The Complex Affective Scene Set (COMPASS): Solving the Social Content Problem in Affective Visual Stimulus Sets

Mariann R. Weierich
University of Nevada, Reno; Hunter College; The City University of New York

Olena Kleshchova
The City University of New York

Jenna Rieder K. Rieder
The City University of New York

Danielle M. Reilly
The University of Nevada, Reno, The City University of New York, Thomas Jefferson University

Follow this and additional works at: https://jdc.jefferson.edu/jchsfp

Let us know how access to this document benefits you

ORIGINAL RESEARCH REPORT

The Complex Affective Scene Set (COMPASS): Solving the Social Content Problem in Affective Visual Stimulus Sets

Mariann R. Weierich[*,†,‡], Olena Kleshchova[†,‡], Jenna K. Rieder[†,‡,§] and Danielle M. Reilly[†]

Social information, including faces and human bodies, holds special status in visual perception generally, and in visual processing of complex arrays such as real-world scenes specifically. To date, unbalanced representation of social compared with nonsocial information in affective stimulus sets has limited the clear determination of effects as attributable to, or independent of, social content. We present the Complex Affective Scene Set (COMPASS), a set of 150 social and 150 nonsocial naturalistic affective scenes that are balanced across valence and arousal dimensions. Participants (n = 847) rated valence and arousal for each scene. The normative ratings for the 300 images together, and separately by social content, show the canonical boomerang shape that confirms coverage of much of the affective circumplex. COMPASS adds uniquely to existing visual stimulus sets by balancing social content across affect dimensions, thereby eliminating a potentially major confound across affect categories (i.e., combinations of valence and arousal). The robust special status of social information persisted even after balancing of affect categories and was observed in slower rating response times for social versus nonsocial stimuli. The COMPASS images also match the complexity of real-world environments by incorporating stimulus competition within each scene. Together, these attributes facilitate the use of the stimulus set in particular for disambiguating the effects of affect and social content for a range of research questions and populations.

1. Introduction

In daily life, people encounter vast arrays of stimuli that compete for visual attention and cognitive resources. Selection of some types of information over others is the end result of a complicated algorithmic process that integrates immediate perceptual salience with the viewer's prior experience and current state. The phenomena underlying and influencing selection are the subject of research focused on how affective information is processed in typical daily life and in more extreme circumstances. From the simplest and most elegant behavioral tasks to the rapidly developing technology of brain imaging, a common essential element is the use of visual stimuli that allow valid measurement of the mechanisms of interest. It follows that stimuli that correspond well to the visual and affective complexity of the physical world, while controlling for the attributes that are most likely to confound results, are necessary for investigation of the interaction of affect with visual selection. Within this framework, visual social (i.e., human) content, such as human faces and bodies, has special status in the competition for prioritized processing. For example, faces or bodies are fixated first in naturalistic scenes (e.g., Fletcher-Watson, Findlay, Leekam, & Benson, 2008; Rosler, End, & Gamer, 2017), faces attract gaze in experimental tasks even at a cost (Cerf, Frady, & Koch, 2009), and only responses to social stimuli reflected the effects of anhedonia in people with schizophrenia compared to healthy controls (e.g., Bodapati & Herbener, 2014). However, unbalanced representation of social and nonsocial information in affective stimulus sets has limited the clear determination of effects as attributable to, or independent of, social content. For example, neutral social images are underrepresented in some sets, which can result in lower power for that category, or the need to repeat images, which, given novelty is a factor in affective processing, can weaken the magnitude of the results. We developed the Complex Affective Scene Set (COMPASS), a novel set of social and nonsocial naturalistic affective scenes, or combinations of valence and arousal, to fill a major gap among existing sets by specifically balancing social content across affect categories, and also by incorporating visual complexity and human diversity.

* Department of Psychology, The University of Nevada Reno, Reno, NV, US

† Department of Psychology, Hunter College, The City University of New York, New York, NY, US

‡ The Graduate Center, The City University of New York, New York, NY, US

§ College of Humanities and Sciences, Thomas Jefferson University, Philadelphia, PA, US

Corresponding author: Mariann R. Weierich (mweierich@unr.edu)

A growing number of visual stimulus sets are available for use in studies of visual and affective processing. Of these, the most well known and well characterized is the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008), which includes images that vary along the dimensions of valence and arousal and span most of the affective space. More recently, the Open Affective Standardized Image Set (OASIS; Kurdi, Lozano, & Banaji, 2017) was introduced as a more current alternative to the IAPS. Both sets include social and nonsocial stimuli, although they are not balanced across affect dimensions. Other stimulus sets have been developed to address specific themes or content. For example, the Geneva Affective Picture Database (GAPED; Dan-Glauser & Scherer, 2011) predominantly includes unpleasant affective stimuli, such as images depicting violations of human and animal rights. The Nencki Affective Pictures System (NAPS; Marchewka, Żurawski, Jednoróg, & Grabowska, 2014) includes specific categories of affective images, as well as erotic (NAPS ERO; Wierzba et al., 2015) and fear-provoking (NAPS SFIP; Michałowski, Droździel, Matuszewski, Koziejowski, Jednoróg, & Marchewka, 2017) subsets. A number of well-developed sets include exclusively social stimuli, most of which are emotional faces (e.g., Karolinska Directed Emotional Face set; Lundqvist, Flykt., & Ohman, 1998; Montreal Set of Facial Displays of Emotion; Beaupre & Hess, 2005; NimStim; Tottenham et al., 2009; Pictures of Facial Affect; Ekman & Friesen, 1976; the Warsaw Set of Emotional Facial Expression Pictures; Olszanowski, Pochwatko, Kuklinski, Scibor-Rylski, Lewinski, & Ohme, 2015), and which offer the capacity to compare responses among human emotional facial expressions.

Depending on the research question, one or several of the previously published stimulus sets could be the most appropriate and useful. However, we suggest that our set represents a unique combination of attributes that make it especially relevant and useful for questions concerning processing of complex daily environments, while controlling for several potentially confounding factors. For example, although, as noted, some sets include both social and nonsocial stimuli, none balance these attributes within affect categories, thereby facilitating direct comparison without the problem of unequal stimulus numbers per category. To meet our objective of creating a set of naturalistic affective scenes that balances social content, we selected images that vary along several salient dimensions and attributes.

1.1. Affect dimensions
COMPASS images vary along two well-established dimensions of affect: valence (unpleasant to pleasant) and arousal (low to high activation; e.g., Barrett, 2006). The COMPASS set is consistent with other affective image sets in that the images fall into six broad combinations of arousal and valence (higher arousal unpleasant, higher arousal pleasant, moderate arousal unpleasant, moderate arousal pleasant, moderate arousal neutral, and lower arousal neutral) that are represented by the boomerang shape of the canonical affective circumplex (e.g., Barrett, 2006; Posner, Russell, & Peterson, 2005). Although we categorize the images in this way to represent much of the

affective space, we also note that these two dimensions do not have objective cutoffs between levels, and the categories thus should be used as a helpful guide rather than an absolute evaluation of image content. Also, as noted earlier, because many questions in affective science center on typical daily affective experience, rather than representing affectively extreme experiences, one of our objectives was to represent a range of affective experiences that people typically encounter in daily life. Given this objective, the COMPASS set does not include affectively extreme stimuli, such as strongly aversive (e.g., mutilated bodies) or strongly erotic (e.g., couple engaged in sexual activity) content.

1.2. Social content
COMPASS scenes are balanced by social content, which we define as representation of humans. Social scenes include clearly discernible people as at least one of the most salient focal points, and nonsocial scenes either do not include people or include people as non-salient percepts (e.g., smaller figures in the background). Social content is a crucial attribute in the context of affective evaluations, because visual and neural processing of affective information differs between social and nonsocial information. For example, pupillometry and eye-tracking studies show that social information preferentially captures visual attention compared to nonsocial information (e.g., Fitzgerald, 1968), and compared to low-level salient features such as contrast and luminance (e.g., End & Gamer, 2017). In addition, the neural regions engaged in affective evaluation of social information differ from those of nonsocial information (e.g., Harris, McClure, Van den Bos, Cohen, & Fiske, 2007).

Within existing affective stimulus sets that include social and nonsocial stimuli, the inclusion of social content often is confounded with arousal or valence (e.g., Colden, Bruder, & Manstead, 2008). For example, social images (e.g., two people hugging or arguing) have more extreme pleasant or unpleasant valence ratings in comparison with nonsocial images (e.g., garbage on the street). In a subset of the IAPS images, images with humans were rated as more arousing, and more unpleasant or pleasant than images with inanimate content. Images with humans also were rated as more unpleasant or less pleasant than images with non-human animal content (Colden et al., 2008). Further, differential processing of images with human content is not limited to downstream top-down processing such as explicit ratings; emotional images were associated with enhanced initial allocation of attention only when the images contained humans (Löw, Bradley, & Lang, 2013). In addition, social images often have greater visual complexity in comparison with nonsocial images, which often consist of simple single objects, and neutral images are more likely to include single non-human objects than social content. COMPASS has equal numbers of social and nonsocial scenes within each affective category. As a result, our stimulus set controls for the potential confound of human content with affect, and similarly can be used to disambiguate the effects of affect and social content, by comparison of data from social and non-social images.

1.3. Stimulus competition

The COMPASS set was designed specifically to represent naturalistic visual arrays, or scenes, rather than discrete single objects or people. We defined a "complex scene" as an image that includes at least two salient points of interest, such that the salient content competes for visual attention. This characteristic is especially important for research questions and methods that require stimulus competition for valid measurement of specific visual mechanisms such as initial allocation of attention to or disengagement of attention from affective content (e.g., Desimone & Duncan, 1995). For example, a valid test of preferential allocation of attention to a specific type of information within a single stimulus array requires that there are alternative targets of attention within the array. The COMPASS set is thus especially useful and appropriate for paradigms that assess covert or overt (e.g., eye tracking) allocation of attention that favors some visual content over other visual content.

1.4. Representation of human diversity

A final distinguishing feature of COMPASS is the representation in the images of people from a variety of racial, ethnic and cultural backgrounds. The impetus for such inclusion was the parallel with the extremely diverse daily environment of the geographic location of our lab in New York City. For example, for studies testing neural or endocrine responses to naturalistic affective information in trauma-exposed participants, it is important that the images reflect the participants' daily experiences. As a result of the inclusion of diverse people and settings, COMPASS can be more reliably applied in a variety of subject populations.

1.5. The influence of sex/gender

There are well-known and documented sex/gender differences in affective processing of visual stimuli (e.g., Andreano, Dickerson, & Barrett, 2014; Cahill, 2006; Soares, Pinheiro, Costa, Frade, Comesana, & Pureza, 2015; Wrase, Klein, Gruesser, Hermann, Flor et al., 2003) and these differences can be particularly pronounced in processing of human faces (e.g., Proverbio, 2016). For these reasons, affective stimulus sets commonly include both overall norms and subdivision by sex/gender. Generally, but not exclusively, the evidence supports that female participants show greater neural responses to unpleasant, higher arousal stimuli, and rate them accordingly, whereas male participants show greater neural responses to pleasant, higher arousal stimuli and rate them accordingly. One potential explanation for differences in affective processing is biology, such as the effects of sex hormones between groups, but also sex hormone differences within women due to fluctuations across the menstrual cycle. Additional explanations implicate gender, whereby societal expectations and experiences of men and women might predispose them to respond differently to affective stimuli. For the purpose of our stimulus set development and norming, we do not make a specific claim regarding the individual or interacting roles of biology or environment, however we did anticipate sex/gender differences in affective ratings, consistent with the preponderance of the literature.

1.6. Stimulus set development goals

We present the Complex Affective Scene Set (COMPASS), a normed set of 300 complex, affectively balanced, naturalistic scenes that include representation of cultural, racial, and ethnic diversity, and that represent visual arrays that are reasonably likely to be encountered in daily life. These images were selected to accomplish the overall goal of creating an affective scene set that balances social (human) and non-social content and covers the canonical affective space, or the combinations of valence and arousal (i.e., affect categories). Within that broad goal, our first aim was to develop a set of complex scenes that approximate daily life experiences and therefore can be used to estimate the magnitude of typical everyday affective responses. In the interest of capturing affective processing of typical, everyday scenes, our set does not include the valence and arousal extremes such as mutilated bodies or strongly erotic content. Our second aim was to distinguish between the influences of affect category and social content on valence and arousal ratings by including equal numbers of social and nonsocial scenes within each affective category. Because it is not possible to entirely remove the affective qualities of human images, this balance is the best strategy to facilitate direct comparisons between social and non-social affective content.

Along with the major aims of the development of the stimulus set, for which we predicted only that the end result would cover the affective space as comprehensively as possible while also balancing relevant attributes, we had two evidence-driven hypotheses. First, given the known special status of social information over non-social information, although during each phases of stimulus set development we sought to achieve equivalence in affective ratings, we hypothesized that the social stimuli nonetheless would continue to be processed differently. Because we used participant ratings in early development phases to select stimuli with approximate affective equivalence for the final stimulus set, comparison of subsequent affect ratings between social and non-social scenes would be circular. Instead, to address this question we conceptualized response time for initial affective ratings as a proxy for processing time. We hypothesized that longer response times for social versus non-social ratings would provide an index of the persistent special status of social content, even when the affective ratings themselves were roughly equivalent. In addition, given the extensively documented sex/gender differences in affective processing, we hypothesized that our data would replicate the previous pattern of sex/gender differences in ratings; men would rate pleasant images as more pleasant and more arousing than would women, whereas women would rate unpleasant images as more unpleasant and more arousing than would men.

2. Method

2.1. Participants

The COMPASS scenes were rated by 847 participants (71% women, 29% men); age $M = 20.5$, $SD = 4.6$, range = 18–53; **Table 1**). An a priori power analysis showed that power to detect a small effect with alpha at .05 and power at .80 would require 230 participants per group.

Table 1: Participants.

Variable	Statistics
Women, n (%)[a]	597 (70.5%)
Men, n (%)	245 (28.9%)
Age in years, M (SD), range	20.5 (4.6), 18–53
Race/ethnicity, n (%)[b]	
Asian	328 (38.7%)
Black	71 (8.4%)
Latinx	159 (18.8%)
White	198 (23.4%)
Other	22 (2.6%)
Multiple	59 (7.0%)
Born outside of the US, n (%)	323 (38.1%)
Years in the US, M (SD)[c]	11.7 (6.4)
English as the 1st language, n (%)	483 (57.0%)
Speak additional language(s), n (%)	633 (74.7%)

[a] Two participants (0.2%) self-reported as transgender men, and three participants (0.4%) did not indicate their gender.
[b] Ten participants (1.2%) did not indicate their ethnicity.
[c] For participants who were not born in the US.

It is essential to be adequately powered to detect even small effects that without detection could invalidate the stimulus set and/or tests of sex differences. Given that recruitment population was known to have a female:male ratio of approximately 2:1, we set a recruitment target to fill the male participant n, with the understanding that open enrollment would result in twice as many female participants. Participants were recruited from a large, non-residential urban university. This student population is extremely ethnically and racially diverse and includes a high percentage of non-traditional students. About one third of the participants (38%) were born outside of the US. For these participants, the mean number of years in the US was M = 11.7 (SD = 6.4). English was the first language for 57% of the participants, and 75% also reported additional languages.

2.2. Stimuli
2.2.1. Image selection criteria
Stimuli were selected from non-copyrighted images on the internet and photographs taken by lab members. We selected full-color images of complex scenes with multiple focal points and excluded images of single objects. As our goal was to create a set of naturalistic scenes, we excluded pictures that appeared to be posed or digitally enhanced, as well as pictures of famous people or places. For the same reason, we excluded images at the extreme ends of the arousal dimension, such as those depicting extreme violence or openly erotic content.

2.2.2. Image specifications
We resized all images to 500 × 667 pixels by adding horizontal and/or vertical black bars where necessary. Because written words capture visual attention, we blurred visible logotypes or written words using Adobe Photoshop. For each scene, we calculated mean luminance as the average pixel value of the gray-scale image, and contrast as the standard deviation across all pixels of the gray-scale image (Bex & Makous, 2002). Most COMPASS scenes (276, 92%) are in landscape orientation.

2.2.3. Scene categories
2.2.3.1. Affective dimensions
The final COMPASS set includes 100 unpleasant (50 each higher and moderate arousal), 100 neutral (50 each moderate and lower arousal), and 100 pleasant (50 each higher and moderate arousal) scenes. Because pleasant and unpleasant information typically is rated as more arousing than neutral information (e.g., Libkuman, Otani, Kern, Viger, & Novak, 2007), pleasant and unpleasant scenes ranged from moderate to higher arousal, whereas neutral scenes ranged from lower to moderate arousal.

2.2.3.2. Social content
Scenes that included clearly discernible people in the foreground or as one of the primary focal points were classified as social (67% of all social scenes contain clearly visible faces). COMPASS includes 150 social and 150 nonsocial scenes (within each category: 25 each higher arousal unpleasant, moderate arousal unpleasant, higher arousal pleasant, moderate arousal pleasant, moderate arousal neutral, lower arousal neutral).

2.2.3.3. Human diversity
The social scenes in COMPASS include representation of racial and ethnic diversity. Because human diversity is an important attribute but not a primary set design factor, the race/ethnicity category is not balanced by number of scenes. Thirty-nine percent of the social scenes include White people, 19% mixed ethnic groups, 16% people of unclear ethnicity, 13% Asian people, 10% Black people, and 3% Latinx people. Forty-one percent of the social scenes include both male and female people, 33% only male people, 20% only female people, and in 7% the gender is unclear (e.g., face is not discernible). Most social scenes (81%) include multiple people.

2.2.3.4. Additional scene attributes
Most scenes (81%) are outdoor scenes, 62 scenes (21%) include animals, and 25 scenes (8%) depict some kind of natural disaster.

2.3. Procedure
2.3.1. Set development
We report only data from the final normed stimulus set, however the stimulus selection and norming procedure had four phases. In phases 1–3, we iteratively developed the final set of 300 stimuli (please see the Supplemental Information for additional detail regarding the first 3 phases of scene selection). In each phase, participants (phase 1 n = 496, phase 2 n = 486, and phase 3 n = 723) rated valence and arousal for a set of scenes. At the end of each of the first three phases, we selected the scenes whose ratings were consistent with the assigned valence and arousal categories and included them in the next

phase. We also discarded scenes that showed a bimodal valence distribution, thus indicating affective ambiguity, and replaced them with new scenes. In the fourth phase (the data reported in this paper), all 847 participants rated the final set of 300 images.

2.3.2. Study procedure

Following consent, a researcher explained the procedure. Each participant then completed the computer rating task and a questionnaire. At the end of the study, participants were debriefed and granted course credit. The study protocol was approved by the Institutional Review Board and carried out in accordance with Standard 8 of the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct.

2.3.3. Rating task

Each participant was seated 60 cm from the computer screen. To control for room illumination and prevent screen glare, overhead lights were turned off and a small 60-watt floor lamp provided the only light source besides the screen. The computer task was administered on a Dell PC with a 19" (1280 × 1024 resolution) no-glare display using E-Prime software.

Participants were informed that the purpose of the study was to learn how people respond to pictures that represent different settings and events, and that they would be viewing and rating 300 pictures (see Text S1 for task instructions). Participants were instructed to provide two ratings for each scene according to their initial reactions. The first rating was for how unpleasant or pleasant the scene made them feel (1 = unpleasant to 9 = pleasant), and the second rating was for how arousing or activating they found the scene to be (1 = low arousal to 9 = high arousal). Because the primary goal of the ratings procedure was to create a stimulus set that had social representation in affect categories that have infrequent social representation in other stimulus sets (e.g., neutral), we prioritized valence ratings rather than counterbalancing the response order. Participants were also informed that the task was not timed. After each participant completed three practice trials, the researcher left the room.

Participants rated four blocks of 75 images, with an opportunity to rest and stretch between blocks. The order of blocks and the order of within-block images was randomized for each participant. For each trial, an image was presented on the computer screen. The participant pressed the spacebar to advance to the first response screen, which showed a 9-point rating scale for valence (unpleasant to pleasant). After the participant entered a valence rating using the keyboard, a 9-point rating scale for arousal (low to high) appeared on the screen. After the participant entered an arousal rating, the next image appeared. Most participants completed the ratings task within 30–40 minutes.

2.3.4. Questionnaire

The demographics questionnaire included items about gender, age, race/ethnicity, birthplace, number of years in the US, parents' birthplaces, and first language. The latter items were included to control for known cultural influences on affective ratings.

3. Results

Data were analyzed using Matlab R2017a and SPSS (24). We excluded trials with reaction times slower than 4000 ms or faster than 150 ms, due to the unreliability of very fast or very slow response times for rating tasks. This filter resulted in the exclusion of 18619 (7%) individual valence ratings and 37861 (15%) individual arousal ratings.[1] After exclusions, each image retained valence ratings by an average of 756 participants ($SD = 13$, range 698–786) and arousal ratings by an average of 692 participants ($SD = 17$, range 643–737). We have reported all manipulations, measures, and exclusions.

3.1. Affective scene categorization

Image names reflect their respective valence, arousal, and social content categories (e.g., NeutLowSoc = neutral valence, lower arousal, social scene). We note that the image names utilize "Negative" and "Positive" rather than the more accurate "Unpleasant" and "Pleasant" due to easier readability of the former when abbreviated. Similarly, we use "Mid" in the image names as a proxy abbreviation for "Moderate".

3.2. Summary statistics of scene-wise valence and arousal ratings

The average valence rating across all scenes was 4.87 ($SD = 1.88$). The lowest (most unpleasant) mean valence rating of 1.36 ($SD = 1.02$) was for scene NegHighSoc_22 depicting childhood bullying. The highest (most pleasant) mean valence rating of 8.25 ($SD = 1.24$) was for scene PosHighNonsoc_1 (tropical island). The average scene-wise standard deviation of valence ratings was 1.63 ($SD = 0.26$). Scene NegHighSoc_12 (man assaulting a woman) had the smallest standard deviation of valence ratings ($M = 1.44$, $SD = 0.96$). Scene NegMidSoc_15 (crying man hugging dog) had the largest standard deviation of valence ratings ($M = 5.68$, $SD = 2.56$).

The average arousal rating across all scenes was 4.41 ($SD = 0.98$). Scene NeutMidNonsoc_1 (a parking lot) had the lowest mean arousal rating and the smallest standard deviation of arousal ratings ($M = 2.27$, $SD = 1.74$). Scene NegHighSoc_22 (childhood bullying) had the highest mean arousal rating of 7.05 ($SD = 2.70$). The average scene-wise standard deviation of arousal ratings was 2.46 ($SD = 0.23$). Scene NegHighNonsoc_7 (severed buffalo heads) had the largest standard deviation of arousal ratings ($M = 6.41$, $SD = 2.91$).

Figure 1 shows distributions of scene-wise means and standard deviations of valence and arousal ratings. The distribution of mean valence ratings is bimodal with one peak near 2 and the other one near 6 (skewness = –0.22). The distribution of mean arousal ratings approaches normality (skewness = 0.14). The distributions of scene-wise standard deviations of valence (skewness = 0.36) and arousal (skewness = –0.47) ratings also approach normality, although the standard deviations of arousal ratings are larger than the standard deviations of valence ratings. Summary statistics for COMPASS and IAPS norms
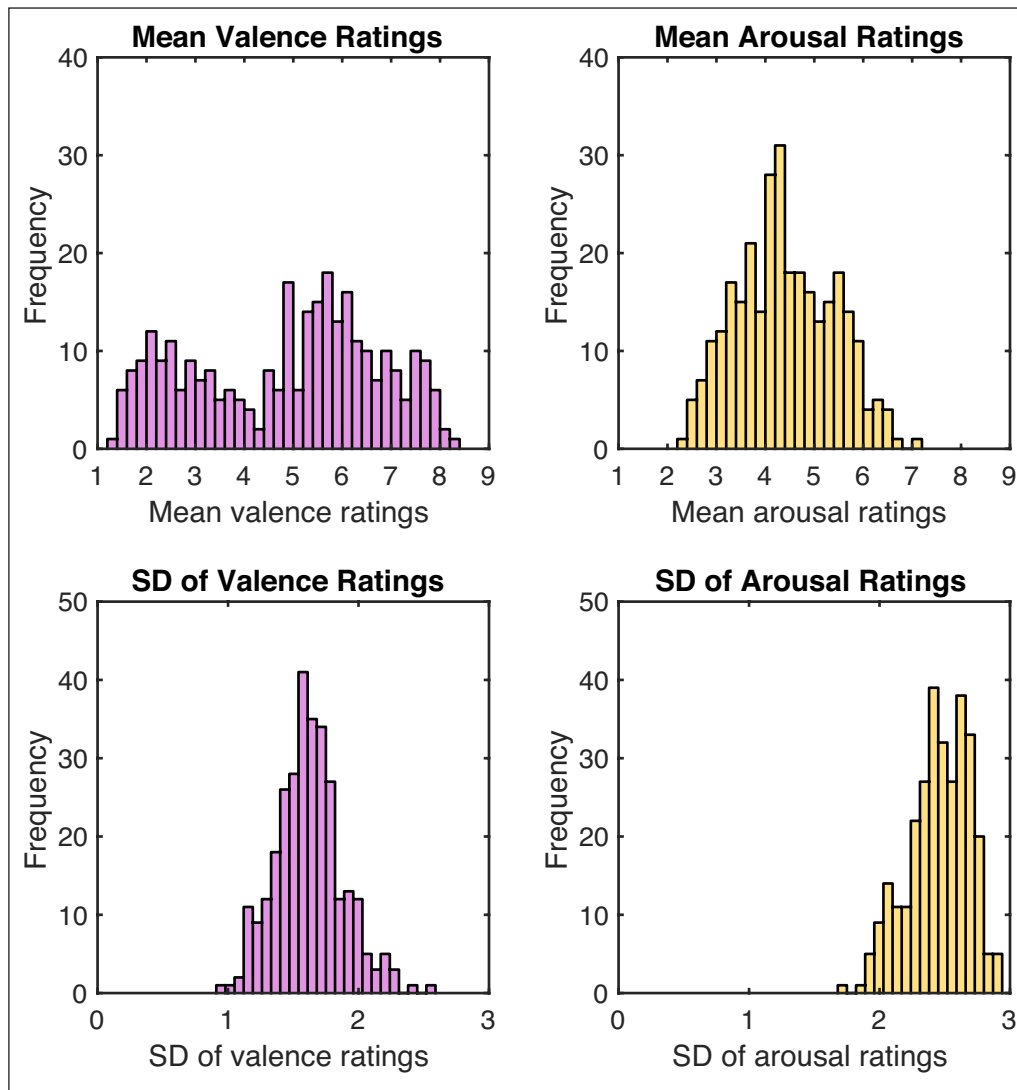
Figure 1: Distributions of scene-wise means and standard deviations of valence and arousal ratings.

by affective category (from Grühn & Scheibe, 2008) are presented in Table S1.

3.3. Valence and arousal ratings

Consistent with other stimulus sets (e.g., Libkuman et al., 2007), COMPASS valence and arousal ratings showed a boomerang-shaped relationship, such that scenes at the extremes of the valence dimension were rated as more arousing than scenes in the middle of the dimension (**Figure 2**). Bivariate distributions of valence and arousal ratings for each image are presented in Figure S2. Also consistent with previous reports (e.g., Kurdi et al., 2017), there was an M-shaped relationship between the means and standard deviations of valence ratings (**Figure 3**). This result indicates that standard deviations tend to be smaller for scenes with mean ratings closer to the three anchor points (1 = unpleasant, 5 = neutral, 9 = pleasant), and larger for scenes with valence means between the anchor points. In contrast, there was a linear relationship between the means and standard deviations of arousal ratings (**Figure 3**), indicating greater variability in arousal ratings for higher arousal scenes.

3.4. Affect ratings by scene category

3.4.1. Valence

Mean valence ratings by scene category are presented in **Table 2**. For each scene, we calculated mean valence ratings across all participants. We then calculated mean ratings across all the scenes within each Valence × Arousal × Social Content Category and conducted a Valence Category × Arousal Category × Social Content repeated-measures ANOVA with valence ratings as the dependent variable. There was a main effect of Valence Category ($F(2,1692) = 7595$, $p < .001$, $\eta_p^2 = .90$). Post-hoc pairwise comparisons with Bonferroni correction showed that pleasant scenes had higher (more pleasant) valence ratings than neutral scenes, which had higher valence ratings than unpleasant scenes (all $ps < .001$). There was a main effect of Arousal Category ($F(1,846) = 2565$, $p < .001$, $\eta_p^2 = .75$), such that lower arousal scenes had higher (more pleasant) valence ratings than higher arousal scenes. There was also a main effect of Social Content ($F(1,846) = 560$, $p < .001$, $\eta_p^2 = .40$), such that nonsocial scenes had higher valence ratings than social scenes. Finally, there was a Valence Category × Arousal Category × Social Content interaction
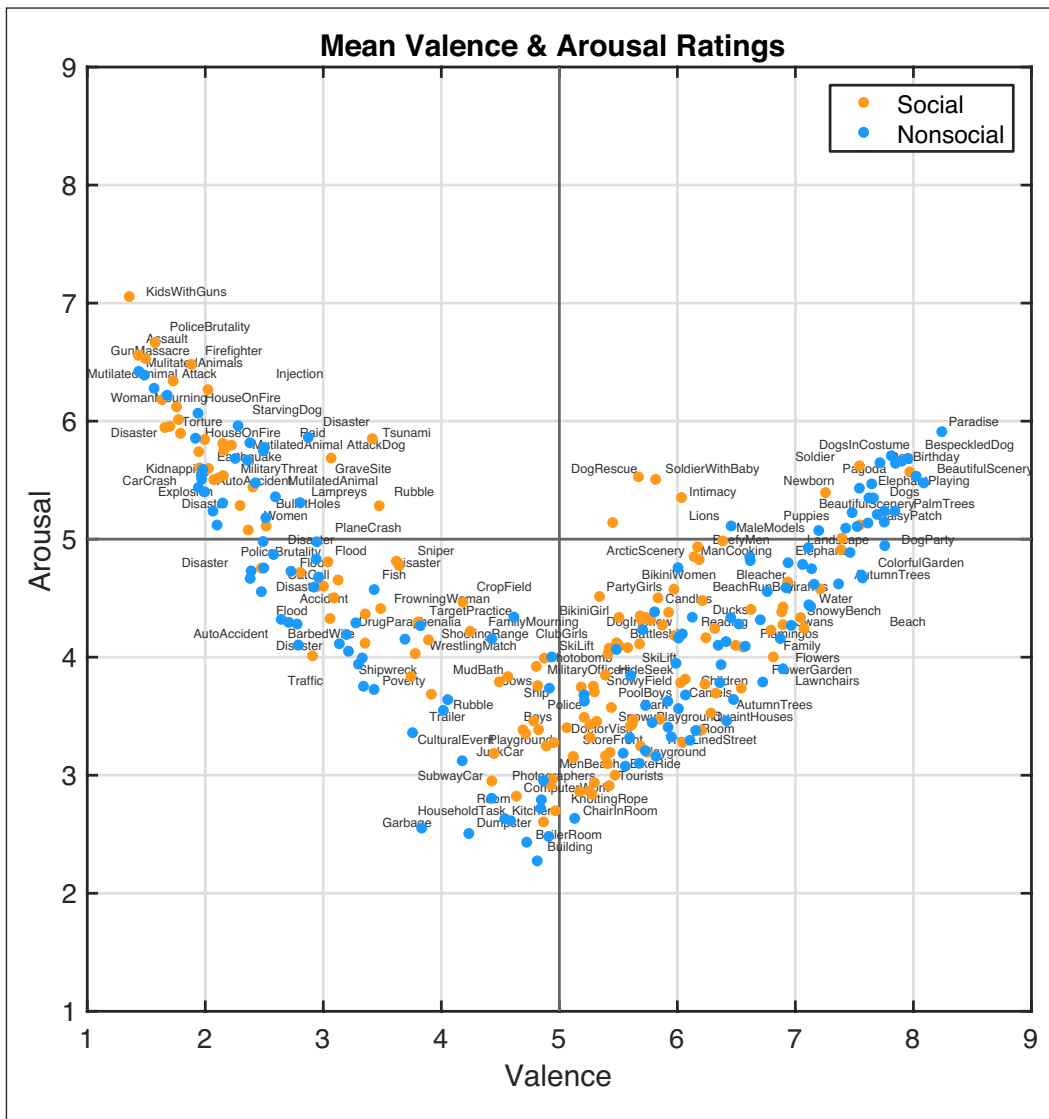
Figure 2: The relation between COMPASS valence (1 = unpleasant, 9 = pleasant) and arousal (1 = low, 9 = high) ratings.
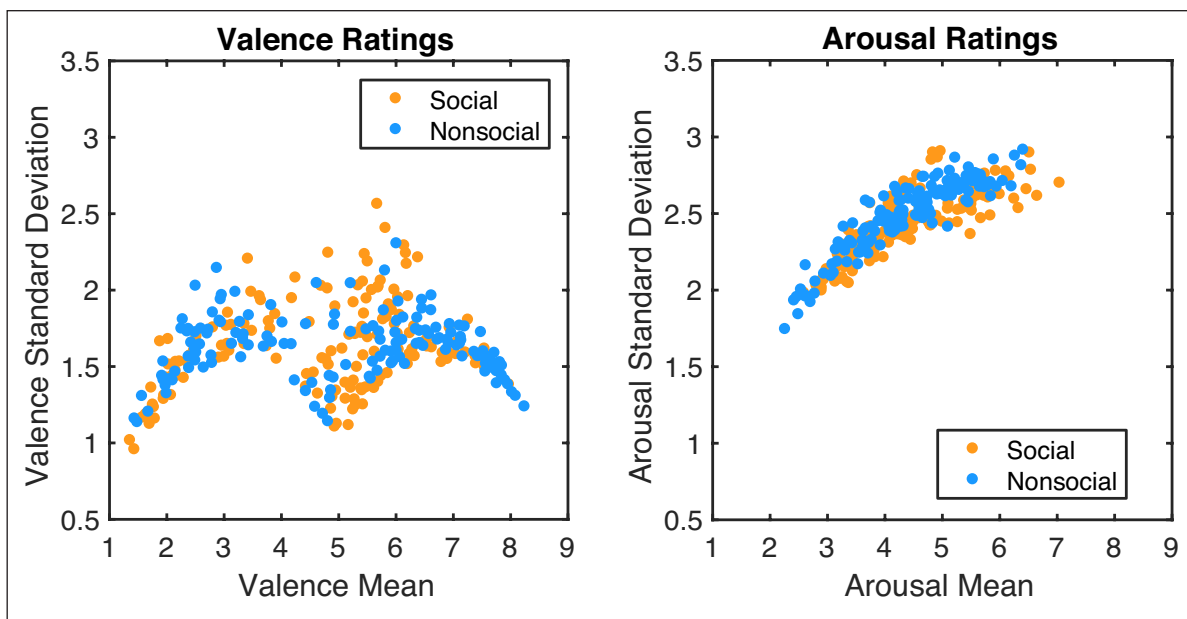

Figure 3: The relations between scene-wise means and standard deviations of COMPASS valence and arousal ratings.

Table 2: Valence and arousal ratings by assigned scene category.

Scene category	Valence ratings, M (SD)	Arousal ratings, M (SD)
Negative Higher Social	1.96 (1.45)	5.95 (2.68)
Negative Higher Nonsocial	2.24 (1.65)	5.49 (2.75)
Negative Moderate Social	3.29 (1.91)	4.68 (2.52)
Negative Moderate Nonsocial	2.96 (1.79)	4.58 (2.60)
Neutral Moderate Social	5.32 (1.64)	3.56 (2.27)
Neutral Moderate Nonsocial	5.06 (1.90)	3.52 (2.36)
Neutral Lower Social	5.37 (1.62)	3.34 (2.22)
Neutral Lower Nonsocial	5.60 (1.83)	3.51 (2.42)
Positive Higher Social	5.64 (2.05)	4.23 (2.64)
Positive Higher Nonsocial	7.47 (1.67)	5.15 (2.70)
Positive Moderate Social	6.64 (1.85)	4.46 (2.52)
Positive Moderate Nonsocial	6.85 (1.79)	4.45 (2.59)
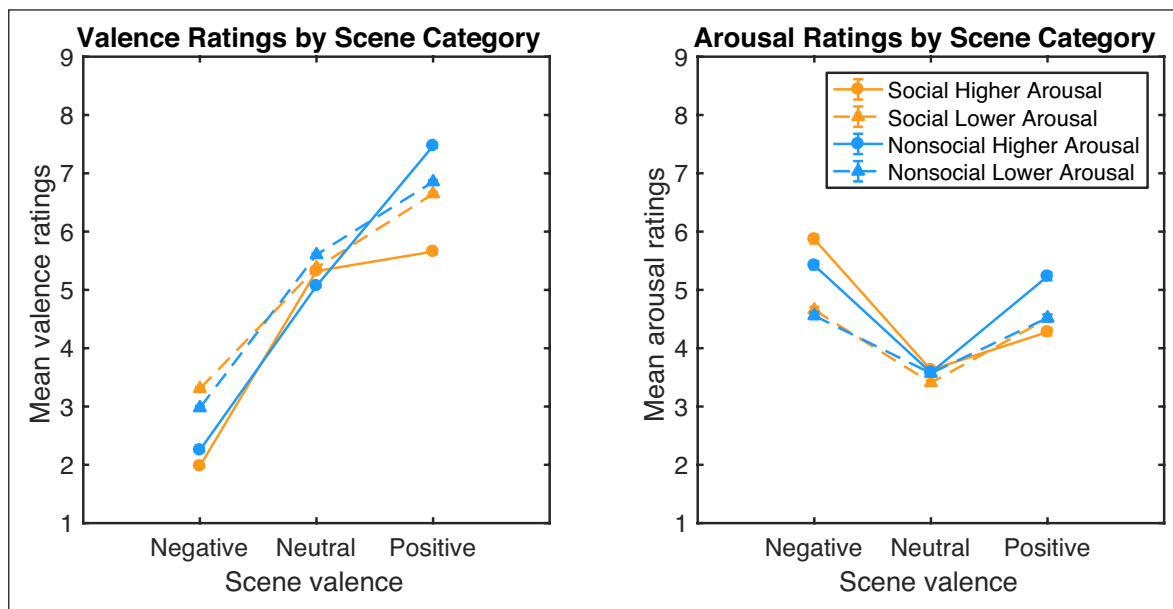
Figure 4: Mean valence and arousal ratings for each affective scene category (lower arousal = low to moderate; higher arousal = moderate to high).

(F(2,1692) = 1150, $p < .001$, $\eta_p^2 = .58$), driven by a larger effect of Social Content on higher arousal pleasant scenes compared to other categories. Higher arousal nonsocial pleasant scenes had higher valence ratings than higher arousal social pleasant scenes (see **Figure 4**).

3.4.2. Arousal

Mean arousal ratings by scene category are presented in **Table 2**. For each scene, we calculated mean arousal ratings across all participants. We then calculated mean ratings across all the scenes within each Valence × Arousal × Social Content Category and conducted a Valence Category × Arousal Category × Social Content repeated-measures ANOVA with arousal ratings as the dependent variable. There was a main effect of Valence Category (F(2,1692) = 454, $p < .001$, $\eta_p^2 = .35$). Post-hoc pairwise

comparisons using Bonferroni correction showed that unpleasant scenes had higher arousal ratings than pleasant scenes, which had higher arousal ratings than neutral scenes (all $ps < .001$). There was a main effect of Arousal Category (F(1,846) = 796, $p < .001$, $\eta_p^2 = .48$), such that higher arousal scenes had higher arousal ratings than lower arousal scenes. There was also a main effect of Social Content (F(1,846) = 24.9, $p < .001$, $\eta_p^2 = .03$), such that nonsocial scenes had higher arousal ratings than social scenes. Finally, there was a Valence Category × Arousal Category × Social Content interaction (F(2,1692) = 269, $p < .001$, $\eta_p^2 = .24$), driven by a greater effect of Social Content on higher arousal pleasant scenes, compared to other affective categories. Higher arousal nonsocial pleasant scenes were rated as more arousing than higher arousal social pleasant scenes (see **Figure 4**).

3.5. Participant gender and affect ratings

Given the well-documented gender differences in affective processing of visual stimuli (e.g., Cahill, 2006; Wrase, Klein, Gruesser, Hermann, Flor et al., 2003), we calculated scene-wise valence and arousal ratings for men and women separately (see **Table 3** and **Figure 5**).

We conducted a Valence Category × Arousal Category × Social Content repeated-measures ANOVA with Participant Gender as a between-subjects factor and valence ratings as the dependent variable. There was a main effect of
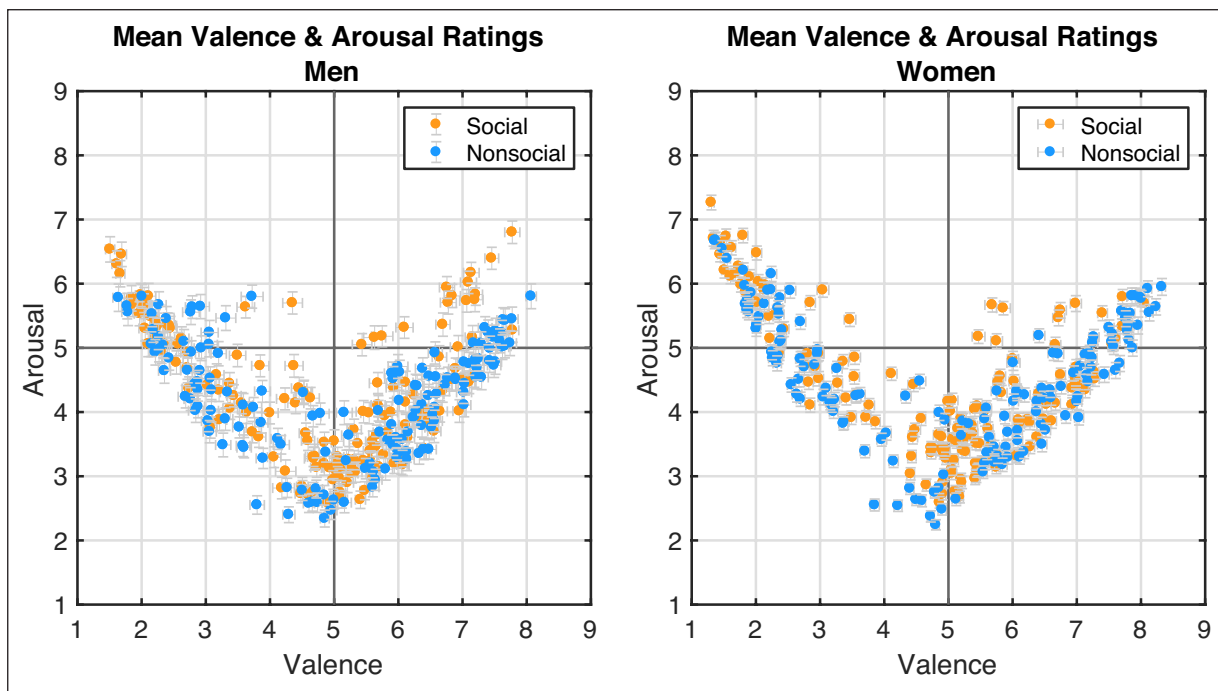
Participant Gender on valence ratings ($F(1,840) = 9.76$, $p = .002$, $\eta_p^2 = .01$), with men providing higher valence ratings on average than women. There was also a Valence Category × Arousal Category × Social Content × Participant Gender interaction ($F(2,1680) = 65.1$, $p < .001$, $\eta_p^2 = .07$): women rated nonsocial higher arousal pleasant scenes as more pleasant than did men and social higher arousal pleasant scenes as less pleasant than did men (**Figure 6**).

We also conducted a Valence Category × Arousal Category × Social Content repeated-measures ANOVA

Table 3: Valence and arousal ratings by participant gender.

Scene category	Valence ratings, *M* (*SD*)		Arousal ratings, *M* (*SD*)	
	Men	**Women**	**Men**	**Women**
Negative Higher Social	2.22 (1.53)***	1.85 (1.40)	5.53 (2.77)***	6.13 (2.63)
Negative Higher Nonsocial	2.56 (1.72)***	2.11 (1.61)	5.13 (2.74)**	5.63 (2.74)
Negative Moderate Social	3.54 (1.89)***	3.18 (1.91)	4.44 (2.50)**	4.78 (2.52)
Negative Moderate Nonsocial	3.25 (1.81)***	2.84 (1.78)	4.31 (2.55)**	4.70 (2.61)
Neutral Moderate Social	5.32 (1.61)	5.35 (1.65)	3.50 (2.21)	3.60 (2.29)
Neutral Moderate Nonsocial	5.10 (1.83)	5.05 (1.93)	3.44 (2.30)	3.55 (2.39)
Neutral Lower Social	5.42 (1.56)	5.35 (1.64)	3.24 (2.15)	3.39 (2.24)
Neutral Lower Nonsocial	5.64 (1.76)	5.59 (1.86)	3.46 (2.38)	3.53 (2.43)
Positive Higher Social	5.97 (2.17)***	5.52 (1.99)	4.61 (2.79)***	4.08 (2.56)
Positive Higher Nonsocial	7.21 (1.64)***	7.58 (1.67)	4.90 (2.60)**	5.26 (2.74)
Positive Moderate Social	6.45 (1.80)***	6.73 (1.87)	4.28 (2.43)*	4.54 (2.56)
Positive Moderate Nonsocial	6.68 (1.73)***	6.92 (1.82)	4.28 (2.47)*	4.52 (2.64)

Note: Asterisks denote significant gender differences in ratings. See Table S2 for the t-test statistics.
* $p < .05$; ** $p < .01$; *** $p < .001$.

Figure 5: The relation between COMPASS valence (1 = unpleasant, 9 = pleasant) and arousal (1 = low, 9 = high) ratings by participant gender. Error bars represent standard errors of the mean.
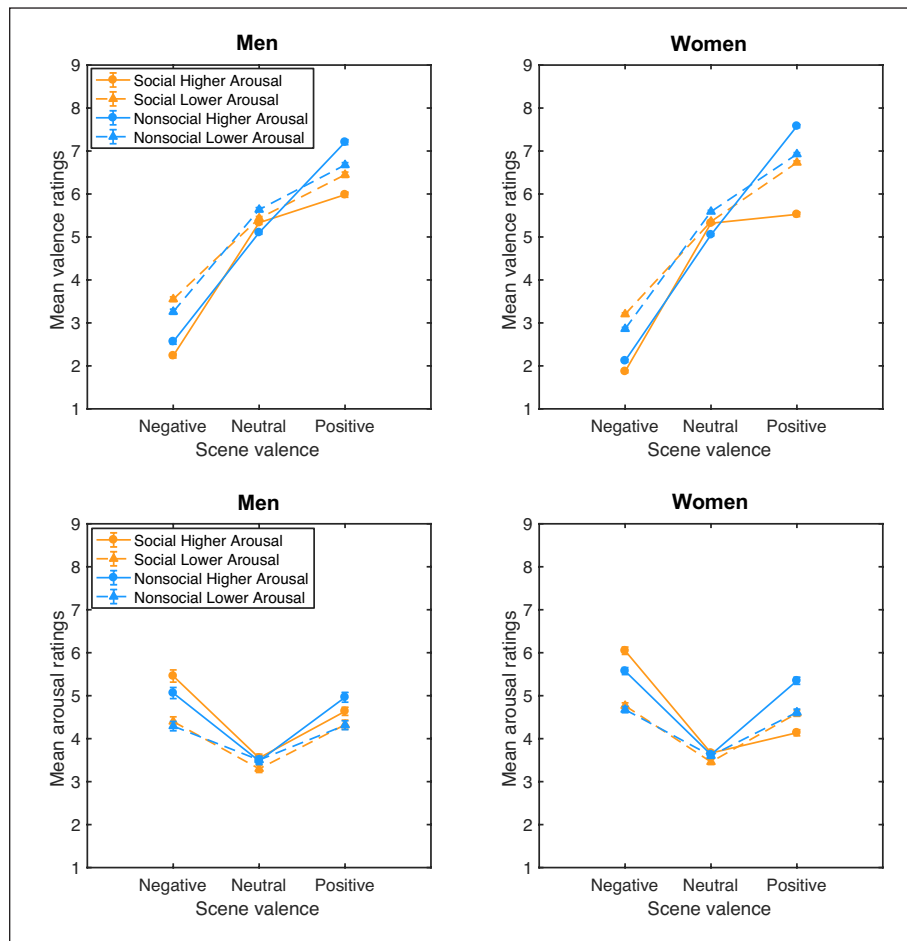
Figure 6: Mean valence (top panels) and arousal (bottom panels) ratings for each affective scene category by participant gender (lower arousal = low to moderate; higher arousal = moderate to high). Error bars represent standard errors of the mean.

with Participant Gender as a between-subjects factor and arousal ratings as the dependent variable. There was a main effect of Participant Gender on arousal ratings ($F(1,840) = 5.60$, $p = .018$, $\eta_p^2 = .01$), with women providing higher arousal ratings than men. There was also a Valence Category × Arousal Category × Social Content × Participant Gender interaction ($F(2,1680) = 28.1$, $p < .001$, $\eta_p^2 = .03$): women rated nonsocial higher arousal pleasant scenes as more arousing than did men and social higher arousal pleasant scenes as less arousing than did men (**Figure 6**).

To identify scene content for which valence and arousal ratings differed by participant gender, we conducted scene-wise independent samples t-tests on valence and arousal ratings. To correct for multiple comparisons, we used a Bonferroni-corrected alpha of $0.05/300 = 0.000167$. For each of four generally gender-discrepant content categories, we tested mean valence and arousal ratings by participant gender using independent samples t-tests.

Men rated higher arousal pleasant scenes depicting scantily dressed women as more pleasant (Men $M = 6.81$, $SD = 1.28$; Women $M = 5.27$, $SD = 1.33$; $t(840) = 15.5$, $p < .001$, $d = 1.18$) and more arousing (Men $M = 5.72$, $SD = 1.94$; Women $M = 3.81$, $SD = 1.85$; $t(840) = 13.4$, $p < .001$, $d = 1.00$) than did women, whereas women rated scenes depicting scantily dressed men as more pleasant (Men $M = 4.77$, $SD = 1.70$; Women $M = 6.05$, $SD = 1.49$; $t(840) = 10.9$, $p < .001$, $d = 0.80$) and more arousing

(Men $M = 3.21$, $SD = 2.11$; Women $M = 4.74$, $SD = 1.99$; $t(839) = 10.0$, $p < .001$, $d = 0.75$) than did men. Women rated pleasant and neutral scenes depicting children and animals as more pleasant (Men $M = 6.82$, $SD = 1.00$; Women $M = 7.40$, $SD = 0.91$; $t(840) = 8.14$, $p < .001$, $d = 0.61$) and more arousing (Men $M = 4.59$, $SD = 1.78$; Women $M = 5.10$, $SD = 1.91$; $t(840) = 3.52$, $p < .001$, $d = 0.27$) than did men. Women rated scenes of destruction, dead or mutilated animals, human suffering or violence as more unpleasant (Men $M = 3.11$, $SD = 0.89$; Women $M = 2.40$, $SD = 0.82$; $t(840) = 11.1$, $p < .001$, $d = 0.83$) and more arousing (Men $M = 4.91$, $SD = 1.85$; Women $M = 5.34$, $SD = 1.87$; $t(840) = 2.98$, $p = .003$, $d = 0.23$) than did men.

3.6. Image specifications by affective scene category

To rule out any confounding effects of low-level features on scene ratings, we calculated mean luminance and contrast across all the scenes within each Valence × Arousal × Social Content Category and tested scene category differences in luminance and contrast. We conducted a Valence Category × Arousal Category × Social Content ANOVA with image luminance as the dependent variable. There was no main effect of Valence Category ($F(2,288) = 1.05$, $p = .350$, $\eta_p^2 < .01$), Arousal Category ($F(2,288) = 0.02$, $p = .978$, $\eta_p^2 < .01$), or Social Content ($F(1,288) = 0.92$, $p = .337$, $\eta_p^2 < .01$) on luminance. In addition, there were no Valence Category × Arousal Category ($F(1,288)$

= .931, p = .335, η_p^2 < .01), Valence Category × Social Content ($F(2,288)$ = 2.40, p = .093, η_p^2 = .02), Arousal Category × Social Content ($F(2,288)$ = 2.40, p = .092, η_p^2 = .02), or Valence Category × Arousal Category × Social Content ($F(1,288)$ = .18, p = .672, η_p^2 < .01) interaction effects on luminance. We also conducted a Valence Category × Arousal Category × Social Content ANOVA with image contrast as the dependent variable. There was no main effect of Valence Category ($F(2,288)$ = 1.90, p = .152, η_p^2 = .01), Arousal Category ($F(2,288)$ = 1.77, p = .173, η_p^2 = .01) or Social Content ($F(1,288)$ = 3.67, p = .056, η_p^2 = .01) on image contrast. In addition, there were no Valence Category × Arousal Category ($F(1,288)$ = .150, p = .699, η_p^2 < .01), Valence Category × Social Content ($F(2,288)$ = .406, p = .667, η_p^2 < .01), Arousal Category × Social Content ($F(2,288)$ = 2.36, p = .096, η_p^2 = .02), or Valence Category × Arousal Category × Social Content ($F(1,288)$ = .822, p = .346, η_p^2 < .01) interaction effects on image contrast. The similarity in luminance and contrast between social and nonsocial scenes, higher and lower arousal scenes, and positive, negative, and neutral scenes was further confirmed via equivalence testing (Lakens, 2017; please see Supplemental Information for the detailed results).

We also tested differences in image complexity between social and non-social scenes using two common measures of image complexity: JPEG compressibility and entropy (Donderi, 2006; Machado et al., 2015). Overall, social content had a small effect on COMPASS image complexity, with nonsocial COMPASS scenes being somewhat more complex than social scenes. However, the effect of social content on image complexity depended on the measure of image complexity (please see Supplemental Information for the detailed results).

3.7. Rating response times by scene category
Due to the special status accorded to social information over nonsocial information, we tested response times (RTs) for valence ratings by category. Mean RTs for valence ratings by scene category are presented in **Table 4**. For each scene, we calculated mean RTs across all participants.
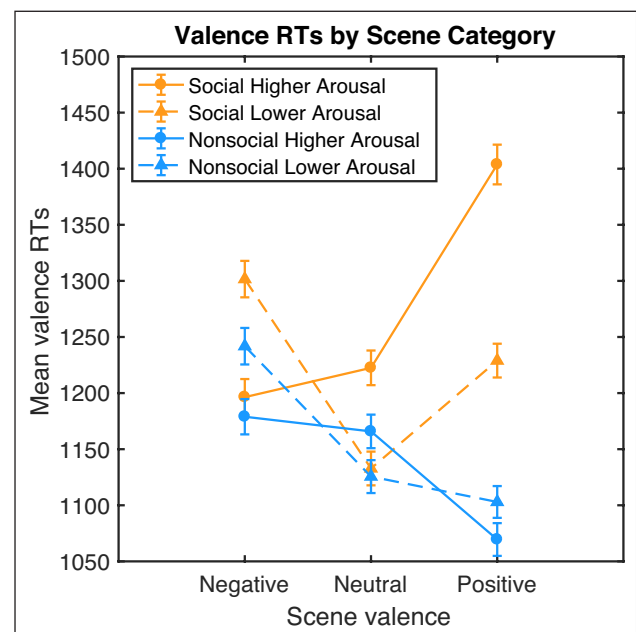
Table 4: Response times for valence ratings by assigned scene category.

Scene category	Response time in ms, *M (SD)*
Negative Higher Social	1176 (818)
Negative Higher Nonsocial	1162 (819)
Negative Moderate Social	1277 (876)
Negative Moderate Nonsocial	1219 (858)
Neutral Moderate Social	1204 (851)
Neutral Moderate Nonsocial	1151 (836)
Neutral Lower Social	1118 (825)
Neutral Lower Nonsocial	1111 (826)
Positive Higher Social	1372 (933)
Positive Higher Nonsocial	1056 (764)
Positive Moderate Social	1213 (825)
Positive Moderate Nonsocial	1090 (785)

We then calculated mean RTs across all the scenes within each Valence × Arousal × Social Content Category and conducted a Valence Category × Arousal Category × Social Content repeated-measures ANOVA with RTs for valence ratings as the dependent variable. There was a main effect of Valence Category ($F(2,1692)$ = 38.5, p < .001, η_p^2 = .04). Post-hoc pairwise comparisons with Bonferroni correction showed that participants had slower rating RTs for unpleasant compared to pleasant scenes, which had slower rating RTs than neutral scenes (all ps < .001). There was a main effect of Arousal Category ($F(1,846)$ = 21.2, p < .001, η_p^2 = .02), such that rating RTs were slower for higher arousal scenes compared to lower arousal scenes. There was also a main effect of Social Content ($F(1,846)$ = 536, p < .001, η_p^2 = .39): rating RTs were slower for social compared to nonsocial scenes.

There was also a Valence Category × Social Content interaction ($F(2,1692)$ = 307, p < .001, η_p^2 = .27), driven by a greater effect of social content on RTs for pleasant scenes compared to other affect categories. Rating RTs were slower for pleasant social scenes compared to pleasant nonsocial scenes. There was also an Arousal Category × Social Content interaction ($F(1,846)$ = 113, p < .001, η_p^2 = .12), such that rating RTs were slower for higher arousal social compared to higher arousal nonsocial scenes.

Finally, there was a Valence Category × Arousal Category × Social Content interaction ($F(2,1692)$ = 116, p < .001, η_p^2 = .12): for social scenes, the effect of valence category on RTs was moderated by arousal category. For higher arousal social scenes, RTs were fastest for negative scenes and slowest for positive scenes. For lower arousal social scenes, RTs were fastest for neutral scenes and slowest for negative scenes. In contrast, for nonsocial scenes, there was no rating RT difference by arousal category for neutral and positive scenes. However, for negative nonsocial scenes, RTs were slower for lower arousal scenes (**Figure 7**).

Figure 7: Mean response times (RTs) in milliseconds for valence ratings by affective scene category (lower arousal = low to moderate; higher arousal = moderate to high). Error bars represent standard errors of the mean.

4. Discussion

We present the Complex Affective Scene Set (COMPASS), a novel set of 300 social and nonsocial complex, naturalistic affective scenes normed on the dimensions of valence (unpleasant to pleasant) and arousal (low to high activation). This set achieves our primary goals and contributes to existing measurement tools in the following ways.

4.1. Coverage of the canonical affective space

Our primary goal was to develop a set of complex scenes that capture daily life experiences, and we did not include affectively extreme stimuli that were less likely to represent daily experience. Consequently, for the arousal dimension, most COMPASS scenes had mean ratings near the midpoint of the arousal scale (i.e., between 4 and 5 on the 1–9 scale). In comparison with published IAPS norms (i.e., Grühn & Scheibe, 2008; Lang et al., 1999), unpleasant and pleasant COMPASS scenes have on average lower arousal ratings, whereas neutral COMPASS scenes have similar arousal ratings. For valence, it is important to note that our division of stimuli into discrete unpleasant, neutral, and pleasant categories was designed to provide coverage of the affective space as much as possible, and that the boundaries for categorization were somewhat arbitrary with respect to the nature and definition of the continuous valence dimension. The "neutral" category covers the middle range of the scale from unpleasant to pleasant, however there is no possible absolute determination of the scale number at which an image is neither pleasant nor unpleasant. With this caveat, the ratings demonstrate good coverage of the space but without the extremes. Most of the unpleasant COMPASS scenes had valence ratings that corresponded to the middle of the generally unpleasant range (i.e., between 2 and 3 on the 9-point scale with 1 as most unpleasant), and most of the pleasant COMPASS scenes had mean valence ratings that corresponded to the middle of the generally pleasant range (i.e., between 6 and 7 on the 9-point scale with 9 as most pleasant). Unpleasant IAPS and COMPASS scenes had similar valence ratings, whereas pleasant IAPS scenes were rated as more pleasant than pleasant COMPASS scenes.

Consistent with the norms for other affective stimulus sets (Kurdi et al., 2017; Lang et al., 1999; Libkuman et al., 2007), the distribution of valence and arousal ratings of COMPASS scenes is shaped like a boomerang, such that highly pleasant and highly unpleasant scenes were rated as more arousing than neutral scenes. However, unpleasant COMPASS scenes were rated as more arousing than pleasant scenes. Consistent with previous reports (e.g., Kurdi et al., 2017), there was less variability in valence ratings for scenes with means near the three anchor points (1 = highly unpleasant, 5 = neutral, 9 = highly pleasant), compared to scenes with mean valence ratings between the anchor points. In contrast, variability in arousal ratings was directly related to the direction of arousal ratings, such that the most arousing scenes also had the greatest variability in arousal ratings.

Although our data are not intended to address the conceptual debates regarding the structure of affect or the affect-emotion distinction, our measurement model favors the bipolar valence-arousal approach outlined by the circumplex model of affect (e.g., Barrett & Russell, 1999). We utilized the bipolar valence-arousal model because we were most interested in a person's initial affective response to visual information, as when a trauma-exposed person first encounters a trauma-relevant stimulus in the environment. We concur that on a longer timescale people can experience some degree of pleasant and unpleasant affect alternatingly (e.g., Kron, Pilkiw, Banaei, Goldstein, & Anderson, 2015), however the literature also supports that only one affective state will predominate initially. Relatedly, we agree with the perspective that valence and arousal constitute the basic affective units experienced by humans, whereas identification and categorization of an emotion requires application of a conceptual label, which by then is one step removed from the initial experience. We were interested primarily in the former, which is why we did not focus on emotion. Our intent is that this stimulus set should be appropriate for testing additional sets of questions, however, and we encourage the use of the set to further test the dual unipolar model of valence (e.g., Kron, Goldstein, Lee, Gardhouse, & Anderson, 2013; Kron et al., 2015), and to further test the interdependence of valence and arousal ratings (e.g,. Larsen, Norris, & Caciopp, 2003). We support an approach whereby researchers are careful about their own questions and about the risks of imposing universal claims about the nature of affective experience where individual differences not only exist (e.g., Kuppens, Tuerlinckx, Russell, & Barrett, 2012), but are vital for a clearer understanding of the fundamental mechanisms of affect.

4.2. Solution for affectively unbalanced social content

Our second primary goal was to distinguish between the influence of affect category and social content on valence and arousal ratings by including equal numbers of social and nonsocial scenes within each affective category. On average, nonsocial scenes were rated as more pleasant and higher in arousal than social scenes, and this effect was driven largely by the higher arousal pleasant category. Specifically, whereas social and nonsocial scenes had similar valence and arousal ratings within the unpleasant and neutral categories, nonsocial higher arousal pleasant scenes were rated as more pleasant and arousing than social higher arousal pleasant scenes. These results are consistent with previously reported confounding effects of social content on valence and arousal ratings (e.g., Colden et al., 2008), supporting the importance of controlling for the social content of affective stimuli. Because the COMPASS images intentionally exclude the highest arousal pleasant (e.g., erotic) and unpleasant (e.g., mutilated bodies) content due to the goal of representing more everyday experiences, researchers who wish to equate arousal and valence extremes between social and nonsocial stimuli might choose to add images from sets such as the IAPS to fit that purpose.

4.3. Demonstration of the persistence of the special status of social content

We designed the COMPASS set to provide a set of social and nonsocial images that are balanced across affect categories, and our development process resulted in the elimination of scenes that did not contribute to this goal. However, this methodological contribution does not eliminate the actual special status effect of social information. Once we had a balanced set of 300 images, we also sought to demonstrate the persistence of the social content effect. Because participants were instructed to respond quickly and in accord with their initial impression of each scene, the most efficient way to complete the 300 image ratings was to respond quickly to each image. We reasoned that slower RTs for the first rating for each image (i.e., the valence rating) would provide evidence for the persistence of a special effect of social content. Consistent with our expectations, initial ratings for social scenes took longer than initial ratings for nonsocial scenes. Although it is not possible with simple rating data to isolate the precise mechanism or mechanisms that account for this effect, the slower response time is consistent with greater attentional capture by social information, for example. In addition, this effect was moderated by image valence, with faster RTs for unpleasant and slower RTs for pleasant social information, suggesting more efficient processing of depicted negative affect relative to depicted positive affect. These results are consistent with prior evidence of more distributed brain network activation in response to the mere presence of social information (e.g., Tso, Rutherford, Fang, Angstadt, & Taylor, 2018), and greater relevance detection for social information (e.g., Schacht & Vrticka, 2018; Vrticka, Sander, & Vuilleumier, 2013). Regardless of mechanism, the persistence of the special status of social information, when controlling for valence and arousal, is clear.

4.4. Replication of rater gender effects

Consistent with prior evidence of gender differences in affective ratings and physiological reactivity to unpleasant stimuli (e.g., Bradley, Codispoti, Sabatinelli, & Lang, 2001; Lithari et al., 2010), women rated unpleasant COMPASS scenes as more unpleasant and more arousing than did men. In addition, women rated higher arousal pleasant nonsocial scenes as more pleasant and more arousing than did men, and higher arousal pleasant social scenes as less pleasant and less arousing than did men. Together, these results are consistent with previously reported gender effects on affect ratings of specific content categories, such as erotica and highly unpleasant scenes (e.g., Kurdi et al., 2017; Marchewka et al., 2014).

4.5. Stimulus competition and additional scene characteristics

In addition to the primary attributes of affect dimensions and social content, the COMPASS scenes also incorporate additional characteristics that position the set well for certain types of research questions. First, the scenes feature stimulus competition in the form of two or more visually salient points of interest. This characteristic is important for questions addressing allocation of visual attention. Combined with careful placement of pre-stimulus fixation points and selection of presentation timing parameters, these scenes can be used to test initial fixation, shifts of attention, and disengagement of attention (e.g., Weierich, Treat, & Hollingworth, 2008) within a single image. In addition, because the set includes representation of human diversity, subsets of the scenes can be used to test interactions of affect with race perception or culture-specific visual information.

4.6. Potential constraints on generality

Two characteristics of our sample might constrain the generality of our results. First, the public, non-residential, urban university sample from which we recruited has a very large proportion of non-traditional age students (sample age range 18–53), and the vast majority of these students have had a much broader and less privileged variety of life experiences than the canonical "WEIRD" (i.e., Western, Educated, Industrialized, Rich, Democratic; Henrich, Heine, & Norenzayan, 2010) undergraduate samples. Nonetheless, the sample had a relatively young mean age (i.e., 20.5, SD 4.6), such that normative ratings provided by younger (i.e., adolescent) or older samples might differ. In addition, our sample was comprised of participants who live in or very near New York City, and the daily experience of life in a large, densely populated, racially and ethnically diverse urban area might have influenced ratings of some of the images, and in particular the social images, which included representation of a range of races and ethnicities. We welcome researchers to conduct norming studies with this stimulus set in additional populations. We also note that although our sample was predominantly female, our male subsample was large enough to adequately power between and within group tests, as reported, and thus this imbalance is not likely to have affected generality with regard to sex or gender. In addition, although the absolute whole sample means represent twice as many ratings from women as from men, in our view the absolute means are less important than the coverage of the affective space as well as the expected within group (i.e., within-gender) patterns that are consistent with the affective space. Together, our whole sample data and the analyses by gender both support the achievement of a stimulus set that covers the affective space.

In addition to further norming in additional populations, due to its unique attributes, including visual complexity, human diversity, and naturalistic everyday-life content, the COMPASS stimulus set can be used to study affective processing in complex daily environments using a variety of methods, including eye-tracking, psychophysiology, and neuroimaging (e.g., Mauss & Robinson, 2009), while controlling for potential confounds, and in particular social content. We provide the basic affective norms for the COMPASS set, however future research will be necessary to characterize COMPASS scenes along other dimensions that might influence affective processing, such as memorability and distinctiveness. Similarly, our

strategy of collecting valence ratings before arousal ratings, although important for our stimulus set development goals, also might have constrained generality; rating order could have influenced the valence and/or arousal ratings, and future work counterbalancing or switching the order will address that question.

4.7. Usage
COMPASS images and image usage rules are available without cost to researchers upon request at the following link: www.compass-scenes.com. In addition to the images, the downloadable content includes scene-wise norms for the total sample and separately by participant gender, and scene-wise attributes (e.g., affective category, social content category, scene content, image orientation, luminance, and contrast).

Data Accessibility Statement
The stimuli, presentation materials, participant data, and analysis scripts can be found on this paper's project page on the www.compass-scenes.com.

Note
[1] To estimate the impact of trial exclusion on scene ratings, we also calculated mean valence and arousal ratings for each scene without excluding trials based on RTs (see Figure S1). The largest absolute difference in mean scene-wise valence ratings before and after trial exclusion was 0.13 (possible range: 0–8), whereas the largest absolute difference in mean scene-wise arousal ratings was 0.26 (possible range: 0–8), suggesting that exclusion of potentially unreliable trials did not have a significant impact on mean scene ratings.

Additional Files
The additional files for this article can be found as follows:

- **Figure S1.** Mean scene-wise valence and arousal ratings before and after exclusion of trials with RTs <150 ms or >4000 ms. DOI: https://doi.org/10.1525/collabra.256.s1
- **Figure S2.** Bivariate distributions of valence and arousal ratings for each image. DOI: https://doi.org/10.1525/collabra.256.s2
- **Table S1.** Valence and arousal summary statistics for COMPASS and IAPS images. DOI: https://doi.org/10.1525/collabra.256.s3
- **Table S2.** Gender differences in valence and arousal ratings (men vs. women) by affective scene category. DOI: https://doi.org/10.1525/collabra.256.s4
- **Text S1.** Task instructions. DOI: https://doi.org/10.1525/collabra.256.s5
- **Text S2.** Stimulus set development. DOI: https://doi.org/10.1525/collabra.256.s6
- **Text S3.** Additional statistical analyses. DOI: https://doi.org/10.1525/collabra.256.s7

Acknowledgements
The authors wish to thank the members of the Cognition, Affect, & Psychopathology Lab.

Competing Interests
The authors have no competing interests to declare.

Author Contributions
- Contributed to conception and design: MRW
- Contributed to acquisition of data: MRW, OK, JKR, DMR
- Contributed to analysis and interpretation of data: MRW, OK, JKR, DMR
- Drafted and/or revised the article: MRW, OK, JKR, DMR
- Approved the submitted version for publication: MRW, OK, JKR, DMR

References
Andreano, J. M., Dickerson, B. C., & **Barrett, L. F.** (2014). Sex differences in the persistence of the amygdala response to negative material. *Social, Cognitive, and Affective Neuroscience*, *9*, 1388–1394. DOI: https://doi.org/10.1093/scan/nst127

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1), 28–58. DOI: https://doi.org/10.1111/j.1745-6916.2006.00003.x

Barrett, L. F., & **Russell, J. A.** (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science*, *8*, 10–14. DOI: https://doi.org/10.1111/1467-8721.00003

Beaupré, M. G., & **Hess, U.** (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-cultural Psychology*, *36*, 355–370. DOI: https://doi.org/10.1177/0022022104273656

Bex, P. J., & **Makous, W.** (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America A*, *19*(6), 1096–1106. DOI: https://doi.org/10.1364/JOSAA.19.001096

Bodapati, A. S., & **Herbener, E. S.** (2014). The impact of social content and negative symptoms on affective ratings in schizophrenia. *Psychiatry Research*, *218*(1–2), 25–30. DOI: https://doi.org/10.1016/j.psychres.2014.03.039

Bradley, M. M., Codispoti, M., Sabatinelli, D., & **Lang, P. J.** (2001). Emotion and motivation II: Sex differences in picture processing. *Emotion*, *1*(3), 300–319. DOI: https://doi.org/10.1037/1528-3542.1.3.300

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience*, *7*(6), 477–484. DOI: https://doi.org/10.1038/nrn1909

Cerf, M., Frady, E. P., & **Koch, C.** (2009). Faces and text attract gaze independent of the task: Experimental

data and computer model. *Journal of Vision, 9*(12), 1–15. DOI: https://doi.org/10.1167/9.12.10

Colden, A., Bruder, M., & **Manstead, A. S.** (2008). Human content in affect-inducing stimuli: A secondary analysis of the International Affective Picture System. *Motivation and Emotion, 32*(4), 260–269. DOI: https://doi.org/10.1007/s11031-008-9107-z

Dan-Glauser, E. S., & **Scherer, K. R.** (2011). The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods, 43*(2), 468–477. DOI: https://doi.org/10.3758/s13428-011-0064-1

Desimone, R., & **Duncan, J.** (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*(1), 193–222. DOI: https://doi.org/10.1146/annurev.ne.18.030195.001205

Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin, 132*(1), 73–97. DOI: https://doi.org/10.1037/0033-2909.132.1.73

Ekman, P., & **Friesen, W. V.** (1976). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychologists Press.

End, A., & **Gamer, M.** (2017). Preferential processing of social features and their interplay with physical saliency in complex naturalistic scenes. *Frontiers in Psychology, 8,* 418. DOI: https://doi.org/10.3389/fpsyg.2017.00418

Fitzgerald, H. E. (1968). Autonomic pupillary reflex activity during early infancy and its relation to social and nonsocial visual stimuli. *Journal of Experimental Child Psychology, 6*(3), 470–482. DOI: https://doi.org/10.1016/0022-0965(68)90127-6

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & **Benson, V.** (2008). Rapid detection of person information in a naturalistic scene. *Perception, 37*(4), 571–583. DOI: https://doi.org/10.1068/p5705

Harris, L. T., McClure, S. M., Van den Bos, W., Cohen, J. D., & **Fiske, S. T.** (2007). Regions of the MPFC differentially tuned to social and nonsocial affective evaluation. *Cognitive, Affective, & Behavioral Neuroscience, 7*(4), 309–316. DOI: https://doi.org/10.3758/CABN.7.4.309

Henrich, J., Heine, S. J., & **Norenzayan, A.** (2010). Most people are not WEIRD. *Nature, 466,* 29. DOI: https://doi.org/10.1038/466029a

Kron, A., Goldstein, A., Lee, D. H.-J., Gardhouse, K., & **Anderson, A. K.** (2013). How are you feeling? Revisiting the quantification of emotional qualia. *Psychological Science, 24,* 1503–1511. DOI: https://doi.org/10.1177/0956797613475456

Kron, A., Pilkiw, M., Banaei, J., Goldstein, A., & **Anderson, A. K.** (2015). Are valence and arousal separable in emotional experience? *Emotion, 15,* 35–44. DOI: https://doi.org/10.1037/a0038474

Kuppens, P., Tuerlinckx, F., Russell, J. A., & **Barrett, L. F.** (2012). The relation between valence and arousal in subjective experience. *Psychological Bulletin, 139,* 917–940. DOI: https://doi.org/10.1037/a0030811

Kurdi, B., Lozano, S., & **Banaji, M. R.** (2017). Introducing the open affective standardized image set (OASIS). *Behavior Research Methods, 49*(2), 457–470. DOI: https://doi.org/10.3758/s13428-016-0715-3

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355–362. DOI: https://doi.org/10.1177/1948550617697177

Lang, P. J., Bradley, M. M., & **Cuthbert, B. N.** (1999). International Affective Picture System (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention.*

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.*

Larsen, J. T., Norris, C. J., & **Cacioppo, J. T.** (2003). Effects of positive and negative affect on electromyographic activity of zygomaticus major and corrugated supercilii. *Psychophysiology, 40,* 776–785. DOI: https://doi.org/10.1111/1469-8986.00078

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*(2), 326–334. DOI: https://doi.org/10.3758/BF03193164

Lithari, C., Frantzidis, C. A., Papadelis, C., Vivas, A. B., Klados, M. A., Kourtidou-Papadeli, C., Pappas, C., Ioannides, A. A., & **Bamidis, P. D.** (2010). Are females more responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions. *Brain Topography, 23*(1), 27–40. DOI: https://doi.org/10.1007/s10548-009-0130-5

Löw, A., Bradley, M. M., & **Lang, P. J.** (2013). Perceptual processing of natural scenes at rapid rates: Effects of complexity, content, and emotional arousal. *Cognitive, Affective, & Behavioral Neuroscience, 13*(4), 860–868. DOI: https://doi.org/10.3758/s13415-013-0179-1

Lundqvist, D., Flykt, A., & **Öhman, A.** (1998). The Karolinska Directed Emotional Faces – KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9. DOI: https://doi.org/10.1037/t27732-000

Machado, P., Romero, J., Nadal, M., Santos, A., Correia, J., & **Carballal, A.** (2015). Computerized measures of visual complexity. *Acta Psychologica, 160,* 43–57. DOI: https://doi.org/10.1016/j.actpsy.2015.06.005

Marchewka, A., Żurawski, Ł., Jednoróg, K., & **Grabowska, A.** (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods, 46*(2), 596–610. DOI: https://doi.org/10.3758/s13428-013-0379-1

Mauss, I. B., & **Robinson, M. D.** (2009). Measures of emotion: A review. *Cognition and Emotion, 23*(2), 209–237. DOI: https://doi.org/10.1080/02699930802204677

Michałowski, J. M., Droździel, D., Matuszewski, J., Koziejowski, W., Jednoróg, K., & **Marchewka,**

A. (2017). The Set of Fear Inducing Pictures (SFIP): Development and validation in fearful and nonfearful individuals. *Behavior Research Methods*, *49*(4), 1407–1419. DOI: https://doi.org/10.3758/s13428-016-0797-y

Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., & Ohme, R. K. (2015). Warsaw set of emotional facial expression pictures: A validation study of facial display photographs. *Frontiers in Psychology*, *5*. DOI: https://doi.org/10.3389/fpsyg.2014.01516

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3), 715–734. DOI: https://doi.org/10.1017/S0954579405050340

Proverbio, A. M. (2016). Sex differences in social cognition: The case of face processing. *Journal of Neuroscience Research*, *95*, 222–234. DOI: https://doi.org/10.1002/jnr.23817

Rosler, L., End., A., & Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. *PLoS ONE*, *12*(7), e0182017. DOI: https://doi.org/10.1371/journal.pone.0182037

Schacht, A., & Vrtička, P. (2018). Spatiotemporal pattern of appraising social and emotional relevance: Evidence from event-related brain potentials. *Cognitive, Affective, and Behavioral Neuroscience*. DOI: https://doi.org/10.1101/230961

Soares, A. P., Pinheiro, A. P., Costa, A., Frade, C. S., Comesana, M., & Pureza, R. (2015). Adaptation of the International Affective Picture System (IAPS) for European Portuguese. DOI: https://doi.org/10.3758/s13428-014-0535-2

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. DOI: https://doi.org/10.1016/j.psychres.2008.05.006

Tso, I. F., Rutherford, S., Fang, Y., Angstadt, M., & Taylor, S. F. (2018). The "social brain" is highly sensitive to the mere presence of social information: An automated meta-analysis and an independent study. *PLoS ONE*, *13*(5), e0196503. DOI: https://doi.org/10.1371/journal.pone.0196503

Vrtička, P., Sander, D., & Vuilleumier, P. (2013). Lateralized interactive social content and valence processing within the human amygdala. *Frontiers in Human Neuroscience*, *6*, 358. DOI: https://doi.org/10.3389/fnhum.2012.00358

Weierich, M. R., Treat, T. A., & Hollingworth, A. (2008). Theories and measurement of visual attentional processing in anxiety. *Cognition and Emotion*, *22*(6), 985–1018. DOI: https://doi.org/10.1080/02699930701597601

Wierzba, M., Riegel, M., Pucz, A., Leśniewska, Z., Dragan, W. Ł., Gola, M., Jednoróg, K., & Marchewka, A. (2015). Erotic subset for the Nencki Affective Picture System (NAPS ERO): Cross-sexual comparison study. *Frontiers in Psychology*, *6*, 1336. DOI: https://doi.org/10.3389/fpsyg.2015.01336

Wrase, J., Klein, S., Gruesser, S. M., Hermann, D., Flor, H., Mann, K., Braus, D. F., & Heinz, A. (2003). Gender differences in the processing of standardized emotional visual stimuli in humans: A functional magnetic resonance imaging study. *Neuroscience Letters*, *348*(1), 41–45. DOI: https://doi.org/10.1016/S0304-3940(03)00565-2

UNIVERSITY of CALIFORNIA PRESS | **Collabra: Psychology**   *Collabra: Psychology* is a peer-reviewed open access journal published by University of California Press.   **OPEN ACCESS**