Supplementary Material: When and Why Static Images Are More Effective Than Videos

Shaojing Fan, Zhiqi Shen, Bryan L. Koenig, Tian-Tsong Ng, Mohan S. Kankanhalli, *Fellow, IEEE*

1 Supplementary Methods

1.1 Overview of NUS Video Emotion Dataset and Validation Set.

We constructed the NUS Video Emotion Dataset (*i.e.*, NVED set) and validation set for our study. We applied three criteria for video selection for NVED set, as described below.

First, we kept the duration of the video clips between 2 to 6 seconds (see Supplementary Fig. S1 for an overall distribution of the durations). We chose this duration range for two reasons: (1) previous related studies [1], [2], [3] used moving stimuli of 6 seconds; (2) two seconds are enough to complete a meaningful action or motion, and are enough for human brain to complete action recognition motion and emotion elicitation [4], [5].

Second, we required each video clip to contain a complete and meaningful action, motion, or event. Moreover, the video clip should be emotion eliciting. For controlling for content, the video clips should be almost constant in semantics across all frames so that observers of isolated frames would see the same content as observers of the whole clip. We excluded videos that have dramatic background change. We further excluded those videos with content familiar to most people (*e.g.*, video clips from famous movies or contain celebrities) or containing captions.

We collected 304 video clips in total based on the above criteria, among which 22 video clips were later discarded due to low consistency among participants' ratings on emotional valence. This results in 282 video clips in total. We further collected 136 emotional video clips using the same criteria as we collected NVED set, but from a totally different pool of sources. We discarded 14 video clips in the data analyses due to low consistency in emotional valence ratings, resulting 122 video clips in total (*i.e.*, validation set). For the in house Experiment, we randomly selected 142 video clips from NVED dataset, and 68 video clips from the validation set, and mixed them together to form a 210 stimulus set.

To ensure data diversity, the video clips in the two sets are collected from multiple sources, ranging from online platforms to publicly available video datasets. Supplementary Fig. S2 shows stimulus examples in validation set. Besides scene diversity, we also kept the data diversity for other factors, such as emotion type and stimulus content types. All videos were resized to approximately two sizes: 854×480 pixels and 854×640 pixels.

We recruited three students from National University of Singapore to label three types of motion details in each video clip: (1) the carrier of the main motion in the video; (2) whether it is biological or non-biological motion [4], and (3) whether the motion contains dynamic facial emotion. We also asked the students to label the camera settings for each video, such as whether the camera was moving while the video was being captured; if the camera was moving, whether the movement was along z-axis (*i.e.*, zoom in or zoom out), or whether it was of wide or narrow angle. These factors are known to influence human sentiments [6]. We further asked the students to tag the video clips with no more than three words (*i.e.*, key words that they think are most representative for the video content). The detailed statistics are reported in Supplementary Figs. S3-S6.

1.2 Participants

Our participants have diverse backgrounds. Supplementary Figs. S7-S9 show the demographics of separate groups of participants in Experiments 1, 3 and in house experiment. All participants were financially remunerated for their time. For Experiments 1-3, participants were paid an average of US$5 per hour. In our in house experiment, participants were paid S$70 for completing the entire task.

1.3 Stimuli

Below are the details on how we collect our data, generate F_{DNN} and the correlation of the three types of static frames.

Data collection. For Experiment 1 and 2, we used the NUS Video Emotion Dataset (NVED), which consists of 304 video clips. Among which 22 video clips were later discarded due to low consistency among participants' ratings on emotional valence. This results in 282 video clips in total (referred to as NVED set). For Experiment 3, we collected 136 emotional video clips using the same criteria as we collected NVED set, but from a totally different pool of sources. We discarded 14 video clips in the data analyses due to low consistency in emotional valence ratings, resulting 122 video clips in total (hereafter validation set, see Supplementary Tables 1 and 2 and Supplementary Figs. 2 and 3 for detailed statistics of both sets). Clips are almost constant in semantics across all frames so that observers of isolated frames would see the same content as observers of the whole clip (see Supplementary Information for detailed video collection criteria). All videos were

- *S. Fan, Z. Shen and M. Kankanhalli are with the School of Computing, National University of Singapore, Singapore 119613. E-mail: {dcsfs, dcsshenz}@nus.edu.sg, mohan@comp.nus.edu.sg.*
- *T. Ng is with ASB Bank, New Zealand. E-mail: tiantsong@gmail.com.*
- *B. Koenig is with the Department of Psychology, Southern Utah University, Cedar City, UT 84720, United States. E-mail: bryankoenig@suu.edu.*

Manuscript received 05 Nov. 2019.

resized to approximately two sizes: 854 × 480 pixels and 854 × 640 pixels.

Distribution of Three Static Stimulus Versions. To explore if the three static stimuli versions ($F_{entropy}$, $F_{optical}$, F_{DNN}) are representative enough for the distinct field, we computed their distribution in the Entropy-Optical flow-DNN response space (*i.e.*, EOD space), which has been reported in the main paper. We further explored the difference of the basic image statistics among the three frame types, which includes mean and variance of the values in the hue-saturation-value (HSV) color space, image contrast, degree of blur, and color distribution, as described below:

Saturation, hue, and illumination: We computed features defined in the HSV space. Saturation indicates chromatic purity. Pure colors in a photo tend to be more appealing than dull or impure ones [7]. We computed the average saturation $f_s = \frac{1}{XY}\Sigma_{x=0}^{X-1}\Sigma_{y=0}^{Y-1} I_S(x,y)$ as the saturation indicator. Hue and illumination were similarly computed by averaging over I_H and I_V separately. Although the interpretation of such features is not as clear as saturation, they were found to be predictive of image aesthetics [7], [8]. We also calculated their variances and got a six-dimension feature in total.

Contrast: We used the similar contrast quality measure as Ke et al. [8], except that we computed the gray-scale level histogram of each image on R, G, B channels separately, and measured the width of the middle 98% gray level mass on each channel.

Blur: The degree of blur of an image is a strong indication for its quality and influences human sentiments. For blur prediction, we estimated the maximum frequency of the image I_b by taking its two dimensional Fourier transform and counting the number of frequencies whose power was greater than some threshold θ. We then normalized it by the size of the image [8]. We set $\theta = 5$ in our algorithm.

Color distribution: We used a similar method as Ke et al. [8] for computing the color distribution. For each image, we quantized the red, green, and blue channels into 16 values. A $4096 = 16^3$ bin histogram was created with the count of each quantized color present in the image. We used the $L1$ metric to calculate the distance between the normalized histograms with the average histogram of the whole image set.

As shown in Fig. 2 (b-c) in the main paper, the three types of static frames do not differ from each other significantly in terms of low-level image statistics. This is further corroborated by non-significant one-way ANOVA results ($F_{2,1209}$s ≤ 0.40, ps ≥ 0.669, η_p^2s < 0.001).

1.4 Overview of Experimental Design

The detailed experiment procedure is shown in Supplementary Fig. S10.

For Experiment 2, we tested three different methods to measure suspense elicitation. First we used a single-item measure—participants were asked to rate the suspense of the stimuli directly ("How suspenseful was the video/image?"). We then applied a multi-item measure of suspense proposed in [9]. Participants were asked to rate four questions related to suspense ("I was on the edge of my seat.", "It was a nail-biting experience.", "It was a gripping experience.", "It was a tension-filled experience."). Suspense is recognized as closely relate to uncertainty [10], [11]. Thus we further asked participants' perception of uncertainty by rating the question ("The image/video made me feel a lot of uncertainty."). The ratings of the single-item measure strongly correlate with the multi-item measure (ρs > 0.92), and both measures correlate strongly with uncertainty rating (ρs > 0.88), suggesting that all three types of measurements are effective in measuring suspense. To make it simple and to save cost, we select the single-item measure as our suspense index. Previous studies have used similar single-item measure with success [12], [13], [14], [15]. We used bootstrapping to evaluate the consistency of suspense perception for the single-item measure (refer to the next sub-section for details). We got an average Spearman's rank correlation of 0.50 collapsed over all three stimulus versions, indicating that human are moderately consistent in suspense perception.

1.5 Measures for Enhancing Human Data Reliability

Below we describe how we ensured the reliability of the human data before, during, and after our human experiments.

First, we performed a pilot study to determine how many observers we needed to reliably (*i.e.*, in an acceptable consistency level) represent human sentiments of a visual stimuli, for both motion and static versions. We used $F_{entropy}$ to represent all static versions. We randomly picked 10 video clips from our stimulus set eliciting various emotions (*e.g.*, joy, excitement, anger, disgust, fear), and had 100 MTurk workers ($> 95\%$ approval rating in Amazon's system) to rate the 13 sentiments and 2 perceptions for each video clip. We recruited another 100 MTurk workers ($> 95\%$ approval rating in Amazon's system) to collect their ratings of the corresponding frame ($F_{entropy}$).

For both video and $F_{entropy}$, we used bootstrapping to evaluate how consistent were human sentiments for various numbers of participants. More specifically, for multiple group sizes, we randomly split the participants into two equal-sized groups and calculated the Spearman's rank correlation (ρ) between the two groups' average ratings. We did so 25 times per group size. We also calculated the root mean square errors (RMSE) of each sentiment rating in the similar way, using the data of all 100 participants as ground truth. Based on [16], [17], for studies of implicit human mental status, we generally consider correlations above 0.4 to be relatively strong, correlations between 0.2 and 0.4 to be moderate. As shown in Fig. S11, when the number of participants was over 15, ρ of both video and $F_{entropy}$ were above 0.4 and RMSE was below 0.075, suggesting that around 15 observers per stimuli is sufficient to reliably represent human sentiments. We further computed Fleiss' kappa to evaluate the reliability of agreement among raters. The Fleiss' kappa values are greater than .03 for all sentiments. We used 15 participants for each stimuli. Our sample sizes are larger to those reported in previous publications [18], [1], [2].

During the experiment, we only invited participants with approved rate higher than 95% from MTurk's system (this applied to Experiments 1 to 3). We added a human verification step at the beginning of the experiment, to ensure only humans (not robots) participate in our tasks. In the middle of the experiment, we inserted two questions in the questionnaire to check if the participants are clicking the buttons randomly. The two questions are "Are you serious in doing this survey?" and "Are you providing answers randomly?".

We recorded the time used by each participant to complete the experiment. We filtered out the submissions that were completed too fast or too slow (trimming the top and bottom 5% based on task completion time) after the experiment. Previous work has

shown that filtering based on completion velocity filters out low quality submissions [19], [20]. We analysed the consistency of human sentiments in Experiments 1. For each stimulus version, we randomly split the participants into two equal-sized groups and calculated the Spearman's rank order correlation (ρ) between the two groups' ratings for each sentiment. We repeated the process for 25 random splits and computed the mean correlations. Results show that participants had moderate consistency across all versions (averaged Spearman's $\rho > 0.54$), suggesting that the comparison of human sentiments on various stimulus versions is meaningful. We further evaluated human consistency on each stimuli. We excluded 22 stimuli on which the standard deviation of participants ratings of emotional valence was larger than 0.3. This filtering threshold is based a previous study by Schmidt & Stock [17].

2 Supplementary Results

The overall ANOVA results of all sentiments for each human experiment are summarized in Supplementary Table S2. The homogeneous subsets tables of post hoc Tukey tests are reported in Supplementary Tables S3 - S52.

2.1 Supplementary Results for Experiment 1

Table S1 shows the raw ratings (before normalization) for each attribute on different stimulus versions.

TABLE S1
Raw ratings for each attribute.

	video	F	F	F
valence	4.78	4.74	4.69	4.67
arousal	4.60	4.53	4.64	4.40
dynamic	4.49	4.43	4.28	4.50
exciting	4.22	4.32	4.32	4.11
happy	4.24	4.24	4.25	4.18
surprising	4.27	4.18	4.31	4.12
awe	4.12	4.30	4.32	4.11
content	4.14	4.29	4.24	4.08
amusing	4.06	4.22	4.25	3.94
provoking	3.92	4.20	4.15	4.01
unusual	3.87	4.10	4.07	3.73
frightening	3.65	3.91	3.90	3.64
disgusting	3.67	3.84	3.85	3.55
sad	3.59	3.76	3.79	3.47
angry	3.50	3.64	3.68	3.33

2.2 Supplementary Results for Experiment 3

For all stimuli in Experiment 3, $F_{optical}$ had higher ratings on all negative sentiments (Tukey's HSD tests, ps < 0.003, Cohen's ds ≥ 2.17), followed in effect size by $F_{entropy}$ and F_{DNN}.

As shown in Supplementary Fig. S13, both $F_{optical}$ and F_{DNN} elicited higher suspense among observers. Consistent with Experiment 2, the suspense increase was most significant on motions from human and nature. Different from Experiment 2, here suspense was enhanced for other types of motion as well.

2.3 Supplementary Results for In House Experiment

Whereas the first three experiments were completed using online crowdsourcing platform, to further validate our findings we carried out an extra experiment in-house. We recruited 55 participants from the National University of Singapore (24 male, age: 21.90 ± 2.63) to repeat Experiments 1-3. We randomly selected 142 video clips from NVED dataset, and 68 video clips from the validation set, and mixed them together to form a 210 stimulus set. We deployed our experiment platform in the university intranet. Participants logged on to our platform at their own convenient time and premises inside the campus. Each participant was assigned to view a total of 210 stimuli of one stimulus version. The stimuli were grouped into 7 sessions with 30 video clips per session. On average, each participant spent 4 hours within 7 days to complete the whole task. Student ID was used to track the performance for reimbursement purpose.

Mixed ANOVA analyses show that stimulus version was significant for all sentiments (Fs ≥ 6.34, ps ≤ 0.0002, η_p^2s ≥ 0.034). As shown in Supplementary Fig. S14, despite of an overall trend of lower absolute values than Experiments 1 and 3, the main findings were replicated—static versions were as effective as videos in eliciting most sentiments. All static versions ($F_{optical}$, $F_{entropy}$ and F_{DNN}) elicited stronger sentiments than original videos on all negative sentiments (Tukey's HSD tests, ps < 0.005, Cohen's ds ≥ 0.82), and most positive sentiments (Tukey's HSD tests, ps < 0.004, Cohen's ds $\geq .37$). All static versions were more unusual as compared to original videos (Tukey's HSD tests, ps < 0.00001, Cohen's ds ≥ 0.43). $F_{optical}$ and F_{DNN} were more provoking than videos (Tukey's HSD tests, ps < 0.024, Cohen's ds ≥ 0.32). Again F_{DNN} had ratings closest to original videos—not different for 7 out of 15 sentiments and perceptions (Tukey's HSD tests, ps ≥ 0.024, Cohen's ds < 0.20).

Stimulus version significantly affected suspense ($F_{3,632} = 57.83, p < 0.00001$, $\eta_p^2 = 0.259$). As shown in Supplementary Fig. S15, the increase in suspense was more significant here than previous experiments—all static versions ($F_{optical}$, $F_{entropy}$, and F_{DNN}) elicited higher suspense than original videos. Static versions ($F_{optical}$ and F_{DNN}) elicited most significantly stronger suspense on motions of humans (Tukey's HSD tests, ps < 0.0003, Cohen's ds ≥ 0.40), followed by motions of nature for $F_{optical}$ and $F_{entropy}$ (Tukey's HSD tests, ps < 0.0002, Cohen's ds ≥ 0.18). No significant enhancement of suspense was found on motions of tools.

To summarize, the in-house Experiment replicates Experiments 1-3 using a mixed stimulus set with locally-recruited participants. Static versions were more effective than videos in eliciting negative sentiments, corroborating our finding from the previous online experiments.

2.4 Supplementary Results for Experiment 4

The detailed statistics for Experiment 4 are reported in Supplementary Tables S53-S74. Videos ending with $F_{optical}$ elicit similar intensity of emotions with original videos. However, videos starting with $F_{optical}$ and ending with the original last frame elicit weaker emotions. We observe similar but weaker trend for $F_{entropy}$. For F_{DNN}, the trend is different. Videos ending with F_{DNN} elicit similar or weaker negative sentiments than videos starting with F_{DNN}. Both of them elicit weaker intensity of negative sentiments than original videos.

3 Supplementary Figures

Fig. S2. Stimulus examples from validation set. The columns represent motions of different types (left to right: human, animal, nature, vehicle, tools). Each row represents the different versions of the same stimuli. **a**, Screen shots of the original video clips (first frame). **b**, Frames with the largest optical flow ($F_{optical}$). **c**, Frames with the largest entropy ($F_{entropy}$). **d**, Frames with the maximum response to DNN model for video valence prediction (F_{DNN}).

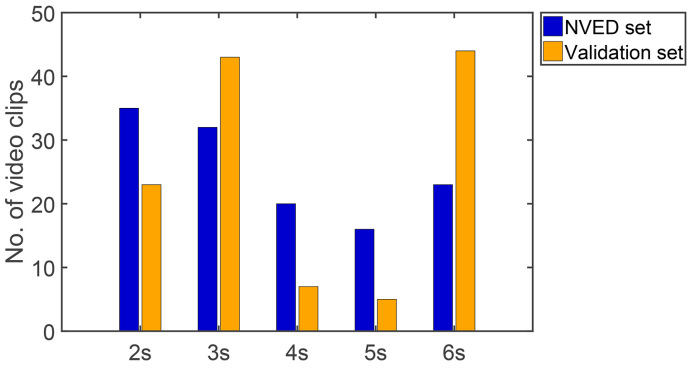


Fig. S1. Distribution histogram of durations of the video clips in NVED and validation stimulus sets.

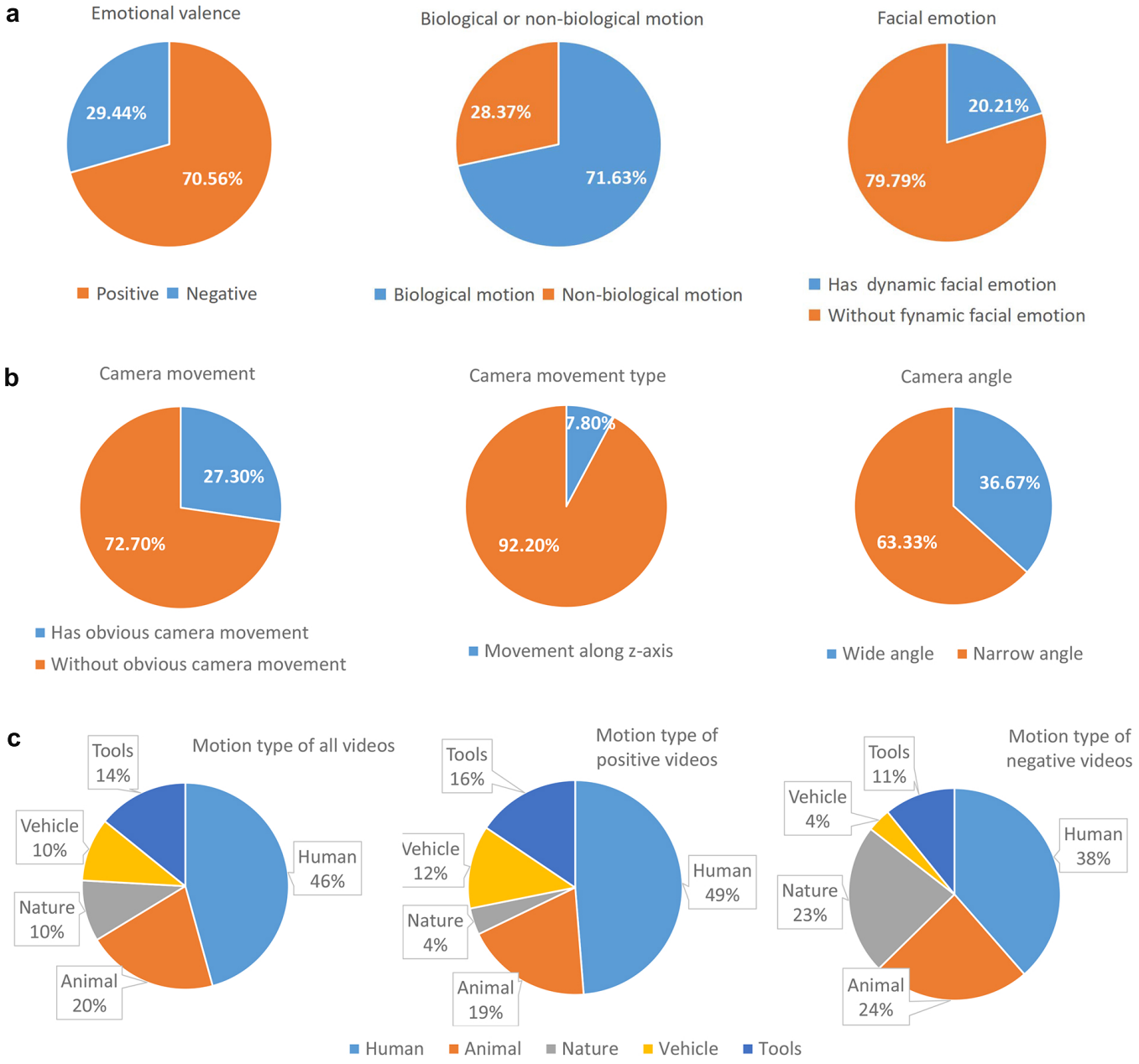


Fig. S3. Statistics of NUS Video Emotion Dataset (NVED) stimulus set. We keep the video clips as diversified as possible for a comprehensive and extendable dataset. **a**, Visualization of diversity in terms of sentiment and semantics. **b**, Visualization of diversity in terms of camera settings. **c**, Visualization of diversity in terms of different stimulus content types.

Fig. S4. Key words for NVED stimulus set. Bigger font of the words represent higher frequency of appearance. **a**, Key words of all video clips. **b**, Key words of positive video clips. **c**, Key words of negative video clips. Videos were grouped into positive and negative by dichotomizing the rating of emotional valence in experiment 1 (averaged across all participants) with a threshold of 0.5.

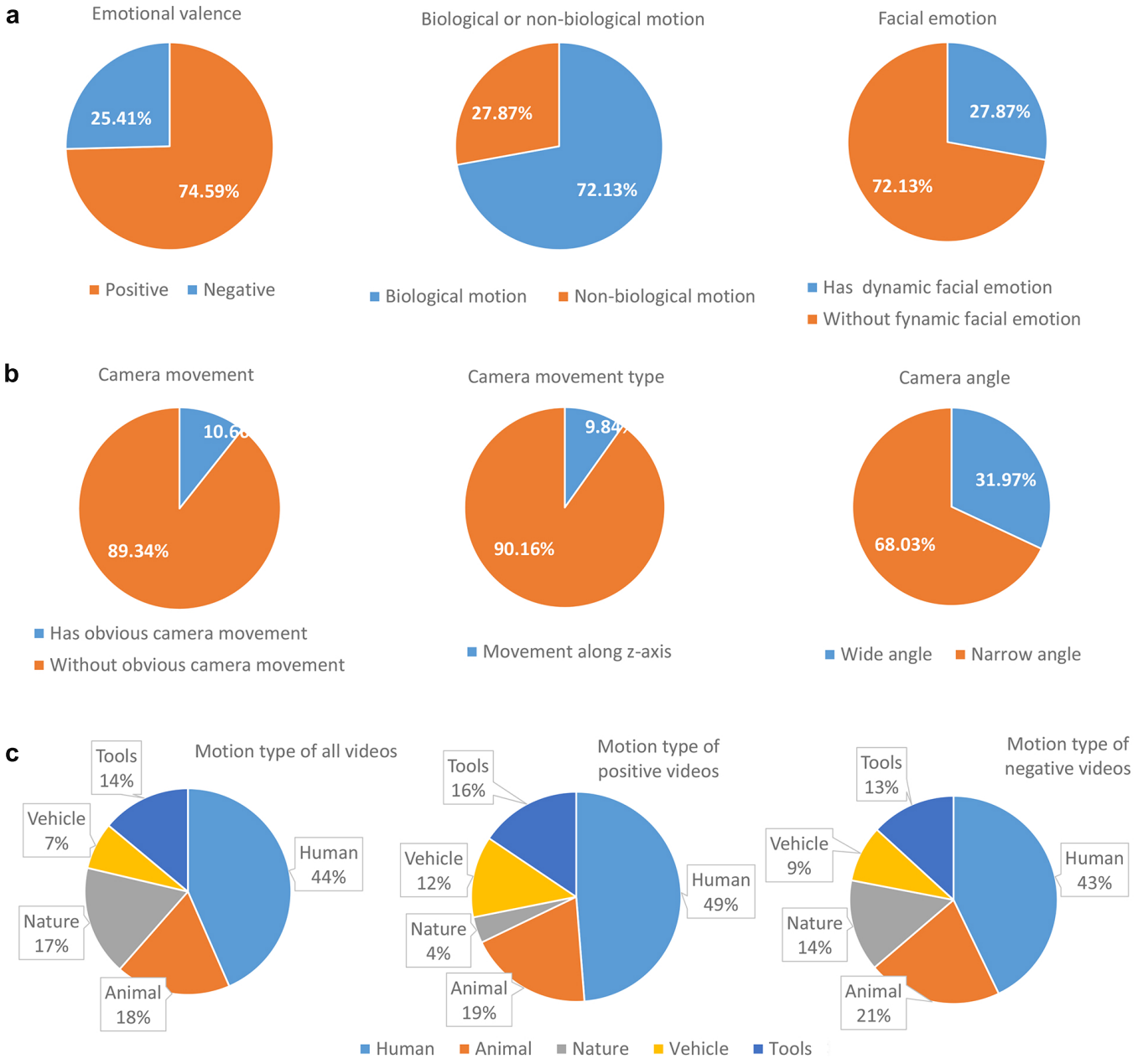


Fig. S5. Statistics of validation stimulus set. The video clips in the validation set are also diversified, which contributes to the venerability of our findings. **a**, Visualization of diversity in terms of sentiment and semantics. **b**, Visualization of diversity in terms of camera settings. **c**, Visualization of diversity in terms of different stimulus content types.

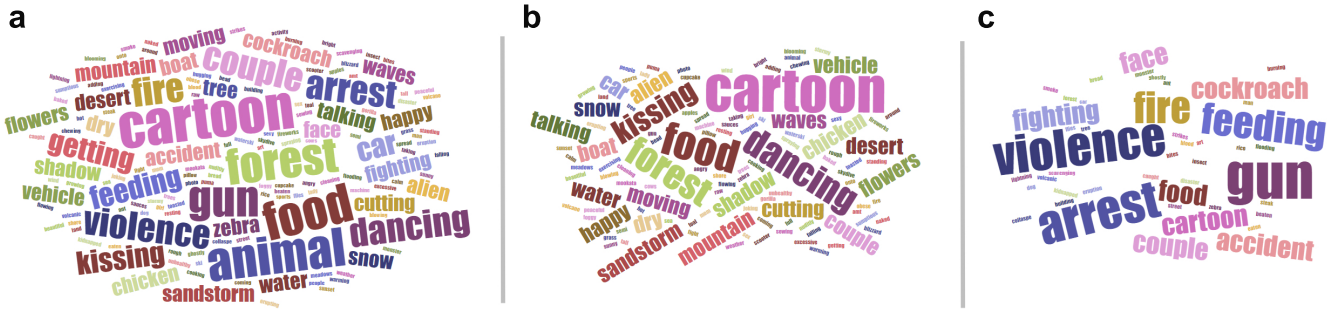


Fig. S6. Key words for validation stimulus set. Higher font of the words represent higher frequency of appearance. **a**, Key words of all video clips. **b**, Key words of positive video clips. **c**, Key words of negative video clips.

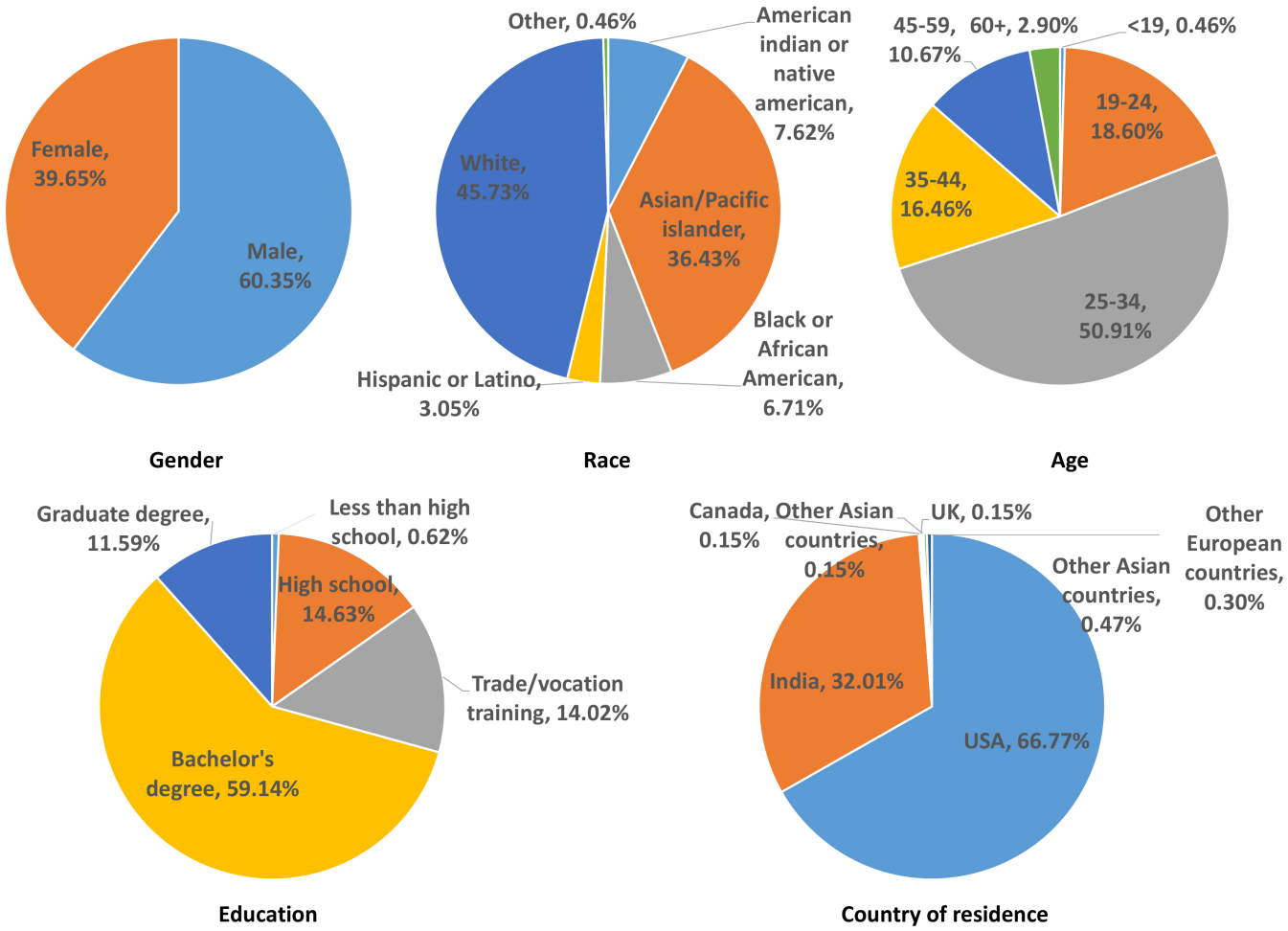


Fig. S7. Demographic statistics of the MTurk participants in Experiment 1. The online participants have a diversified background in terms of gender, race, age, education, and country of residence.

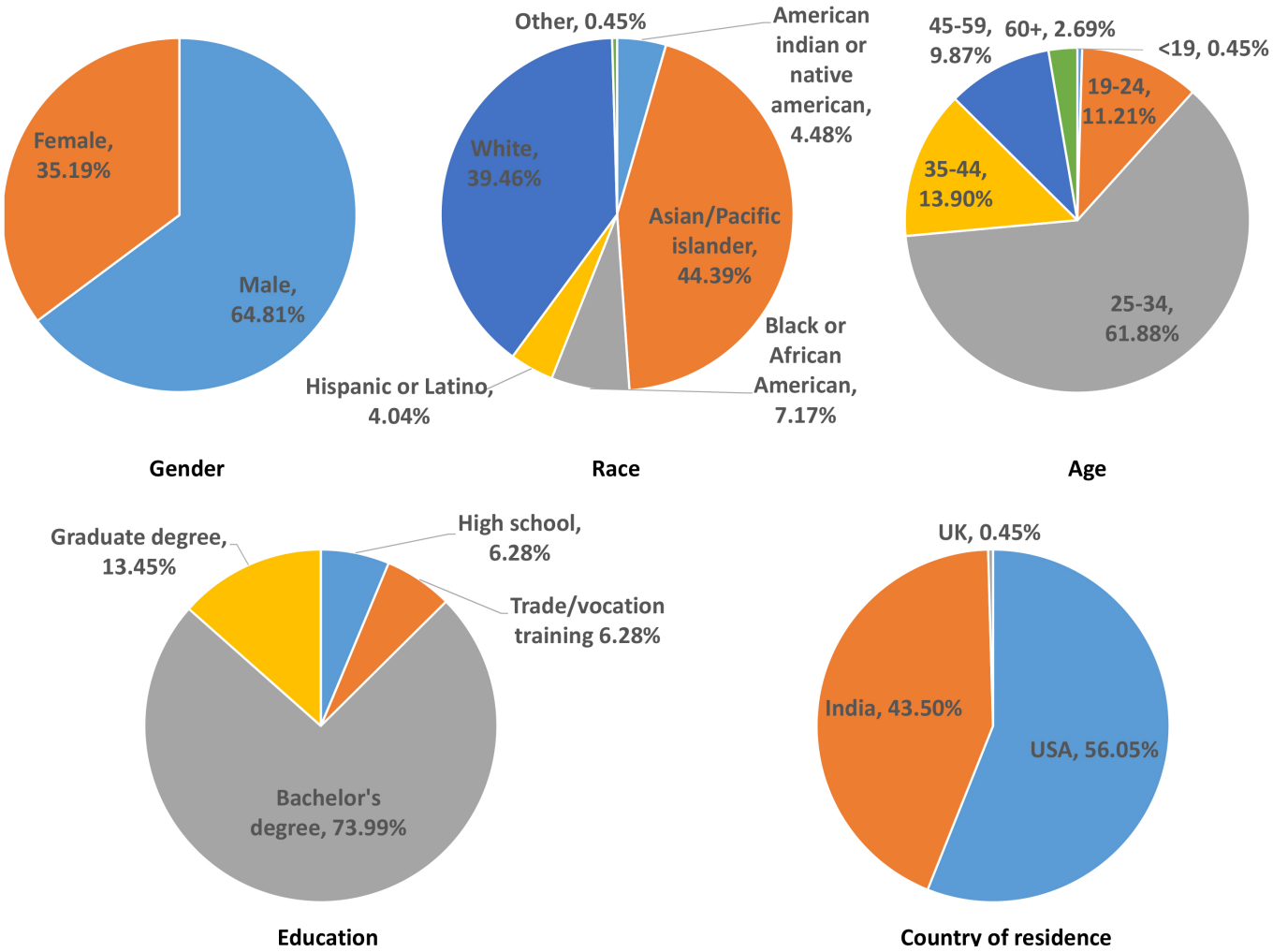


Fig. S8. Demographic statistics of the MTurk participants in Experiment 3 for validation stimulus set.

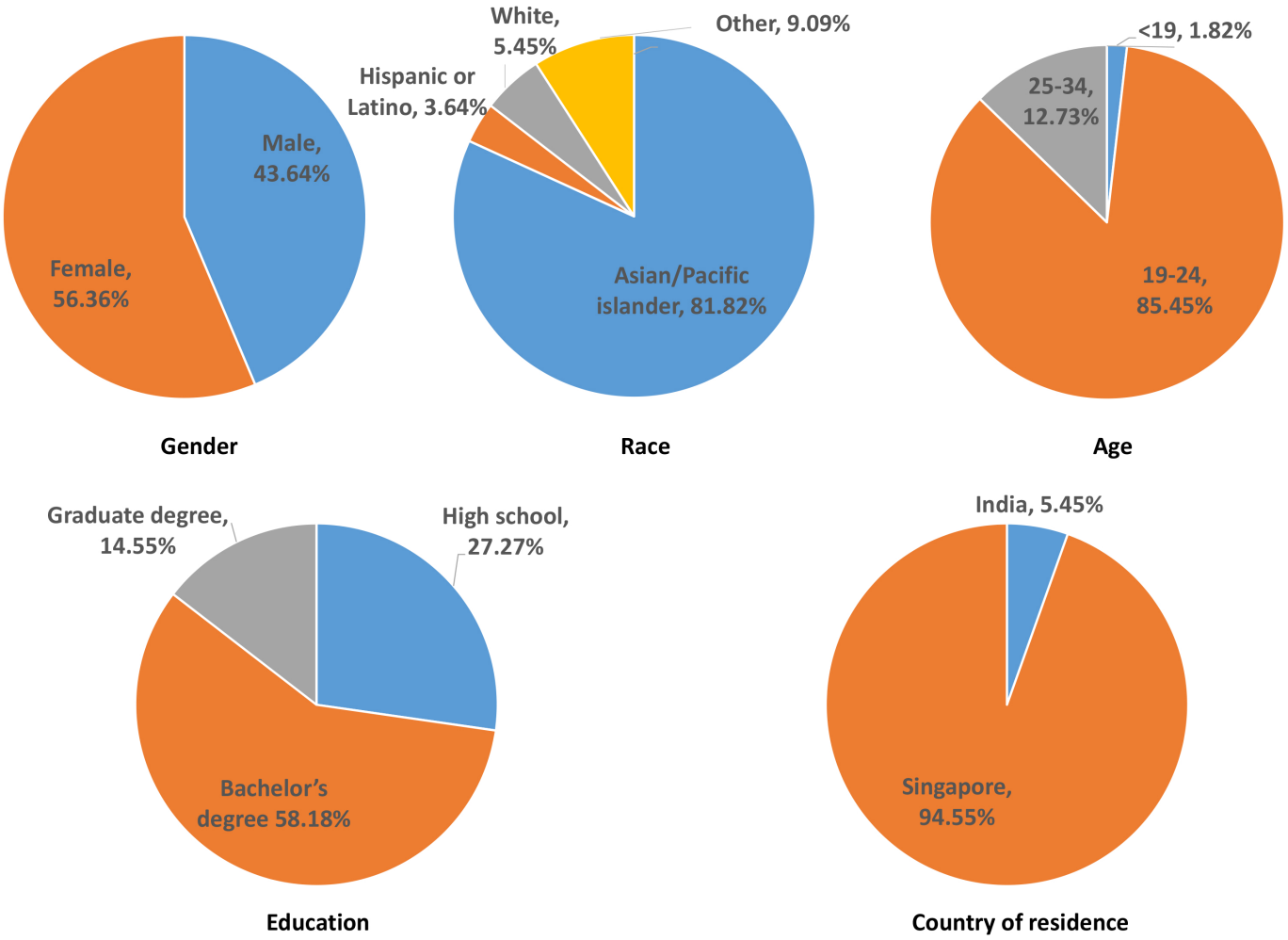


Fig. S9. Demographic statistics of the locally recruited participants in the in house Experiment. Although locally recruited, our participants still have a mixed background in terms of gender, race, age, education, and country of residence.

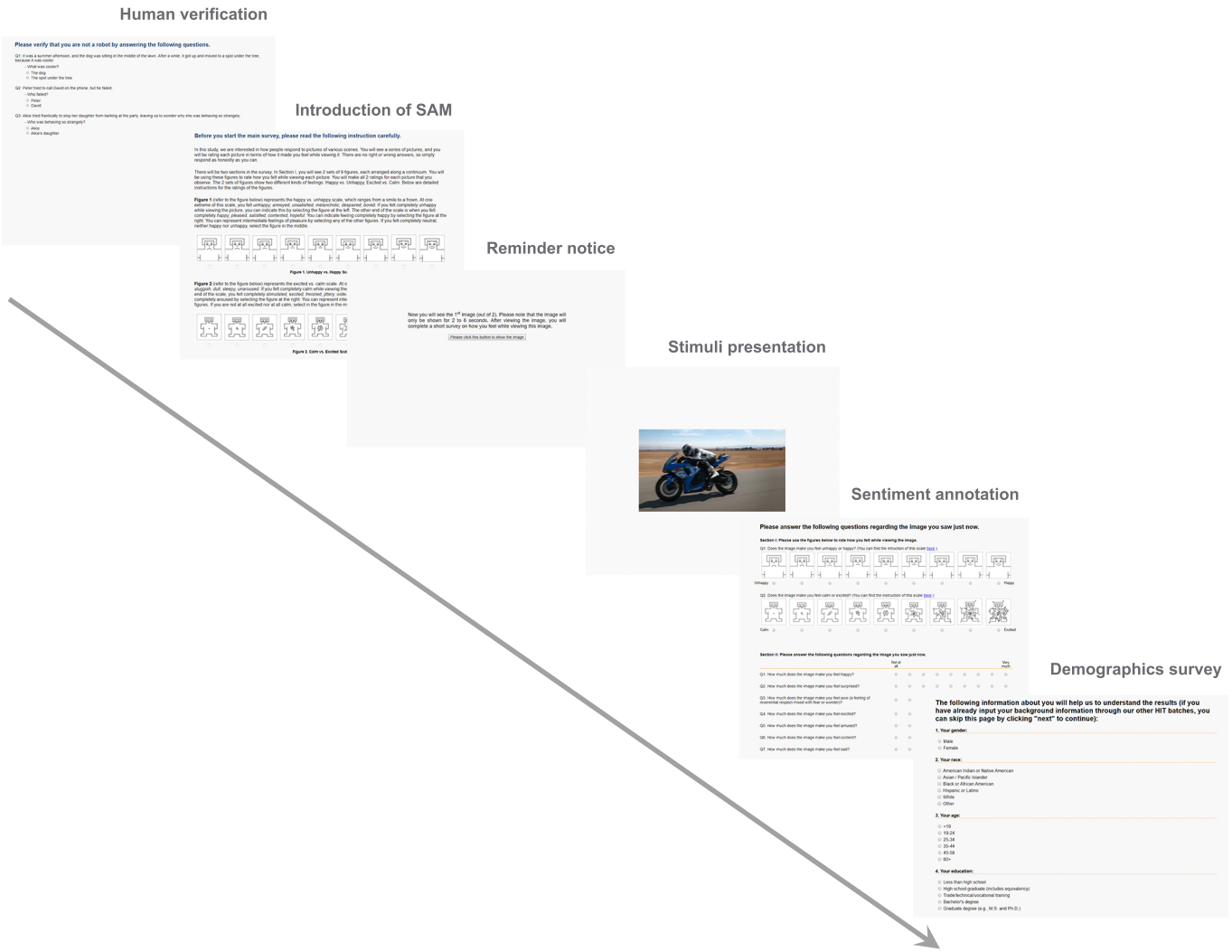


Fig. S10. Experimental procedures for all human experiments in this study. First, participants need to pass human verification. Next, they will be given an introduction of the Self-Assessment Manikin (SAM) rating. They will then be reminded that each stimuli will be shown for only 2 to 6 seconds. They can click the button in the reminder notice when they are ready, and the stimuli will be shown in the center of the screen. After 2 to 6 seconds (the time is dependent on the duration of the original video clip of the stimuli), participants will need to complete a questionnare to indicate how they feel about the stimuli they saw just now. Finally, participants are encouraged to indicate their demographic background.

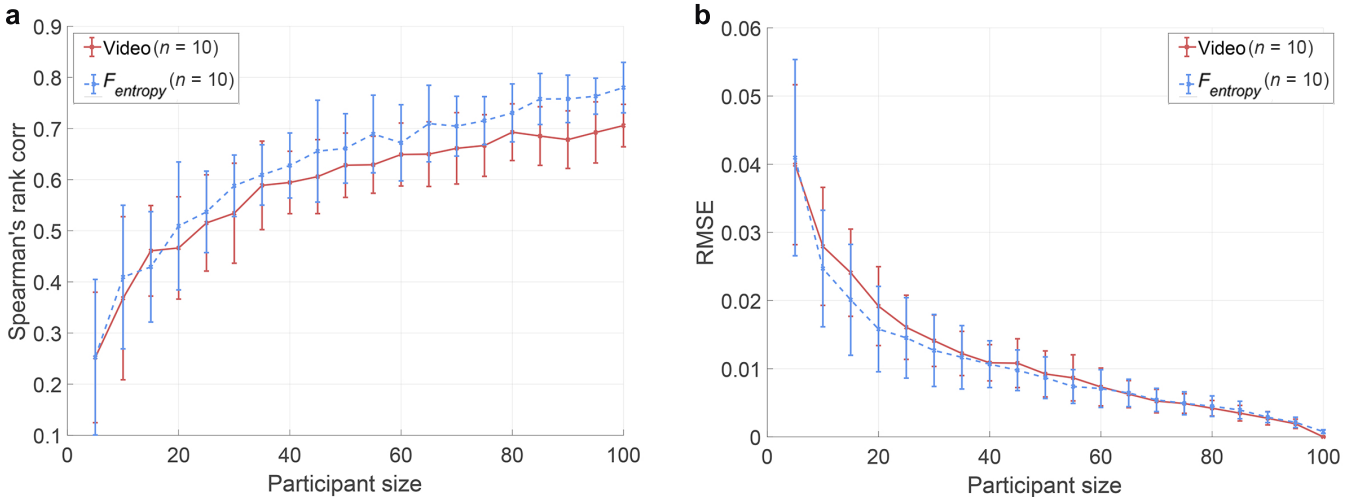


Fig. S11. Human performance analysis on pilot study. We performed a pilot study to determine how many observers we needed to reliably (*i.e.*, at an acceptable consistency level) represent human sentiments of a visual stimuli, for both motion and static versions. Our pilot results suggest that around 15 observers per stimuli is sufficient to reliably represent human sentiments. **a**, Spearman's rank correlation between two random splits of participants as a function of participants size. **b**, Root mean square error of image scores as a function of participants size. All results are averaged over 25 random splits. For all figures in this paper, error bars indicate the standard error of the mean.

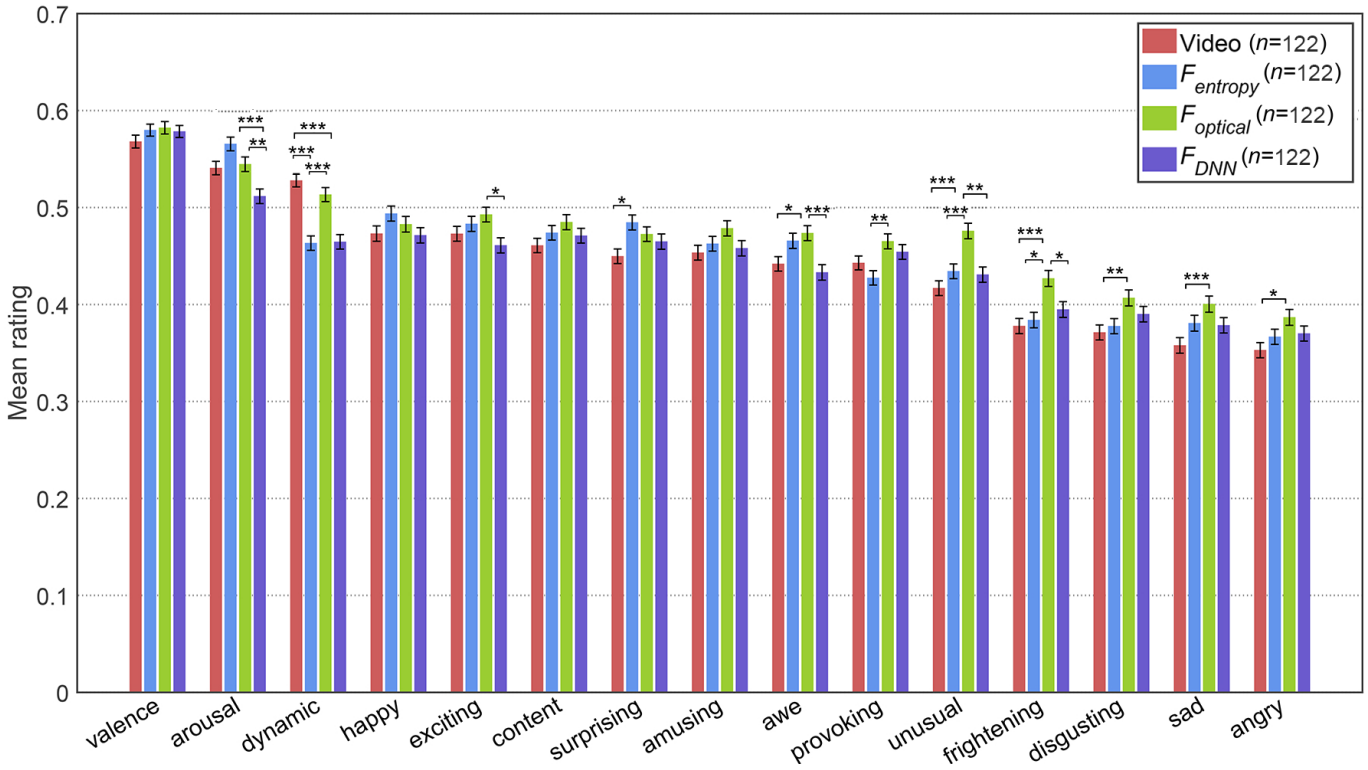


Fig. S12. Results of Experiment 3. Replicating Experiment 1, in Experiment 3 with a different stimulus set and separate group of participants, static frames elicited many sentiments to the same degree, and elicited negative sentiments more strongly, more so for $F_{optical}$. For "Valence" and "Arousal" we used the 9-point scale Self-Assessment Manikin Test [21]. We used a 9-point scale (1 = completely disagree, 9 = completely agree) for the rest sentiments. All ratings are normalized between 0 and 1. Due to multiple dependent variables, we used Bonferroni correction with a reduced significant level α of 0.003. The asterisks are denoted as following: * $p \leq 0.003$, ** $p \leq 0.0006$, *** $p \leq 0.00006$.

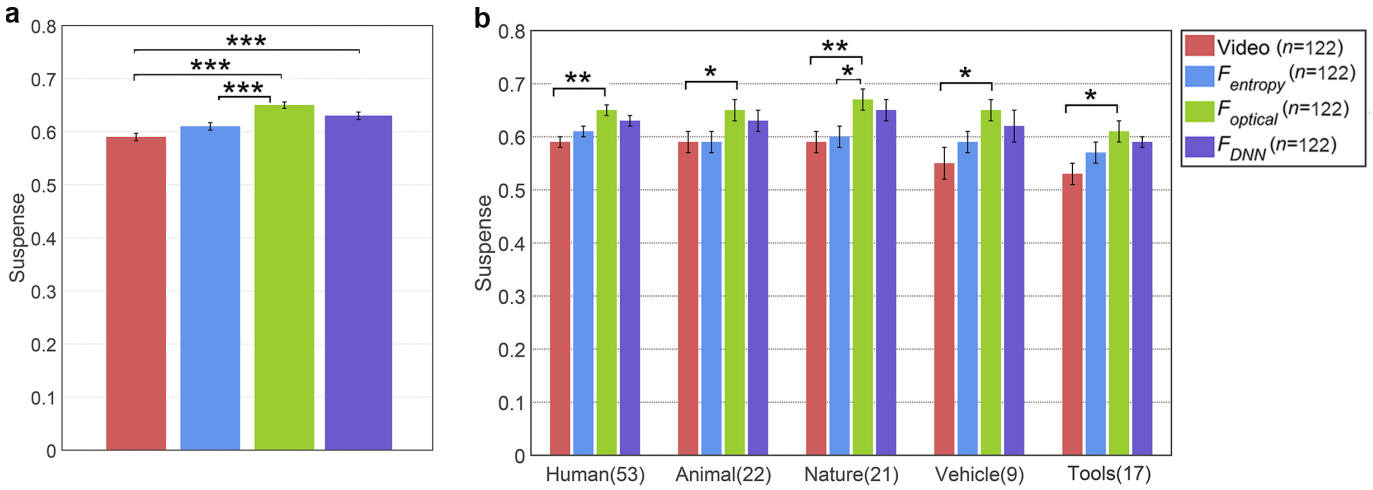


Fig. S13. Results of suspense perception on the validation set in Experiment 3. **a**, $F_{optical}$ and F_{DNN} elicit higher suspense than video on all stimuli. **b**, Results on stimuli grouped by motion categories. Replicating Experiment 2, the increase in suspense is more significant for human and natural motions. The numbers in parentheses show the number of stimuli in each category. The asterisks are denoted as following: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Those without asterisks indicate non-significant results.

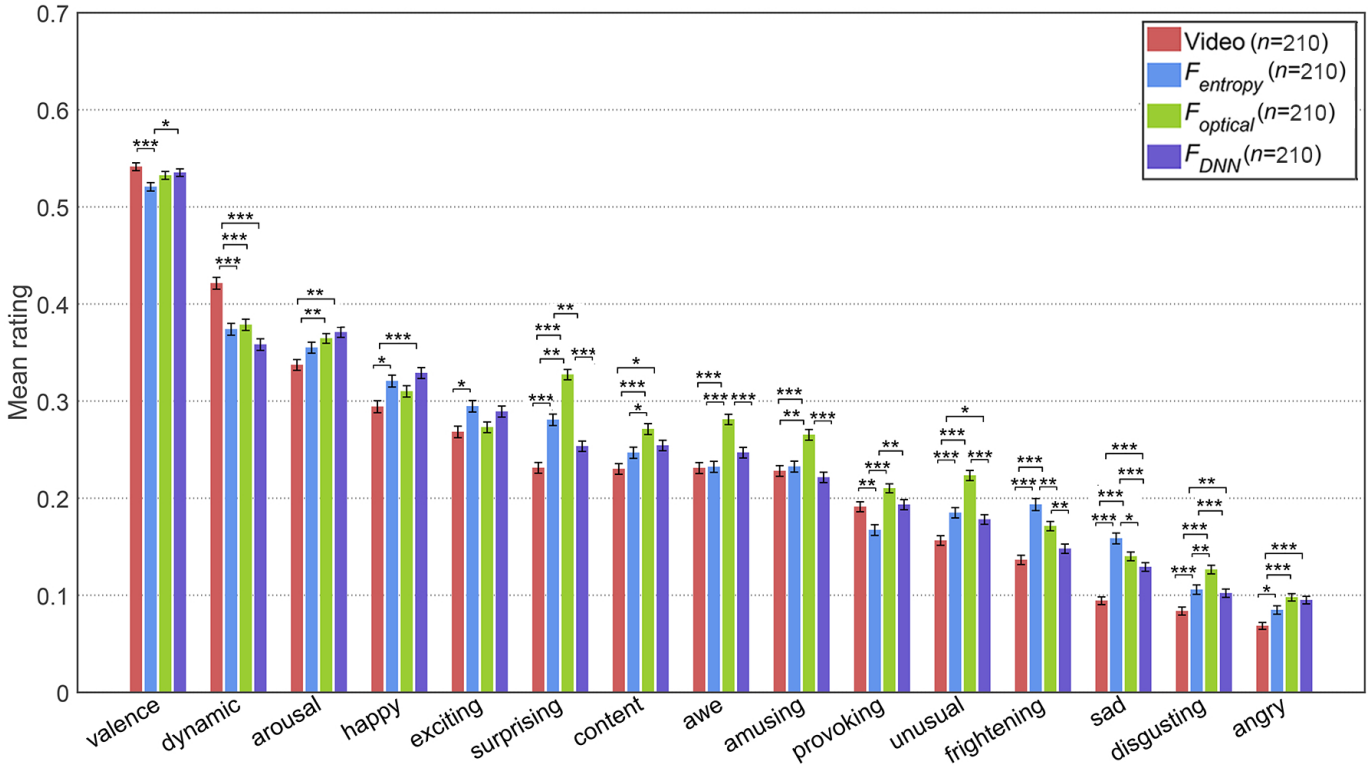


Fig. S14. Results of the in house Experiment. Replicating Experiments 1 and 3, in the in house Experiment with locally recruited participants, videos and static frames had similar effects on most sentiments. Static frames, especially $F_{optical}$, elicited stronger negative sentiments than videos. Compared with Experiments 1 and 3, the average ratings were lower in the in house Experiment. This might be because in Experiment 1 and 3, the number of stimulus each MTurk viewed ranges between 30 to 90. Whereas in the in house Experiment, each participant viewed a total of 210 stimuli from a specific version. The repetitive emotion elicitation within a same participant might lower his/her self-reported ratings. For "Valence" and "Arousal" we used the 9-point scale Self-Assessment Manikin Test [21]. We used a 9-point scale (1 = completely disagree, 9 = completely agree) for the rest sentiments. All ratings are normalized between 0 and 1. Due to multiple dependent variables, we used Bonferroni correction with a reduced significant level α of 0.003. The asterisks are denoted as following: * $p \leq 0.003$, ** $p \leq 0.0006$, *** $p \leq 0.00006$.

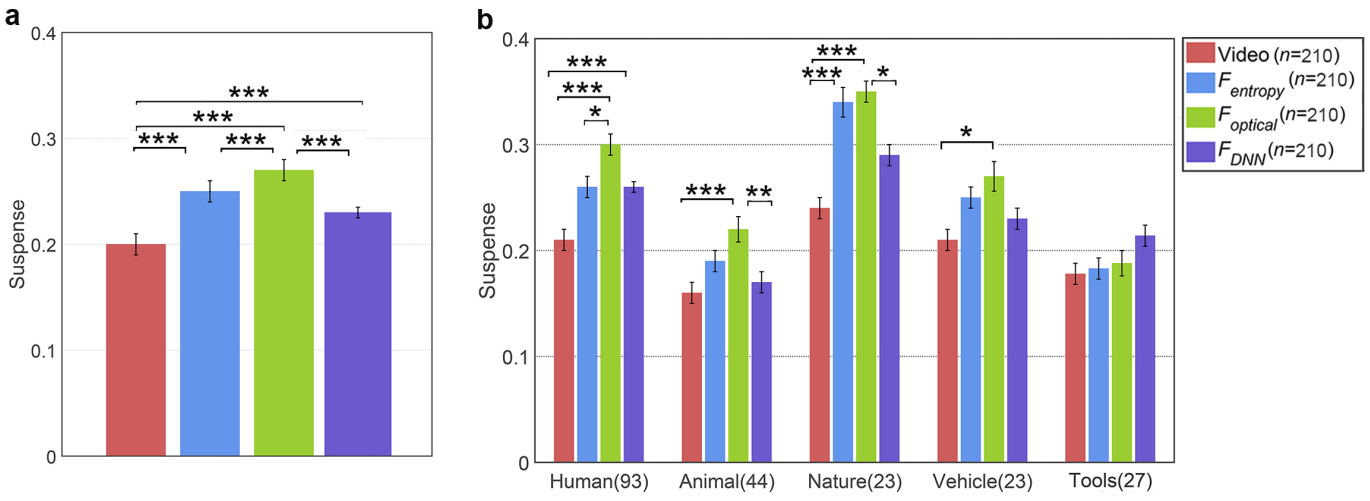


Fig. S15. Results for suspense perception in the in house Experiment. **a**, $F_{optical}$ and F_{DNN} elicit higher suspense than video on all stimuli. **b**, Results on stimuli grouped by motion categories. Replicating Experiment 2 and 3, stimulus version significantly affected suspense, and the effect was most significant for human and natural motions. The numbers in parentheses show the number of stimuli in each category. The asterisks are denoted as following: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Those without asterisks indicate non-significant results.

SUPPLEMENTARY TABLES

TABLE S2

Results on stimulus version of Experiments 1-3 and In house Experiment. The average participant-level rating on each sentiment was set as the dependent variable, stimulus version as a fixed factor and stimuli index (representing different content) as a random factor. The results of suspense perception in Experiment 2 is integrated in Experiment 1.

Sentiments	Experiment 1	Experiment 3	In House Experiment
Valence	$F_{3,860} = 1.03, p = 0.381, \eta_p^2 = 0.004$	$F_{3,346} = 1.97, p = 0.119, \eta_p^2 = 0.017$	$F_{3,635} = 8.68, \quad p = 0.0001, \eta_p^2 = 0.034$
Arousal	$F_{3,863} = 6.26, p = 0.0003, \eta_p^2 = 0.021$	$F_{3,350} = 8.06, p < 0.0001, \eta_p^2 = 0.065$	$F_{3,643} = 14.71, p < 0.00001, \eta_p^2 = 0.102$
Exciting	$F_{3,861} = 4.64, p = 0.003, \eta_p^2 = 0.016$	$F_{3,348} = 4.45, p = 0.004, \eta_p^2 = 0.037$	$F_{3,640} = 6.34, \quad p < 0.0002, \eta_p^2 = 0.108$
Happy	$F_{3,860} = 0.35, p = 0.787, \eta_p^2 = 0.001$	$F_{3,348} = 1.63, p = 0.182, \eta_p^2 = 0.011$	$F_{3,643} = 10.90, p < 0.00001, \eta_p^2 = 0.108$
Surprising	$F_{3,863} = 3.32, p = 0.019, \eta_p^2 = 0.011$	$F_{3,348} = 1.59, p = 0.191, \eta_p^2 = 0.014$	$F_{3,639} = 76.16, p < 0.00001, \eta_p^2 = 0.279$
Awe	$F_{3,861} = 6.60, p = 0.005, \eta_p^2 = 0.022$	$F_{3,350} = 6.22, p = 0.0004, \eta_p^2 = 0.051$	$F_{3,641} = 29.46, p < 0.00001, \eta_p^2 = 0.100$
Content	$F_{3,860} = 4.10, p = 0.007, \eta_p^2 = 0.014$	$F_{3,351} = 3.84, p = 0.011, \eta_p^2 = 0.032$	$F_{3,640} = 19.36, p < 0.00001, \eta_p^2 = 0.094$
Amusing	$F_{3,860} = 10.07, p < 0.00001, \eta_p^2 = 0.034$	$F_{3,349} = 3.88, p = 0.009, \eta_p^2 = 0.032$	$F_{3,640} = 21.98, p < 0.00001, \eta_p^2 = 0.076$
Provoking	$F_{3,860} = 10.07, p < 0.00001, \eta_p^2 = 0.040$	$F_{3,351} = 13.68, p < 0.00001, \eta_p^2 = 0.105$	$F_{3,639} = 18.52, p < 0.00001, \eta_p^2 = 0.034$
Unusual	$F_{3,864} = 16.96, p < 0.00001, \eta_p^2 = 0.056$	$F_{3,357} = 19.30, p < 0.00001, \eta_p^2 = 0.139$	$F_{3,638} = 43.33, p < 0.00001, \eta_p^2 = 0.165$
Frightening	$F_{3,862} = 10.95, \ p < 0.00001, \eta_p^2 = 0.011$	$F_{3,355} = 13.97, p < 0.00001, \eta_p^2 = 0.106$	$F_{3,636} = 35.99, p < 0.00001, \eta_p^2 = 0.204$
Disgusting	$F_{3,862} = 10.29, \ p < 0.00001, \eta_p^2 = 0.035$	$F_{3,355} = 5.83, \quad p = 0.001, \eta_p^2 = 0.047$	$F_{3,639} = 30.76, p < 0.00001, \eta_p^2 = 0.158$
Sad	$F_{3,861} = 9.77, \ p < 0.00001, \eta_p^2 = 0.011$	$F_{3,353} = 6.08, \quad p = 0.001, \eta_p^2 = 0.048$	$F_{3,636} = 51.96, p < 0.00001, \eta_p^2 = 0.305$
Angry	$F_{3,861} = 11.05, \ p < 0.00001, \eta_p^2 = 0.037$	$F_{3,354} = 4.11, \quad p = 0.007, \eta_p^2 = 0.034$	$F_{3,640} = 19.36, p < 0.00001, \eta_p^2 = 0.168$
Dynamic	$F_{3,866} = 10.90, \ p < 0.00001, \eta_p^2 = 0.036$	$F_{3,345} = 36.18, p < 0.00001, \eta_p^2 = 0.239$	$F_{3,636} = 22.16, p < 0.00001, \eta_p^2 = 0.065$
Suspenseful	$F_{3,851} = 3.50, \quad p = 0.015, \eta_p^2 = 0.037$	$F_{3,371} = 14.64, p < 0.00001, \eta_p^2 = 0.106$	$F_{3,632} = 57.83, p < 0.00001, \eta_p^2 = 0.259$

TABLE S3
Homogeneous subsets of "valence" in Experiment 1.

Stimulus type	subset
	1
$F_{entropy}$	0.59
F_{DNN}	0.59
$F_{optical}$	0.59
video	0.60
Sig.	0.149

TABLE S4
Homogeneous subsets of "arousal" in Experiment 1.

Stimulus type	subset	
	1	2
F_{DNN}	0.55	
$F_{optical}$	0.57	0.57
video		0.57
$F_{entropy}$		0.58
Sig.	0.157	0.521

TABLE S5
Homogeneous subsets of "exciting" in Experiment 1.

Stimulus type	subset	
	1	2
F_{DNN}	0.52	
video	0.53	0.53
$F_{optical}$	0.54	0.54
$F_{entropy}$		0.54
Sig.	0.010	0.085

TABLE S6
Homogeneous subsets of "happy" in Experiment 1.

Stimulus type	subset
	1
F_{DNN}	0.52
$F_{optical}$	0.53
video	0.53
$F_{entropy}$	0.53
Sig.	0.823

TABLE S7
Homogeneous subsets of "surprising" in Experiment 1.

Stimulus type	subset
	1
F_{DNN}	0.52
$F_{optical}$	0.52
video	0.53
$F_{entropy}$	0.54
Sig.	0.012

TABLE S8
Homogeneous subsets of "awe" in Experiment 1.

Stimulus type	subset		
	1	2	3
video	0.51		
F_{DNN}	0.52	0.52	
$F_{optical}$		0.54	0.54
$F_{entropy}$			0.54
Sig.	0.843	0.035	0.970

TABLE S9
Homogeneous subsets of "content" in Experiment 1.

Stimulus type	subset
	1
F_{DNN}	0.52
video	0.52
$F_{entropy}$	0.53
$F_{optical}$	0.53
Sig.	0.017

TABLE S10
Homogeneous subsets of "amusing" in Experiment 1.

Stimulus type	subset		
	1	2	3
F_{DNN}	0.50		
video	0.51	0.51	
$F_{optical}$		0.53	0.53
$F_{entropy}$			0.53
Sig.	0.423	0.011	0.878

TABLE S11
Homogeneous subsets of "provoking" in Experiment 1.

Stimulus type	subset	
	1	2
video	0.49	
F_{DNN}	0.51	0.51
$F_{entropy}$		0.52
$F_{optical}$		0.52
Sig.	0.074	0.017

TABLE S12
Homogeneous subsets of "unusual" in Experiment 1.

Stimulus type	subset	
	1	2
F_{DNN}	0.47	
video	0.48	
$F_{entropy}$		0.51
$F_{optical}$		0.51
Sig.	0.284	0.880

TABLE S13
Homogeneous subsets of "frightening" in Experiment 1.

Stimulus type	subset	
	1	2
video	0.46	
F_{DNN}	0.46	
$F_{entropy}$		0.48
$F_{optical}$		0.49
Sig.	0.985	1.000

TABLE S14
Homogeneous subsets of "disgusting" in Experiment 1.

Stimulus type	subset	
	1	2
F_{DNN}	0.45	
video	0.46	
$F_{optical}$		0.48
$F_{entropy}$		0.48
Sig.	0.399	1.000

TABLE S15
Homogeneous subsets of “sad” in Experiment 1.

Stimulus type	subset	
	1	2
F_{DNN}	0.44	
video	0.45	
$F_{optical}$		0.47
$F_{entropy}$		0.47
Sig.	0.457	0.938

TABLE S16
Homogeneous subsets of “angry” in Experiment 1.

Stimulus type	subset		
	1	2	3
F_{DNN}	0.42		
video	0.44	0.44	
$F_{entropy}$		0.45	0.45
$F_{optical}$			0.46
Sig.	0.084	0.039	0.917

TABLE S17
Homogeneous subsets of “dynamic” in Experiment 1.

Stimulus type	subset	
	1	2
$F_{entropy}$	0.54	
$F_{optical}$	0.55	0.55
video		0.56
F_{DNN}		0.57
Sig.	0.025	0.068

TABLE S18
Homogeneous subsets of “suspenseful” in Experiment 2.

Stimulus type	subset	
	1	2
video	0.56	
F_{DNN}	0.57	0.57
$F_{entropy}$	0.57	0.57
$F_{optical}$		0.58
Sig.	0.847	0.037

TABLE S19
Regression results for Experiment 2 for stimuli of human and nature scenes. For each sentiment as the dependent variable, two regression models were run. In model 1, predictors were $F_{entropy}$, $F_{optical}$, and F_{DNN} (dummy coded with full video as the reference stimulus). Model 2 added suspense as a predictor. F_{DNN} is excluded from the table as it was never significant. Numbers in model columns are beta coefficients (*p $< .05$).

	$F_{entropy}$		$F_{optical}$		Suspense	R2-change
DV	Model 1	Model 2	Model 1	Model 2	Model 2	
Valence	.01	.04	.01	.07	-.49*	.24
Arousal	.05	.07	.03	.00	-.26*	.19
Happy	.03	.06	-.01	.04	-.44*	.19
Surprising	.05	.07	-.04	-.01	-.24*	.06
Awe	.14*	.16*	.09	.12*	-.26*	.06
Exciting	.07	.09	.03	.07	-.31*	.09
Amusing	.12*	.14*	.08	.12*	-.36*	.13
Content	.09	.11*	.08	.12*	-.35*	.12
Sad	.12*	.10*	.10*	.08	.24*	.06
Angry	.13*	.12*	.12*	.10*	.17*	.03
Frightening	.15*	.13*	.19*	.16*	.25*	.06
Disgusting	.14*	.14*	.14*	.14*	.00	.00
Provoking	.16*	.15*	.19*	.19*	.10	.02
Unusual	.13*	.13*	.16*	.15*	.06	.00

TABLE S20
Homogeneous subsets of "valence" in Experiment 3.

Stimulus type	subset
	1
video	0.57
F_{DNN}	0.58
$F_{entropy}$	0.58
$F_{optical}$	0.58
Sig.	0.191

TABLE S21
Homogeneous subsets of "arousal" in Experiment 3.

Stimulus type	subset	
	1	2
F_{DNN}	0.51	
video		0.54
$F_{optical}$		0.54
$F_{entropy}$		0.57
Sig.	1.000	0.030

TABLE S22
Homogeneous subsets of "exciting" in Experiment 3.

Stimulus type	subset	
	1	2
F_{DNN}	0.46	
video	0.47	0.47
$F_{entropy}$	0.48	0.48
$F_{optical}$		0.49
Sig.	0.081	0.146

TABLE S23
Homogeneous subsets of "happy" in Experiment 3.

Stimulus type	subset
	1
F_{DNN}	0.47
video	0.47
$F_{optical}$	0.48
$F_{entropy}$	0.49
Sig.	0.064

TABLE S24
Homogeneous subsets of "surprising" in Experiment 3.

Stimulus type	subset	
	1	2
video	0.45	
F_{DNN}	0.46	0.46
$F_{optical}$	0.47	0.47
$F_{entropy}$		0.48
Sig.	0.071	0.151

TABLE S25
Homogeneous subsets of "awe" in Experiment 3.

Stimulus type	subset		
	1	2	3
F_{DNN}	0.43		
video	0.44	0.44	
$F_{entropy}$		0.47	0.47
$F_{optical}$			0.47
Sig.	0.786	0.055	0.834

TABLE S26
Homogeneous subsets of "content" in Experiment 3.

Stimulus type	subset
	1
video	0.46
F_{DNN}	0.47
$F_{entropy}$	0.47
$F_{optical}$	0.48
Sig.	0.044

TABLE S27
Homogeneous subsets of "amusing" in Experiment 3.

Stimulus type	subset
	1
video	0.45
F_{DNN}	0.46
$F_{entropy}$	0.46
$F_{optical}$	0.48
Sig.	0.039

TABLE S28
Homogeneous subsets of "provoking" in Experiment 3.

Stimulus type	subset	
	1	2
$F_{entropy}$	0.43	
video	0.44	0.44
F_{DNN}	0.45	0.45
$F_{optical}$		0.47
Sig.	0.022	0.078

TABLE S29
Homogeneous subsets of "unusual" in Experiment 3.

Stimulus type	subset	
	1	2
video	0.42	
F_{DNN}	0.43	
$F_{entropy}$	0.43	
$F_{optical}$		0.48
Sig.	0.273	1.000

TABLE S30
Homogeneous subsets of "frightening" in Experiment 3.

Stimulus type	subset	
	1	2
video	0.38	
$F_{entropy}$	0.38	
F_{DNN}	0.39	
$F_{optical}$		0.43
Sig.	0.323	1.000

TABLE S31
Homogeneous subsets of "disgusting" in Experiment 3.

Stimulus type	subset	
	1	2
video	0.37	
$F_{entropy}$	0.38	0.38
F_{DNN}	0.39	0.39
$F_{optical}$		0.41
Sig.	0.232	0.018

TABLE S32
Homogeneous subsets of “sad” in Experiment 3.

Stimulus type	subset	
	1	2
video	0.36	
F_{DNN}	0.38	0.38
$F_{entropy}$	0.39	0.38
$F_{optical}$		0.40
Sig.	0.101	0.128

TABLE S33
Homogeneous subsets of “angry” in Experiment 3.

Stimulus type	subset	
	1	2
video	0.35	
$F_{entropy}$	0.37	0.37
F_{DNN}	0.37	0.37
$F_{optical}$		0.39
Sig.	0.311	0.184

TABLE S34
Homogeneous subsets of “dynamic” in Experiment 3.

Stimulus type	subset	
	1	2
$F_{entropy}$	0.46	
F_{DNN}	0.46	
$F_{optical}$		0.51
video		0.53
Sig.	0.999	0.354

TABLE S35
Homogeneous subsets of “suspenseful” in Experiment 3.

Stimulus type	subset		
	1	2	3
video	0.59		
$F_{entropy}$	0.61	0.61	
F_{DNN}		0.63	0.63
$F_{optical}$			0.65
Sig.	0.212	0.092	0.060

TABLE S36
Regression results for Experiment 3 for stimuli of human and nature scenes. For each sentiment as the dependent variable, two regression models were run. In model 1, predictors were $F_{entropy}$, $F_{optical}$, and F_{DNN} (dummy coded with full video as the reference stimulus). Model 2 added suspense as a predictor. F_{DNN} is excluded from the table as it was not significant for most sentiments. Numbers in model columns are beta coefficients (*p $< .05$).

	$F_{entropy}$		$F_{optical}$		Suspense	R2-change
DV	Model 1	Model 2	Model 1	Model 2	Model 2	
Valence	-.12	-.08	-.02	-.02	-.39*	.15
Arousal	-.05	-.03	.00	.00	-.24*	.05
Happy	-.05	-.01	.00	.00	-.36*	.13
Surprising	.07	.10	.10	.09	-.26*	.06
Awe	.00	.04	.11	.11	-.33*	.11
Exciting	.00	.04	.09	.09	-.29*	.09
Amusing	.04	.08	.12	.12	-.30*	.09
Content	-.01	.02	.19*	.19*	-.31*	.10
Sad	.11	.11	.10	.10	-.03	.00
Angry	.14*	.15*	.12	.12	-.08	.01
Frightening	.17*	.17*	.16*	.16*	-.01	.00
Disgusting	.17*	.17*	.22*	.22*	-.05	.00
Provoking	.11	.12	.19*	.19*	.11	.01
Unusual	-.13	-.11	.00	.00	-.15*	.02

TABLE S37
Homogeneous subsets of “valence” in the in house Experiment.

Stimulus type	subset	
	1	2
$F_{entropy}$	0.52	
$F_{optical}$	0.53	0.53
F_{DNN}		0.53
video		0.54
Sig.	0.035	0.168

TABLE S38
Homogeneous subsets of “arousal” in the in house Experiment.

Stimulus type	subset	
	1	2
video	0.34	
$F_{entropy}$	0.36	0.36
$F_{optical}$		0.36
F_{DNN}		0.37
Sig.	0.099	0.172

TABLE S39
Homogeneous subsets of “exciting” in the in house Experiment.

Stimulus type	subset	
	1	2
video	0.27	
$F_{optical}$	0.27	0.27
F_{DNN}	0.29	0.29
$F_{entropy}$		0.29
Sig.	0.048	0.038

TABLE S40
Homogeneous subsets of “happy” in the in house Experiment.

Stimulus type	subset	
	1	2
video	0.29	
$F_{optical}$	0.31	0.31
$F_{entropy}$	0.29	0.32
F_{DNN}		0.33
Sig.	0.149	0.055

TABLE S41
Homogeneous subsets of “surprising” in the in house Experiment.

Stimulus type	subset		
	1	2	3
video	0.23		
F_{DNN}	0.25		
$F_{entropy}$		0.28	
$F_{optical}$			0.33
Sig.	0.026	1.000	1.000

TABLE S42
Homogeneous subsets of “awe” in the in house Experiment.

Stimulus type	subset	
	1	2
video	0.23	
$F_{entropy}$	0.23	
F_{DNN}	0.25	
$F_{optical}$		0.28
Sig.	0.177	1.000

TABLE S43
Homogeneous subsets of “content” in the in house Experiment.

Stimulus type	subset		
	1	2	3
video	0.23		
$F_{entropy}$	0.25		
F_{DNN}		0.25	0.25
$F_{optical}$			0.27
Sig.	0.106	0.738	0.100

TABLE S44
Homogeneous subsets of “amusing” in the in house Experiment.

Stimulus type	subset	
	1	2
F_{DNN}	0.22	
video	0.23	
$F_{entropy}$	0.23	
$F_{optical}$		0.27
Sig.	0.449	1.000

TABLE S45
Homogeneous subsets of “provoking” in the in house Experiment.

Stimulus type	subset	
	1	2
$F_{entropy}$	0.17	
video		0.19
F_{DNN}		0.19
$F_{optical}$		0.21
Sig.	1.000	0.023

TABLE S46
Homogeneous subsets of “unusual” in the in house Experiment.

Stimulus type	subset		
	1	2	3
video	0.16		
F_{DNN}		0.17	
$F_{entropy}$		0.18	
$F_{optical}$			0.22
Sig.	1.000	0.745	1.000

TABLE S47
Homogeneous subsets of “frightening” in the in house Experiment.

Stimulus type	subset		
	1	2	3
video	0.14		
F_{DNN}	0.15		
$F_{optical}$		0.17	
$F_{entropy}$			0.19
Sig.	0.250	1.000	1.000

TABLE S48
Homogeneous subsets of “frightening” in the in house Experiment.

Stimulus type	subset		
	1	2	3
video	0.08		
F_{DNN}		0.10	
$F_{entropy}$		0.11	
$F_{optical}$			0.13
Sig.	1.000	0.898	1.000

TABLE S49
Homogeneous subsets of “sad” in the in house Experiment.

Stimulus type	subset		
	1	2	3
video	0.09		
F_{DNN}		0.13	
$F_{optical}$		0.14	
$F_{entropy}$			0.16
Sig.	1.000	0.184	1.000

TABLE S50
Homogeneous subsets of “angry” in the in house Experiment.

Stimulus type	subset	
	1	2
video	0.07	
$F_{entropy}$		0.08
F_{DNN}		0.10
$F_{optical}$		0.10
Sig.	1.000	0.050

TABLE S51
Homogeneous subsets of “dynamic” in the in house Experiment.

Stimulus type	subset	
	1	2
F_{DNN}	0.36	
$F_{entropy}$	0.37	
$F_{optical}$	0.38	
video		0.42
Sig.	0.059	1.000

TABLE S52
Homogeneous subsets of “suspenseful” in the in house Experiment.

Stimulus type	subset		
	1	2	3
video	0.20		
$F_{entropy}$		0.26	
F_{DNN}		0.23	
$F_{optical}$			0.27
Sig.	1.000	0.155	1.000

TABLE S53
Results on video type of Experiment 4 for videos starting and ending with three different static frames. The average participant-level rating on each sentiment was set as the dependent variable, and stimulus version as an independent variable.

Sentiments	$F_{optical}$	$F_{entropy}$	F_{DNN}
Valence	$F_{2,239} = 1.38, p = 0.254, \eta_p^2 = 0.012$	$F_{2,137} = 0.32, p = 0.730, \eta_p^2 = 0.005$	$F_{2,206} = 0.44, \quad p = 0.645, \eta_p^2 = 0.004$
Arousal	$F_{2,239} = 4.02, p = 0.019, \eta_p^2 = 0.033$	$F_{2,137} = 2.86, p = 0.061, \eta_p^2 = 0.041$	$F_{2,206} = 2.40, p = 0.094, \eta_p^2 = 0.023$
Exciting	$F_{2,239} = 1.92, p = 0.148, \eta_p^2 = 0.016$	$F_{2,137} = 1.43, p = 0.244, \eta_p^2 = 0.021$	$F_{2,206} = 3.11, \quad p = 0.047, \eta_p^2 = 0.030$
Happy	$F_{2,239} = 3.37, p = 0.036, \eta_p^2 = 0.028$	$F_{2,137} = 0.29, p = 0.748, \eta_p^2 = 0.004$	$F_{2,206} = 0.37, p = 0.689, \eta_p^2 = 0.004$
Surprising	$F_{2,239} = 1.00, p = 0.369, \eta_p^2 = 0.008$	$F_{2,137} = 2.98, p = 0.054, \eta_p^2 = 0.042$	$F_{2,206} = 2.90, p = 0.057, \eta_p^2 = 0.028$
Awe	$F_{2,239} = 2.73, p = 0.067, \eta_p^2 = 0.023$	$F_{2,137} = 0.96, p = 0.385, \eta_p^2 = 0.014$	$F_{2,206} = 0.10, p = 0.907, \eta_p^2 = 0.001$
Content	$F_{2,239} = 4.44, p = 0.013, \eta_p^2 = 0.036$	$F_{2,137} = 1.94, p = 0.148, \eta_p^2 = 0.028$	$F_{2,206} = 0.79, p = 0.455, \eta_p^2 = 0.008$
Amusing	$F_{2,239} = 1.48, p = 0.012, \eta_p^2 = 0.004$	$F_{2,137} = 0.16, p = 0.849, \eta_p^2 = 0.002$	$F_{2,206} = 0.99, p = 0.374, \eta_p^2 = 0.010$
Dynamic	$F_{2,239} = 1.16, p = 0.315, \eta_p^2 = 0.019$	$F_{2,137} = 15.72, p < 0.00001, \eta_p^2 = 0.189$	$F_{2,206} = 0.57, p = 0.567, \eta_p^2 = 0.006$
Provoking	$F_{2,239} = 17.93, p < 0.00001, \eta_p^2 = 0.199$	$F_{2,137} = 5.97, \quad p < 0.00001, \eta_p^2 = 0.081$	$F_{2,206} = 3.96, p = 0.021, \eta_p^2 = 0.037$
Unusual	$F_{2,239} = 9.45, \quad p < 0.0002, \eta_p^2 = 0.119$	$F_{2,137} = 0.28, \quad p = 0.759, \eta_p^2 = 0.004$	$F_{2,206} = 21.77, p < 0.00001, \eta_p^2 = 0.176$
Frightening	$F_{2,239} = 7.59, \quad p < 0.001, \eta_p^2 = 0.084$	$F_{2,137} = 7.21, \quad p < 0.002, \eta_p^2 = 0.097$	$F_{2,206} = 26.61, p < 0.00001, \eta_p^2 = 0.207$
Disgusting	$F_{2,239} = 9.22, \quad p < 0.0002, \eta_p^2 = 0.111$	$F_{2,137} = 2.77, \quad p = 0.066, \eta_p^2 = 0.039$	$F_{2,206} = 14.49, p < 0.00001, \eta_p^2 = 0.124$
Sad	$F_{2,239} = 8.14, \quad p < 0.0004, \eta_p^2 = 0.103$	$F_{2,137} = 24.94, \ p < 0.00001, \eta_p^2 = 0.270$	$F_{2,206} = 54.96, p < 0.00001, \eta_p^2 = 0.350$
Angry	$F_{2,239} = 7.30, \quad p < 0.001, \eta_p^2 = 0.081$	$F_{2,137} = 18.65, \ p < 0.00001, \eta_p^2 = 0.216$	$F_{2,206} = 47.41, p < 0.00001, \eta_p^2 = 0.317$
Suspenseful	$F_{2,239} = 95.60, \ p < 0.00001, \eta_p^2 = 0.447$	$F_{2,137} = 61.51, \ p < 0.00001, \eta_p^2 = 0.477$	$F_{2,206} = 64.76, p < 0.00001, \eta_p^2 = 0.388$

TABLE S54
Homogeneous subsets of "sad" in Experiment 4 for $F_{optical}$.

Stimulus type	subset	
	1	2
video starting with $F_{optical}$	0.38	
video ending with $F_{optical}$	0.42	0.42
original video		0.45
Sig.	0.085	0.141

TABLE S55
Homogeneous subsets of "angry" in Experiment 4 for $F_{optical}$.

Stimulus type	subset	
	1	2
video starting with $F_{optical}$	0.38	
video ending with $F_{optical}$	0.41	0.42
original video		0.44
Sig.	0.110	0.172

TABLE S56
Homogeneous subsets of "frightening" in Experiment 4 for $F_{optical}$.

Stimulus type	subset	
	1	2
video starting with $F_{optical}$	0.40	
video ending with $F_{optical}$	0.43	0.43
original video		0.46
Sig.	0.052	0.280

TABLE S57
Homogeneous subsets of "disgusting" in Experiment 4 for $F_{optical}$.

Stimulus type	subset	
	1	2
video starting with $F_{optical}$	0.39	
video ending with $F_{optical}$		0.45
original video		0.46
Sig.	1.000	0.962

TABLE S58
Homogeneous subsets of "provoking" in Experiment 4 for $F_{optical}$.

Stimulus type	subset	
	1	2
video starting with $F_{optical}$	0.42	
video ending with $F_{optical}$		0.48
original video		0.50
Sig.	1.000	0.457

TABLE S59
Homogeneous subsets of "unusual" in Experiment 4 for $F_{optical}$.

Stimulus type	subset	
	1	2
video starting with $F_{optical}$	0.44	
video ending with $F_{optical}$		0.49
original video		0.50
Sig.	1.000	0.820

TABLE S60
Homogeneous subsets of "suspenseful" in Experiment 4 for $F_{optical}$.

Stimulus type	subset		
	1	2	3
video starting with $F_{optical}$	0.41		
video ending with $F_{optical}$		0.50	
original video			0.60
Sig.	1.000	1.000	1.000

TABLE S61
Homogeneous subsets of "sad" in Experiment 4 for $F_{entropy}$.

Stimulus type	subset	
	1	2
video starting with $F_{entropy}$	0.29	
video ending with $F_{entropy}$		0.40
original video		0.44
Sig.	1.000	0.139

TABLE S62
Homogeneous subsets of "angry" in Experiment 4 for $F_{entropy}$.

Stimulus type	subset	
	1	2
video starting with $F_{entropy}$	0.31	
video ending with $F_{entropy}$		0.39
original video		0.43
Sig.	1.000	0.145

TABLE S63
Homogeneous subsets of "frightening" in Experiment 4 for $F_{entropy}$.

Stimulus type	subset	
	1	2
video starting with $F_{entropy}$	0.38	
video ending with $F_{entropy}$	0.41	0.41
original video		0.46
Sig.	0.458	0.034

TABLE S64
Homogeneous subsets of "disgusting" in Experiment 4 for $F_{entropy}$.

Stimulus type	subset
	1
video ending with $F_{entropy}$	0.41
video starting with $F_{entropy}$	0.42
original video	0.45
Sig.	0.095

TABLE S65
Homogeneous subsets of "provoking" in Experiment 4 for $F_{entropy}$.

Stimulus type	subset	
	1	2
video starting with $F_{entropy}$	0.43	
video ending with $F_{entropy}$	0.46	0.46
original video		0.48
Sig.	0.076	0.445

TABLE S66
Homogeneous subsets of "unusual" in Experiment 4 for $F_{entropy}$.

Stimulus type	subset
	1
video starting with $F_{entropy}$	0.47
video ending with $F_{entropy}$	0.48
original video	0.48
Sig.	0.749

TABLE S67
Homogeneous subsets of "suspenseful" in Experiment 4 for $F_{entropy}$.

Stimulus type	subset	
	1	2
video starting with $F_{entropy}$	0.45	
video ending with $F_{entropy}$	0.45	
original video		0.60
Sig.	0.996	1.000

TABLE S68
Homogeneous subsets of "sad" in Experiment 4 for F_{DNN}.

Stimulus type	subset	
	1	2
video ending with F_{DNN}	0.30	
video starting with F_{DNN}	0.35	
original video		0.46
Sig.	0.011	1.000

TABLE S69
Homogeneous subsets of "angry" in Experiment 4 for F_{DNN}.

Stimulus type	subset		
	1	2	3
video ending with F_{DNN}	0.31		
video starting with F_{DNN}		0.35	
original video			0.45
Sig.	1.000	1.000	1.000

TABLE S70
Homogeneous subsets of "frightening" in Experiment 4 for F_{DNN}.

Stimulus type	subset	
	1	2
video ending with F_{DNN}	0.35	
video starting with F_{DNN}	0.38	
original video		0.46
Sig.	0.109	1.000

TABLE S71
Homogeneous subsets of "disgusting" in Experiment 4 for F_{DNN}.

Stimulus type	subset	
	1	2
video ending with F_{DNN}	0.37	
video starting with F_{DNN}	0.40	
original video		0.46
Sig.	0.158	1.000

TABLE S72
Homogeneous subsets of "provoking" in Experiment 4 for F_{DNN}.

Stimulus type	subset
	1
video ending with F_{DNN}	0.45
original video	0.49
video starting with F_{DNN}	0.50
Sig.	0.022

TABLE S73
Homogeneous subsets of "unusual" in Experiment 4 for F_{DNN}.

Stimulus type	subset		
	1	2	3
video starting with F_{DNN}	0.41		
video ending with F_{DNN}		0.46	
original video			0.50
Sig.	1.000	1.000	1.000

TABLE S74
Homogeneous subsets of "suspenseful" in Experiment 4 for F_{DNN}.

Stimulus type	subset	
	1	2
video starting with F_{DNN}	0.45	
video ending with F_{DNN}	0.46	
original video		0.60
Sig.	0.860	1.000

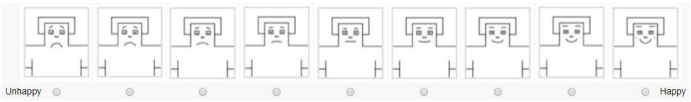
SUPPLEMENTARY QUESTIONNAIRE

Following are the questions we used for sentiments ratings in Experiments 3 and 4. The questionnaire in Experiment 1 is almost the same except without question 19.

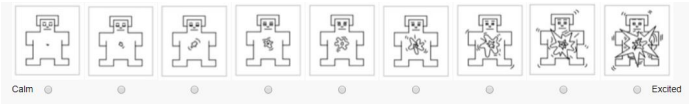
Please answer the following questions regarding the image you saw just now.

Section I: Please use the figures below to rate how you felt while viewing the image.

Q1. Does the image make you feel unhappy or happy?



Q2. Does the image make you feel calm or excited?



Section II: Please answer the following questions regarding the image you saw just now.

Q1. How much does the image make you feel happy?

Q2. How much does the image make you feel surprised?

Q3. How much does the image make you feel awe (a feeling of reverential respect mixed with fear or wonder)?

Q4. How much does the image make you feel excited?

Q5. How much does the image make you feel amused?

Q6. How much does the image make you feel content?

Q7. How much does the image make you feel sad?

Q8. How much does the image make you feel angry?

Q9. How much does the image make you feel frightened?

Q10. How much does the image make you feel disgusted?

Q11. Is the scene provoking?

Q12. Is the scene dynamic/in motion?

Q13. Is the scene unusual or strange?

Q14. Are you serious in doing this survey?

Q15. Are you providing answers randomly?

Q16. Is the scene suspenseful?

REFERENCES

[1] B. H. Detenber, R. F. Simons, and G. G. Bennett Jr, "Roll em!: The effects of picture motion on emotional responses," *Journal of Broadcasting & Electronic Media*, vol. 42, no. 1, pp. 113–127, 1998.

[2] M. K. Uhrig, N. Trautmann, U. Baumgärtner, R.-D. Treede, F. Henrich, W. Hiller, and S. Marschall, "Emotion elicitation: A comparison of pictures and films," *Frontiers in psychology*, vol. 7, p. 180, 2016.

[3] M. Redi, N. OHare, R. Schifanella, M. Trevisiol, and A. Jaimes, "6 seconds of sound and vision: Creativity in micro-videos," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 4272–4279, IEEE, 2014.

[4] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception & psychophysics*, vol. 14, no. 2, pp. 201–211, 1973.

[5] J. E. LeDoux, "Emotion circuits in the brain," *Annual review of neuroscience*, vol. 23, no. 1, pp. 155–184, 2000.

[6] H. Zettl, *Sight, sound, motion: Applied media aesthetics*. Cengage Learning, 2013.

[7] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *ECCV 2006*, pp. 288–301, Springer, 2006.

[8] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," in *CVPR*, vol. 1, pp. 419–426, IEEE, 2006.

[9] J. G. Moulard, M. W. Kroff, and J. A. G. Folse, "Unraveling consumer suspense: The role of hope, fear, and probability fluctuations," *Journal of Business Research*, vol. 65, no. 3, pp. 340–346, 2012.

[10] W. Hubert and R. de Jong-Meyer, "Autonomic, neuroendocrine, and subjective responses to emotion-inducing film stimuli," *International Journal of Psychophysiology*, vol. 11, no. 2, pp. 131–140, 1991.

[11] P. Vorderer, H. J. Wulff, and M. Friedrichsen, *Suspense: Conceptualizations, theoretical analyses, and empirical explorations*. Routledge, 2013.

[12] P. Comisky and J. Bryant, "Factors involved in generating suspense," *Human Communication Research*, vol. 9, no. 1, pp. 49–58, 1982.

[13] P. Vorderer, S. Knobloch, and H. Schramm, "Does entertainment suffer from interactivity? the impact of watching an interactive tv movie on viewers' experience of entertainment," *Media Psychology*, vol. 3, no. 4, pp. 343–363, 2001.

[14] L. F. Alwitt, "Suspense and advertising responses," *Journal of Consumer Psychology*, vol. 12, no. 1, pp. 35–49, 2002.

[15] S. Abuhamdeh, M. Csikszentmihalyi, and B. Jalal, "Enjoying the possibility of defeat: Outcome uncertainty, suspense, and intrinsic motivation," *Motivation and Emotion*, vol. 39, no. 1, pp. 1–10, 2015.

[16] T. Shortell, "An introduction to data analysis & presentation," *World Wide Web: http://academic. brooklyn. cuny. edu/soc/courses/712/chap18. html*, 2001.

[17] S. Schmidt and W. G. Stock, "Collective indexing of emotions in images. a study in emotional information retrieval," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 5, pp. 863–876, 2009.

[18] B. H. Detenber and B. Reeves, "A bio-informational theory of emotion: Motion and image size effects on viewers," *Journal of Communication*, vol. 46, no. 3, pp. 66–84, 1996.

[19] M. Allahbakhsh, B. Benatallah, A. Ignjatovic, H. R. Motahari-Nezhad, E. Bertino, and S. Dustdar, "Quality control in crowdsourcing systems: Issues and directions," *IEEE Internet Computing*, vol. 17, no. 2, pp. 76–81, 2013.

[20] X. Ma, J. T. Hancock, K. L. Mingjie, and M. Naaman, "Self-disclosure and perceived trustworthiness of airbnb host profiles.," in *CSCW*, pp. 2397–2409, 2017.

[21] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.