



# A new type of pictorial database: The Bicolor Affective Silhouettes and Shapes (BASS)

Claudia Kawai<sup>1</sup> · Gáspár Lukács<sup>1</sup> · Ulrich Ansorge<sup>1,2,3</sup>

Accepted: 19 February 2021 / Published online: 7 May 2021

© The Author(s) 2021

## Abstract

We introduce the Bicolor Affective Silhouettes and Shapes (BASS): a set of 583 normed black-and-white silhouette images that is freely available via <https://osf.io/anej6/>. Valence and arousal ratings were obtained for each image from US residents as a Western population ( $n = 777$ ) and Chinese residents as an Asian population ( $n = 869$ ). Importantly, the ratings demonstrate that, notwithstanding their visual simplicity, the images represent a wide range of affective content (from very negative to very positive, and from very calm to very intense). In addition, speaking to their cultural neutrality, the valence ratings correlated very highly between US and Chinese ratings. Arousal ratings were less consistent between the two samples, with larger discrepancies in the older age groups inviting further investigation. Due to their simplistic and abstract nature, our silhouette images may be useful for intercultural studies, color and shape perception research, and online stimulus presentation in particular. We demonstrate the versatility of the BASS by an example online experiment.

**Keywords** Affective rating · Silhouette · Emotion · Database · Valence · Arousal · Online research

Images are an excellent medium to convey information, simple or complex. They serve as stimuli in a wide range of research areas, such as emotion, attention, or aesthetics research, among others (e.g., Bradley & Lang, 2007; Huston et al., 2015; Lindsay, 2020). In the present article, we introduce an open access database of normed Bicolor Affective Silhouettes and Shapes (BASS).<sup>1</sup> Each of the BASS images

consists of  $300 \times 300$  black and white pixels, providing a computationally economical and visually uniform layout.<sup>2</sup>

New freely available normed images enrich research by providing greater diversity of choices for research, including potential replications of findings with different kinds of images to ensure generalizability. Additionally, participants may get habituated to the images in existing databases, meaning they may process them differently on repeated exposures (Baker et al., 2010; Foa & Kozak, 1986; Ramaswami, 2014). The present BASS database may be used, for instance, for conceptual replications and novel research in such diverse areas as space-valence congruence effects (e.g., Meier & Robinson, 2004), priming of evaluations (e.g., Fazio, 2001),

<sup>1</sup> Some of the BASS images are not *silhouettes* in the strictest sense, but rather *silhouette-like shapes* because they also illustrate hollow parts or outlines (e.g., the edges and dots on dice). Nonetheless, for simplicity, in the present article, we refer to all the BASS images as “silhouettes.”

✉ Claudia Kawai  
claudia.kawai@univie.ac.at

<sup>1</sup> Department of Cognition, Emotion, and Methods in Psychology, University of Vienna, Vienna, Austria

<sup>2</sup> Vienna Cognitive Science Hub, University of Vienna, Vienna, Austria

<sup>3</sup> Research Platform Mediatized Lifeworlds, University of Vienna, Vienna, Austria

<sup>2</sup> Several databases of shapes or objects do exist, like MultiPic (Duñabeitia et al., 2018), ALOI (Geusebroek et al., 2005), the sets by Nishimoto et al. (2012), by Brady et al. (2013), or by Snodgrass and Vanderwart (Rossion & Pourtois, 2004; Snodgrass & Vanderwart, 1980). However, these were not created (let alone verified or normed) with the intention of eliciting affect: The majority of these images depict everyday objects or animals. With a few exceptions, like dangerous versus cute animals, or weapons, it is unlikely that they would elicit strong affective responses or emotions. In the OASIS, for example, people and scene pictures were rated as more emotional than objects (Kurdi et al., 2017, p. 463), which suggests that these databases are not ideal for emotion research purposes.

affective priming (e.g., Hermans et al., 2001), or emotional facilitation (e.g., Schupp et al., 2003).

More importantly, the BASS also has very specific advantages compared to other available stimulus sets: Silhouette images allow for the study of meaning-related processing without many of the confounds that are present in words and pictures. Whereas words carry inherent differences, such as their length and their phonological transparency (or orthographic depth, see Aro & Wimmer, 2003; Frost et al., 1987; Schmalz et al., 2016) for which one would have to control, these confounding differences are absent in silhouettes. Likewise, whereas pictures differ from one another in terms of color heterogeneity, depth cues, spatial perspective, or feature complexity, silhouettes are more uniform and less complex, and can therefore be more easily equated for these factors, without corrupting their meaning entirely.

The main purpose of the present research article is to demonstrate that the BASS images evoke a very wide range of affective representations on the side of the participants—despite the images' relative simplicity. We demonstrate this primarily by collecting valence and arousal ratings in a large and nationally representative US (Western) sample. We provide further evidence from analogous ratings from an extensive (Eastern) sample from Mainland China and from demonstrating characteristic valence-color congruence effects in an example response time experiment using BASS images (in an Austrian sample).

This methodological approach validates several important characteristics of the BASS images, relevant to but often not investigated in image sets used in psychological research. First, typically, normative ratings of affective picture databases like the Geneva Affective Picture Database (GAPED), the Open Affective Standardized Image Set (OASIS), the International Affective Picture System (IAPS), or the Nencki Affective Picture System (NAPS) (Dan-Glauser & Scherer, 2011; Kurdi et al., 2017; Lang et al., 1997; Marchewka et al., 2014, respectively) are largely based on mono-cultural samples, with ratings predominantly collected in Western populations only. Moreover, within Western populations, samples were often restricted to student participants.<sup>3</sup> Yet, ideally, behavioral image or stimulus sets should be applicable in different cultures in order to allow investigation of universal psychological phenomena and topics. The IAPS, for example, has been generally found to be valid among Western cultures (Deák et al., 2010; Grünh & Scheibe, 2008; Verschuere et al., 2001, to name a few), but less so among Asian cultures (Gong & Wang, 2016; Huang et al., 2015). Therefore, from the start, for the current BASS images, we intended to collect normative rating data from at least two large, representative samples of different cultural origin: one Western (US) and one Eastern (Mainland China) sample. Silhouette

images theoretically carry potential for more culturally neutral affective representations than photographic images, due to the latter's culture-specific details (such as depictions of regionally different vegetation, landscapes, food, people, architecture, traffic etc.), which are often subtle and difficult to control for. We tested this conceptual notion by directly comparing the consistency of the ratings from our Western and Asian samples.

Second, the bicolor silhouettes tackle one particular virulent problem posed by online studies: accurate color rendition of heterogeneous colors on different physical displays. By implication of their high realism, color photos of natural scenes carry substantial color heterogeneity and, in fact, their realism relies in part on this color heterogeneity (cf. Hansen & Gegenfurtner, 2009). It has indeed been shown that differences in the display of photographs, such as varying brightness or resolution, can substantially affect the evaluation of emotional content (e.g., Lakens et al., 2013; Mould et al., 2012). However, the huge variety of physical devices used to display images (e.g., smartphones, tablets, and laptops by different manufacturers) plus an increased variance in idiosyncratic monitor settings and lighting conditions on the side of the viewer compromise accurate color rendition, especially in online studies (Anwyl-Irvine et al., 2020)—a problem which increases with the complexity of the image's color palette. In contrast, the black-and-white silhouettes of the BASS database were already created and selected (from existing silhouettes) under the premise of having to convey their affective meanings largely by their contours alone and less by their exact color rendition. In this way, the BASS images reduce the heterogeneity of the necessary color palette for their rendition and the problems associated with this color heterogeneity. We demonstrated this by collecting participants' affective ratings using the BASS images in inverse colors (pilot study).

However, this is not to say that colors could not be manipulated and used to study their effects (e.g., on affective ratings) in studies with BASS images. For photographs, manipulations of color rendition can easily corrupt image meaning altogether (cf. Oliva & Schyns, 2000; Mould et al., 2012). Such manipulation requires complicated technical procedures (e.g., Orzan et al., 2007), and, even more fundamental, one can hardly be sure what manipulation of colors in a photography affects emotional image content in what manner: Setting up a proper balance of hue, saturation, and brightness is difficult enough when calibrating the display of a single color (Wilms & Oberfeld, 2018), let alone for a photograph with thousands of interacting colors. In contrast, the BASS images were created with the intention of keeping an unambiguous meaning by way of their contours, even if presented in varying colors. Thus, BASS images offer an extremely easy way to manipulate and study or control color: They can be converted effortlessly to any other two colors<sup>4</sup>—and, as long as the two colors are discernible from each other, the

<sup>3</sup> IAPS: US undergraduate students, OASIS: more diverse but predominantly white US participants, GAPED: South/West-European psychology students, NAPS: Polish and other mostly European students.

<sup>4</sup> An R script for easy instant conversion of any number of images to any chosen color is available via <https://gasparl.github.io/BASS/>.

content remains unambiguous. Below, we also demonstrate how such color manipulations of the BASS images could be used in research by investigating known color-valence congruence effects using variously colored BASS images for an example experiment.

Third, large photographic images are generally not optimal for fast decision tasks based on their diverse visual characteristics and complex semantics: Various depicted details and their many subtle visual properties might affect each participant very differently, influencing processing time and potentially confounding results (especially where small effects are expected and processing-time differences in fractions of seconds can affect outcomes). In contrast, the contents of the less ambiguous (few or single) objects in each BASS image is easy to grasp quickly and better suited for a task involving fast decisions, unaffected by visual or semantic noise. In addition, their simplistic and abstract nature makes BASS silhouettes more resistant than photographs to incidental but potentially meaning-corrupting influences such as display size or viewing distance (Anwyl-Irvine et al., 2020).<sup>5</sup> Big file sizes of photographic images used in experiments typically pose technical problems too, especially in connection with online research, which is becoming more and more important in psychological science. For one, common browsers are not optimized for precisely timing image presentation, and physical display times can be strongly affected by loading large size images (e.g., Garaizar & Reips, 2019). For another, some participants might be reluctant (or even unable, where infrastructure is weak) to download many large files. It is not unusual for an experiment to require several hundreds of different stimuli (especially when each stimulus is presented repeatedly in different forms, e.g., varying hue, etc.; e.g., Huang et al., 2015; Kawai et al., 2020; Oliva & Schyns, 2000). Calculating with the average file size of IAPS images (262 KB; cf. averages of 92 KB for OASIS and 789 KB for GAPED), just 100 images take 26 MB—already a notable hurdle for potential participants. Participant dropout due to inadequate internet speed could even introduce confounds via selective attrition (Zhou & Fishbach, 2016). With an extremely economical average file size of 2.2 KB (max. 6.0 KB), the BASS images minimize the issues of display timing and download. Our follow-up experiment on color-valence congruence effects also serves to demonstrate these usability advantages in fast decision tasks online.

Fourth, silhouettes, contours, and shapes themselves constitute a field of scientific interest (e.g., Bar, 2007; Biederman, 1987; Marr, 1982), and the affective qualities associated with geometric contour properties may be of interest to future researchers (e.g., Leder et al., 2011), although they have not been a major focus of research on contours so far. This could be particularly interesting when comparing human object recognition with artificial object recognition (e.g., Rajalingham

et al., 2018): While objects as simple as silhouettes may be processed by machines similarly to humans, affective responses are (currently) only evoked in humans.

## Creating the BASS

Before carrying out the large-scale ratings for the BASS, we assessed several issues in a pilot study ( $n = 180$ , students of the University of Vienna). Detailed information on the pilot study can be found in the online supplementary material available on <https://osf.io/anej6/>. Firstly, researchers suspected, but never tested, possible contamination of affective ratings when participants had to rate valence and arousal in one experiment (Kurdi et al., 2017), since prior studies showed that question order in surveys can influence judgements (Lau et al., 1990; Wilcox & Wleziem, 1993). We put this assumption to the test and split participants into two groups: one rating only valence (or only arousal) and another rating images on both valence and arousal. Results indeed confirmed contamination in affective ratings when two affective ratings were required in the same experiment. Thus, we decided to use only one rating measure (either valence or arousal) per participant in the main study. In the pilot study, we simultaneously tested for the relative-color independence of the silhouette ratings by studying possible effects of color-inversion of silhouettes (black-on-white vs. white-on-black) on ratings. High positive correlations between the (original) black-on-white images and the inverted white-on-black images suggested that there is no apparent preferred color mode for ratings of our silhouettes (clarity ratings:  $r[618] = .866$ , 95% CI [.845, .885]; valence:  $r[618] = .918$ , 95% CI [.905, .930]; arousal:  $r[618] = .719$ , 95% CI [.678, .755]). The ratings from the pilot study, given by Austrian psychology students, also helped us to decide on a reasonable exclusion threshold for intrarater correlation, which we subsequently used as an individual reliability measure for data quality control in the main study. Since intrarater correlation turned out to be lower in the arousal rating than in the valence rating condition, we determined different exclusion thresholds depending on the rating condition (see Data Exclusion).

Clarity ratings provided by the first as well as a second online prestudy ( $n = 50$ , Chinese participants) helped to identify unclear or ambiguous images in the original set of 620. After excluding those with the lowest clarity ratings, the subset that was used in the large-scale ratings in the US and China comprised a total of 583 silhouettes.

## Method

### Participants

For the US rating, a total of 806 participants was recruited via Prolific ([www.prolific.co](http://www.prolific.co)). The study was divided into two

<sup>5</sup> The contents of most BASS images can be easily recognized even when they are strongly blurred or scaled to icon size.

parts: one sub-study comprised the valence rating ( $n = 402$ ), the other the arousal rating ( $n = 404$ ) task. The sample was collected as a “representative sample for the United States of America” (sex, age, ethnicity, according to Simplified US Census; excluding people under the age of 18) and participants received £2.17 upon study completion.

The experiment had to be completed at a desktop computer. Participants could read all relevant information on the welcome page and consented to participate by clicking the consent button at the bottom.

After exclusion (see Data Exclusion), valid data from 777 participants remained (arousal rating: 386; valence rating: 391), out of which 402 participants were female, 375 were male. Ages ranged from 18 to 78 years, with a mean age of 45.19 years ( $SD = 16.09$ ).<sup>6</sup> Completion took an average of 17 min ( $SD = 4$  min, median = 15 min).

For the Chinese sample, the recruitment (and compensation) of online participants in China was managed by Dynata. In the supplementary materials, we provide more information on Dynata’s incentive and compensation system. Chinese participants were asked to provide demographic data after consenting to the study. We only allowed participation for those who indicated (1) their age being between 18 and 95 years, (2) Mainland China as their country of residence, and (3) any gender information. A total of 1341 registered participants submitted their rating data, of which 472 had to be excluded due to poor data quality (for more details see Data Exclusion). We aimed for a representative sample (sex; age, excluding people under the age of 18 years) for the People’s Republic of China, but since our study required PC and Internet access, relatively younger age groups are over-represented.

After exclusion, valid data from 869 participants remained (arousal rating: 455; valence rating: 414), out of which 419 participants were female and 450 were male. Ages ranged from 18 to 81 years, with a mean age of 34.5 years ( $SD = 9.4$ ). Completion took on average 25 min ( $SD = 116$  min, median = 14 min).<sup>7</sup>

An overview of the demographic data from all included participants (US and China) is provided in Table 1 (Appendix A).

## Materials

The majority of the silhouettes were acquired using Google Images ([images.google.com](https://images.google.com)) under the creative commons license, by searching either directly for silhouettes or for photographic pictures that we edited into black-and-white silhouettes using GNU Image Manipulation Program

<sup>6</sup> For two US participants, we do not know their definite ages, only which age group they belong to. Mean age and  $SD$  are calculated without these two participants.

<sup>7</sup> Duration ranged from 10 min to 53 hours. The latter was an extreme outlier that is no doubt due to a measurement error: For example, the participant may have opened the link to the task but only began completing it after a very long pause.

(The GIMP Development Team, 2019) and R (R Core Team, 2020).

We restricted our search to images labeled “available for reuse with modification.” Most silhouettes are from [pixabay.com](https://pixabay.com) ( $n = 444$ ), [cleanpng.com](https://cleanpng.com) ( $n = 33$ ), and [svgsilh.com](https://svgsilh.com) ( $n = 29$ ). Eighteen additional silhouettes were created for this project by a professional illustrator. We collected a wide range of images, depicting humans, animals, objects, and scenes, which we indicated by a category column in the database for easier stimulus selection. Special focus was put on collecting images of a wide range of valence (positive, negative, and neutral) as well as arousal (low, medium, and high) levels. Every image was scaled and/or cropped to a size of  $300 \times 300$  pixels. All images were checked (or converted) to contain only fully black (RGB: 0, 0, 0) and fully white (RGB: 255, 255, 255) pixels.

## Procedure

**Online Rating Study – US** After giving their informed consent, participants saw the instruction page with an explanation of the terms valence and/or arousal (see Figs. 11 and 12 in Appendix A), depending on the participants’ rating condition, and three example images (selected based on the pilot studies as average in both arousal and valence; none of these images were used in the subsequent task for the given participant). The experiment started as soon as the “Start” button on the bottom of the instruction page was pressed. Each black-and-white silhouette stimulus ( $300 \times 300$  px) was presented on a grey background (RGB: 128, 128, 128) for 2 s. When this time had elapsed, the image disappeared and was replaced with the rating scale, a line ranging from “very low” on the left to “very high” on the right, with nine equally spaced tick marks (see Fig. 13 in Appendix A). Participants could enter their ratings by clicking on the scale(s) and submit their rating by pressing a button that appeared below, or they could skip the rating. After confirming their choice, the next image appeared on the screen, and so on. A total of 145–146 silhouettes were presented to each participant. Additionally, we incorporated two attention checks. Hyperlinks to the original experimental websites are available via <https://osf.io/anej6/>.

**Attention checks** After rating every stimulus in the list, a black-and-white image was shown after which the participant had to indicate what was depicted on it out of a list of answers (this was the same item for all participants, a clearly discernible black-and-white image of a car). In addition, we assessed participants’ attention via intrarater reliability. Following the first attention check, the participant was presented with five of their most highly and five of their most lowly rated images (in randomized order). We calculated Spearman’s correlation coefficient between the first and second ratings for these ten items.

**Online Rating Study – China** The procedure of the experiment was identical to that used for the US sample, with the exception

that Chinese participants were asked to provide demographic data (gender, age, country of residence) after giving consent to the study.<sup>8</sup> The experimental website including the informed consent and instruction pages was translated into Mandarin Chinese by a native speaker (a psychologist) and independently double-checked by another native speaker (a linguist).

### Data exclusion

Participants were excluded according to the following criteria: (1) the rate of skipped answers was above 25%; (2) no intrarater correlation could be calculated;<sup>9</sup> (3) for arousal ratings, a failed Attention Check 1 and an intrarater correlation of less than .77 or otherwise a passed Attention Check 1 but an intrarater correlation of less than .67; (4) for valence ratings, a failed Attention Check 1 and an intrarater correlation of less than .80; or otherwise a passed Attention Check 1 but an intrarater correlation of less than .70.<sup>10</sup> In addition, in the Chinese rating, data from participants was excluded if they did not provide a personal identification code (generated by Dynata).

Surprisingly, exclusion rates for the China data (472/1341 = 35.2%) were much larger than that for the US data (29/806 = 3.6%). This may have happened for various reasons, but a major difference is that Prolific (US data) specializes in collecting data for scientific purposes, while Dynata serves mainly as a provider of market research data.

## Results

Each image received between 91 and 119 ratings. For each country, we computed mean valence and mean arousal ratings per image.

### US rating

**Reliability** For valence and arousal ratings, separately, we generated 1000 random split halves of our sample and calculated the Spearman–Brown reliability coefficient of the mean ratings. The mean of the correlation coefficients in the valence dimension was  $\bar{R}_{val} = .986$  ( $SD = 5.51 \times 10^{-5}$ , range:  $R_{min} = .985$  and  $R_{max} = .986$ ), and in the arousal dimension  $\bar{R}_{aro} = .921$  ( $SD = 3.50 \times 10^{-4}$ , range:  $R_{min} = .920$  and  $R_{max} = .929$ ), demonstrating extremely high reliability in both affective dimensions.

<sup>8</sup> The demographic data for people who participated in the representative US rating was provided by Prolific.

<sup>9</sup> This can either happen when the repetition items needed to calculate the correlation were skipped or when the same rating, for example, “5”, was given on every trial.

<sup>10</sup> Exclusion criteria 3–4 were empirically determined by pilot testing in the pilot study, which is described in more detail in the supplementary online material.

**Relationship between valence and arousal** Mean valence ratings from the US participants lie between 1.11 and 8.35 ( $M = 5.64$ ,  $SD = 1.76$ ); mean arousal ratings lie between 2.51 and 7.23 ( $M = 5.00$ ,  $SD = 1.02$ ). The correlation between valence and arousal ratings was significant, with  $r(581) = -.323$ , 95% CI  $[-.394, -.248]$ ,  $p < .001$ ,  $BF_{10} = 5.76 \times 10^{12}$ . However, the Pearson correlation coefficient shows a relatively weak linear relationship and a look at the distributions of the US mean ratings, as shown in Fig. 1, suggests a U-shaped relationship between valence and arousal in the BASS images, where arousal dips for neutral images but then increases with rising valence.<sup>11</sup>

**Effects of demographic variables** Prior research has identified differences in emotion processing based on factors like gender (Bradley et al., 2001; Proverbio et al., 2009; Sabatinelli et al., 2004; Wrase et al., 2003) and age (Bradley & Lang, 2007; Grünh & Scheibe, 2008; Pörto et al., 2010). To investigate potential gender-related differences in our affective ratings, we calculated the individual means for valence and arousal ratings for male and female US participants.

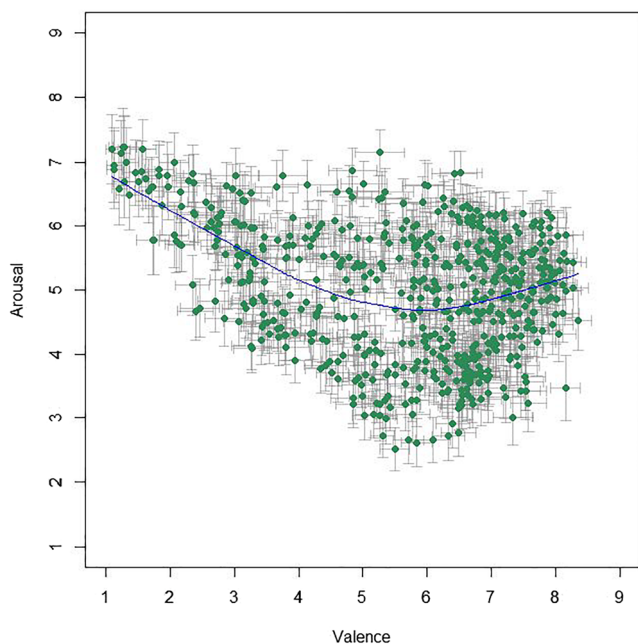
The correlation between the mean ratings made by male and female participants was very high, for both valence,  $r(581) = .972$ , 95% CI  $[.967, .976]$ ,  $p < .001$ ,  $BF_{10} > 10^{200}$ , and arousal,  $r(581) = .877$ , 95% CI  $[.857, .895]$ ,  $p < .001$ ,  $BF_{10} = 2.31 \times 10^{182}$ .

For age-wise comparison, we performed a median split for the demographic variable age on both of our US subsamples (valence, arousal) and correlated the respective mean ratings of the halves. For the valence rating group (median age = 46 years), correlation between the younger (< 46 years) and the older ( $\geq 46$  years) age group was very high, with  $r(581) = .976$ , 95% CI  $[.972, .979]$ ,  $p < .001$ ,  $BF_{10} > 10^{200}$ . The same was true for the arousal rating group (median age = 44.5 years), with a correlation between the younger (< 45 years) and the older ( $\geq 45$  years) age group of  $r(581) = .906$ , 95% CI  $[.891, .920]$ ,  $p < .001$ ,  $BF_{10} > 10^{200}$ . Finally, Fig. 2 shows that the valence-arousal pattern is very similar for all age groups (divided into five brackets following Prolific’s age stratification, see Table 1; see also Fig. 14 in Appendix A).

### Chinese rating

**Reliability** The mean of the correlation coefficients in the valence dimension was  $\bar{R}_{val} = .980$  ( $SD = 7.64 \times 10^{-5}$ , range:  $R_{min} = .979$  and  $R_{max} = .980$ ), and in the arousal dimension  $\bar{R}_{aro} = .921$  ( $SD = 2.87 \times 10^{-4}$ , range:  $R_{min} = .917$  and  $R_{max} = .923$ ). Interrater reliability for both affective dimensions was again extremely high and almost identical to the US sample.

<sup>11</sup> The same U-shaped relationship was found in the IAPS and OASIS. Our observed valence-arousal correlation is very similar to that of the IAPS, which shows  $r(1192) = -.289$ , 95% CI  $[-.340, -.236]$ ,  $p < .001$ ,  $BF_{10} = 2.16 \times 10^{21}$ , while the correlation in the OASIS was not significant, with  $r(898) = -.058$ , 95% CI  $[-.123, .007]$ ,  $p = .082$ ,  $BF_{01} = 2.84$ .



**Fig. 1** Mean ratings (95% CI error bars) per image for valence and arousal from US participants. To illustrate the tendency of the relation, the blue line depicts locally weighted regression (LOWESS)

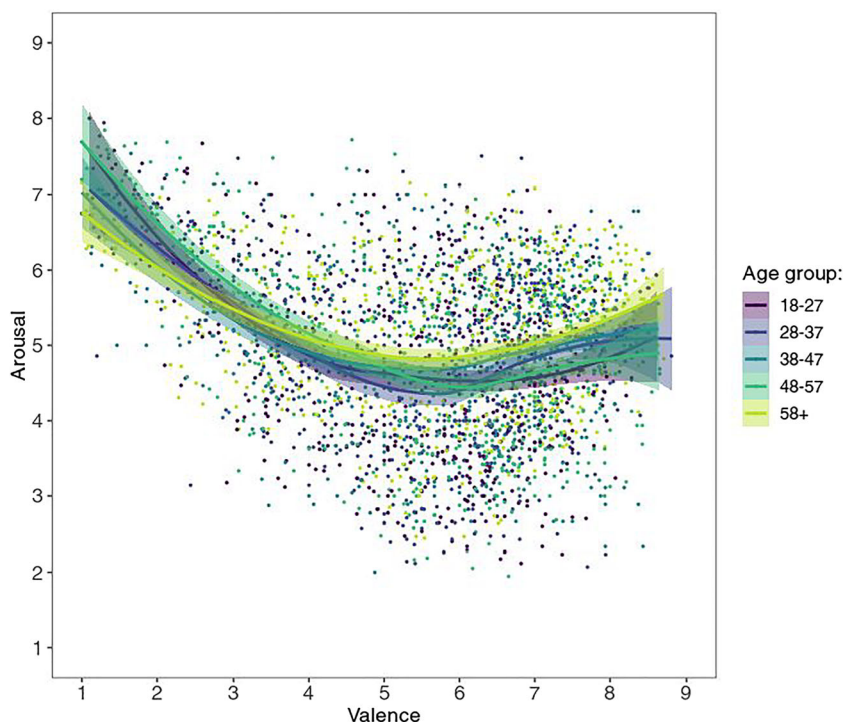
**Relationship between valence and arousal** Mean valence ratings from the Chinese participants lie between 1.47 and 7.84 ( $M = 5.46$ ,  $SD = 1.37$ ); mean arousal ratings lie between 3.65 and 6.75 ( $M = 5.31$ ,  $SD = 0.64$ ). Unexpectedly, and different from the results in the US sample, mean arousal and mean valence ratings showed a strong positive correlation, with  $r(581) = .805$ ,

95% CI [.774, .832],  $p < .001$ ,  $BF_{10} = 2.33 \times 10^{129}$ . The distributions of the Chinese mean valence and arousal ratings are shown in Fig. 3, which illustrates that negative images received relatively low arousal ratings while positive images received relatively high arousal ratings (as opposed to the US ratings, where arousal ratings were distributed fairly evenly between positive and negative images).

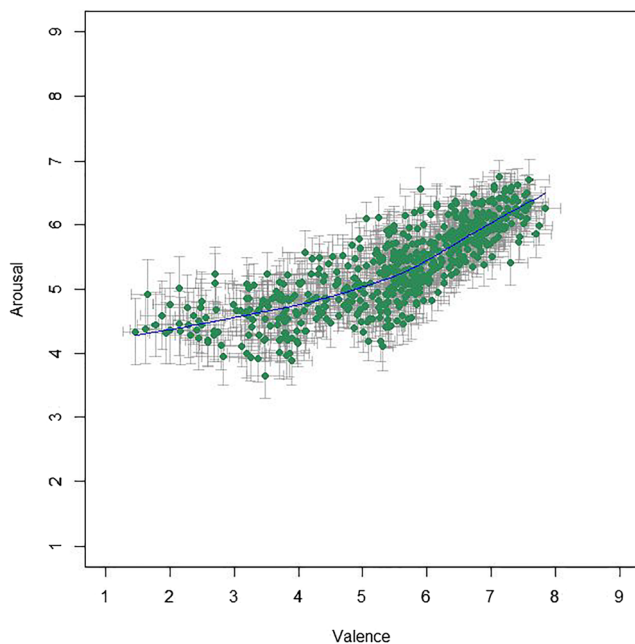
**Effects of demographic variables** Similarly to the US sample, mean ratings between Chinese male and female participants were highly correlated for valence,  $r(581) = .979$ , 95% CI [.975, .982],  $p < .001$ ,  $BF_{10} > 10^{200}$ , and arousal,  $r(581) = .856$ , 95% CI [.832, .876],  $p < .001$ ,  $BF_{10} = 4.11 \times 10^{163}$ .

For age-wise comparison, we again performed a median split for age on both Chinese subsamples (valence, arousal). For the valence rating group (median age = 32 years), correlation between the younger (< 32 years) and the older ( $\geq 32$  years) age group was extremely high, with  $r(581) = .974$ , 95% CI [.970, .978],  $p < .001$ ,  $BF_{10} > 10^{200}$ . For the arousal rating group (median age = 33 years), the agreement between the younger (< 33 years) and the older ( $\geq 33$  years) age group was still very high, although somewhat lower,  $r(581) = .785$ , 95% CI [.752, .814],  $p < .001$ ,  $BF_{10} = 3.41 \times 10^{118}$ . Note that the median age was much lower in the Chinese compared to the US sample: The Chinese sample was split in the early 30s while the US sample was split in the mid-40s.

As for the US sample, five age brackets were created—however, the ratings by Chinese age group “58+” are hardly reliable since this sample consists of a mere 14



**Fig. 2** Mean valence by mean arousal ratings per picture by age group from US participants, with corresponding LOWESS lines



**Fig. 3** Mean ratings (95% CI error bars) per image for valence and arousal from Chinese participants with the LOWESS line in blue

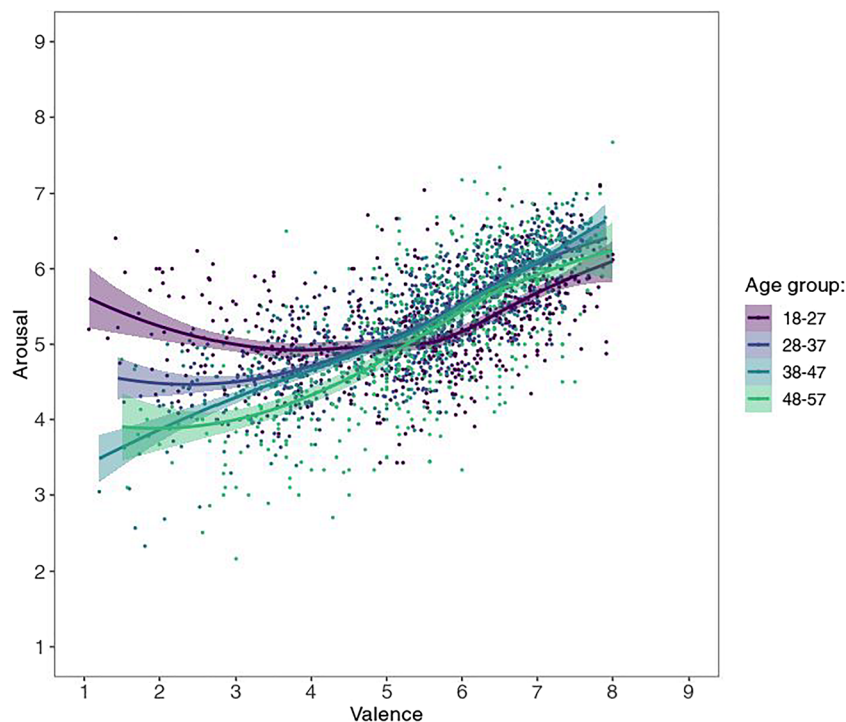
participants—therefore they are omitted from the present figures. (They are nonetheless largely in line with the rest of the data; see supplementary figures at <https://osf.io/anej6/>). Most interestingly, the different age groups show somewhat different patterns (see Fig. 4), with ratings from younger age groups, especially the youngest (ages 18–27),

tending to resemble more to the US sample and the U-shape often observed in Western samples.

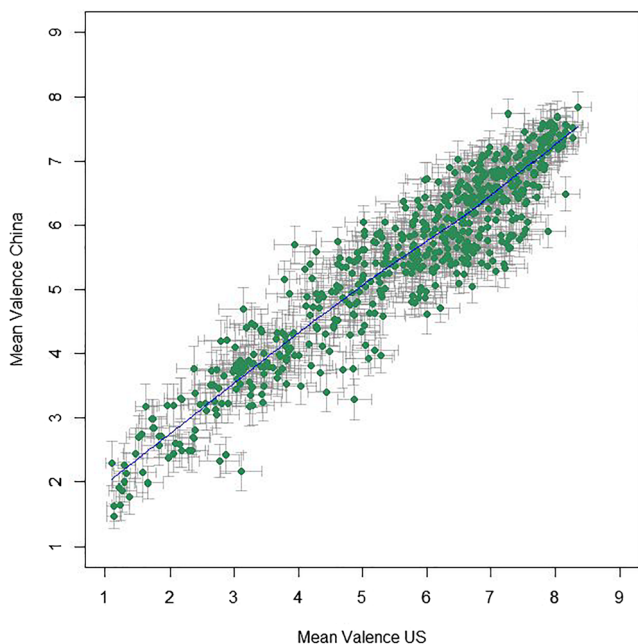
### Comparison of US and Chinese ratings

Mean ratings in the valence category from the US sample showed extremely high positive correlation with those from the Chinese rating,  $r(581) = .936$ , 95% CI [.925, .946],  $p < .001$ ,  $BF_{10} > 10^{100}$  (see Fig. 5). A paired  $t$  test showed that Americans rated images more positive than Chinese, with a statistically significant but practically negligible difference of 0.18, 95% CI [0.12, 0.23] (US:  $5.64 \pm 1.76$ , China:  $5.46 \pm 1.37$ ),  $t(582) = 6.35$ ,  $p < .001$ ,  $d = 0.26$ , 95% CI [0.18, 0.35],  $BF_{10} = 1.21 \times 10^7$ . Both these findings are very similar for all age groups and both sexes (see Fig. 6; Table 2).

The mean ratings for arousal showed only a weak, albeit significant, correlation,  $r(581) = .282$ , 95% CI [.206, .355],  $p < .001$ ,  $BF_{10} = 2.30 \times 10^9$  (see Fig. 7). The lower correlation in the arousal (as compared to the valence) dimension relates to the fact that, as mentioned above, the data from US raters showed a U-shaped valence-arousal-relationship, but the Chinese average affective ratings displayed a linear relationship. However, just as the ratings by younger age groups showed more visually similar valence-arousal patterns (see Fig. 4), they also demonstrated substantially higher correlation with US arousal ratings, as most clearly reflected in the youngest age group (see Fig. 8;  $r = .53$  as opposed to .22, .17, and  $-.07$  for the other three age groups, see Table 2;  $p < .001$  for all  $z$ -test comparisons; Diedenhofen & Musch, 2015). This



**Fig. 4** Mean valence by mean arousal ratings per picture by age group from Chinese participants, with corresponding LOWESS lines



**Fig. 5** Relationship between US and Chinese mean valence ratings (95% CI error bars) per image, with the LOWESS line in blue

pattern of age group influence is similar for both males and females, though the correlation was generally higher for males ( $r = .39$  vs  $.20$ ; Table 2; see also all figures via <https://osf.io/anej6/> with male and female ratings in separate panels).

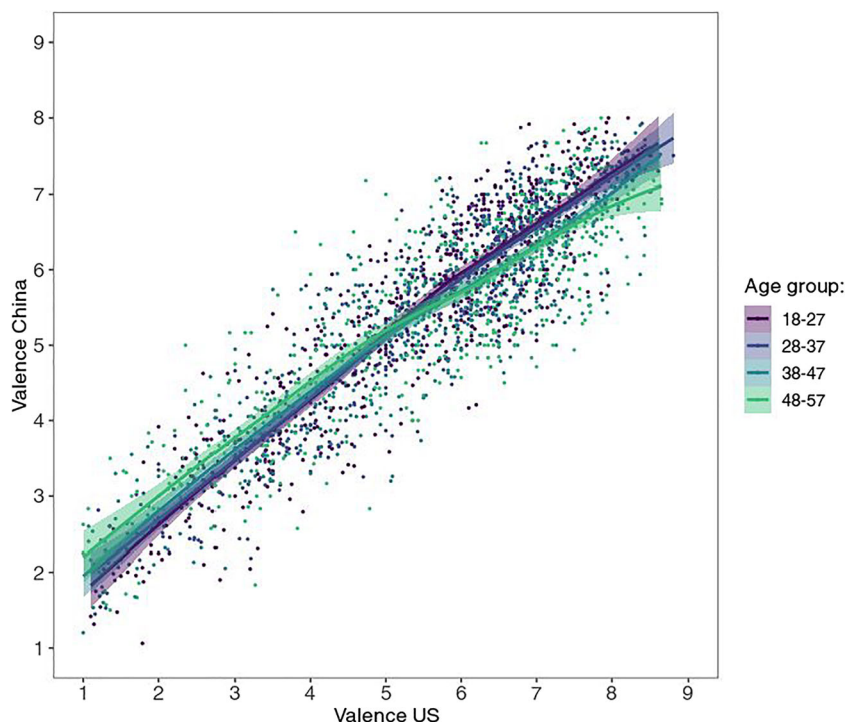
A paired  $t$  test showed that arousal ratings by Americans were lower than those by Chinese, again a statistically

significant yet very small difference,  $-0.31$ , 95% CI  $[-0.40, -0.23]$  (mean rating US:  $5.00 \pm 1.02$ , China:  $5.31 \pm 0.64$ ),  $t(582) = -7.30$ ,  $p < .001$ ,  $d = -0.30$ , 95% CI  $[-0.39, -0.22]$ ,  $BF_{10} = 4.95 \times 10^9$ . This relation is similar for all age groups and both sexes (Table 2).

All in all, we may conclude that the consensus for valence ratings was generally higher than for arousal ratings, both within a sample of a given culture (see correlation tests for interrater reliability, gender, and age) and between the two cultures tested here.

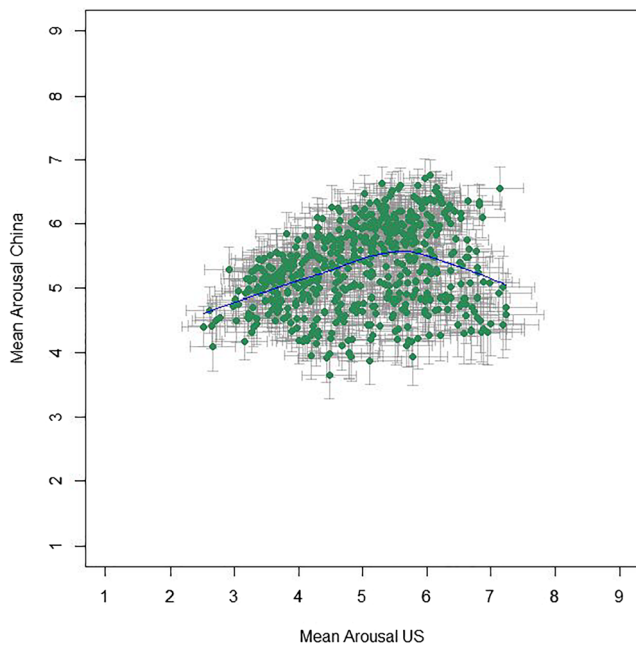
### Example experiment: Color-valence congruence effects in the BASS

We also included an example experiment, in which we tested the suitability of the BASS images for online research. Here, we manipulated color to demonstrate one major advantage of the BASS images: their easy rendition in a different color. Content-wise, it has been repeatedly shown that the stimulus valence of words interacts with perceptual features like color and brightness: Positive stimuli are associated with green (rather than red) and with white (rather than black; Kawai et al., 2020; Kuhbandner & Pekrun, 2013; Lakens et al., 2012; Meier et al., 2004, 2015; Moller et al., 2009). As a validation of the suitability of the BASS for online research, in a web-based fast decision task, we tested if the BASS silhouettes can induce similar effects. Details about the



**Fig. 6** Relationship between US and Chinese mean valence ratings per image by age group, with corresponding LOWESS lines





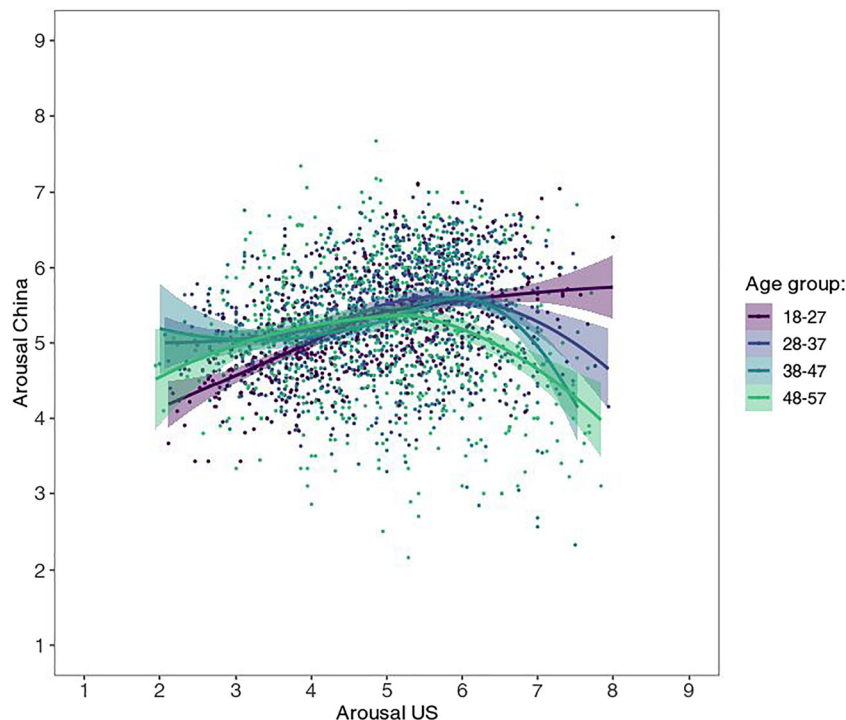
**Fig. 7** Relationship between US and Chinese mean arousal ratings (95% CI error bars) per image, with the LOWESS line in blue

experiment can be found in the online supplementary material. Below we provide only a brief summary.

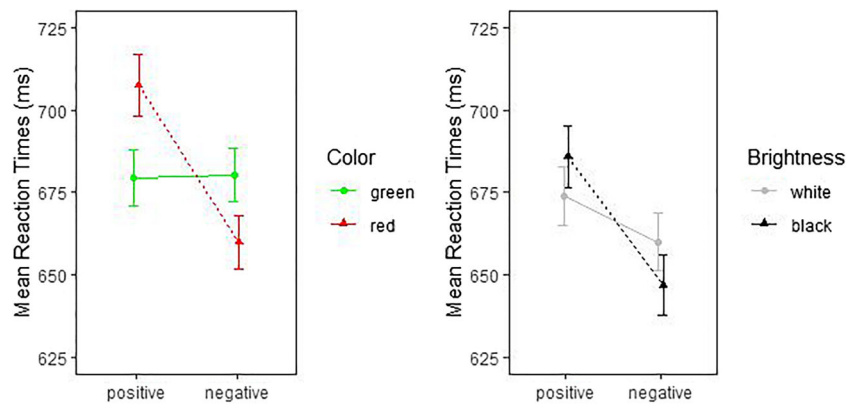
**Method** We selected a subset of 60 positive and 60 negative silhouettes from the BASS (a list of filenames is provided in Table 4 in Appendix C). These 120 images were presented to

90 students (age =  $23.0 \pm 3.5$ ; 27 male) from the University of Vienna in two experimental blocks: a “color” block, containing each silhouette once in red and once in green color; and a “brightness” block, containing each silhouette once in black and once in white (with silhouette background always in gray, see Appendix C, Fig. 15, for examples). The task was to categorize the valence of each silhouette via key press as either positive or negative. For each block, we ran a repeated-measures analysis of variance (ANOVA) on the correct mean response times (RTs) and another on the mean error rates (ERs).

**Results** According to expectations, the two (Color: red vs. green; within)  $\times$  two (Valence: positive vs. negative; within) ANOVA in the color block showed a significant interaction between color and valence for RTs,  $F(1, 89) = 91.46, p < .001, \eta_p^2 = .507, 90\% \text{ CI } [.383, .595], \eta_G^2 = .022, BF_{10} = 3.10 \times 10^9$ , and also for ERs,  $F(1, 89) = 83.67, p < .001, \eta_p^2 = .485, 90\% \text{ CI } [.359, .576], \eta_G^2 = .106, BF_{10} = 2.30 \times 10^{15}$ . Positive silhouettes were categorized faster and more accurately when presented in green rather than red, and, inversely, negative silhouettes were categorized faster and more accurately when presented in red rather than green. Also, in the brightness block, the silhouettes elicited the expected brightness–valence interactions in both RTs,  $F(1, 89) = 18.68, p < .001, \eta_p^2 = .173, 90\% \text{ CI } [.068, .286], \eta_G^2 = .005, BF_{10} = 65.74$ , and ERs,  $F(1, 89) = 8.76, p = .004, \eta_p^2 = .090, 90\% \text{ CI } [.017, .191], \eta_G^2 = .011, BF_{10} = 7.55$ . Positive silhouettes were categorized faster and more



**Fig. 8** Relationship between US and Chinese mean arousal ratings per image by age group, with corresponding LOWESS lines



**Fig. 9** Mean reaction times for positive and negative silhouettes in the color block (*left*) and the brightness block (*right*). Error bars indicate SEMs

accurately when presented in white rather than black, and inversely, negative silhouettes were categorized faster and more accurately when presented in black rather than white.

Means and *SDs* for RTs and ERs are illustrated in Figs. 9 and 10, and the means can be found in Table 5.

## Discussion

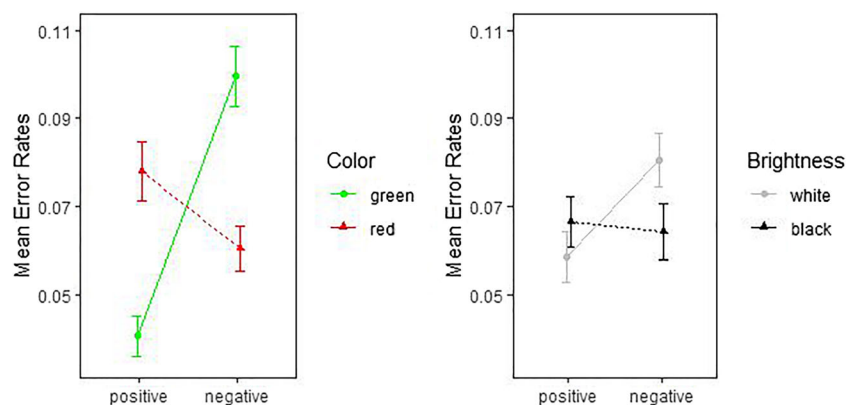
For the current study, we created an affective silhouette database with the specific goal of providing an easy and openly accessible lightweight stimulus pool capable of reliable elicitation of affective representations on a wide range of valence and arousal levels that was validated by extensive samples of two of the biggest culture groups studied in psychology. In the US sample, we found that the images cover a very wide range of affective representations from very negative to very positive (valence), and from very calm to very intense (arousal), indicating various potential applications (see Introduction). The Chinese sample was remarkably similar in valence ratings, covering a wide range with an extremely high correlation with the US sample.

However, the Chinese *arousal* ratings were only weakly correlated, and comparatively limited in range. This relates to the fact that data from the US participants demonstrated a

U-shaped relationship between valence and arousal, while data from the Chinese participants, instead, displayed a more linear valence–arousal relationship (strong positive correlation between valence and arousal ratings) on average.

Studies that look at affective ratings for emotional *words* in Chinese find a U-shaped valence-arousal relation similar to the one apparent in the US ratings (Ho et al., 2015; Liu et al., 2018; Yao et al., 2017). It is noteworthy, that participants in the affective word studies were high school or university students. Our Chinese sample was more diverse: The age-group-wise analysis shows that younger Chinese display a more quadratic relationship which becomes more linear with increasing age. This provides a plausible reconciliation of these previous findings with our results in that it is only the relatively older Chinese generations whose perception substantially differs from those in Western populations, while the younger generations perceive images more in accordance with the Western population. We can only speculate about the reasons, but the more similar ratings of young US and Chinese participants could reflect a larger shared sphere of visual experiences, for instance, through digital social media (e.g., TikTok).

However, two studies that compared affective ratings for emotional *photographs* (a subset of the IAPS) between Chinese and Western samples interestingly seem to indicate



**Fig. 10** Mean error rates for positive and negative silhouettes in the color block (*left*) and the brightness block (*right*). Error bars indicate the SEMs

the opposite phenomenon: Gong and Wang (2016) found a high correlation of valence and arousal ratings between Chinese and German older adults<sup>12</sup>, and Huang et al. (2015) argued that Chinese young adults display a different rating behavior from US young adults for IAPS pictures.

Generally, more empirical data on cultural differences in emotion representation or judgments is necessary to establish how and when Chinese rating behavior differs from that of a Western culture. One possibility might be, for instance, that the lower arousal for negative affective silhouettes, which explains most of the difference in our BASS ratings (and the lack of U-shape among Chinese participants, especially with increasing age), might be a result of conflict avoidance or emotion control as a more generally accepted strategy among Chinese than Westerners (Gong & Wang, 2016; Tjosvold & Sun, 2002).

We also demonstrated how the BASS images could be used in online experiments. In our example experiment, we showed the typical color-congruence effects of stimulus color on speed and accuracy of judgments about image valences: Positive images were rated faster and more accurately if presented in green than in red, and if presented in white than in black, while the opposite held true of negative images.<sup>13</sup> This result replicated prior research and already proves at least one particular use of the BASS images in scientific research.

All in all, the quality of our samples' data (as measured by the respective interrater correlation) is outstanding and, where comparable, its general characteristics are in line with effects observed in notable affective databases like IAPS, OASIS, ANEW and BAWL-R (Vö et al., 2009), to name a few.

The BASS, however, aims only to complement, and not to replace, realistic picture databases, for research for which realistic images do not provide ideal stimuli for a given experiment for any of the reasons detailed in the Introduction (e.g., complications of color manipulation, file size, etc.).

While the simplicity of silhouettes can be advantageous for some research designs, their innate lack of context information may simultaneously lead to more (or

less) ambiguity in their interpretations between participants than is the case for pictures that are rich in detail. Heterogeneous cultural backgrounds between participant groups could add to differences in content evaluation if contexts interact with objects in photorealistic image sets, but not or at least less so with silhouettes where context is sparse. However, while the direct interpretation of some of the silhouettes may, thus, be more or less ambiguous than that of a photographic image of the corresponding content (e.g., some participants may perceive a coyote's silhouette simply as a dog when presented without a fitting habitat as context), in the current study, the affective ratings of the silhouettes generally demonstrated very high consistencies (very high inter-rater correlation, low *SDs*, and, in case of valence, very high intercultural correlation), which attest to the silhouettes' emotional unambiguousness. Furthermore, several numeric measures in our BASS database, such as the rating differences between US and China per image, the corresponding *BFs* or residuals, or simply the *SDs* of the ratings in either sample, allow researchers to choose images of emotional clarity or culture-neutrality suiting their needs best.

Nonetheless, and although the rating data of our BASS images correspond well with the data obtained for photographic databases, we cannot rule out that the silhouettes evoke smaller emotional responses than contextually richer photographic pictures when using other measures (e.g., skin conductance, heart rate, facial muscles activities). Do photorealistic pictures elicit stronger emotion-related physiological responses than silhouettes, and if so, what are the driving forces in such a case (object detail, color, depth and perspective, etc.)? We hope that future research, facilitated by the BASS, will help to address these questions.

## Conclusions

The BASS offers a novel database of visual silhouettes for psychological research optimally suited for the study of image-elicited affective representations in online experiments and validated by normative ratings by representative Western and Asian samples.

**Acknowledgements** Claudia Kawai and Gáspár Lukács are recipients of DOC Fellowships of the Austrian Academy of Sciences at the Department of Cognition, Emotion, and Methods in Psychology at the University of Vienna. The research was further supported by Dr. Franziska Pinsker, for which we thank her. We want to acknowledge the assistance of Yang Zhang, who translated the experiment to Mandarin and collected the online pilot data in China. We also thank Wenyi Chu for her feedback on clarity of the experimental procedure, and Sophie Hanke, Tobias Greif, and Bence Szaszko for making browsing the BASS easier. The meticulous

<sup>12</sup> Gong and Wang (2016) compared their Chinese participant sample with a mean age of  $67.3 \pm 4.96$  years to a German participant sample with a mean age of  $69.61 \pm 3.58$  years (Grühn & Scheibe, 2008). Thus, in this comparison, participants were older than our oldest Chinese age group (58+) which had a mean age of  $62.36 \pm 5.70$  years and included only 14 participants.

<sup>13</sup> The finding may seem to contradict the pilot study results showing that black-on-white and (inverted) white-on-black images are rated very similarly, demonstrating equal recognizability. However, in the pilot study, each participant was presented with only one image type (either all images black-on-white, or all white-on-black). Color-valence congruence effects can depend on dichotomous decision alternatives within the same task (cf. Kawai et al., 2020; Lakens et al., 2012), and, the present results are in line with this possibility and previous research showing that such congruence effects were only elicited in the example experiment (where each task contained contrasting color types alternating trial by trial), not during the affective rating study.

comments Anna Walker provided are greatly appreciated. Special thanks goes to the talented Anna Lumaca ([allhailthesnail.com](mailto:allhailthesnail.com)) for creating and contributing 20 much needed “ghastly” silhouettes to the BASS.

**Availability of data, materials, and code** The BASS and its accompanying data table are available for download from the OSF project page under <https://osf.io/anej6/>. The project page further contains all preregistrations, collected raw data, experimental and analysis scripts, a useful R-script to convert the colors of the silhouettes, as well as detailed reports on the procedures involved in the creation of the BASS and its follow-up experiment. A browsable and downloadable version of the complete database (silhouette images and data table) is also available via <https://gasparl.github.io/BASS/>.

**Authors' contributions** Concept by C.K. and U.A., images by C.K., study design and analysis by C.K. and G.L., presentation software by G.L. and C.K., manuscript by C.K., G.L., and U.A.

**Funding** Open access funding provided by University of Vienna. This research was funded by the Austrian Academy of Sciences (C.K.'s personal grant number 25068, G.L.'s personal grant number 24945).

## Declarations

**Ethics, consent to participate, and consent for publication** The research was conducted according to the principles expressed in the Declaration of Helsinki. During online participation, no potentially identifying information such as IP addresses was collected and participants were informed that (i) their data would be treated anonymously, and (ii) they could unconditionally stop their participation. Participants were informed that their anonymized data would be published and were given the opportunity to prohibit the use of their data. Informed consent was obtained from all subjects.

**Conflict of interest** The authors have no competing interests.

## Appendix A

**Table 1** Demographics of the included participants from the US and Chinese ratings with corresponding means and *SD* for valence and arousal

	US	Mean Valence Rating ( <i>SD</i> )	Mean Arousal Rating ( <i>SD</i> )	China	Mean Valence Rating ( <i>SD</i> )	Mean Arousal Rating ( <i>SD</i> )
<b>Sex</b>						
Female	402	5.70 (2.39)	5.00 (2.25)	419	5.49 (1.97)	5.25 (1.84)
Male	375	5.58 (2.17)	4.99 (2.28)	450	5.45 (1.94)	5.38 (1.82)
<b>Sum</b>	<b>777</b>			<b>869</b>		
<b>Age</b>						
18–27	138	5.41 (2.23)	4.85 (2.24)	186	5.42 (1.99)	5.23 (1.85)
28–37	147	5.55 (2.13)	4.90 (2.33)	419	5.48 (1.97)	5.38 (1.9)
38–47	132	5.66 (2.36)	5.04 (2.21)	176	5.47 (1.94)	5.32 (1.68)
48–57	134	5.70 (2.26)	4.96 (2.39)	74	5.50 (1.85)	5.20 (1.75)
58+	226	5.78 (2.37)	5.17 (2.16)	14	5.47 (1.78)	5.45 (1.74)
<b>Ethnicity</b>						
Asian	58					
Black	106					
Mixed	35					
Other	30					
White	548					

**Table 2** US and Chinese valence and arousal rating correlation per sex and age group

	Valence correlation [95% CI]	Arousal correlation [95% CI]
<b>Sex</b>		
Female	.913 [.899, .926]	.198 [.119, .275]
Male	.942 [.933, .951]	.387 [.316, .454]
<b>Age</b>		
18–27	.910 [.895, .923]	.530 [.469, .586]
28–37	.921 [.908, .933]	.219 [.140, .295]
38–47	.899 [.882, .913]	.170 [.090, .248]
48–57	.856 [.833, .876]	–.069 [–.150, .012]
58+ *	.689 [.644, .730]	.251 [.173, .326]

\* This oldest age group includes only 14 Chinese participants (cf. Table 1), hence the related correlations are not reliable.

## Information and Consent

### Aim:

A picture is worth a thousand words - and can be understood around the globe. With your help, we are creating a silhouette database. You will be shown black and white images, and you will need to rate each image. Your participation enables researchers and others interested to use standardized and representative research material.

### Payment:

The task takes around 15-20 minutes. Completed and valid participation will be rewarded with 2.17 GBP via Prolific. (Your Prolific ID was identified as \_\_\_\_\_. You will receive a corresponding completion link at the end of the experiment.) Please note that we may refuse payment if you pay no attention to the task, give random responses and fail on attention checks.

### Health Risks:

Some of the images we are about to show you may be explicit and can depict violence or erotica. However, there are no known physical or psychological risks associated with participating in this study.

### Your Rights:

You can stop participating in the study at any time without giving a reason (by closing or refreshing this website). In this case, all recorded data will be fully deleted. You can also skip answers during the test, if, for example, you were not sure what you saw on a given image. However, please do not skip too many; we reserve the right to refuse payment in case of a large amount of skipped images. If you would like to know the results of the study, please contact the experimenter (claudia.kawai@univie.ac.at).

### Anonymity and Privacy:

The results of the study are used and published for research purposes. The data do not provide any information about you personally. You will not be named and your identity will be kept strictly confidential. Even after you have participated (until publication), you can prohibit the use of your data.

### Consent:




By clicking the "Consent & Continue" button, you declare that you have read and understood the information above. You confirm that you will be concentrating on the tasks and complete them to the best of your abilities. You agree to participating in this study and the subsequent analysis of your data by authorized personnel.

Consent & Continue

Fig. 11 Screenshot of the experiment's welcome page

### Instructions

We will show you around 160 silhouettes in black and white, like, for example, the following:

Each silhouette will be presented for only a short period of time (2 seconds), so please pay close attention! After the 2 seconds have passed, the silhouette will be replaced with a rating scale.

Use the scale to rate the arousal (low affective/high affective) of the image. Low arousal means that something is calm, tranquil or unexciting. High arousal means that something is intense, exciting or exhilarating.

To give your rating on the scale, move the mouse cursor over the respective scale and click for the slider to appear. Now you can move the slider to the position that you think fits your evaluation best, via clicking or dragging. After selecting a value on each scale, the "Submit" button appears below. When you are satisfied with your choice, click the "Submit" button. Then the next image will appear, and so on. If you do not want to give a rating for an image, click "Continue without rating" and then "Submit".

**You can give your ratings quickly and intuitively. Do not overthink it!**

Fig. 12 Screenshot of the instruction page after consenting (example from the arousal substudy)

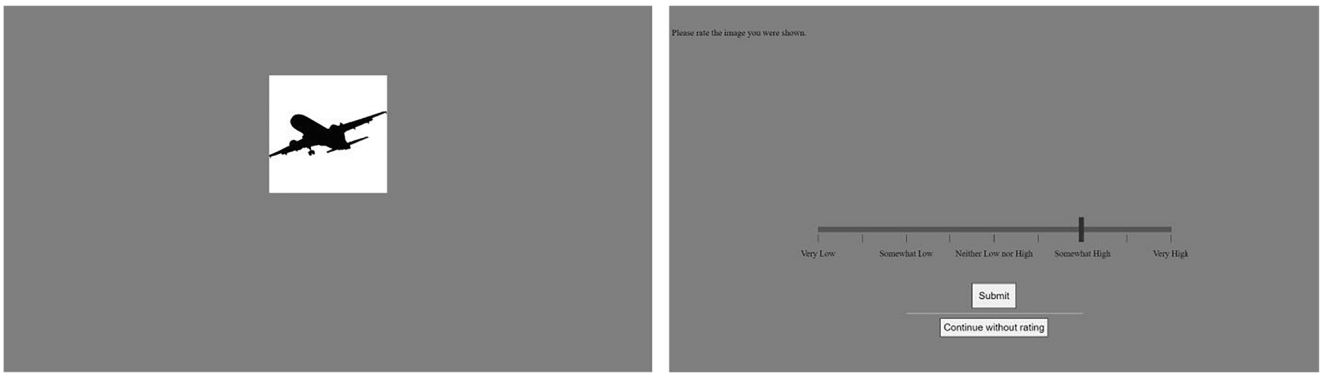


Fig. 13 Screenshot of the presentation of the silhouette stimulus (start of the trial), followed by the appearance of the rating scale (example from the arousal substudy)

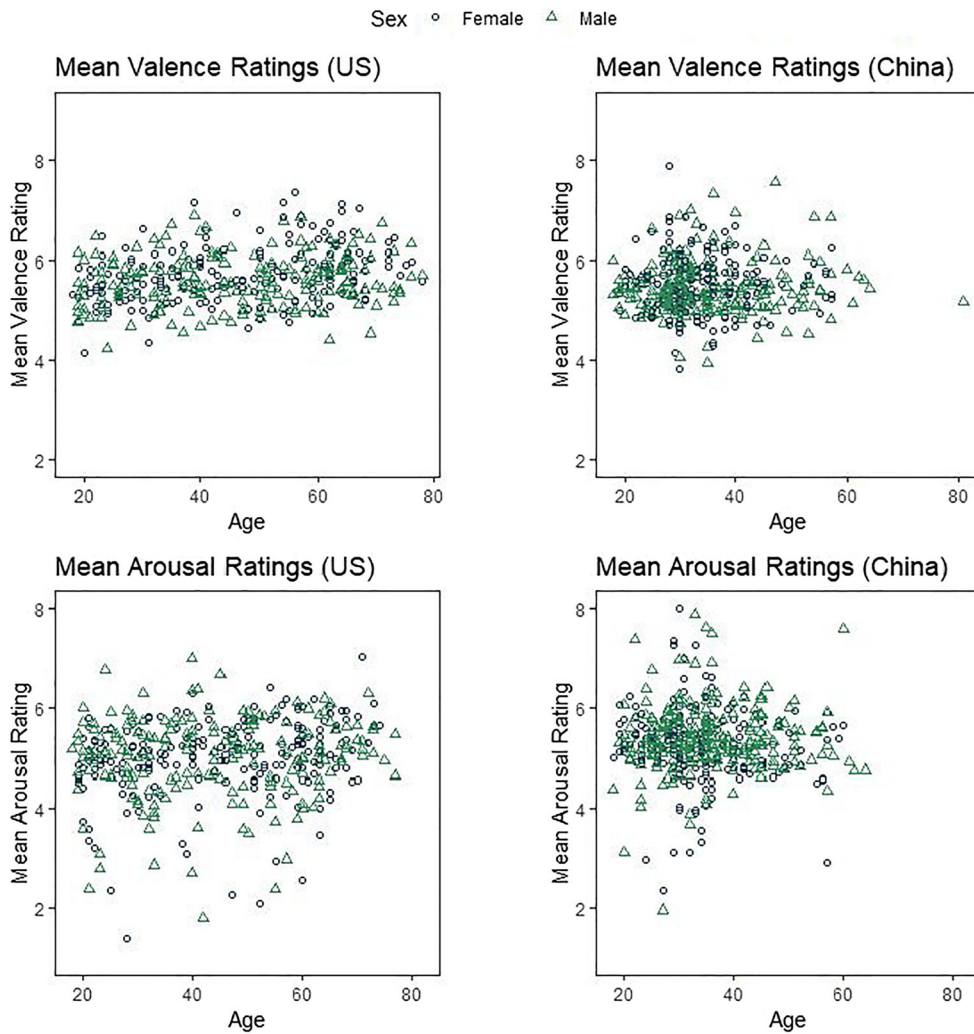


Fig. 14 Mean valence and mean arousal ratings per participant according to their country, sex, and age group

## Appendix B

**Table 3** Bicolor Affective Silhouettes and Shapes (BASS) datatable manual

Column	Explanation
file_name	The image file name.
val_mean_us	The mean of all valence ratings.
aro_mean_us	The mean of all arousal ratings.
val_sd_us	The <i>SD</i> of all valence ratings.
aro_sd_us	The <i>SD</i> of all arousal ratings.
val_n_us	The number of valence ratings.
aro_n_us	The number of arousal ratings.
val_cilo_us	Lower limit of 95% CI for the mean of all valence ratings.
val_ciup_us	Upper limit of 95% CI for the mean of all valence ratings.
aro_cilo_us	Lower limit of 95% CI for the mean of all arousal ratings.
aro_ciup_us	Upper limit of 95% CI for the mean of all arousal ratings.
val_diff	The difference between the means of US and Chinese valence ratings. (val_mean_us minus val_mean_ch).
aro_diff	The difference between the means of US and Chinese arousal ratings. (aro_mean_us minus aro_mean_ch).
val_bf_10	Wilcoxon Bayes factor in support of difference between US and Chinese. valence ratings.
val_bf_01	Wilcoxon Bayes factor in support of equivalence between US and Chinese. valence ratings.
aro_bf_10	Wilcoxon Bayes factor in support of difference between US and Chinese. arousal ratings.
aro_bf_01	Wilcoxon Bayes factor in support of equivalence between US and Chinese. arousal ratings.
val_resid	Linear regression residuals for valence ratings: observed values minus fitted. values for the US-Chinese mean ratings relations for the entire sample. A larger value means larger US mean valence rating than Chinese mean valence rating, relative to what is expected when values are perfectly correlated (linear fitted values). The closer this value is to zero, the more likely it is that the given image is suitable for cross-cultural measurement of valence. (Nonetheless, this value is quite close to val_diff, which is more straightforward to interpret.)
aro_resid	Linear regression residuals for arousal ratings (same as for valence, see above).
black_px	Number of black pixels in the image.
white_px	Number of white pixels in the image.
keywords	Keywords for the given image.
source	Source of the (original) image the silhouette is based on.

*Note.* The columns listed above with “\_us” suffixes refer to rating data from the US sample. The data referring to the Chinese ratings is indicated with “\_ch” suffixes – here omitted for brevity. (E.g., val\_mean\_ch: The mean of all valence ratings from China, etc.)

## Appendix C

**Table 4** Filenames of the positive and negative silhouettes used in the verification study

Positive silhouettes		Negative silhouettes	
trapeze4	badminton	injury	handinjury
wedding2	jump	attack2	grenade
love	highjump4	execution3	robber
singer3	highjump	abuse2	bazooka
couple6	acrobat3	abuse	pollution2
romance	couple5	hanging	gunman
volleyball	skiing2	drunkard	drunk
champion	dance2	vomit2	tick
euphoria	ferriswheel	argument4	axe3
iceskater3	concert	handcuff2	scorpion2
success2	iceskater	noose	mosquito
affection2	skateboard	argument3	gasmask2
cardio	trapeze3	argument2	handcuff
handball	tennis	dispute	knife2
kiss4	handball2	argument6	pollution
ballet	winner	hazard	mosquito2
basketball4	rockstar	hitman	knife4
propose	carefree	sorrow2	snake2
kiss3	basketball	murder2	wasp
joy	breakdance	gunman2	machete
acrobat	skiing3	beggar2	tank2
trapeze	soccer	skull2	soldier2
jump2	aerobics	scolding2	gun
happiness	highjump2	gunman3	kalashnikov
surfer2	snowboard2	knife	tank
skiing	couple7	axe	wildwest
soccer2	gymnastics3	accident2	sniper
father2	highjump3	roach2	rifle
propose3	dog3	vomit	skull
skateboard2	poledance	falling6	snake



**Fig. 15** Example for color-manipulated silhouettes (from left to right): original, red-gray, green-gray, black-gray, and white-gray (as used in the example experiment)

**Table 5** Means and SDs for reaction times (RTs) and error rates (ERs) per color and valence categories in the Bicolor Affective Silhouettes and Shapes (BASS) validation experiment

Color	Valence	RT $M \pm SD$	ER $M \pm SD$
Red	Negative	659.96 $\pm$ 200.61	0.06 $\pm$ 0.24
Red	Positive	708.00 $\pm$ 233.20	0.08 $\pm$ 0.27
Green	Negative	680.91 $\pm$ 207.14	0.10 $\pm$ 0.30
Green	Positive	679.24 $\pm$ 217.30	0.04 $\pm$ 0.20
Black	Negative	647.60 $\pm$ 202.37	0.06 $\pm$ 0.25
Black	Positive	686.14 $\pm$ 217.78	0.07 $\pm$ 0.25
White	Negative	660.16 $\pm$ 202.40	0.08 $\pm$ 0.27
White	Positive	674.30 $\pm$ 210.28	0.06 $\pm$ 0.24

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2020). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-020-01501-5>
- Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, 24(4), 621–635. <https://doi.org/10.1017/S0142716403000316>
- Baker, A., Mystkowski, J., Culver, N., Yi, R., Mortazavi, A., & Craske, M. G. (2010). Does habituation matter? Emotional processing theory and exposure therapy for acrophobia. *Behaviour Research and Therapy*, 48(11), 1139–1143. <https://doi.org/10.1016/j.brat.2010.07.009>
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7), 280–289. <https://doi.org/10.1016/j.tics.2007.05.005>
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147. <https://doi.org/10.1037/0033-295X.94.2.115>
- Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, 24(6), 981–990. <https://doi.org/10.1177/0956797612465439>
- Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001). Emotion and motivation II: Sex differences in picture processing. *Emotion*, 1(3), 300–319. <https://doi.org/10.1037/1528-3542.1.3.300>
- Bradley, M. M., & Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In *Handbook of emotion elicitation and assessment* (pp. 29–46). Oxford University Press.
- Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, 43(2), 468–477. <https://doi.org/10.3758/s13428-011-0064-1>
- Deák, A., Csenki, L., & Révész, G. (2010). Hungarian ratings for the International Affective Picture System (IAPS): A cross-cultural comparison. *Empirical Text and Culture Research*, 4(8), 90–101.
- Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE*, 10(4), e0121945. <https://doi.org/10.1371/journal.pone.0121945>
- Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, 71(4), 808–816. <https://doi.org/10.1080/17470218.2017.1310261>
- Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion*, 15(2), 115–141. <https://doi.org/10.1080/0269993004200024>
- Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin*, 99(1), 20–35. <https://doi.org/10.1037/0033-2909.99.1.20>
- Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 13(1), 104–115. <https://doi.org/10.1037/0096-1523.13.1.104>
- Garaizar, P., & Reips, U.-D. (2019). Best practices: Two web-browser-based methods for stimulus presentation in behavioral experiments with high-resolution timing requirements. *Behavior Research Methods*, 51(3), 1441–1453. <https://doi.org/10.3758/s13428-018-1126-4>
- Geusebroek, J.-M., Burghouts, G. J., & Smeulders, A. W. M. (2005). The Amsterdam Library of Object Images. *International Journal of Computer Vision*, 61(1), 103–112. <https://doi.org/10.1023/B:VISI.0000042993.50813.60>
- Gong, X., & Wang, D. (2016). Applicability of the International Affective Picture System in Chinese older adults: A validation study: Cross-cultural validity of the IAPS. *PsyCh Journal*, 5(2), 117–124. <https://doi.org/10.1002/pchj.131>



- Grühn, D., & Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): Do ratings become more extreme with age? *Behavior Research Methods*, 40(2), 512–521. <https://doi.org/10.3758/BRM.40.2.512>
- Hansen, T., & Gegenfurtner, K. R. (2009). Independence of color and luminance edges in natural scenes. *Visual Neuroscience*, 26(1), 35–49. <https://doi.org/10.1017/S0952523808080796>
- Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition & Emotion*, 15(2), 143–165. <https://doi.org/10.1080/02699930125768>
- Ho, S. M. Y., Mak, C. W. Y., Yeung, D., Duan, W., Tang, S., Yeung, J. C., & Ching, R. (2015). Emotional valence, arousal, and threat ratings of 160 Chinese words among adolescents. *PLOS ONE*, 10(7), e0132294. <https://doi.org/10.1371/journal.pone.0132294>
- Huang, J., Xu, D., Peterson, B. S., Hu, J., Cao, L., Wei, N., Zhang, Y., Xu, W., Xu, Y., & Hu, S. (2015). Affective reactions differ between Chinese and American healthy young adults: A cross-cultural study using the international affective picture system. *BMC Psychiatry*, 15(1), 60. <https://doi.org/10.1186/s12888-015-0442-9>
- Huston, J. P., Nadal, M., Mora, F., Agnati, L. F., & Conde, C. J. C. (2015). *Art, aesthetics, and the brain*. OUP Oxford.
- Kawai, C., Lukács, G., & Ansorge, U. (2020). Polarities influence implicit associations between colour and emotion. *Acta Psychologica*, 209, 103143. <https://doi.org/10.1016/j.actpsy.2020.103143>
- Kuhbandner, C., & Pekrun, R. (2013). Joint effects of emotion and color on memory. *Emotion*, 13(3), 375–379. <https://doi.org/10.1037/a0031821>
- Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, 49(2), 457–470. <https://doi.org/10.3758/s13428-016-0715-3>
- Lakens, D., Fockenberg, D. A., Lemmens, K. P. H., Ham, J., & Midden, C. J. H. (2013). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion*, 27(7), 1225–1246. <https://doi.org/10.1080/02699931.2013.781501>
- Lakens, D., Semin, G. R., & Foroni, F. (2012). But for the bad, there would not be good: Grounding valence in brightness through shared relational structures. *Journal of Experimental Psychology: General*, 141(3), 584–594. <https://doi.org/10.1037/a0026468>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 1, 39–58.
- Lau, R. R., Sears, D. O., & Jessor, T. (1990). Fact or artifact revisited: Survey instrument effects and pocketbook politics. *Political Behavior*, 12(3), 217–242. <https://doi.org/10.1007/BF00992334>
- Leder, H., Tinio, P. P. L., & Bar, M. (2011). Emotional valence modulates the preference for curved objects. *Perception*, 40(6), 649–655. <https://doi.org/10.1068/p6845>
- Lindsay, G. W. (2020). Attention in psychology, neuroscience, and machine learning. *Frontiers in Computational Neuroscience*, 14, 29. <https://doi.org/10.3389/fncom.2020.00029>
- Liu, P., Li, M., Lu, Q., & Han, B. (2018). Norms of valence and arousal for 2,076 Chinese 4-character words. In K. Hasida & W. P. Pa (Eds.), *Computational Linguistics* (Vol. 781, pp. 88–98). Springer Singapore. [https://doi.org/10.1007/978-981-10-8438-6\\_8](https://doi.org/10.1007/978-981-10-8438-6_8)
- Marchewka, A., Żurawski, L., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, 46(2), 596–610. <https://doi.org/10.3758/s13428-013-0379-1>
- Marr, D. (1982). *Vision. A computational investigation into the human representation and processing of visual information*. W. H. Freeman and Company.
- Meier, B. P., Fetterman, A. K., & Robinson, M. D. (2015). Black and white as valence cues: A large-scale replication effort of Meier, Robinson, and Clore (2004). *Social Psychology*, 46(3), 174–178. <https://doi.org/10.1027/1864-9335/a000236>
- Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science*, 15(4), 243–247. <https://doi.org/10.1111/j.0956-7976.2004.00659.x>
- Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white: Automatic inferences about stimulus valence based on brightness. *Psychological Science*, 15(2), 82–87. <https://doi.org/10.1111/j.0963-7214.2004.01502002.x>
- Moller, A. C., Elliot, A. J., & Maier, M. A. (2009). Basic hue-meaning associations. *Emotion*, 9(6), 898–902. <https://doi.org/10.1037/a0017811>
- Mould, D., Mandryk, R. L., & Li, H. (2012). Emotional response and visual attention to non-photorealistic images. *Computers & Graphics*, 36(6), 658–672. <https://doi.org/10.1016/j.cag.2012.03.039>
- Nishimoto, T., Ueda, T., Miyawaki, K., Une, Y., & Takahashi, M. (2012). The role of imagery-related properties in picture naming: A newly standardized set of 360 pictures for Japanese. *Behavior Research Methods*, 44(4), 934–945. <https://doi.org/10.3758/s13428-011-0176-7>
- Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41, 176–210. <https://doi.org/10.1006/cogp.1999.0728>
- Orzan, A., Bousseau, A., Barla, P., & Thollot, J. (2007). Structure-preserving manipulation of photographs. *Proceedings of the 5th International Symposium on Non-Photorealistic Animation and Rendering - NPAR '07*, 103. <https://doi.org/10.1145/1274871.1274888>
- Pôrto, W. G., Bertolucci, P. H. F., & Bueno, O. F. A. (2010). The paradox of age: An analysis of responses by aging Brazilians to International Affective Picture System (IAPS). *Revista Brasileira de Psiquiatria*, 33(1), 10–15. <https://doi.org/10.1590/S1516-44462010005000015>
- Proverbio, A. M., Adorni, R., Zani, A., & Trestianu, L. (2009). Sex differences in the brain response to affective scenes with or without humans. *Neuropsychologia*, 47(12), 2374–2388. <https://doi.org/10.1016/j.neuropsychologia.2008.10.030>
- Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., & DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *The Journal of Neuroscience*, 38(33), 7255–7269. <https://doi.org/10.1523/JNEUROSCI.0388-18.2018>
- Ramaswami, M. (2014). Network plasticity in adaptive filtering and behavioral habituation. *Neuron*, 82(6), 1216–1229. <https://doi.org/10.1016/j.neuron.2014.04.035>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, 33(2), 217–236. <https://doi.org/10.1068/p5117>
- Sabatinielli, D., Fleisch, T., Bradley, M. M., Fitzsimmons, J. R., & Lang, P. J. (2004). Affective picture perception: Gender differences in visual cortex? *NeuroReport*, 15(7), 1109–1112. <https://doi.org/10.1097/00001756-200405190-00005>
- Schmalz, X., Beyersmann, E., Cavalli, E., & Marinus, E. (2016). Unpredictability and complexity of print-to-speech correspondences increase reliance on lexical processes: More evidence for the orthographic depth hypothesis. *Journal of Cognitive Psychology*, 28(6), 658–672. <https://doi.org/10.1080/20445911.2016.1182172>

- Schupp, H. T., Markus, J., Weike, A. I., & Hamm, A. O. (2003). Emotional facilitation of sensory processing in the visual cortex. *Psychological Science, 14*(1), 7–13. <https://doi.org/10.1111/1467-9280.01411>
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174–215.
- The GIMP Development Team. (2019). GIMP. Retrieved from <https://www.gimp.org>
- Verschuere, B., Crombez, G., & Koster, E. (2001). The International Affective Picture System: A Flemish Validation Study. *Psychologica Belgica, 41*(4), 205–217. <https://doi.org/10.5334/pb.981>
- Tjosvold, D., & Sun, H. F. (2002). Understanding conflict avoidance: Relationship, motivations, actions, and consequences. *International Journal of Conflict Management, 13*(2), 142–164. <https://doi.org/10.1108/eb022872>
- Vö, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods, 41*(2), 534–538. <https://doi.org/10.3758/BRM.41.2.534>
- Wilcox, N., & Wlezien, C. (1993). The contamination of responses to survey items: Economic perceptions and political judgments. *Political Analysis, 5*, 181–213. <https://doi.org/10.1093/pan/5.1.181>
- Wilms, L., & Oberfeld, D. (2018). Color and emotion: Effects of hue, saturation, and brightness. *Psychological Research, 82*(5), 896–914. <https://doi.org/10.1007/s00426-017-0880-8>
- Wrase, J., Klein, S., Gruesser, S. M., Hermann, D., Flor, H., Mann, K., Braus, D. F., & Heinz, A. (2003). Gender differences in the processing of standardized emotional visual stimuli in humans: A functional magnetic resonance imaging study. *Neuroscience Letters, 348*(1), 41–45. [https://doi.org/10.1016/S0304-3940\(03\)00565-2](https://doi.org/10.1016/S0304-3940(03)00565-2)
- Yao, Z., Wu, J., Zhang, Y., & Wang, Z. (2017). Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behavior Research Methods, 49*(4), 1374–1385. <https://doi.org/10.3758/s13428-016-0793-2>
- Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111*(4), 493–504. <https://doi.org/10.1037/pspa0000056>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.