Stimuli. We used pictures from the IAPS (Lang et al, 1997)**.** Our aim was to sample equally from as much of the valence and arousal space available from the IAPS to limit sampling biases. To this end we constructed an 8 x 8 stimulus selection grid. Of the 64 possible cells, 40 cells had at least one entry. If there was more than one picture in a cell, then the one in the left most corner was selected. In addition, as extremes have been found to affect EDA activity more than midrange or low values (e.g. Bradley, et al., 2001), each combination of extreme values of valence and arousal were included, resulting in 8 combinations of low and high positive valence, negative valence and arousal. Four pictures were selected from each of these eight combinations, yielding a total of 72 pictures in the stimulus set. This full and balanced representation of the IAPS space had no linear correlation between valence and arousal, r = -.08 with a marginal quadratic component, R^2 = .06.

Design. Trials began with a blank screen presented randomly for 10 to 21 seconds, to avoid anticipatory responses and for recording of physiological signals. Then, a picture was presented for 6 seconds followed by the rating scales. On each trial participants rated their feelings according to one of the two self report models, either bipolar valence and arousal or unipolar pleasant (PL) and unpleasant (UN) valence. As such, each participant rated half of the pictures according to pleasant and unpleasant unipolar model and half of the pictures according to bipolar-valence arousal model. We chose this approach so participants would not confuse the 4 rating scales, keeping them maximally independent, and more likely to show differences. To avoid confusion ratings scale were presented in fixed order for each participant along the entire experiment (for example, in the bipolar valence model: arousal scale was always appeared first and bipolar valence scale always second). The order of scale was counter balanced between participants. The 72 pictures were presented in three blocks of 24 pictures each with one minute breaks in between. Each block was divided into 3 sub-blocks. For each sub-block (6 consequtive pictures) rating scale were of the same model. That is, type of ratings was change every 6 pictures and participant could anticipate which rating scale would come next. The order of pictures in each sub-block was designed in such a way as to not include pictures from two successive squares in the 8 × 8 grid (e.g. the closest values of valence and arousal in the sample)

and to ensure that at least two of the pictures were from negative and two were from the positive wings of valence. Order of blocks, order of sub-blocks and order of pictures within sub-block was counter balance across participants in such a way that each picture was presented at first, second and third block and different locations within a block. At the same way, ratings according the two models (valence-arousal or pleasant/unpleasant) were counter balance so that across participants, a specific sequence of 72 pictures was rated the same amount of times by valence arousal and by pleasant/unpleasant. Pictures were presented full screen onto a 19-inch monitor situated 0.5 m away from the participant. Four pseudorandom presentation orders were counterbalanced between participants.

Data reduction. Reactions in corrugator EMG were determined by dividing activity 1 sec. before picture presentation with the average change over the 6s picture period (Lang et al., 1993; Larsen et al., 2003). EDA was estimated by log transformation (log[EDA+1]) of the maximum change occurring between 1 and 4 seconds after picture onset (Bradley, Codispoti, Cuthbert et al., 2001; Lang et al., 1993).

Physiological Response Measurement. Physiological data was acquired using the 2004 BIOPAC System. Signals were sampled at 1000 Hz. Left eye corrugator EMG activity was measured with 4mm electrodes which were placed according to Fridlund and Cacioppo (1986). Using ANSLAB 2.4, frequencies below 20 Hz were filtered, and then EMG signals were notched at 60 Hz and rectified. To measure EDA, two skin conductance electrodes were placed on the hypothenar eminence of the left index and middle fingers. The signal was acquired with a GSR100C BIOPAC amplifier and calibrated before each session.

Results. Bipolar valence ratings showed a high positive correlation with IAPS norms, $r = .95$ $p < 0.0001$; ratings of arousal were also highly correlated with the IAPS norms, $r = .86$, $p < 0.0001$. These correlation coefficients are identical to previous reports using pictures covering all of the IAPS space (Ito et al., 1998). As predicted based on previous studies (Larsen et al, 2003), (PL – UN) scores from the separate valence scales were highly correlated with bipolar valence, $r = .96$, $p < .0001$, suggesting that participants employed the separate and bipolar valence scales similarly.

Model specification for experiment 1, relation between self reports and EDA:

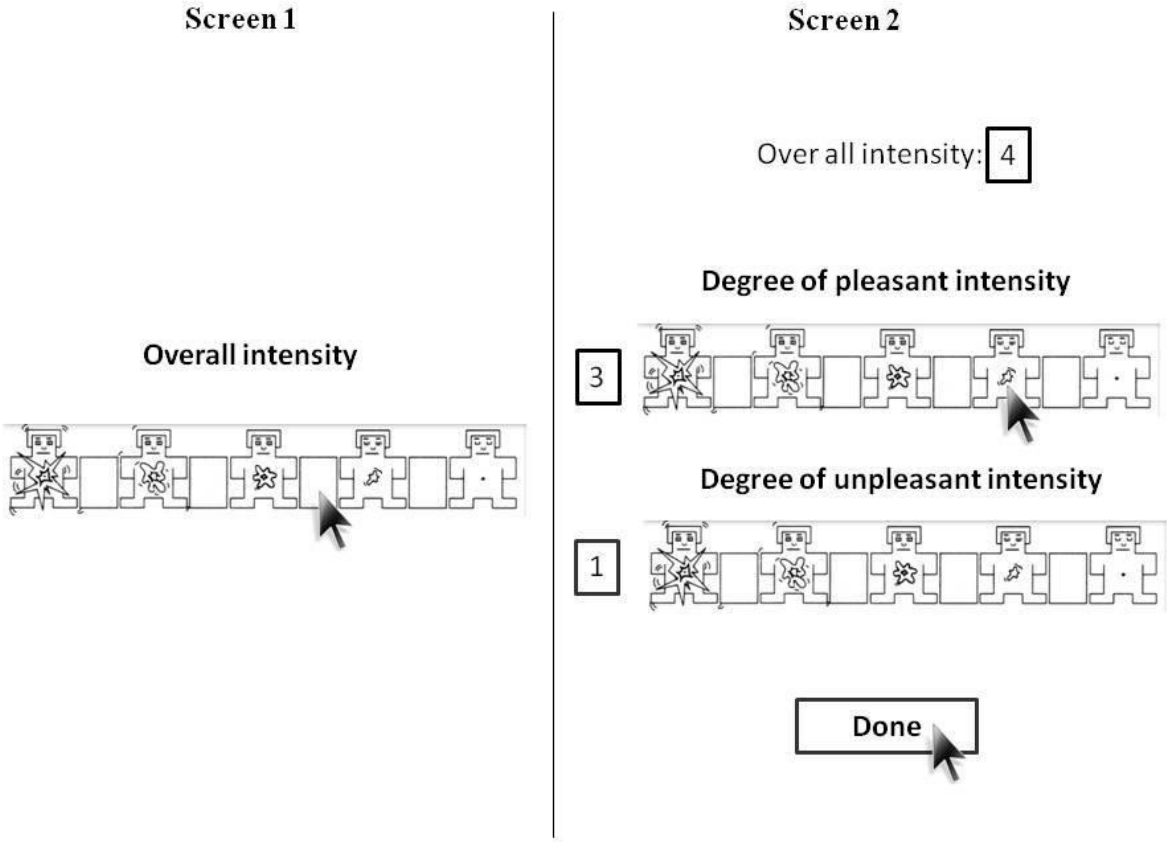
Here we compare the relation between arousal ratings and EDA to the relation of PL+UN ratings with EDA via hierarchical linear modeling. EDA scores were entered as dependent variable (criterion). Centered arousal scores and centered PL+UN scores for each trail (72 points per participant) were entered in *one* vector as *one* predictor; we term this predictor 'arousal/PL+UN ratings scores'. At the same way, a second, continuous vector consists of centered bipolar valence and PL-UN ratings scores enter as a second predictor and is termed 'valence/PL-UN scores'. A third, effect-coding vector (-1 and 1), was added to model. This vector coded for each of the scores in vectors 1 and 2, the model they are belong to: bipolar valence and arousal ratings were coded as -1 and PL + UN, and PL-UN were coded 1. We term this predictor "type of scale". Subject intercept was entered as random effect.

Relation between self reports and EMG activity – model specification and results:

EMG corrugator activity was enter as the dependent (criterion) variable. The predictors were: participants' intercepts (random variable), ratings type (bipolar vs unipolar valence) as a classification predictor (effect coding), and centered continuous predictors of self reported valence (bipolar valence vs. PL-UN) and arousal (arousal vs. PL+UN).

Replicating the traditional dissociation between arousal and valence in predicting facial EMG, together, ratings of bipolar valence and PL-UN scores were significantly related to currogator EMG, b = -.01, t (2119) = -3.65, p < .0003, while arousal and PL+UN were not, t (2119) = 0.5. In contrast with the EDA results, the interaction of valence ratings with ratings type was significant, t (2119) = -2.26, p = .02, revealing a small advantage of PL- UN over bipolar valence in predicting EMG. Separate simple-effect analysis demonstrated that both PL-UN (b = -0.01, t (1049) = -8.31, p <. 0001) and bipolar valence (b = -0.03, t (1049) = -6.17, p <. 0001) were significantly related to currogator EMG.

Fig. S1: Intensity Scale



References

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, *23*, 567–589.