

Text S2.

*Stimulus Set Development.*

In the paper 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.

*Selection of Images.* The teams who selected the initial set first were trained in our approach to affect (i.e., arousal and valence) and calibrated loosely to the IAPS stimuli, such that they (1) understood the constructs very well, and (2) had a large assortment of exemplars to “train” on. Team members then each searched for images that they thought would fit the 6 canonical combinations of arousal and valence (or for images that they thought might fit the other three combinations that are less common) from non-copyrighted images on the internet and photographs taken by lab members. All images also were required to be full-color images of complex scenes with multiple focal points, and excluded images of single objects. At this initial step, lab personnel reviewed images that were selected by other lab personnel and only images with 100% agreement for inclusion criteria were used at each step. As our goal was to create a set of naturalistic, everyday 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 excessively erotic content, as although naturalistic, such representations are less frequently encountered in regular daily life. We also excluded images that were immediately likely to have bimodal valence distributions; for example, some images of scantily clad people clearly were considered pleasant by some, but vulgar (and unpleasant) by others, and thus were discarded.

*Stimulus Categorization Check.* For each of the first three phases of norming (phase 1  $n = 496$ , phase 2  $n = 486$ , and phase 3  $n = 723$ ), participants rated valence and arousal for a set of 300 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 discarded scenes with mean standard deviations greater than 3, due to the probability of crossing multiple affect categories. We also discarded scenes that showed a bimodal valence distribution, thus indicating affective ambiguity, and replaced them with new scenes.

*Last pre-norming steps.* Finally, we added additional scenes to complete unfilled cells (e.g., neutral social). In the fourth phase (the data reported in this paper), all 847 participants rated the final set of 300 images.