CrossMark

Introducing the Open Affective Standardized Image Set (OASIS)

Benedek Kurdi[1] · **Shayn Lozano**[1] · **Mahzarin R. Banaji**[1]

Abstract We introduce the Open Affective Standardized Image Set (OASIS), an open-access online stimulus set containing 900 color images depicting a broad spectrum of themes, including humans, animals, objects, and scenes, along with normative ratings on two affective dimensions—*valence* (i.e., the degree of positive or negative affective response that the image evokes) and *arousal* (i.e., the intensity of the affective response that the image evokes). The OASIS images were collected from online sources, and valence and arousal ratings were obtained in an online study (total $N = 822$). The valence and arousal ratings covered much of the circumplex space and were highly reliable and consistent across gender groups. OASIS has four advantages: (a) the stimulus set contains a large number of images in four categories; (b) the data were collected in 2015, and thus OASIS features more current images and reflects more current ratings of valence and arousal than do existing stimulus sets; (c) the OASIS database affords users the ability to interactively explore images by category and ratings; and, most critically, (d) OASIS allows for free use of the images in online and offline research studies, as they are not subject to the copyright restrictions that apply to the International Affective Picture System. The OASIS images, along with normative valence and arousal ratings, are available for download from www.benedekkurdi.com/#oasis or https://db.tt/yYTZYCga.

Images represent both the physical and social worlds. Using a few thousand pixels, they can depict an unlimited array of people, objects, and scenes and can evoke a range of affective responses, such as happiness, excitement, contentment, sadness, anger, or disgust. Many studies in the behavioral and brain sciences require images that elicit varied emotions associated with social and nonsocial phenomena (Quigley, Lindquist, & Barrett, 2014). To facilitate such research, the Center for the Study of Emotion and Attention developed the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008), an internationally available normative set of emotional stimuli. The latest version of the IAPS contains 1,195 color photographs that have been assigned normative ratings on three dimensions—valence, arousal, and dominance.[1] Since its inception, the IAPS has been widely used in psychological, psychophysiological, and neuroscience research covering a broad array of topics, including affective processing in children (Hajcak & Dennis, 2009), psychophysiological reactions to sleep-related images in insomnia (Baglioni et al., 2010), fear conditioning (Wessa & Flor, 2007), the emotional modulation of attention (Cohen, Henik, & Mor, 2011), moral cognition (Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005), and implicit attitudes (Payne, Cheng, Govorun, & Stewart, 2005). To date, several thousand research studies have been published using IAPS images, making the IAPS one of the most frequently used stimulus sets in behavioral research today. Its contribution to advancing research cannot be overstated.

✉ Mahzarin R. Banaji
  mahzarin_banaji@harvard.edu

1 Department of Psychology, Harvard University,
   Cambridge, MA 02138, USA

---

[1] Categorical affective ratings for the IAPS images are also available (Mikels et al., 2005).

Springer

However, IAPS was produced in pre-Internet research times, and as such the images are subject to copyright restrictions that prohibit their use in online research studies. In fact, the copyright agreement accompanying the IAPS clearly stipulates that users are not allowed "to place them [i.e., IAPS images] on any internet or computer-accessible websites" [sic]. Given the increased reliance on online samples in behavioral research, this restriction poses an unnecessary constraint on research progress. The Internet has massively changed how psychological research is being conducted, both by expanding the range of phenomena under study and by providing new and more robust tools for implementing projects. Data collection has become faster, less expensive, and more efficient, because researchers can easily post research studies online for data collection from large and diverse pools of participants without geographic and other boundaries (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011; Kraut et al., 2004; Mason & Suri, 2012; Paolacci, Chandler, & Ipeirotis, 2010), with the exception of digital literacy. It is hardly a surprise, then, that hundreds of behavioral studies are being conducted online at any given time—a number that increases daily (Krantz, 2015). For instance, Project Implicit's data collection and educational website on topics of implicit group attitudes and beliefs (http://implicit.harvard.edu) has been in use since 1998 and has gathered data from over 15 million tests. However, the issues addressed by online research studies are not restricted to implicit social bias (Nosek, Banaji, & Greenwald, 2002); they range from voters' competence at assessing incumbent politicians (Huber, Hill, & Lenz, 2012) to life satisfaction (Peterson, Park, & Seligman, 2005), from emotion in decision making (Seo & Barrett, 2007) to personality (Buchanan, Johnson, & Goldberg, 2005), and from compulsive hoarding (Frost, Tolin, Steketee, Fitch, & Selbo-Bruns, 2009) to the prevention of sexually transmitted diseases (Pequegnat et al., 2007).

As the volume and complexity of online research increases, a greater number of behavioral researchers require access to pretested images that vary in affective valence and intensity. Although several general and specialized visual stimulus sets are available to facilitate behavioral research on emotion, such as the IAPS (Lang et al., 2008), the Karolinska Directed Emotional Faces (Lundqvist, Flykt, & Öhman, 1998), and the Geneva Affective Picture Database (GAPED; Dan-Glauser & Scherer, 2011), there is a lack of standardized, open-access, and widely available stimulus sets with corresponding normative affective ratings. Therefore, as of now, researchers conducting online studies are forced to create visual stimuli in an ad hoc manner, which is time-consuming, inefficient, and limits the comparability and generalizability of research findings.

The goal of the present project was to create an open-access standardized stimulus set containing affective images, which—similarly to the IAPS—includes a broad spectrum of themes. We sought to create an independent novel stimulus set containing high-quality contemporaneous images with the broadest possible coverage of circumplex space, rather than any direct content mapping to IAPS images. We collected 900 images depicting a wide range of categories, including humans, animals, scenes, and objects, from open-access online sources, and recruited a diverse sample of participants for a norming study to gauge their affective responses to the images. Although affective responses can be conceptualized and measured in a number of different ways, we relied on the circumplex model of affect (Russell, 1980) and collected self-reported subjective ratings on *valence* (i.e., the positivity or negativity of the affective response) and *arousal* (i.e., the level of excitement that an observer experiences).

Method

Participants

The participants were recruited through Amazon's Mechanical Turk (MTurk) to "rate images of everyday objects and scenes" in exchange for $0.75. MTurk offers a participant pool that is more diverse than undergraduate samples (Berinsky et al., 2012; Buhrmester et al., 2011), such as the sample that provided normative ratings for the IAPS images (Lang et al., 2008). Moreover, it has been demonstrated repeatedly that studies conducted via MTurk offer valid and reliable data whose quality is comparable to that of data collected in lab studies (Buhrmester et al., 2011; Mason & Suri, 2012; Paolacci et al., 2010; Rand, 2012). To make sure that participants were sufficiently attentive and motivated, we restricted participation to workers with an approval rate of at least 90 % on previous human intelligence tasks (HITs) completed on MTurk, and at least 50 HITs completed. Moreover, to prevent intercultural differences in terms of valence and arousal judgments or response styles from contaminating the results, the HIT was displayed only to workers from the United States. Potential participants received a disclaimer that they might find some of the images displayed in the study disturbing, due to "sexually explicit, violent, or traumatic" content. The original target number for participants was 800; however, because some participants failed to submit their HIT on Amazon MTurk after completing the study, we ended up with usable data from 822 participants. On average, it took participants 25.13 min (SD = 6.68) to complete the study.

The participants exhibited considerable variability in terms of age, gender, geographic location, and socioeconomic background. Participants' ages ranged from 18 to 74 years, with a mean of 36.63 years (SD = 11.91). With 420 female and 398 male participants, the gender distribution was balanced. Participants' ideological self-placement, race, highest level of education, and household income also varied considerably

(see Fig. 1), although, relative to the national average, liberal, White, highly educated, and high-income participants were overrepresented in our sample, as they are in most online research studies (Berinsky et al., 2012). Detailed demographic data have not been reported for the participants who rated the original IAPS images. However, given that the IAPS images were rated exclusively by introductory psychology students (Lang et al., 2008), one can reasonably assume that the sample used in the present study was considerably more diverse in terms of age and socioeconomic background.

Materials

The 900 images included in the study were obtained from a variety of online sources, most notably Pixabay (https://pixabay.com/en/; $N = 646$) and Wikipedia (https://www.wikipedia.org; $N = 172$). The images were found using Google Images (https://images.google.com). We collected a broad range of images that we expected to cover the largest possible area in circumplex space. The image search was restricted to images labeled as being available for reuse with modification, thus ensuring that the images could be edited and redistributed without any restrictions and free of charge. We standardized the size of the images by scaling and/or cropping them to 500 × 400 pixels. The images were then randomly assigned to four lists containing 225 images each.

Prior to the study, each image was placed into one of four mutually exclusive categories—animals, objects, people, and scenes. The images in the animal category ($N = 134$) depict various animals, including dogs, snakes, insects, birds, spiders, sharks, lions, monkeys, and cats. The images in the objects category ($N = 200$) depict a wide range of objects, including human-made objects such as fences, jewelry, cars, bottles, and balls of yarn, and natural objects such as leaves, rocks, pebbles, and flowers. The images in the people category ($N = 346$) depict humans alone, in dyads, and in groups in various situations of daily life. The images in the scenes category ($N = 220$) depict urban and rural spaces, as well as weather phenomena such as lightning or earthquakes. It should be noted that the category assignments were made by the first and second authors and had no basis in empirical measurement. They serve merely to facilitate the use of the stimulus set.

Procedure

Prestudy We considered whether to instruct participants to focus on the valence and intensity of each image or to ask about the feelings that each image evoked in them. We thus created two sets of instructions: *image-focused* instructions asking participants to indicate the level of valence or arousal intrinsic to each image, and *internal*

state-focused instructions asking participants to indicate the level of valence and arousal that each image has evoked in them[2] (Quigley et al., 2014). To determine whether these two sets of instructions led to different assessments, and if so, in what way, we conducted a prestudy with 184 participants, also recruited from Amazon MTurk.

Participants were randomly assigned to one of the two sets of instructions (see Appendixes 1 and 2) and to either the valence or the arousal dimension. We tested the effects of the instruction manipulation on the valence and arousal ratings by fitting a mixed-effects linear regression to the data, with random intercepts for images and participants, and a fixed effect for instruction focus (image-focused vs. internal state-focused). No significant effect of instruction focus was found, $t(90) = 0.789$, $p = .432$, suggesting that participants rated the images similarly irrespective of whether they had been assigned to the image-focused or the internal state-focused instruction condition. Therefore, in the main study, we dropped this manipulation and used only the image-focused instructions.

Main study The 900 images were randomly assigned to four lists of 225 images each. Each list was tested in a separate study. Four lists were created in order to prevent participant fatigue. Using a mixed-effects model including random effects for participants and images and a fixed effect for list, we found that ratings did not differ across the lists, $\chi^2(3) = 1.01$, $p = .798$. Therefore, all results are reported by collapsing across lists. To avoid contamination between the two dimensions of valence and arousal (Bishop, Oldendick, & Tuchfarber, 1984; Lau, Sears, & Jessor, 1990; Schuman, Kalton, & Ludwig, 1983; Wilcox & Wlezien, 1993), each participant was randomly assigned to rate the images on only the valence or only the arousal dimension. First, participants received a general description of the study (Screen 1) and detailed instructions on the meaning of the dimension that they would be asked to rate (Screen 2; for the full set of instructions, see Appendix 1). We expected that the meaning of the valence dimension might be more intuitively clear to participants than the meaning of the arousal dimension. To prevent the valence dimension from interfering with judgments of arousal, the participants in the arousal condition received an additional set of instructions (Screen 3) explaining the difference between valence and arousal and the orthogonal nature of the two dimensions.

After reading the instructions, participants were presented with the 225 images in an individually randomized order and were asked to rate the images using a 7-point Likert scale.

---

[2] The IAPS is based on internal state-focused instructions. For instance, the arousal instructions say, "[a]t one extreme of the scale you felt stimulated, excited, frenzied, jittery, wide-awake, aroused" (Lang et al., 2008).
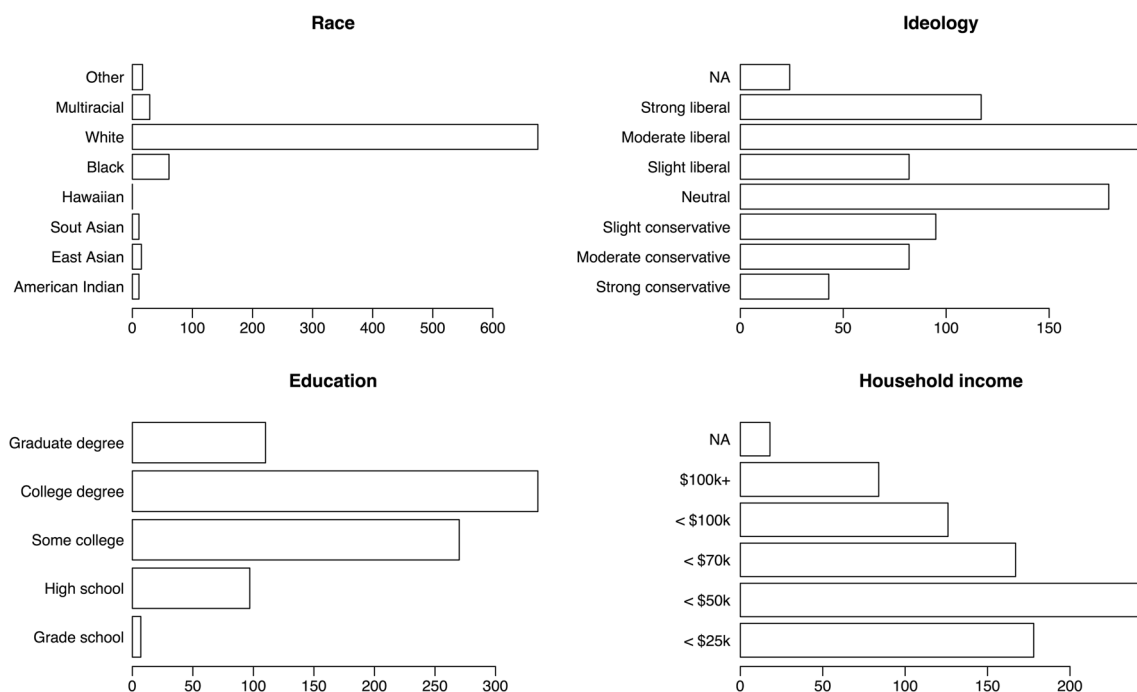
Fig. 1 Distributions of some key demographic variables (race, ideological self-placement, highest educational attainment, and annual household income) in the sample

Even though the IAPS images were originally normed using a so-called Self-Assessment Manikin (SAM; Hodes, Cook, & Lang, 1985) rather than a verbal scale, it has been shown that, at least for the valence and arousal dimensions assessed in the present study, SAM and corresponding verbal scales are very highly correlated (Bradley & Lang, 1994). Each image was displayed in a size of 500 × 400 pixels on a separate screen, and the rating scale was placed below the image. For the valence dimension, the word *Valence* was displayed above the rating scale and the points of the scale were labeled as *Very negative, Moderately negative, Somewhat negative, Neutral, Somewhat positive, Moderately positive,* and *Very positive.* For the arousal dimension, the word *Arousal* was displayed above the rating scale and the points of the scale were labeled as *Very low, Moderately low, Somewhat low, Neither low nor high, Somewhat high, Moderately high,* and *Very high.*

Valence and *arousal* are technical terms; however, they were used as labels because their meanings were clearly explained to participants in the instructions and we did not want to impose any new, potentially misleading, labels on the scales. After rating each image, participants clicked a button to proceed to the next screen. The survey was set up to proceed in a forward only direction; that is, participants were not allowed to return to previous screens.

After providing ratings for all 225 images, participants were asked to complete a standard demographic questionnaire including items on gender, age, ethnicity, race, ideological self-placement, annual household income, highest educational attainment, current ZIP code, and ZIP code where they had

lived longest. On the following screen, participants received a 6-digit code with which they were able to claim compensation on Amazon MTurk. On the last screen of the study, participants were thanked and debriefed.

Results

Univariate distributions

The number of valence ratings provided for each image ranged from 101 to 108, with a mean of 103.25 ratings ($SD = 2.77$) per image. The number of arousal ratings provided for each image ranged from 100 to 104, with a mean of 102.23 ratings ($SD = 1.30$) per image.

The mean valence and arousal ratings and the corresponding standard deviations were calculated for each image. The distribution of the imagewise means and standard deviations is shown in Fig. 2. Valence ratings ranged from 1.11 to 6.49, showing good usage of the entire range of the scale. The mean valence rating was 4.33, somewhat above the theoretical midpoint of the scale, and the median valence standard deviation was 1.09. Overall, the distribution of valence ratings was fairly uniform, although a Kolmogorov–Smirnov test for uniformity did not formally confirm this impression, $D = 0.17$, $p < .001$. Arousal ratings ranged from 1.69 to 5.72, and thus the range of arousal ratings was more restricted than the range of valence ratings. The mean arousal rating was 3.67, somewhat below the theoretical midpoint of the scale, and the median
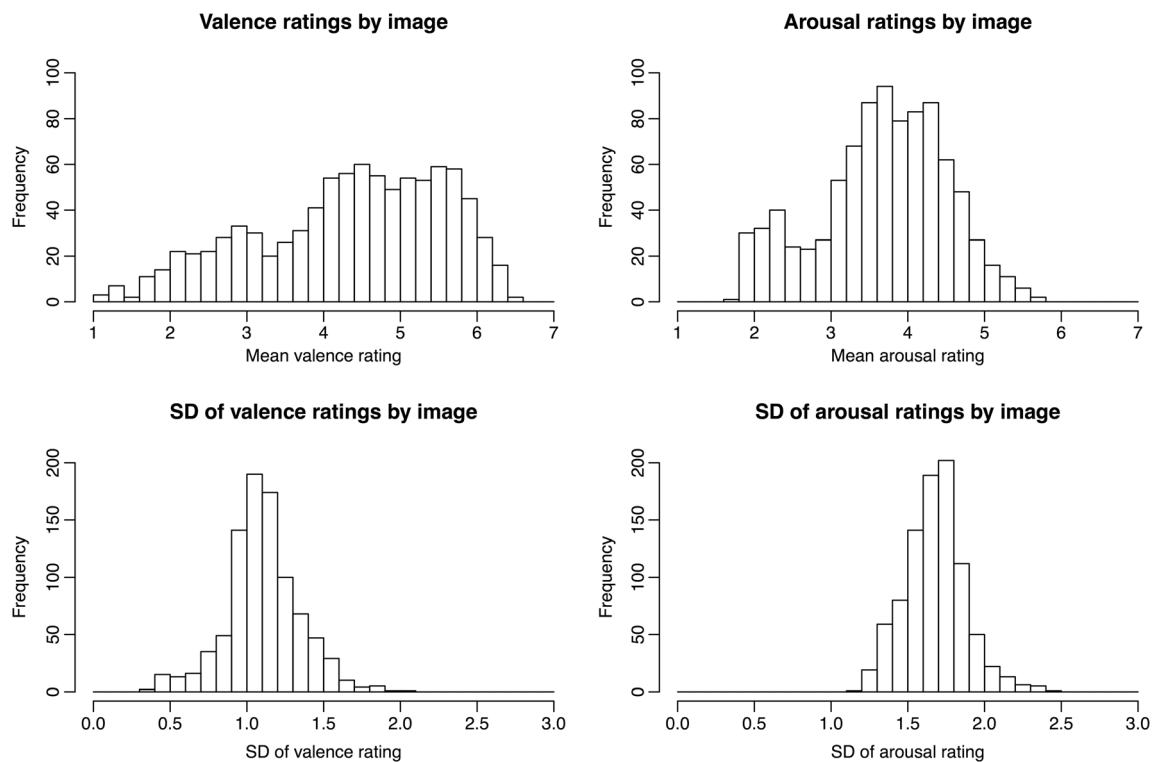
Fig. 2 Univariate distributions of imagewise mean valence and arousal ratings (top row) and distributions of imagewise valence and arousal standard deviations (bottom row)

arousal standard deviation was 1.68 (and thus higher than that for valence ratings). From visual inspection, the distribution of arousal ratings seemed fairly normal, although a Kolmogorov–Smirnov test for normality did not formally confirm this impression, $D = 0.96$, $p < .001$.

The soundness of our measures is underscored by the striking visual similarity between the valence and arousal distributions in the present study and in the IAPS (Lang et al., 2008). In fact, submitting standardized valence and arousal ratings from both datasets to Kolmogorov–Smirnov tests confirmed that the valence scores had been sampled from the same underlying distributions across both studies, $D = 0.04$, $p = .257$, and that the arousal scores had been sampled from similar, if not exactly the same, distributions, $D = 0.07$, $p = .023$. Moreover, in the IAPS, just as in this study, the range of arousal ratings (from 1.72 to 7.35 out of the theoretically possible range of 1 to 9) was smaller than the range of valence ratings (from 1.31 to 8.34), and the median arousal standard deviation (2.15) was higher than the median valence standard deviation (1.57).

We evaluated the face validity of these valence and arousal ratings by probing which images received the most highly positive and negative valence ratings and the highest and lowest arousal ratings, and which images had the highest and lowest valence and arousal standard deviations, indicating low and high levels of agreement, respectively. The most positive valence rating ($M = 6.49$, $SD = 0.78$) was obtained for

image I256, which depicts a young puppy in a polka-dotted coffee pot, and the most negative valence rating was obtained for image I496 ($M = 1.11$, $SD = 0.42$), which depicts an emaciated person in a concentration camp. Image I679, which depicts a rooftop, had the lowest valence standard deviation (0.34, $M = 4.06$), and image I540, which depicts heterosexual oral intercourse, had the highest valence standard deviation (2.03, $M = 4.84$).

The highest arousal rating ($M = 5.72$, $SD = 1.67$) was obtained for image I537, which depicts heterosexual intercourse, and the lowest arousal rating ($M = 1.69$, $SD = 1.24$) was obtained for image I860, which depicts a concrete wall. Image I597, which depicts a pile of blank paper, had the lowest arousal standard deviation (1.19, $M = 1.82$), and image I208, which depicts dead bodies lying on the ground, had the highest arousal standard deviation (2.48, $M = 4.51$). These values demonstrate face validity and confirm the soundness of our valence and arousal measures.

Reliability

Because every participant rated only a subset of the images, we calculated interrater reliabilities for the valence and arousal scales using a resampling method. For each dimension, we randomly generated 1,000 split halves, calculated the correlation between the two halves, and took the mean of the correlation distribution as our reliability measure. For the valence

dimension, the interrater reliability was excellent, R_{val} = .984 (SD = 0.002, range: R_{min} = .974 and R_{max} = .989). For the arousal dimension, the interrater reliability was somewhat lower but still outstanding, R_{aro} = .929 (SD = 0.015, range: R_{min} = .833 and R_{max} = .958).

We hypothesized that in comparison to the valence scale, the relatively lower reliability of the arousal scale might have been due to differences in judgments between men and women. In other words, the ratings might have been more consistent within than across gender groups. To test this possibility, we calculated separate reliability measures for our male and female participants. For the female participants, we obtained $R_{aro/w}$ = .930 (SD = 0.014, range: R_{min} = .864 and R_{max} = .958), and for the male participants, we obtained $R_{aro/m}$ = .964 (SD = 0.015, range: R_{min} = .862 and R_{max} = .957). However, interrater reliability depends not only on the internal consistency of a measure but also on the sample size. To adjust for the fact that the female and male subsamples were smaller than the overall sample, we used the Spearman–Brown prophecy formula[3] to calculate the expected reliability of the measure if the sizes of the female and male subsamples were equal to that of the overall sample. For female participants, we obtained an adjusted reliability score of $\overline{R}_{aro/w}$ = .961, and for male participants, we obtained an adjusted reliability score of $\overline{R}_{aro/m}$ = .964, each of which is higher than the original reliability estimate of R_{aro} = .929. The fact that the interrater reliabilities for each gender exceeded the interrater reliability for the sample as a whole suggests that, as hypothesized, the relatively lower reliability of the arousal scale was in part due to a lack of internal consistency across, rather than within, gender groups.

Relationship between means and standard deviations

In the next step, we investigated the relationship between the imagewise means and imagewise standard deviations for each of the two affective dimensions, by fitting linear, quadratic, and cubic regressions to the data with the imagewise means as predictors and the imagewise standard deviations as criterion variables.

For the valence dimension, the scatterplot (see Fig. 3, left pane) shows an M-shaped relationship between the means and standard deviations. In other words, the standard deviations were lowest at both ends and at the midpoint of the valence scale. This kind of relationship seems quite reasonable, considering that the valence scale used in the study had three meaningful anchor points—its low end (highly negative images), its midpoint (completely neutral images), and its high end (highly positive images). The visual impression of an M-shaped relationship was further confirmed by the fact that a cubic regression provided the best fit to the data, although the

relationship between the valence means and standard deviations still remained fairly weak, R^2 = .036. Participants exhibited fairly high levels of agreement, overall. The weak cubic trend might have nevertheless have arisen because levels of agreement were especially high—and thus standard deviations were especially low—for some images at both extremes and at the midpoint of the valence scale.

For the arousal dimension, the scatterplot (see Fig. 3, right pane) displays an inverted U-shaped relationship between the means and standard deviations. In other words, standard deviations were lowest at the low end of the arousal scale, became higher as the mean increased, and leveled off again at the very high end of the scale. This kind of relationship seems quite reasonable, considering that the arousal scale used in the study did not have a meaningful midpoint. Moreover, the highest arousal ratings were typically obtained for images depicting nudity, which (a) were rated differently by men and women and (b) might have given rise to socially desirable responding (Crowne & Marlowe, 1960; Paulhus, 1984) for some participants, but not for others. The visual impression of an inverted U-shaped relationship was further confirmed by the fact that a quadratic regression provided the best fit to the data, with a fairly strong relationship between the arousal means and standard deviations, R^2 = .411.[4]

Relationship between valence and arousal

The relationship between the valence and arousal ratings is shown in Fig. 4. As expected, we found no significant linear relationship between the valence and arousal ratings, Pearson's r = −.06 [−.12; .01], t(898) = −1.74, p = .081.[5] In addition to the lack of a strong correlation, the fact that all four quadrants of the circumplex space contain a fair number of images (positive–high arousal, N = 197; positive–low arousal, N = 392; negative–high arousal, N = 145; and negative–low arousal, N = 166) provides further evidence for a fairly balanced bivariate distribution of valence and arousal ratings. However, as with the IAPS, the valence and arousal ratings show a boomerang-shaped bivariate distribution, such that arousal

---

[3] $\rho_{KK} = \frac{K\rho_{tt}}{1+(K-1)\rho_{tt}}$, where ρ_{tt} is the interrater reliability of the measure given the current sample size, and K is the lengthening factor.

[4] These results are highly similar to the corresponding results from the IAPS. In the IAPS, just as in the present study, we found an M-shaped relationship between the valence means and standard deviations, with a cubic regression providing the best fit to the data, R^2 = .048. The same applied to the relationship between the arousal means and standard deviations, with an inverted U-shaped relationship between the two and a quadratic regression providing the best fit to the data, R^2 = .188.

[5] The correlation between valence and arousal ratings is similarly negative, but stronger and statistically significant, for the IAPS images, r = −.29 [−0.34; −0.24], t(1192) = −10.43, p < .001. The fact that both dimensions are significantly correlated in the IAPS might have to do with some participants misinterpreting the SAM as representing neutral arousal to high arousal, rather than, as intended, low to high arousal (Kuppens, Tuerlinckx, Russell, & Barrett, 2013).
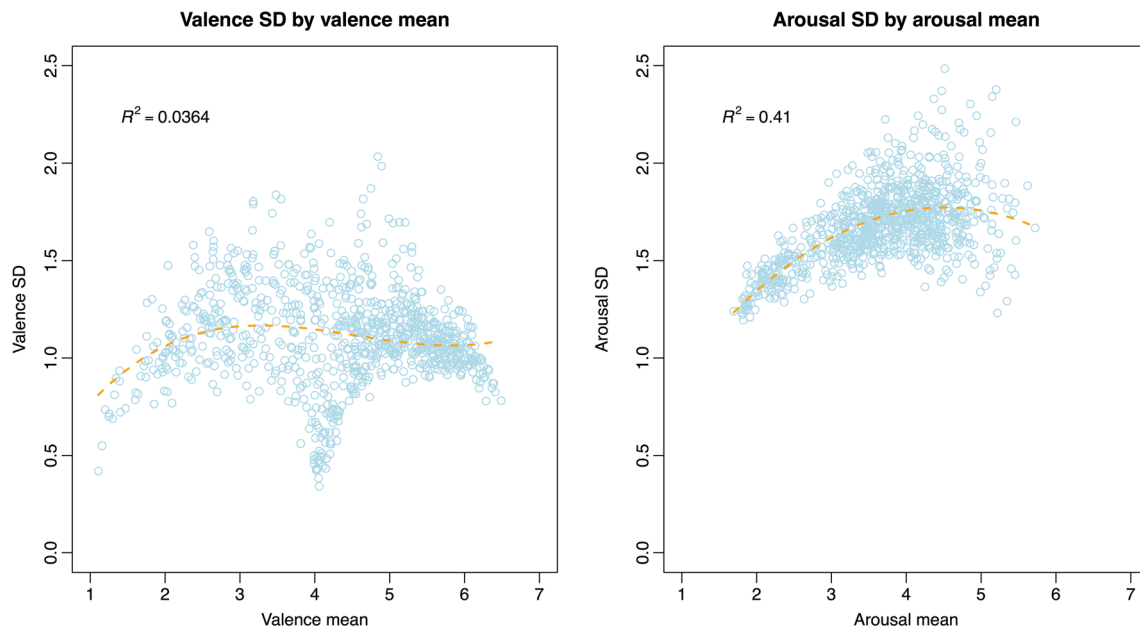
Valence SD by valence mean

$R^2 = 0.0364$

Arousal SD by arousal mean

$R^2 = 0.41$

Fig. 3 (Left) Relationship between imagewise valence means and imagewise valence standard deviations, with the best-fitting cubic regression line. (Right) Relationship between imagewise arousal means and imagewise arousal standard deviations, with the best-fitting quadratic regression line

ratings are highest at the most positive and most negative levels of valence (on the relationship between valence and arousal more generally, see Kuppens et al., 2013; Lang, 1995). Thus, in the future we plan to add further images to OASIS to correct for the relative undersampling of the low-arousal positive and low-arousal negative segments of the circumplex space. At the same time, we note that, unlike IAPS, OASIS contains a reasonable number of mid-arousal and high-arousal neutral images.
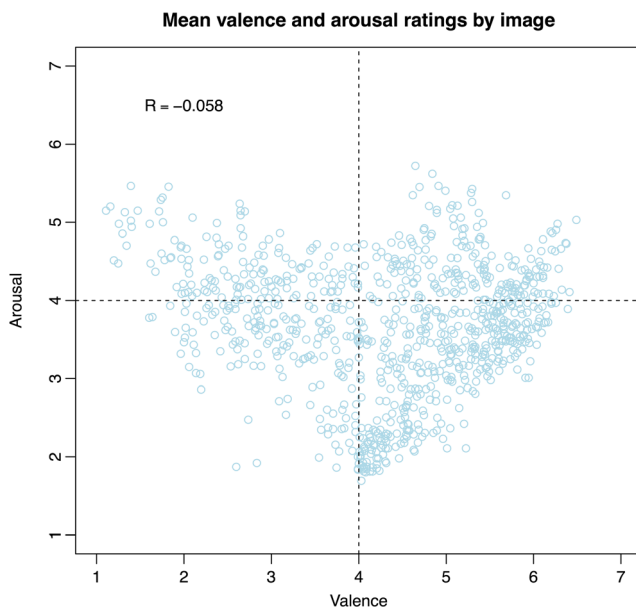
Mean valence and arousal ratings by image

R = −0.058

Fig. 4 Image ratings in circumplex space, with valence (measured on a 1–7 Likert scale) on the *x*-axis and arousal (also measured on a 1–7 Likert scale) on the *y*-axis

Figure 5 illustrates the relationship between valence and arousal ratings by image categories. Descriptively, the image categories modulated univariate valence and arousal distributions. At 4.45, animals had the highest mean valence rating ($SD = 1.24$), followed by people ($M = 4.38$, $SD = 1.20$), scenes ($M = 4.25$, $SD = 1.44$), and objects ($M = 4.22$, $SD = 0.97$). On the arousal dimension, animals were also rated as most highly arousing ($M = 4.00$, $SD = 0.55$), followed by people ($M = 3.89$, $SD = 0.67$), scenes ($M = 3.88$, $SD = 0.78$), and objects ($M = 2.84$, $SD = 0.79$). The correlation between valence and arousal was strongest for images depicting animals, $r = −.27$ [$−.42$; $−.11$], $t(132) = −3.23$, $p = .002$; followed by objects, $r = −.15$ [$−.28$; $−.01$], $t(198) = −2.07$, $p = .039$; scenes, $r = −.11$ [$−.24$; $−.02$], $t(218) = −1.62$, $p = .106$; and people, $r = −.01$ [$−.11$; $.09$], $t(344) = −0.23$, $p = .817$. However, because the assignment of category labels was not based on any rigorous empirical classification and we did not seek to obtain an exhaustive or representative sample of images from each category, these values should merely be understood as characterizing this particular stimulus set rather than any general psychological phenomenon.

Gender differences

On the basis of previous research on gender differences in affective processing (Bellezza, Greenwald, & Banaji, 1986; Bradley, Codispoti, Sabatinelli, & Lang, 2001; Sabatinelli, Flaisch, Bradley, Fitzsimmons, & Lang, 2004; Wrase et al., 2003), and involving the IAPS in particular (Lang et al., 2008), we expected that participant gender might modulate the valence and arousal ratings. Therefore, we calculated
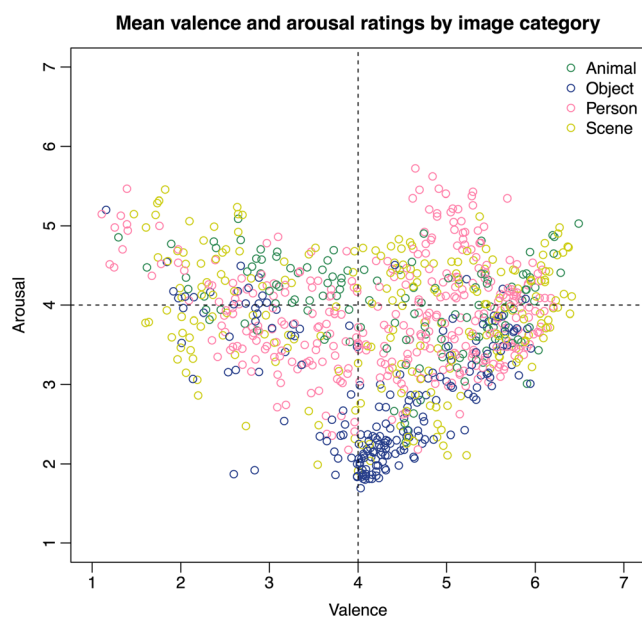
Fig. 5 Image ratings in circumplex space, with valence (measured on a 1–7 Likert scale) on the x-axis and arousal (also measured on a 1–7 Likert scale) on the y-axis. The colors correspond to different image categories

mean valence and arousal ratings for each image, broken down by participant gender. The mean valence ratings provided by women were almost perfectly correlated with the mean valence ratings provided by men, $r = .95$ [.95; .96], $t(898) = 92.62$, $p < .001$. On the arousal dimension, the mean ratings provided by women were also highly, although not perfectly, correlated with the mean ratings provided by men, $r = .83$ [.81; .85], $t(898) = 43.93$, $p < .001$. These correlations are significantly different from each other, $z = 14.21$, $p < .001$. Even though the correlations across genders were high, we still found considerable gender differences for some images, especially in the arousal ratings of sexually explicit images. Therefore, we recalculated the correlation between women and men after removing the arousal ratings for these images from the data. This resulted in a somewhat higher correlation, $r = .88$ [.87; .90], $t(839) = 54.50$, $p < .001$. The improvement was statistically significant, $z = 4.44$, $p < .001$.

Even though we obtained high correlations between the valence and arousal ratings provided by men and women, we also wanted to test for potential gender differences in terms of the mean levels of valence and arousal ratings assigned to the images. We investigated such gender effects by fitting a mixed-effects linear model to the data with random intercepts for images and participants, and an interaction between participant gender (male vs. female) and rating dimension (valence vs. arousal). We obtained a significant main effect for dimension, $t(817.1) = 10.95$, $p < .001$, but no significant gender effect, $t(819.2) = 0.21$, $p = .836$, and no interaction, $t(817.8) = 1.46$, $p = .143$. This confirms that after controlling for image-specific and participant-specific variation, men and

women rated both dimensions similarly. However, to enable researchers to select pictures that either differentiate or do not differentiate on the basis of gender, we provide valence and arousal ratings modulated by participant gender.

Effects of demographic variables

Earlier findings suggested that younger and older adults might differ from each other in terms of their valence and arousal ratings of IAPS images (Grühn & Scheibe, 2008). After applying a median split to the continuous age variable, we found very high correlations between both the valence ratings, $r = .98$ [.97; .98], $t(898) = 133.03$, $p < .001$, and the arousal ratings, $r = .92$ [.91; .93], $t(898) = 72.67$, $p < .001$, provided by younger and older participants. Follow-up analyses indicated that the relatively lower correlation in arousal ratings was primarily due to differences in how younger and older participants rated the sexually explicit images. On average, older participants tended to assign somewhat higher arousal ratings ($M = 4.51$, $SD = 0.71$) to these images than did younger participants ($M = 4.40$, $SD = 0.76$), $t(58) = 2.58$, $p = .012$. This apparent inconsistency between our study and the study conducted by Grühn and Scheibe (2008) might be due to differences in the participant pools. Whereas that study focused specifically on age differences, and thus used two distinct groups of participants (younger adults, 18–31 years; older adults, 63–77 years), we did not restrict participation to any specific age group.

We obtained similarly high split-half correlations after dividing our sample into low-income and high-income groups. On the valence dimension, the ratings provided by low-income and high-income participants were correlated at $r = .96$ [.95; .96], $t(898) = 102.99$, $p < .001$, and on the arousal dimension, they correlated at $r = .95$ [.94; .95], $t(898) = 88.04$, $p < .001$. Thus, income did not seem to significantly modulate the valence and arousal ratings in the present study.

We also probed whether ideological self-placement affected the mean valence and arousal ratings assigned to the OASIS images. After dichotomizing the ideology variable and excluding participants who identified as ideologically neutral, we found reasonably high correlations between both the valence ratings, $r = .87$ [.85; .88], $t(898) = 52.32$, $p < .001$, and the arousal ratings, $r = .78$ [.75; .80], $t(898) = 37.32$, $p < .001$, provided by liberal and conservative participants. Follow-up analyses revealed that the valence and arousal ratings provided by liberals and conservatives were especially inconsistent for a few select themes. Liberal participants tended to assign more positive valence ratings than conservative participants to images depicting nudity and sexuality, whereas the conservative participants tended to assign more positive valence ratings than liberal participants to images depicting guns, soldiers, and police officers. Similarly, conservatives tended to rate images depicting guns, war scenes, and

police officers as more arousing than did liberals, whereas liberals tended to rate sexually explicit images as more arousing than did conservatives. It should be noted, however, that our sample was ideologically unbalanced, containing more liberal ($N = 396$) than conservative ($N = 220$) participants, $\chi^2(1) = 50.29$, $p < .001$.

The dichotomization of continuous variables is useful for ease of understanding, but it produces a loss of power and can lead to spurious results (Altman & Royston, 2006; MacCallum, Zhang, Preacher, & Rucker, 2002). Therefore, we repeated our analyses using age and ideological self-placement as continuous variables. In the first step, we calculated correlations between the ratings provided by each pair of participants in the study (separately for valence and arousal, because the groups of participants providing ratings for each dimension were nonoverlapping). In the second step, we calculated difference scores for each pair of participants in terms of age and ideology. And, finally, we calculated whether pairwise differences in terms of age or ideology predicted levels of agreement. The mean pairwise correlation between participants was $r = .41$ ($SD = .25$, median $= .45$). As indicated by the higher reliability values for the valence dimension obtained above, the mean pairwise correlation was significantly higher for the valence dimension, $r = .57$ ($SD = .16$, median $= .60$), than for the arousal dimension, $r = .24$ ($SD = 0.22$, median $= .25$), $z = 41.39$, $p < .001$. Crucially, however, both the relationship between age differences and the pairwise correlations, $r = .07$, and the relationship between ideological distance and the pairwise correlations, $r = -.04$, remained weak, almost nonexistent, suggesting that neither age nor ideology had any moderating effect on our valence and arousal ratings.

Discussion

We collected and normed images online in order to create the OASIS dataset. OASIS contains 900 open-access images depicting a variety of themes that have been rated by a fairly diverse sample of American adults on two affective dimensions, valence (i.e., the positivity or negativity of an image) and arousal (i.e., the intensity of the affective response that the image evokes). Similar to IAPS, the mean ratings on the valence dimension formed a nearly uniform distribution, whereas the mean ratings on the arousal dimension formed a nearly normal distribution. Also in line with the IAPS, the variability of the valence ratings was higher than the variability of the arousal ratings. The images with the most negative and positive valence and the highest and lowest arousal ratings suggest that the ratings obtained for the OASIS images have good face validity. The interrater reliability of the valence dimension was excellent, whereas the interrater reliability of the arousal dimension was somewhat lower but still outstanding.

Also in line with the IAPS, the valence and arousal dimensions had a negative linear relationship with each other, although in the present stimulus set this relationship was not statistically significant. Moreover, the OASIS images showed the same boomerang-shaped distribution characterizing the IAPS images, such that extremely positive and extremely negative images tended to have the highest arousal ratings. The mean valence ratings had an M-shaped cubic relationship with valence standard deviations, such that standard deviations were lower at both extremes and at the midpoint of the valence scale. This relationship is most probably due to the fact that the valence scale has three meaningful anchor points (both extremes and the midpoint). The relationship between the mean arousal ratings and arousal standard deviations exhibited a quadratic trend, with standard deviations being higher toward the high end of the scale, due to gender differences in terms of assigning arousal ratings to the sexually explicit images, and possibly due to socially desirable responding that affected some participants more than others. The proclivity for socially desirable responding is a fairly stable personality trait (Furnham, 1986), and it seems reasonable to assume that a person high in socially desirable responding might be motivated to report relatively low arousal ratings for sexually arousing images.

We also investigated the effects of gender and certain demographic variables. We showed that the relatively low interrater reliability of the arousal ratings was due to a lack of homogeneity across genders. However, the reliability of the arousal dimension came close to that of the valence dimension once gender differences were controlled for. Moreover, the imagewise mean valence ratings were almost perfectly correlated across genders, whereas the correlation between women and men was somewhat lower for the arousal dimension. The correlation became higher after the images with sexually explicit content were removed from consideration. The effects of demographic variables, including age, income, and ideological self-placement, were negligible, although some systematic differences emerged between liberals and conservatives in rating images involving sexuality and violence.

Future directions

The analyses presented in this report obviously do not exhaust the possibilities that this stimulus set offers. Future work should investigate (1) low-level visual properties of the OASIS images in order to enable researchers to control for such properties when selecting subsets of images for inclusion in a particular study; (2) ratings of the OASIS images based on alternative conceptualizations of affect, including a third dimension (dominance or control) and discrete emotional labels, as well as additional dimensions such as distinctiveness and memorability; (3) the validity and reliability of OASIS across cultural contexts and social

groups; and (4) affective responses to the OASIS images that go beyond self-reported subjective ratings.

First, even though valence and arousal are uncorrelated with low-level visual properties across the entire IAPS stimulus set, it has been shown that specific subsets of images can systematically differ from each other in terms of spatial frequencies—that is, in the level of visual detail present in a stimulus per visual angle (Delplanque, N'diaye, Scherer, & Grandjean, 2007). Therefore, to be able to eliminate the effects of this low-level confound on neural processing, spatial-frequency norms should be created for the OASIS images. Such norms would allow researchers to control for any possible differences in terms of this property across subsets of images selected for use in a particular study.

Second, affective responses need not be conceptualized as points in a two-dimensional space. Future work should investigate where the OASIS images are located in a three-dimensional space of affect that includes a third dimension, such as dominance or control (Fontaine, Scherer, Roesch, & Ellsworth, 2007). Moreover, not all negative images are created equal. The behavioral and neural responses to negative stimuli differ, depending on whether they are perceived as immediately threatening to the individual or as merely being negative and nonthreatening (Kveraga et al., 2015). Importantly, affect can also be conceptualized as a set of discrete states rather than as dimensional (Barrett, 1998; Ekman, 1992; Izard, 1992), and, similarly to the IAPS images (Libkuman, Otani, Kern, Viger, & Novak, 2007; Mikels et al., 2005), the images included in the OASIS dataset could also be rated using discrete emotion labels rather than two- or three-dimensional scales. Furthermore, even though the OASIS images were selected to represent various segments of circumplex space, they are subject to variation on other, not necessarily affective, dimensions, such as consequentiality, memorability, meaningfulness, and familiarity. In some studies, researchers may want to control for these variables, whereas in other studies they may want to systematically probe their effects on other variables. Therefore, it would be desirable to have an additional set of norms on these properties of the OASIS images (for relevant ratings of the IAPS images, see Libkuman et al., 2007).

Third, studies conducted with samples from such varied societies as Bosnia (Drace, Efendic, Kusturica, & Landzo, 2013), Brazil (Lasaitis, Ribeiro, & Bueno, 2008), Chile (Dufey, Fernandez, & Mayol, 2011), Hungary (Deák, Csenki, & Révész, 2010), and India (Lohani, Gupta, & Srinivasan, 2013) have revealed major similarities, but also subtle cultural differences, in terms of the valence and arousal ratings assigned to IAPS images. Therefore, future studies should verify the cross-cultural validity of OASIS. Moreover, even though in the present study we did not find any strong age effects, future work involving larger-scale samples of older adults or non-self-report measures might reveal

age-related differences in the processing of affectively relevant images included in OASIS (Grühn & Scheibe, 2008; Moriguchi et al., 2011).

Fourth, social desirability (Crowne & Marlowe, 1960; Paulhus, 1984) and lack of introspective access to the contents of one's mind (Nisbett & Wilson, 1977) can distort responses on self-report measures. Thus, affective responses to the OASIS images should also be measured using less controlled and more automatic implicit measures (Banaji, 2001), psychophysiological reactions (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000), and functional magnetic resonance imaging (Moriguchi et al., 2011; Weierich, Wright, Negreira, Dickerson, & Barrett, 2010). At the same time, it should be noted that a reasonably stable relationship has been observed between psychophysiological responses such as heart rate, skin conductance, facial electromyography, and the startle reflex, on the one hand, and subjective affective ratings of IAPS images, on the other (Lang, Bradley, & Cuthbert, 1990). We have no reason to suspect that this would be otherwise with the OASIS images.

Researchers from laboratories around the world are invited to contribute to the effort of norming the OASIS images. As we discussed above, such studies might (1) address further perceptual, cognitive, or affective dimensions not included in the present norming study; (2) use a range of novel samples (e.g., non-American samples or samples from specific social or demographic groups); or (3) use tools other than self-report, such as the methods offered by psychophysiology or neuroimaging. Such additional ratings will be included in the OASIS dataset and made widely available to researchers if the ratings are sent to oasis@fas.harvard.edu.

Usage

The OASIS stimulus set, containing 900 color images at a size of 500 × 400 pixels, is available at www.benedekkurdi.com/#oasis or https://db.tt/yYTZYCga. The images can be downloaded, used, and modified free of charge for research purposes. Along with the images, we provide an accompanying data file listing the unique identifier, theme, category, source, valence mean, valence standard deviation, valence sample size, arousal mean, arousal standard deviation, and arousal sample size for each image. Because—at least for some image categories—we found considerable gender differences in terms of valence, and especially arousal, ratings, the same information is also provided broken down by gender in a separate data file.

At the first URL, we also offer an online tool that displays an interactive scatterplot of the valence and arousal ratings, broken down by image category. The tool

lets users restrict the scatterplot to one or more image categories (e.g., only objects, or people and animals). Furthermore, when the user clicks on a point in the scatterplot, the tool displays a thumbnail of the corresponding image, along with its unique identifier and its valence (mean and standard deviation) and arousal (mean and standard deviation) ratings. On the basis of the unique identifier displayed by the tool, the user can easily find the image in the downloaded stimulus set and include it in a research study.

Conclusion

In this article, we have introduced OASIS, an open-access online stimulus set containing color images normed on two affective dimensions, valence and arousal. OASIS offers four distinct advantages. First, it contains a large number of images spanning a wealth of different themes that cover much of circumplex space (including mid- to high-arousal images that are neutral on the valence dimension) and, unlike the images from IAPS, have been assigned to four broad categories with large numbers of stimuli in each. Second, because the OASIS ratings were obtained in 2015 from a diverse sample of American adults, rather than an under-graduate sample, they offer a valid reflection of self-reported contemporary assessments of valence and arousal. Each image was rated by a relatively high number of participants (ranging from 101 to 108), and valence and arousal remained unconfounded because each participant provided ratings on only one dimension. Third, unencumbered by the copyright restrictions that apply to similar stimulus sets, such as the IAPS, the OASIS images allow for free reuse and modification. Thus, they may be used freely in both online and offline research studies. Finally, at www. benedekkurdi.com/#oasis we offer an online tool that enables users to freely download the OASIS images along with their normative valence and arousal ratings, and to interactively explore them by category and by valence and arousal ratings. Our hope is that the OASIS stimulus set will prove to be a useful resource for researchers studying any aspect of affective responding in research studies conducted over the Internet or in the lab.

Appendix 1

Table 1 Image-focused instructions

Screen	Valence	Arousal
Screen 1	Pictures communicate. They depict people, objects, and scenes. In this study, we are interested in how various pictures make people feel. As you answer, please remember: There are no right or wrong answers. We are interested in your individual opinion. Please do not take too long over any picture. Your first response is as good as any.	
Screen 2	We will ask you to rate a series of pictures in terms of how positive or negative they are. If the picture represents something good or positive, please use the right side of the scale to mark your answer. Positive images are those that represent things that make us happy, satisfied, competent, proud, contented, delighted, and so on. It doesn't matter what the specific picture is about, as long as it represents something positive or good. If the picture represents something bad or negative, please use the left side of the scale to mark your answer. Negative images are those that represent things that make us unhappy, upset, irritated, angry, sad, depressed, and so on. It doesn't matter what the specific picture is about, as long as it is something negative or bad. Use the middle of the scale to indicate that the picture makes you feel neutral, that is, that you think that the picture is neither positive nor negative.	We will ask you to rate a series of pictures in terms of the amount of emotion that they evoke. In other words, we would like to know how much emotional intensity the picture creates; whether the picture captures something good or bad doesn't matter. We are interested only in the degree of excitement, energy, or intensity of feeling it represents. Use the right side of the scale to mark your answer if the picture represents something that is strongly emotional. The words that we might use to describe the emotional state that the picture creates are aroused, alert, activated, charged, or energized. Use the left side of the scale to mark your answer if the picture represents something that is not strongly emotional. The words that we might use to describe the emotional state that the picture creates are unaroused, slow, still, de-energized, calm, or peaceful. Use the middle of the scale to indicate an image that is moderately arousing or halfway through the two extremes.

Table 1 (continued)

Screen	Valence	Arousal
	Please use the full range of the scale to make your responses rather than relying on only a few points.	Please use the full range of the scale to make your responses rather than relying on only a few points.
Screen 3		When making your ratings, you need to forget about whether the picture is about something positive or negative, good or bad. Instead, we ask you to rate the picture only in terms of the emotional intensity that it evokes.
		For instance, you might have two pictures—one depicting an athlete winning a gold medal (something good) and another depicting an athlete getting injured and losing the race (something bad). You can give both images the same rating because they are both creating a similar level of emotion, even though the feelings are not the same. Your answers should be based only on how much emotional intensity the picture captures.

Appendix 2

Table 2   Internal state-focused instructions

Screen	Valence	Arousal
Screen 1	Pictures communicate. They depict people, objects, and scenes. In this study, we are interested in how various pictures make people feel. As you answer, please remember: There are no right or wrong answers. We are interested in your individual opinion. Please do not take too long over any picture. Your first response is as good as any.	
Screen 2	We will ask you to rate a series of pictures in terms of how positive or negative they make you feel. If the picture creates a positive feeling, please use the right side of the scale to mark your answer. Positive feelings include feeling happy, satisfied, competent, proud, contented, delighted, and so on. It doesn't matter what the specific feeling is, as long as it is something positive. If the picture creates a negative feeling, please use the left side of the scale to mark your answer. Negative feelings include feeling unhappy, upset, irritated, angry, sad, depressed, and so on. It doesn't matter what the specific feeling is, as long as it is something negative. Use the middle of the scale to indicate that the picture makes you feel neutral, that is, neither positive nor negative. Please use the full range of the scale to make your responses rather than relying on only a few points.	We will ask you to rate a series of pictures in terms of how excited or aroused they make you feel. In other words, we would like to know how intense the feeling is that the picture evokes; whether it is good or bad doesn't matter. Use the right side of the scale to mark your answer if the picture makes you feel emotionally aroused, alert, activated, charged, or energized. Use the left side of the scale to mark your answer if the picture makes you feel no emotional arousal at all, that is, if it makes you feel unaroused, slow, still, de-energized, calm, or peaceful. Use the middle of the scale to indicate that you are moderately aroused, that is, halfway through the two extremes. Please use the full range of the scale to make your responses rather than relying on only a few points.
Screen 3		When making your ratings, you need to forget about whether the picture makes you feel good or bad. We ask you to rate the picture only in terms of emotional intensity and not in terms of goodness or badness. For instance, you might have two pictures—one depicting an athlete winning a gold medal (something good) and another depicting an athlete getting injured and losing the race (something bad). You can give both images the same rating because they are both creating a similar level of emotion, even though the feelings are not the same. Your answers should be based only on how much emotional intensity the picture creates.

References

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ, 332,* 1080. doi:10.1136/bmj.332.7549.1080

Baglioni, C., Lombardo, C., Bux, E., Hansen, S., Salveta, C., Biello, S., … Espie, C. A. (2010). Psychophysiological reactivity to sleep-related emotional stimuli in primary insomnia. *Behaviour Research and Therapy, 48,* 467–475. doi:10.1016/j.brat.2010.01.008

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger III, J. S. Nairne, & I. Neath (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–149). Washington: American Psychological Association.

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion, 12,* 579–599. doi:10.1080/026999398379574

Bellezza, F. S., Greenwald, A. G., & Banaji, M. R. (1986). Words high and low in pleasantness as rated by male and female college students. *Behavior Research Methods, Instruments, & Computers, 18,* 299–303. doi:10.3758/BF03204403

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20,* 351–368. doi:10.1093/pan/mpr057

Bishop, G. F., Oldendick, R. W., & Tuchfarber, A. J. (1984). Interest in political campaigns: The influence of question order and electoral context. *Political Behavior, 6,* 159–169. doi:10.1007/BF01207760

Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001). Emotion and motivation II: Sex differences in picture processing. *Emotion, 1,* 300–319. doi:10.1037/1528-3542.1.3.300

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. doi:10.1016/0005-7916(94)90063-9

Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a Five-Factor Personality Inventory for use on the Internet. *European Journal of Psychological Assessment, 21,* 115–127. doi:10.1027/1015-5759.21.2.115

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. doi:10.1177/1745691610393980

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 173–191). New York: Guilford Press.

Cohen, N., Henik, A., & Mor, N. (2011). Can emotion modulate attention? Evidence for reciprocal links in the attentional network test. *Experimental Psychology, 58,* 171–179. doi:10.1027/1618-3169/a000083

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349–354. doi:10.1037/h0047358

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,* 468–477. doi: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,* 90–101.

Delplanque, S., N'diaye, K., Scherer, K., & Grandjean, D. (2007). Spatial frequencies or emotional effects? *Journal of Neuroscience Methods, 165,* 144–150. doi:10.1016/j.jneumeth.2007.05.030

Drace, S., Efendic, E., Kusturica, M., & Landzo, L. (2013). Cross-cultural validation of the "International Affective Picture System" (IAPS) on a sample from Bosnia and Herzegovina. *Psihologija, 46,* 17–26. doi:10.2298/PSI1301017D

Dufey, M., Fernandez, A. M., & Mayol, R. (2011). Adding support to cross-cultural emotional assessment: Validation of the International Affective Picture System in a Chilean sample. *Universitas Psychologica, 10,* 521–533.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6,* 169–200. doi:10.1080/02699939208411068

Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science, 18,* 1050–1057. doi:10.1111/j.1467-9280.2007.02024.x

Frost, R. O., Tolin, D. F., Steketee, G., Fitch, K. E., & Selbo-Bruns, A. (2009). Excessive acquisition in hoarding. *Journal of Anxiety Disorders, 23,* 632–639. doi:10.1016/j.janxdis.2009.01.013

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*(3), 385–400. doi:10.1016/0191-8869(86)90014-0

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,* 512–521. doi:10.3758/BRM.40.2.512

Hajcak, G., & Dennis, T. A. (2009). Brain potentials during affective picture processing in children. *Biological Psychology, 80,* 333–338. doi:10.1016/j.biopsycho.2008.11.006

Hodes, R. L., Cook, E. W., & Lang, P. J. (1985). Individual differences in autonomic response: Conditioned association or conditioned fear? *Psychophysiology, 22,* 545–560. doi:10.1111/j.1469-8986.1985.tb01649.x

Huber, G. A., Hill, S. J., & Lenz, G. S. (2012). Sources of bias in retrospective decision making: Experimental evidence on voters' limitations in controlling incumbents. *The American Political Science Review, 106,* 720–741. doi:10.1017/S0003055412000391

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion–cognition relations. *Psychological Review, 99,* 561–565. doi:10.1037/0033-295X.99.3.561

Krantz, J. H. (2015). Psychological research on the net [Webpage]. Retrieved July 22, 2015, from http://psych.hanover.edu/Research/exponnet.html

Kraut, R. E., Olson, J. S., Banaji, M. R., Bruckman, A. S., Cohen, J. M., & Couper, M. P. (2004). Psychological research online: Report of the Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American Psychologist, 59,* 105–117. doi:10.1037/0003-066X.59.2.105

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin, 139,* 917–940. doi:10.1037/a0030811

Kveraga, K., Boshyan, J., Adams, R. B., Jr., Mote, J., Betz, N., Ward, N., … Barrett, L. F. (2015). If it bleeds, it leads: Separating threat from mere negativity. *Social Cognitive and Affective Neuroscience, 10,* 28–35. doi:10.1093/scan/nsu007

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist, 50,* 372–385. doi:10.1037/0003-066X.50.5.372

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97,* 377–395. doi:10.1037/0033-295X.97.3.377

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual (Technical Report A-8)*. Gainesville: University of Florida, Center for Research in Psychophysiology.

Lasaitis, C., Ribeiro, R. L., & Bueno, O. F. A. (2008). Brazilian norms for the International Affective Picture System (IAPS): Comparison of the affective ratings for new stimuli between Brazilian and North-American subjects. *Jornal Brasileiro de Psiquiatria, 57,* 270–275. doi:10.1590/S0047-20852008000400008

Lau, R. R., Sears, D. O., & Jessor, T. (1990). Fact or artifact revisited: Survey instrument effects and pocketbook politics. *Political Behavior, 12,* 217–242. doi:10.1007/BF00992334

Libkuman, T. M., Otani, H., Kern, R., Viger, S. G., & Novak, N. (2007). Multidimensional normative ratings for the International Affective Picture System. *Behavior Research Methods, 39,* 326–334. doi:10.3758/BF03193164

Lohani, M., Gupta, R., & Srinivasan, N. (2013). Cross-cultural evaluation of the International Affective Picture System on an Indian sample. *Psychological Studies, 58,* 233–241. doi:10.1007/s12646-013-0196-8

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces (KDEF) [CD ROM]*. Solna, Sweden: Karolinska Institutet, Department of Clinical Neuroscience, Psychology section.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19–40. doi:10.1037/1082-989X.7.1.19

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44,* 1–23. doi:10.3758/s13428-011-0124-6

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the international affective picture system. *Behavior Research Methods, 37,* 626–630. doi:10.3758/BF03192732

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). Opinion: The neural basis of human moral cognition. *Nature Reviews Neuroscience, 6,* 799–809. doi:10.1038/nrn1768

Moriguchi, Y., Negreira, A., Weierich, M., Dautoff, R., Dickerson, B. C., Wright, C. I., & Barrett, L. F. (2011). Differential hemodynamic response in affective circuitry with aging: An fMRI study of novelty, valence, and arousal. *Journal of Cognitive Neuroscience, 23,* 1027–1041. doi:10.1162/jocn.2010.21527

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259. doi:10.1037/0033-295X.84.3.231

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6,* 101–115. doi:10.1037/1089-2699.6.1.101

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5,* 411–419.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598–609. doi:10.1037/0022-3514.46.3.598

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89,* 277–293. doi:10.1037/0022-3514.89.3.277

Pequegnat, W., Rosser, B. R. S., Bowen, A. M., Bull, S. S., DiClemente, R. J., Bockting, W. O., … Zimmerman, R. (2007). Conducting Internet-based HIV/STD prevention survey research: Considerations in design and evaluation. *AIDS and Behavior, 11,* 505–521. doi:10.1007/s10461-006-9172-9

Peterson, C., Park, N., & Seligman, M. E. P. (2005). Orientations to happiness and life satisfaction: The full life versus the empty life. *Journal of Happiness Studies, 6,* 25–41. doi:10.1007/s10902-004-1278-z

Quigley, K. S., Lindquist, K. A., & Barrett, L. F. (2014). Inducing and measuring emotion and affect: Tips, tricks, and secrets. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 220–252). New York: Cambridge University Press.

Rand, D. G. (2012). The promise of Mechanical Turk How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology, 299,* 172–179. doi:10.1016/j.jtbi.2011.03.004

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178. doi:10.1037/h0077714

Sabatinelli, D., Flaisch, T., Bradley, M. M., Fitzsimmons, J. R., & Lang, P. J. (2004). Affective picture perception: Gender differences in visual cortex? *NeuroReport, 15,* 1109–1112.

Schuman, H., Kalton, G., & Ludwig, J. (1983). Context and contiguity in survey questionnaires. *Public Opinion Quarterly, 47,* 112–115.

Seo, M.-G., & Barrett, L. F. (2007). Being emotional during decision making—Good or bad? An empirical investigation. *Academy of Management Journal, 50,* 923–940. doi:10.5465/AMJ.2007.26279217

Weierich, M. R., Wright, C. I., Negreira, A., Dickerson, B. C., & Barrett, L. F. (2010). Novelty as a dimension in the affective brain. *NeuroImage, 49,* 2871–2878. doi:10.1016/j.neuroimage.2009.09.047

Wessa, M., & Flor, H. (2007). Failure of extinction of fear responses in posttraumatic stress disorder: Evidence from second-order conditioning. *American Journal of Psychiatry, 164,* 1684–1692. doi:10.1176/appi.ajp.2007.07030525

Wilcox, N., & Wlezien, C. (1993). The contamination of responses to survey items: Economic perceptions and political judgments. *Political Analysis, 5,* 181–213.

Wrase, J., Klein, S., Gruesser, S. M., Hermann, D., Flor, H., Mann, K., … Heinz, A. (2003). Gender differences in the processing of standardized emotional visual stimuli in humans: A functional magnetic resonance imaging study. *Neuroscience Letters, 348,* 41–45.