



Article scientifique

Article

2011

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance

Dan Glauser, Elise; Scherer, Klaus R.

How to cite

DAN GLAUSER, Elise, SCHERER, Klaus R. The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. In: Behavior Research Methods, 2011, vol. 43, n° 2, p. 468–477. doi: 10.3758/s13428-011-0064-1

This publication URL: <https://archive-ouverte.unige.ch/unige:96403>

Publication DOI: [10.3758/s13428-011-0064-1](https://doi.org/10.3758/s13428-011-0064-1)

The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance

Elise S. Dan-Glauser · Klaus R. Scherer

Published online: 24 March 2011
© Psychonomic Society, Inc. 2011

Abstract In emotional research, efficient designs often rely on successful emotion induction. For visual stimulation, the only reliable database available so far is the International Affective Picture System (IAPS). However, extensive use of these stimuli lowers the impact of the images by increasing the knowledge that participants have of them. Moreover, the limited number of pictures for specific themes in the IAPS database is a concern for studies centered on a specific emotion thematic and for designs requiring a lot of trials from the same kind (e.g., EEG recordings). Thus, in the present article, we present a new database of 730 pictures, the Geneva Affective PicturE Database, which was created to increase the availability of visual emotion stimuli. Four specific negative contents were chosen: spiders, snakes, and scenes that induce emotions related to the violation of moral and legal norms (human rights violation or animal mistreatment). Positive and neutral pictures were also included: Positive pictures represent mainly human and animal babies as well as nature sceneries, whereas neutral pictures mainly depict inanimate objects. The pictures were rated according to valence, arousal, and the congruence of the represented scene with internal (moral) and external (legal) norms. The constitution of the database and the results of the picture ratings are presented.

Keywords Emotion induction · Picture database · Snakes · Spiders · Normative significance

E. S. Dan-Glauser · K. R. Scherer
University of Geneva,
Geneva, Switzerland

E. S. Dan-Glauser (✉)
Stanford Psychophysiology Lab, Department of Psychology,
Jordan Hall, 450 Serra Mall,
Stanford, CA 94305–2130, USA
e-mail: elise.danglauser@gmail.com

Many studies in affective science aim to evaluate and understand the dynamics and underlying mechanisms of emotional processes. Numerous frameworks have theorized that emotion episodes start with a cognitive evaluation of a particular situation (Ellsworth & Scherer, 2003; Scherer, 1999, 2001). To study this process, it is essential to be able to efficiently modulate the environment to help understand the vast spectrum of emotional reactions to affectively charged situations. In the majority of the experimental designs used in this field, reaction assessment is performed soon after a confrontation with emotionally salient stimuli. This framework is applied to all kinds of reaction measures, such as judgment tasks, behavioral reactions (e.g., timing and accuracy), facial expressions, as well as physiological reactions that are both centrally and peripherally driven. This vast array of possible measurements, if correctly used, is able to give the researcher a comprehensive view of emotional processes, from a few milliseconds after their onset to the later consequences of the emotional episode. However, despite how accurate, precise, and reliable these measures are, they become scientifically useless if the emotion induction is unsuccessful.

Therefore, emotional induction settings have to be chosen carefully. This choice is driven by the type of measurement required (e.g., an emotion induction using real social situations could be difficult to study with central nervous system measurements), but also by the perceptual modality that seems most appropriate to the design and research questions. Several possibilities have been explored so far to induce emotional reactions, relying on different contexts and various degrees of participant involvement. The most widely used method of emotion induction is the presentation of emotionally salient material, without explicitly asking for a personal contribution from the participant. If stimuli are relevant enough, an appraisal is

automatically executed and will trigger reactions in other measurable components of emotion such as physiological responses, expressivity, action tendencies, and subjective feeling (see Scherer, 1987, 2004, for a comprehensive view of the component process model). Although this kind of induction can target different perceptual modalities, the use of the visual channel remains the most common to convey emotional stimulation. Stimulation of the visual path relies on the presentation of emotional films (dynamic stimulation, see Fredrickson & Levenson, 1998; and Gross & Levenson, 1993, for an illustration of such studies) or pictures (static stimulation, see, e.g., Codispoti, Bradley, & Lang 2001; Lang, Greenwald, Bradley, & Hamm 1993; Stark et al., 2004). The use of visual stimulation is prevalent in research due to its ease and its relation to well-described visual path physiology and processes. Furthermore, temporal reactions are precisely timed with this kind of stimulation, since measurements can be accurately locked to the onset of the stimulations.

If static visual stimulation is to be chosen for a particular study, the type of pictures that are necessary for the setting and the intrinsic value of these stimuli must be defined. Both parameters have to correspond to controlled or manipulated aspects of the study. Finding appropriate static visual stimuli for a particular research question is not an easy task. Two possibilities exist for the researcher: Either all stimuli have to be gathered for each particular study, or databases from other labs/studies can be used (if available). In any case, the collected pictures need to have been tested on several parameters prior to the study in order to provide a well-characterized sample of stimuli that can be reliably used as controlled inducing material.

This is the case of the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert 1999). The IAPS has been created to meet two major goals: first, to permit better control in selecting stimuli, and second, to allow comparison and replication of experiments performed with the same material. The IAPS allows for the selection of pictures according to standardized ratings of valence, arousal, and control, based on the theoretical background of a three-dimensional affective space construct (Bradley & Lang, 1994; Mehrabian & Russell, 1974; Osgood, 1969; Russell & Mehrabian, 1977). Cross-validation studies have been conducted and show reliable induction of expressive and physiological emotion responses with these stimuli (Greenwald, Cook, & Lang 1989; Lang et al., 1993). Since its validation, hundreds of studies have relied on this system, which proves the crucial need for such a database of stimuli in our field.

However, we have noticed two major issues while using the IAPS. The first issue arises after intensive use of any kind of stimuli and occurs when several studies with the IAPS material are performed in one lab within the same

period of time (or in several labs using the same pool of participants). If the same participants are exposed to the exact same material, the material loses part of its emotional power. For each group of participants, it is thus necessary to present a particular set of pictures only once if one wants to obtain reaction to “new” emotional stimuli without having to deal with habituation-related reactions. This concerns most departments in which studies are performed on psychology students, who often participate in many different studies during the same term.

The second issue deals with the large range of themes pictured in the IAPS. If the content of the picture is of particular importance to the research question (e.g., if the research deals with emotional reactions to specific situations or objects, or even with specific emotion categories), stimulus selection will have to focus on very specific picture content, which may not occur with high enough frequency in a broad-topic database such as the IAPS. For example, constructing a study evaluating phobic reactions to spider or snake pictures may require numerous different pictures of snakes and spiders. For this particular case, the IAPS database contains 17 snake pictures and only six spider pictures, which is insufficient to plan a design relying on spider and snake stimuli. This is particularly true for designs requiring presentation of multiple trials per condition in order to obtain detectable results. As an example, EEG studies require specially adapted designs, often relying on multiple presentations of stimuli of the same type, to be able to highlight differences in the event-related potentials (ERPs). Depending on the ERP component that is analyzed, one will need several dozens of trials per condition. As an example, the guidelines by Luck (2005) recommend between 20 and several hundreds of trials. Gathering the material for such analyses is a time- and energy-consuming task. Researchers would greatly benefit from a new pool of static stimulations, complementary to the IAPS, facilitating the work on specific and unique thematics that are of frequent use in psychology and neuroscience research.

The present work

The widespread and constant need for emotional pictures in the field of affective science research, and the necessity of stimuli dealing with specific topics for our studies, led to the constitution, rating, and distribution of a new picture database, the Geneva Affective PicturE Database (GAPED), which we present in the present article. According to appraisal theorists, emotions arise after evaluation of the environment according to several criteria (Frijda, 1986; Roseman, 1984; Scherer, 1984; Smith & Ellsworth, 1985). Valence is one of the most basic features that is evaluated,

supposedly early in the process, and having the important consequence of driving the behavior toward approach or avoidance. In addition, attempts to categorize emotions have led to numerous descriptions of an affective space, mainly composed of the two dimensions of valence and arousal (Lang et al., 1999; Russell, 1980; Watson & Tellegen, 1985). Valence is undoubtedly an important dimension of this categorization (Barrett, 2006; Watson, Wiese, Vaidya, & Tellegen 1999). Valence is strongly linked with the important appraisal of pleasantness (Smith & Ellsworth, 1985), and has also a nearly one-to-one relationship with intrinsic pleasantness (Scherer, Dan, & Flykt 2006), both appraisal criteria evaluating the amount of pleasantness conveyed by a particular stimulus.

Consequently, in order to provide pictures for situations that occur very frequently and that match interest of a wide variety of theoretical backgrounds, valence has been chosen as the primary dimension to select several categories of pictures. Our goal was to construct a database with many images of a given topic. Therefore, four negative categories have been chosen. To follow a similar balanced structure as in previous research databases, such as the IAPS, and to create stimuli that have the power to induce non-negative emotions, neutral and positive picture categories were also included and rated.

Pictures of spiders and snakes were first selected, which represent two negative categories of our database. A review of online scientific libraries showed that more than 300 studies so far have dealt with snake and/or spider phobia, using measurement of reactions to various kinds of spider and/or snake stimuli. A picture database gathering numerous images of these animals could be useful for subsequent research on this topic. This kind of research often relies on the assumption that such stimuli are of high relevance because of their biological and evolutionary threat-related content, and that they therefore induce strong aversive reactions (see, e.g., Öhman, 1993; and Öhman & Mineka, 2001).

To choose the other two negative categories, we turned to the appraisal criterion of normative significance. One can assume that strong emotion induction can occur from an evaluation of changing situations according to a social standpoint (Manstead & Fischer, 2001). Thus, our interest was also to gather pictures that could induce emotion according to social relevance, and, more particularly, norm significance. Checking the acceptability of the stimulus with respect to social standards is part of the appraisal process, with the rationale that for the socially organized species, the behavior of a single individual must comply with the norms of the entire group. For long-lasting social organizations and stable communities, behaviors violating one of these rules are regarded as important by the other members, and thus have the strong

potential to elicit an emotional response (Scherer, 1987). Therefore, we selected pictures corresponding to low compatibility with social norms (external, mainly determined by legality) and personal norms (internal, mainly driven by morality). Evaluation of low acceptability of the stimuli with respect to social norms is particularly relevant in the elicitation of anger (Sander, Grandjean, & Scherer 2005; Scherer, 1999), but could also be associated with sadness, disgust, pity, guilt, shame, and contempt. The appraisal criterion of normative significance has been too often neglected in the emotion induction procedure, and we believe that new research would benefit from investigating the appraisal of such criterion. Not only could a vast array of emotions arise from such stimulations, but they could also have the inbuilt capacity of becoming powerful inductors. On this topic, we were particularly interested in the link between militant attitudes and emotional reactions to current social issues. Pictures focusing on human rights violation and animal mistreatment were therefore selected.

Non-negative categories were separated as neutral or positive. People were generally not represented in images of the neutral category, due to the difficulty of finding scenes with neutral facial expressions. Thus, neutral images were mainly pictures representing objects, as well as buildings and furniture. For selection of positive pictures, a few main topics were chosen, again to make available a set of same-kind stimulations. Because of the more frequent use of negative stimuli than of positive stimuli in psychology and neuroscience research so far, separate positive categories have not been created; thus, fewer same-kind pictures are available in the GAPED for positive than for negative categories. Recently, Brosch, Sander, and Scherer (2007) showed that the judgment of pictures representing babies of several species (human as well as animals) triggers more positive valence ratings as well as higher arousal than does the judgment of pictures of adult individuals. This is congruent with the *Kindchenschema* (baby schema), described by Lorenz (1943), which refers to the increased attention and more positive reactions people have when confronted with specific features of young individuals across species. The positive category of the GAPED is thus composed of pictures representing human babies, and to a lesser extent, young animals (about 60 images in total). The other thematic significantly present in the positive category is nature and landscape (about 50 pictures).

After the selection, the pictures were judged on the basis of the focused criteria used to select the database. Ratings were thus obtained on valence, as well as on the degree of acceptability of the picture according to social norms. This latter judgment evaluates the degree to which the pictures were considered as being immoral and

illegal,¹ that is in violation with the rater's normative standpoint. This judgment was specifically relevant to the characterization of the stimuli referring to human rights violation and animal mistreatment. Given the importance of arousal level in the organization of affective space (Lang et al., 1993), as well as in the characterization of brain circuitry (see, e.g., Anders, Lotze, Erb, Grodd, & Birbaumer, 2004), a rating of arousal for the GAPED pictures was also obtained. This judgment study is described in the present article.

We expected to see clear differences in valence scores among the positive, neutral, and negative categories (snakes, spiders, human concerns, and animal mistreatments). For arousal ratings, we also expected to see differences between categories, since valence ratings are rarely independent from arousal levels (Isen, 1998; Scherer et al., 2006). Thus, negative pictures (snakes, spiders, human concerns, and animal mistreatments) were expected to have greater arousal levels than neutral and positive pictures. Ratings of acceptability with respect to internal (moral) and external (legal) norms concerned only human concern and animal mistreatment categories. We expected low ratings on these two scales for these two categories, indicating that the pictures were judged to be both immoral and illegal, according to the raters.

Method

Pictures were gathered through an extensive online Web search on the various composing themes. All pictures were rated after having been centered and resized.

Stimuli characteristics

Six categories were included: snakes, spiders, human concerns (depicting scenes violating human rights), animal mistreatments (picturing animal mistreatment scenes), neutral, and positive pictures. Each emotional category contained over 100 pictures. The exact number for each category can be found Table 1.

All pictures were resized and cropped to a 640 x 480 pixel size. Texts and comments were removed to leave only the pictorial aspects, centered on the main subject of the represented scene.

Table 1 Number of pictures selected for each category

Category	Number of pictures	
	Before removal of outliers	After removal of outliers
Negative		
Snakes	133	133
Spiders	160	158
Human concerns	108	105
Animal mistreatments	129	124
Neutral	94	89
Positive	130	121

Outlier definition was (a) Negative, score (outlier) > mean (all negative pictures) + 2 *SD*; (b) Neutral, score (outlier) < mean (all neutral pictures) – 2 *SD* or score (outlier) > mean (all neutral pictures) + 2 *SD*; (c) Positive, score (outlier) < mean (all positive pictures) – 2 *SD*

Participants

Sixty participants were recruited from a second-year psychology class and were rewarded with course credit for their participation. Participants had a mean age of 24 years (ranging from 19 to 43 years; *SD* = 5.9), and 52 of them were right-handed. The majority of the participants spoke French as their native language (51), and the rest spoke Spanish and Italian (*n* = 2 for each language), and German, Hungarian, Ruman, Turkish, and Arabic (*n* = 1 for each language). These last nine participants were nevertheless perfectly fluent in French, which was the language in which the study was conducted.

Stimulus presentation and rating scales

The database was initially composed of 754 pictures. Given the time necessary to rate all of the pictures and the low level of interest for such a task, raters were divided in five groups, each rating a subset of the database. To be able to compare groups, 39 pictures were rated by all participants. This small set was composed of six to eight pictures from each category, proportional to the initial number of pictures in each category in the database. The other pictures from each category were distributed in each of the rating groups. In total, each person rated 182 images. This experiment lasted about 1 hr, depending on the participant's rating speed. Picture presentation order was semirandomized, with the condition that no currently viewed picture belonged to the same category as the previously rated one. Each picture was presented in full screen during 4 sec. After the presentation, the rating scales were presented on a new

¹ Since the study was conducted in Switzerland, legal judgments were implicitly related to the Swiss legislation (equivalent to most Western countries regarding child abuse, food deprivation for human being, and animal mistreatments). However, the perception of the gravity of such mistreatment depends mostly on the individual cultural background, no matter in which country the research is conducted.

screen, and the participants could begin the rating. A small-sized exemplar of the picture being rated was still visible in the upper right corner of the screen to remind participants of the current picture.

Five continuous rating scales were presented, each ranging from 0 to 100 points. Participants indicated their ratings by placing a cursor along a gradually colored line. The first scale went from *negative* to *positive* and completed the sentence, “You are judging this image as ...”. This scale gave the rating of valence for the pictures (from 0 = *very negative pictures* to 100 = *very positive pictures*, with 50 = *neutral*). The second and third scales were judging the degree of arousal elicited by the picture with the introductory sentence, “Confronted with this image you are feeling: ...” with two interrelated scales: one going from *Calm* to *Excited*, and the other from *Stimulated* to *Relaxed*. These arousal scales were chosen according to the results of arousal ratings found in a previous study (Scherer et al., 2006). As compared with the ratings performed for the IAPS, only valence and arousal were reproduced in the present study; the level of dominance was not rated. In fact, dominance level doesn’t seem to be considered in most studies selecting stimuli according to the circumplex model of affect (see Russell, 1980). The fourth and fifth questions of this rating study dealt with the picture’s acceptability with respect to internal and external norms. The participants had to rate the compatibility with internal norms by judging whether the contents of the picture were morally and ethically acceptable, and its compatibility with external norms by judging whether the contents of the picture were legally acceptable. Both scales ranged from *Absolutely not acceptable* to *Totally acceptable* and completed the introductory sentences: “From a moral/ethical point of view, how did you find the scene or actions represented?” and “From a legal point of view, how did you find the scene or actions represented?” For these last two questions, participants were also offered the choice of answering by checking a box labeled “irrelevant” if the answer to the question made no sense (e.g., for snake and spider pictures). The participants could take as much time as they needed to complete the rating. Once ratings were completed for one image, a button on the screen permitted the participant to view the next picture.

Procedure

Participants came to the lab in groups of eight to 20. Once comfortably seated in front of a computer screen, they were prompted to start by the experimenter. They were also told that they could choose to stop the experiment at any time. Each participant was personally informed that he or she was going to see many pictures of spiders and snakes, among

other pictures, and that any discomfort should be reported immediately to stop the experiment as soon as possible.² Task instructions were displayed on the screen. The first part of the instructions explained that we were interested in the way visual scenes could affect someone. Participants were asked to look carefully at each picture presented. After explaining that they had to think about their reaction to the pictures and the status of the scene in our society, the rating scales were presented in detail. Finally, the mode of response, which entailed using sliders to be moved along a gradually colored bar (from white to dark blue), was explained. At the end of the experiment, the participants were thanked and received the corresponding course credit.

Results and discussion

This section will first focus on the results of the common picture sample (39 pictures, see the [Method](#) section), rated by all participants. A second part will describe the rating calculations and the outlier removal, a third part will concentrate on the description of the rating mean and range for each category of the database, and, finally, several aspects of the picture categories will be portrayed.

Rater group comparison

For all variables (valence, arousal, internal norms compatibility [moral acceptability], and external norms compatibility [legal acceptability]), judgments of each rater group for the common pictures were strongly correlated. No correlation coefficient was inferior to $r = .835$, $p < .0001$, and the average correlation was found to be $r = .922$, $p < .0001$. Thus, the five rating groups were considered to be from the same sample, and the ratings for the 754 pictures were considered as one database with comparable ratings.

Rating calculations and outlier removal

For each picture, average ratings were obtained for all relevant questions. All ratings ranged from 0 to 100. Valence ratings corresponded to the positivity of the picture. The two arousal ratings were compared for all pictures. The two scales, *Calm–Excited* and *Stimulated–Relaxed* were, as expected, strongly and negatively correlated ($r = -.96$, $p < .01$). The scores for the *Stimulated–Relaxed* question were reversed, and the two arousal scales were averaged to obtain a continuous arousal description

² No participant decided to stop the experiment, and data from the 60 participants were included in the results

going from “insignificantly” to “highly arousing.” Scores of acceptability with respect to internal and external norms were calculated for all pictures belonging to the human concern and the animal mistreatment categories. A score of 0 meant that the participants found a picture to be either immoral (internal norms) or illegal (external norms), whereas a score of 100 points meant that the scene depicted in the picture was totally acceptable either from a moral (internal) or a legal (external) standpoint..

In order to obtain clearly distinguishable categories and to eliminate outliers that may mislead the user about the real emotion induction power of the picture, the valence rating has been used to remove several pictures from the database. The mean plus or minus two standard deviations was used as a criterion. Outliers in negative categories were those for which the valence rating was superior to the mean plus two standard deviations. Outliers in the positive category were those for which valence ratings were below the mean value minus two standard deviations. Outliers in the neutral category were targeted above and below the mean plus and minus two standard deviations. In total, 24 pictures were removed, leaving 730 pictures in the database. The exact number left for each category can be found Table 1.

Rating description

For each category, a summary of the ratings (mean, *SD*, median, and range) is given Table 2.

First, valence and arousal scales were analyzed. Figure 1 shows the distribution of the ratings in the valence and arousal space. For each category, a polygon delimits the space in which all pictures of this same category are found. Then, the two ratings of internal (moral) and external (legal) norms (normative significance) were examined.

Valence

Valence ratings were examined for each category. Mean valence ratings show that pictures preliminarily defined as positive were indeed rated positively (all pictures above 71 points), whereas all of the pictures preliminarily defined as negative were rated as having a low valence (below 50 points in mean). However, neutral pictures were slightly above the scale midpoint. This may be due to the relative comparison with many negative pictures. Results show that there is overlap neither between the ratings of the positive category and those of the neutral category, nor between the ratings of the positive and those of the negative categories. However, a high percentage of negative picture valence ratings overlaps with neutral picture ratings (10% for animal mistreatment pictures, 25% for spider pictures, 29% for human concern pictures, and nearly 53% for snake pictures). The high overlap between snake and neutral pictures is intriguing. We argue that this is due to the implicit relative hierarchy that raters make among the pictures to be rated. With snakes being considered as less negative than the other thematics, their valence gradually shifts toward the positive side. Despite this unexpected outcome, one is still able to find in the database more than 174 snake and neutral pictures (104 and 70, respectively) that are not overlapping with each other, which should be sufficient for most designs needing snake stimuli and that are focused on valence variations.

Arousal

Regarding arousal ratings, neutral and positive pictures elicited relatively low arousal (mean rating below 22 points for positive pictures and below 25 for neutral), whereas the

Table 2 Summary of the ratings (mean, *SD*, median, and range) for each picture category

Category	<i>M(SD)</i>	Median	Range	<i>M(SD)</i>	Median	Range
			Valence			
Negative						Arousal
Snakes	41.52(11.17)	41.54	16.66-63.66	53.63(10.72)	54.09	30.56-72.95
Spiders	35.12(11.37)	35.60	9.52-57.01	58.16(10.32)	57.75	37.38-78.43
Human concerns	27.95(17.54)	24.70	0.72-61.41	58.71(15.02)	57.28	32.31-92.40
Animal mistreatments	21.26(12.40)	17.90	0.41-49.51	60.63(12.01)	61.95	28.41-89.04
Neutral	55.78(6.08)	55.04	40.95-68.85	24.93(7.75)	23.79	10.20-43.79
Positive	89.65(6.20)	91.12	71.86-98.74	21.60(10.72)	19.62	5.85-66.01
			Internal (moral) norms			
Negative						External (legal) norms
Human concerns	29.84(17.76)	27.21	0.58-77.09	35.41(17.44)	34.17	1.02-76.63
Animal mistreatments	26.98(13.94)	23.80	0.36-75.41	36.38(15.57)	33.39	0.45-78.69

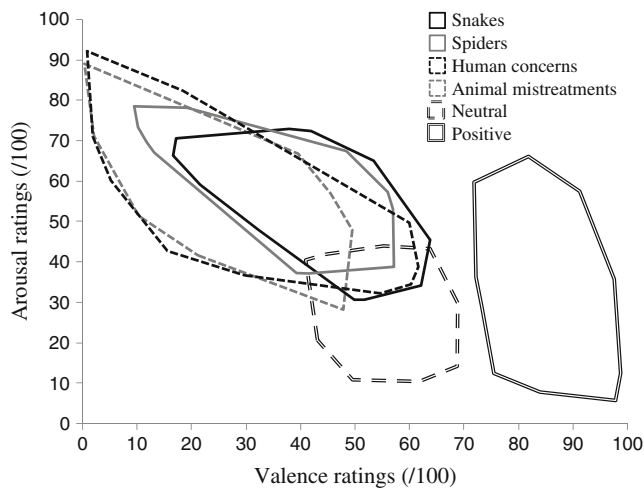


Fig. 1 Representation of the outcome ratings in the valence/arousal space for each category. Polygons represent the surfaces occupied by all the images in a given category

other categories had mildly arousing levels, with mean ratings ranging between 53 and 61 points.

Normative significance

Regarding the ratings of the acceptability of the pictures with respect to internal (moral) and external (legal) norms, means over the two scales and for the two categories ranged between 27 and 36 points, which was congruent with our expectations. Some of the ratings for the human concern and animal mistreatment pictures spread across the scale to reach values up to 75, which clearly show that some pictures do not convey the expected content. Eleven percent of the pictures in these two categories had an internal (moral) norm rating greater than 50 points, whereas 21% of the pictures were rated as being legally acceptable (ratings greater than 50 points). These particular pictures could be used as control images because they have the same kind of content as lower rated pictures, but they are potentially able to trigger different emotions. In any case, the database can be down-sampled to match the prerequisites of future research.

Category comparison

One-way ANOVAs were performed with each rating separately, comparing the level of rating for each picture category. A Type III calculation was used for obtaining the sums of squares.³ Comparative analyses of the categories show strong differences. The mean ratings for each category and significant contrasts are depicted in Fig. 2.

³ Type II sums of squares were also calculated to account for eventual biases due to unbalanced numbers of pictures in each category (Langsrud, 2003). Since the same results were obtained, the Type III calculation has been preferred to report category comparison results.

Valence

Concerning valence, we expected positive pictures to be highly rated and significantly different from neutral pictures. Because of their negative connotations, negative categories were expected to significantly differ in valence ratings from positive and neutral categories. Both expectations were met: A one-way ANOVA showed a significant main effect of the picture category on valence ratings, $F(5, 724) = 558.74$, $p < .01$, $\eta_p^2 = .79$. All picture categories were significantly differentiated by Tukey HSD post hoc tests (see Fig. 2, top left panel). Thus, in addition to what had been predicted, valence differences between each of the negative categories (snakes, spiders, human concerns, and animal mistreatments) were detected, although all these pictures were generally negative (mean rating = 32.24).

Arousal

The one-way ANOVA on arousal ratings also showed a significant main effect of the picture category, $F(5, 724) = 289.04$, $p < .01$, $\eta_p^2 = .66$. Tukey HSD post hoc tests showed a difference between a first group composed of spider, human concern, and animal mistreatment pictures (similar to each other, mean rating = 59.10); a second group composed of snake pictures only (mean rating = 53.63); and a third group composed of neutral and positive pictures (also similar to each other, mean rating = 23.01) ($p < .01$) (see Fig. 2, top right panel).

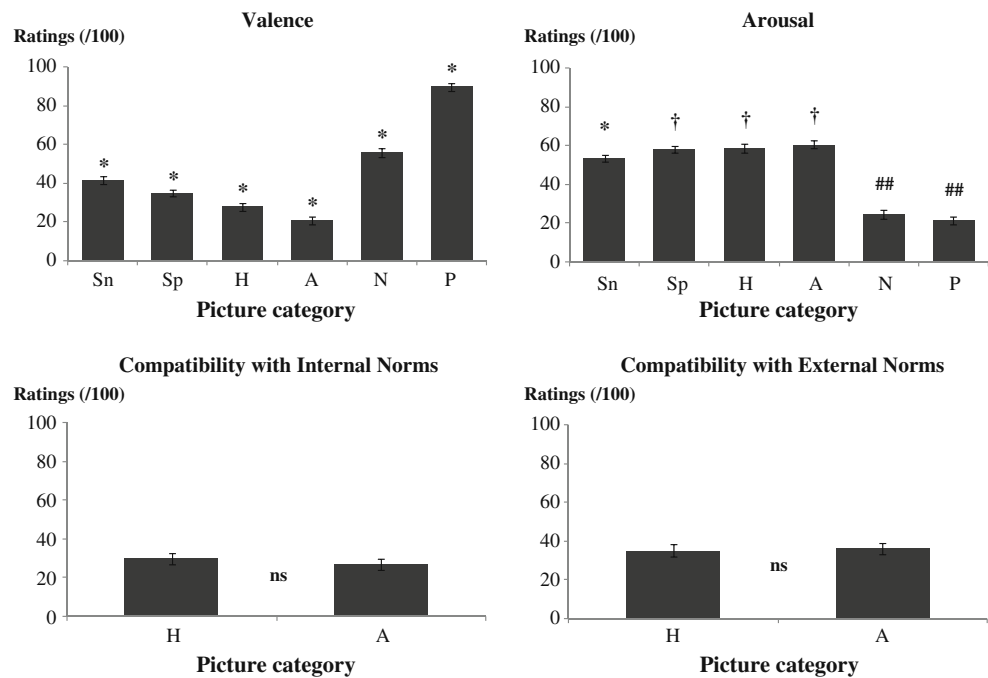
Normative significance

Human concerns and animal mistreatments were not significantly different; neither regarding internal (moral) norms, $F(1, 227) = 1.85$, *ns*, nor regarding external (legal) norms, $F(1, 227) = 0.20$, *ns*, (see Fig. 2, two bottom panels).

Evolution of ratings across time

Results of past research show that repeated or sustained presentation of similar stimuli create a sensitization, resulting in stronger emotional reactions the more (or the longer) negative stimuli are presented (Bradley, Cuthbert, & Lang, 1996; and Smith, Bradley, & Lang, 2005). To examine the effect of time on the GAPED ratings, comparisons of the rating evolution from early in the session to late in the session were performed. For each category, the first and last 33% of the rated images were compared with paired *t* tests. For the great majority of the tests performed (81%), results did not show a significant difference in ratings across time, $t(59)$, from -1.86 to 1.70 , $p = ns$. For these comparisons, ratings do not seem to be influenced by the moment in the session during which

Fig. 2 Representation of mean ratings for each relevant picture category (either six or two, depending on the rating question). Error bars represent the 95% confidence interval for the mean. Sn Snake pictures, Sp Spider pictures, H Human concern pictures, A Animal mistreatment pictures, N Neutral pictures, P Positive pictures. * $p < .01$ with each of the other categories, † $p < .01$ with Sn, N, and P pictures, ## $p < .001$ with each of the negative picture categories (Sn, Sp, H, and A)



pictures of a same category are rated. Globally, this analysis shows that, although many related pictures were evaluated, the sequential nature of the rating process did not influence the resulting GAPED ratings. Three tests were, however, significant. First, valence for animal mistreatment pictures showed a significant decrease in rating, $t(59) = 2.78$, $p < .01$, as ratings shifted from 25 to 20 toward the end of the session. Second, internal norm (moral) ratings also showed a decrease over time for animal mistreatment pictures, $t(59) = 2.77$, $p < .01$ (32 at the beginning to 25 toward the end). These two results are congruent with earlier results of sensitization. A third significant result showed a decrease in arousal ratings for spider pictures, $t(59) = 2.73$, $p < .01$. These images had a mean arousal of 62 at the beginning of the session, and a mean of 57 toward the end. Note that these effects are independent of the particular pictures in each category, the presentation being totally randomized across participants. For this same reason, the individual ratings reported for each particular picture represent a mean between ratings made at different time points during the session, which tempers down the effect of the repetitive presentations of pictures of the same kind. Moreover, since the few significant decreases do not drastically change the specification of images (animal mistreatment pictures are still of low valence and of low internal norm compatibility, and spider pictures are still quite high on arousal), the GAPED category ratings cannot be considered as biased by the duration of the rating process.

Limitations

A few limitations to the GAPED construction and ratings need to be discussed. Positive emotions are often neglected

in the study of emotions. Explanations for this lack of research include a large focus on negative outcomes from the clinical psychology, associated models that are built up only for negative affects, and generally lower methodological challenges for negative than for positive emotion induction (Fredrickson, 2004; Fredrickson & Cohn, 2008). The GAPED has followed this orientation, with attention being put on developing large negative categories and a unique positive category. Consequently, the database is asymmetric, with many more negative than positive pictures. This asymmetry is also present in the contents, which are more specific in the negative pictures. Again, the goal of the database is to have a large pool of images to choose from, and a balanced design between positive and negative pictures can still be obtained by selecting image subsets.

A similar concern arises when considering the dimensions that are tested. Obviously, the categories may differ on many more characteristics than what has been tested, such as complexity or familiarity. Because of time/design constraints, it is rarely possible to obtain ratings on more than a few scales while examining a large number of stimuli. We choose four scales according to the vast interest in valence and arousal variation in affective science research and the relevance of testing norm significance with part of this database. We obviously cannot serve the interest of every research direction, but the GAPED would eventually benefit from further testing on other specific relevant dimensions.

Judging several different kinds of negative pictures at once can bias the negativity obtained for each single category. In fact, judgments are likely to be performed in a classification manner, creating a sort of hierarchy between the negative categories, as shown by the valence differences

observed. This is true for any judgment and rating study, and it is a problem that is hard to circumvent. However, ratings can still be reliably compared within a given category (and between negative and non-negative groups). Moreover, random presentation of pictures helps to counterbalance this effect. If getting close to the absolute assessment of valence is a concern in a particular study, the best option would be to re-rate a subset of the negative images, chosen according to the planned research.

Conclusion

Now that this new database, GAPED, has been created and some relevant ratings have been obtained for each of the pictures, we believe it can be reliably used in research dealing with spider and snake presentations or militant-related exposure, along with positive and neutral pictures. Researchers will be able to select the sample corresponding to their design needs from the entire database, either by using the rating proposed, or by re-rating the sample in their own population, with the same or other rating parameters relevant to their research question.

Contrary to the IAPS database, the goal of the GAPED is not to be able to compare research performed by using the same database. Actually, time has proven that not many laboratories have used this opportunity with the IAPS. The major goal of the present work is to provide researchers with some additional prerated emotional pictures. Simultaneously, the GAPED also supplies material for studies dealing with specific subjects. We hope that this will help lighten the material selection process—a time consuming task that too often hinders the progression of research.

Supplemental Materials Six pictures folders (named *Sn* = snake pictures, *Sp* = spider pictures, *H* = human concern pictures, *A* = animal mistreatment pictures, *N* = neutral pictures, *P* = positive pictures), as well as six text files (same name structure), containing the ratings on the two or four dimensions for the 730 GAPED images, may be downloaded from www.affective-sciences.org/researchmaterial, under number 7. “The Geneva Affective PicturE Database (GAPED): A 730 picture database for emotion induction.” Picture name starts with the code of the category (see above) and is followed by a three-digit number. Please note that numbers are discontinuous.

References

- Anders, S., Lotze, M., Erb, M., Grodd, W., & Birbaumer, N. (2004). Brain activity underlying emotional valence and arousal: A response-related fMRI study. *Human Brain Mapping*, 23, 200–209.
- Barrett, L. F. (2006). Valence is a basic building block of emotional life. *Journal of Research in Personality*, 40, 35–55.
- Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1996). Picture media and emotion: Effects of a sustained affective context. *Psychophysiology*, 33, 662–670.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25, 49–59.
- Brosch, T., Sander, D., & Scherer, K. R. (2007). That baby caught my eye...Attention capture by infant faces. *Emotion*, 7, 685–689.
- Codispoti, M., Bradley, M. M., & Lang, P. J. (2001). Affective reactions to briefly presented pictures. *Psychophysiology*, 38, 474–478.
- Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, H. H. Goldsmith, & K. R. Scherer (Eds.), *Handbook of affective sciences* (pp. 572–595). Oxford: Oxford University Press.
- Fredrickson, B. (2004). The broaden-and-build theory of positive emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 1367–1377.
- Fredrickson, B. L., & Cohn, M. A. (2008). Positive emotions. In M. Lewis, J. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 777–796). New York: Guilford Press.
- Fredrickson, B. L., & Levenson, R. W. (1998). Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. *Cognition and Emotion*, 12, 191–220.
- Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press.
- Greenwald, M. K., Cook, E. W., & Lang, P. J. (1989). Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of Psychophysiology*, 3, 51–64.
- Gross, J., & Levenson, R. W. (1993). Emotional suppression: Physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology*, 64, 970–986.
- Isen, A. (1998). On the relationship between affect and creative problem solving. In S. W. Russ (Ed.), *Affect, creative experience, and psychological adjustment* (pp. 3–17). Ann Arbor: Braun-Brumfield.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International Affective Picture System (IAPS): Instruction manual and affective ratings*. (Tech. Rep. No. A-4). Gainesville, FL: University of Florida, Center for Research in Psychophysiology.
- Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30, 261–273.
- Langsrud, Ø. (2003). ANOVA for unbalanced data: Use Type II instead of Type III sums of squares. *Statistics and Computing*, 13, 163–167.
- Lorenz, K. (1943). Die angeborenen Formen möglicher Erfahrung [The innate forms of potential experience]. *Zeitschrift für Tierpsychologie*, 5, 233–519.
- Luck, S. J. (2005). Ten simple rules for designing ERP experiments. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 17–32). Cambridge: MIT Press.
- Manstead, A. S. R., & Fischer, A. H. (2001). Social appraisal: The social world as object of and influence on appraisal processes. *Appraisal processes in emotion: Theory, methods, research*. (pp. 221–232). New York, NY: Oxford University Press.
- Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. Cambridge: MIT Press.
- Öhman, A. (1993). Fear and anxiety as emotional phenomena: Clinical phenomenology, evolutionary perspectives and information processing mechanisms. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 511–536). New York: Guilford Press.
- Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, 108, 483–522.
- Osgood, C. E. (1969). On the whys and wherefores of E, P, and A. *Journal of Personality and Social Psychology*, 12, 194–199.

- Roseman, I. (1984). Cognitive determinants of emotion: A structural theory. In P. Shaver (Ed.), *Review of Personality & Social Psychology* (Vol. 5, pp. 11–36). Beverly Hills: Sage.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178.
- Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11, 273–294.
- Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks*, 18, 317–352.
- Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Hillsdale: Erlbaum.
- Scherer, K. R. (1987). Toward a dynamic theory of emotion: The component process model of affective states. *Geneva Studies in Emotion and Communication*, 1, 1–98.
- Scherer, K. R. (1999). Appraisal theory. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 637–667). Chichester: Wiley.
- Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York: Oxford University Press.
- Scherer, K. R. (2004). Feelings integrate the central representation of appraisal-driven response organization in emotion. In A. S. R. Manstead, N. H. Frijda, & A. H. Fischer (Eds.), *Feelings and emotions: The Amsterdam symposium* (pp. 136–155). Cambridge: Cambridge University Press.
- Scherer, K. R., Dan, E. S., & Flykt, A. (2006). What determines a feeling's position in affective space? A case for appraisal. *Cognition and Emotion*, 20, 92–113.
- Smith, J. C., Bradley, M. M., & Lang, P. J. (2005). State anxiety and affective physiology: Effects of sustained exposure to affective pictures. *Biological Psychology*, 69, 247–260.
- Smith, C., & Ellsworth, P. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48, 813–838.
- Stark, R., Schienle, A., Walter, B., Kirsch, P., Blecker, C., Ott, U., et al. (2004). Hemodynamic effects of negative emotional pictures —A test-retest analysis. *Neuropsychobiology*, 50, 108–118.
- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219–235.
- Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, 76, 820–838.