; Cambridge University Press: New York, NY, USA, 2007; pp. 581–607. [\[CrossRef\]](#)
- Lakens, D.; Fockenberg, D.A.; Lemmens, K.P.H.; Ham, J.; Midden, C.J.H. Brightness differences influence the evaluation of affective pictures. *Cogn. Emot.* **2013**, *27*, 1225–1246. [\[CrossRef\]](#) [\[PubMed\]](#)
- Russell, J.A. Affective space is bipolar. *J. Pers. Soc. Psychol.* **1979**, *37*, 345–356. [\[CrossRef\]](#)
- Russell, J.A.; Mehrabian, A. Evidence for a three-factor theory of emotions. *J. Res. Personal.* **1977**, *11*, 273–294. [\[CrossRef\]](#)
- Soares, A.P.; Pinheiro, A.P.; Costa, A.; Frade, C.S.; Comesaña, M.; Pureza, R. Adaptation of the International Affective Picture System (IAPS) for European Portuguese. *Behav. Res. Methods* **2015**, *47*, 1159–1177. [\[CrossRef\]](#) [\[PubMed\]](#)

30. Olofsson, J.K.; Nordin, S.; Sequeira, H.; Polich, J. Affective picture processing: An integrative review of ERP findings. *Biol. Psychol.* **2008**, *77*, 247–265. [[CrossRef](#)] [[PubMed](#)]
31. Lang, P.J.; Bradley, M.M.; Cuthbert, B.N. *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual*; Technical Report A-8; University of Florida: Gainesville, FL, USA, 2008.
32. Redies, C.; Grebenkina, M.; Mohseni, M.; Kaduhm, A.; Dobel, C. Global image properties predict ratings of affective pictures. *Front. Psychol.* **2020**, *11*, 953. [[CrossRef](#)] [[PubMed](#)]
33. Branco, D.; Gonçalves, Ó.F.; Bermúdez i Badia, S. A systematic review of International Affective Picture System (IAPS) around the world. *Sensors* **2023**, *23*, 3866. [[CrossRef](#)] [[PubMed](#)]
34. Matsumoto, D.; Hwang, H.S. Culture and emotion: The integration of biological and cultural contributions. *J. Cross-Cult. Psychol.* **2012**, *43*, 91–118. [[CrossRef](#)]
35. Mesquita, B. Emotions in collectivist and individualist contexts. *J. Pers. Soc. Psychol.* **2001**, *80*, 68–74. [[CrossRef](#)] [[PubMed](#)]
36. Matsumoto, D.; Yoo, S.H.; Nakagawa, S. 37 Members of the Multinational Study of Cultural Display Rules. Culture, emotion regulation, and adjustment. *J. Pers. Soc. Psychol.* **2008**, *94*, 925–937. [[CrossRef](#)] [[PubMed](#)]
37. Sheldon, K.M.; Titova, L.; Gordeeva, T.O.; Osin, E.N.; Lyubomirsky, S.; Bogomaz, S. Russians inhibit the expression of happiness to strangers: Testing a display rule model. *J. Cross-Cult. Psychol.* **2017**, *48*, 718–733. [[CrossRef](#)]
38. Lindquist, K.A.; MacCormack, J.K.; Shaback, H. The role of language in emotion: Predictions from psychological constructionism. *Front. Psychol.* **2015**, *6*, 444. [[CrossRef](#)] [[PubMed](#)]
39. Pavlenko, A. *The Bilingual Mind: And What It Tells Us about Language and Thought*; Cambridge University Press: Cambridge, UK, 2014. [[CrossRef](#)]
40. Alvarado, N.; Jameson, K.A. Shared knowledge about emotion among Vietnamese and English bilingual and monolingual speakers. *J. Cross-Cult. Psychol.* **2011**, *42*, 963–982. [[CrossRef](#)]
41. Bidelman, G.M. The role of the auditory brainstem in processing musically relevant pitch. *Front. Psychol.* **2013**, *4*, 264. [[CrossRef](#)] [[PubMed](#)]
42. Virtala, P.; Tervaniemi, M. Neurocognition of major-minor and consonance-dissonance. *Music Percept.* **2017**, *34*, 387–404. [[CrossRef](#)]
43. Kirsanov, A.; Koriakina, M.; Panasenko, D.; Nikishkina, U.; Blagovechtchenski, E. Electrophysiological and behavioural responses to consonant and dissonant piano chords as standardised affective stimuli. *Front. Hum. Neurosci.* **2025**, *19*, 1689067. [[CrossRef](#)] [[PubMed](#)]
44. EmoMadrid. Available online: <https://www.psicologiauam.es/CEACO/EmoMadrid.htm> (accessed on 4 April 2025).
45. EmoMadrid. Contributing Data to EmoMadrid. Available online: <https://www.psicologiauam.es/CEACO/EmoMadrid/EmoMadridContrib.htm> (accessed on 4 April 2025).
46. Saal, F.E.; Downey, R.G.; Lahey, M.A. Rating the ratings: Assessing the psychometric quality of rating data. *Psychol. Bull.* **1980**, *88*, 413–428. [[CrossRef](#)]
47. Bonett, D.G.; Wright, T.A. Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika* **2000**, *65*, 23–28. [[CrossRef](#)]
48. McGraw, K.O.; Wong, S.P. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* **1996**, *1*, 30–46. [[CrossRef](#)]
49. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [[CrossRef](#)] [[PubMed](#)]
50. Bland, J.M.; Altman, D.G. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* **1999**, *8*, 135–160. [[CrossRef](#)] [[PubMed](#)]
51. Rosenthal, R. *Meta-Analytic Procedures for Social Research*, Rev. ed.; Sage: Newbury Park, CA, USA, 1991. [[CrossRef](#)]
52. Nakagawa, S.; Schielzeth, H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **2013**, *4*, 133–142. [[CrossRef](#)]
53. Breusch, T.S.; Pagan, A.R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* **1979**, *47*, 1287–1294. [[CrossRef](#)]
54. Pessoa, L. Emotion and the interactive brain: Insights from comparative neuroanatomy and complex systems. *Emot. Rev.* **2018**, *10*, 204–216. [[CrossRef](#)] [[PubMed](#)]
55. Barrett, L.F. The theory of constructed emotion: An active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* **2017**, *12*, 1–23. [[CrossRef](#)] [[PubMed](#)]
56. Barrett, L.F. Solving the emotion paradox: Categorization and the experience of emotion. *Personal. Soc. Psychol. Rev.* **2006**, *10*, 20–46. [[CrossRef](#)] [[PubMed](#)]
57. Lim, N. Cultural differences in emotion: Differences in emotional arousal level between the East and the West. *Integr. Med. Res.* **2016**, *5*, 105–109. [[CrossRef](#)] [[PubMed](#)]

58. Kuppens, P.; Tuerlinckx, F.; Russell, J.A.; Barrett, L.F. The relation between valence and arousal in subjective experience. *Psychol. Bull.* **2013**, *139*, 917–940. [[CrossRef](#)] [[PubMed](#)]
59. Quinn, N.; Mathews, H.F. Emotional arousal in the making of cultural selves. *Anthropol. Theory* **2016**, *16*, 359–389. [[CrossRef](#)]
60. Tsai, J.L.; Knutson, B.; Fung, H.H. Cultural variation in affect valuation. *J. Pers. Soc. Psychol.* **2006**, *90*, 288–307. [[CrossRef](#)] [[PubMed](#)]
61. Tsai, J.L. Ideal affect in daily life: Implications for affective experience, health, and social behavior. *Curr. Opin. Psychol.* **2017**, *17*, 118–128. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.