Predicting Ad Liking and Purchase Intent: Large-scale Analysis of Facial Responses to Ads

Daniel McDuff, *Student Member, IEEE,* Rana El Kaliouby, Jeffrey F. Cohn, and Rosalind Picard, *Fellow, IEEE*

***Abstract*—Billions of online video ads are viewed every month. We present a large-scale analysis of facial responses to video content measured over the Internet and their relationship to marketing effectiveness. We collected over 12,000 facial responses from 1,223 people to 170 ads from a range of markets and product categories. The facial responses were automatically coded frame-by-frame. Collection and coding of these 3.7 million frames would not have been feasible with traditional research methods. We show that detected expressions are sparse but that aggregate responses reveal rich emotion trajectories. By modeling the relationship between the facial responses and ad effectiveness we show that ad liking can be predicted accurately (ROC AUC=0.85) from webcam facial responses. Furthermore, the prediction of a change in purchase intent is possible (ROC AUC=0.78). Ad liking is shown by eliciting expressions, particularly positive expressions. Driving purchase intent is more complex than just making viewers smile: peak positive responses that are immediately preceded by a brand appearance are more likely to be effective. The results presented here demonstrate a reliable and generalizable system for predicting ad effectiveness automatically from facial responses without a need to elicit self-report responses from the viewers. In addition we can gain insight into the structure of effective ads.**

***Index Terms*—Facial expressions, emotion, market research.**

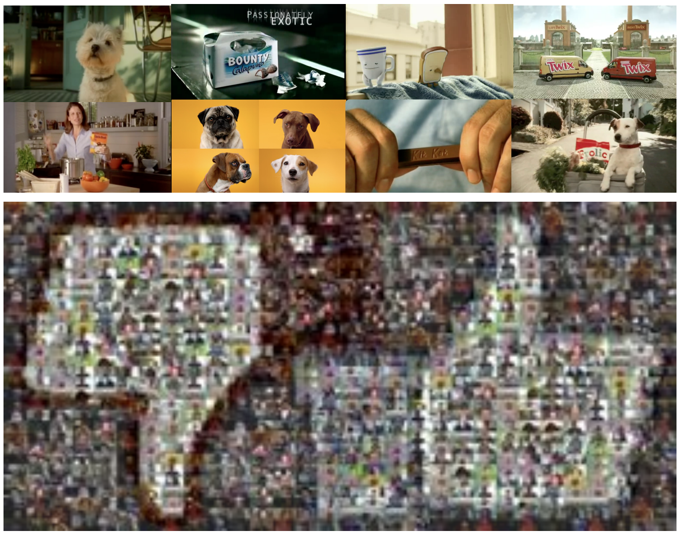


Fig. 1. In this work we present a large-scale analysis of facial responses to online video ads. Top) Example frames from the ads tested. Bottom) Frames from the webcam videos collected of viewer's responses. Permission was given for these images to be used.

I. Introduction

NON-VERBAL signals, such as facial expressions, can communicate highly detailed information about a person's experience. The face, in particular, has been shown to display discriminative valence information. Greater zygomatic major muscle (AU12, occurring in smiles) activity was observed during ads with positive emotional tone and greater corrugator muscle (AU4, brow furrow) activity was observed during ads with negative emotional tone [1].

The Facial Action Coding System (FACS) [2] is a catalog of 44 unique action units (AUs) that correspond to each of the face's 27 muscles. FACS enables objective, reliable and quantitative measurement of facial activity. Action units can combine to create thousands of meaningful facial expressions. However, FACS coding requires specialist training and is a labour intensive task. It can take five to six hours to code a minute of video. Computer vision systems can now reliably code many of these actions automatically [3]. Furthermore, recently it has been shown that these systems can be deployed effectively "in-the-wild" (e.g. online and in public spaces) not just in more controlled settings [4], [5].

Online video is growing fast. In the US in November 2013[1] over 189 million viewers watched videos online and the average viewer watched 19 hours of video online. In the US billions of dollars are spent on video ads each year. A total of nearly 27 billion ads were viewed in November 2013 and this reached more than 50% of the US population. This is almost three times the number of viewings compared to the same month in 2012. Video ads accounted for 36.2% of all videos viewed. Internet TV sites like Hulu and Netflix frequently ask viewers about the relevance or their enjoyment of ads. In addition to viewing videos, more people are sharing video content with others. In 2013 72% of adult Internet users used video-sharing sites.[2] Many companies will place their advertisements on video sharing sites such as YouTube in order to promote social sharing of their content. The Internet not only allows advertisers to reach more people it also allows for more precisely targeted advertising. Evidence has shown that targeting of advertisements can be beneficial to consumers and raise the profits for all involved [6].

Understanding the relationship between emotional re-

D. J. McDuff is with the Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. (e-mail: djmcduff@media.mit.edu).

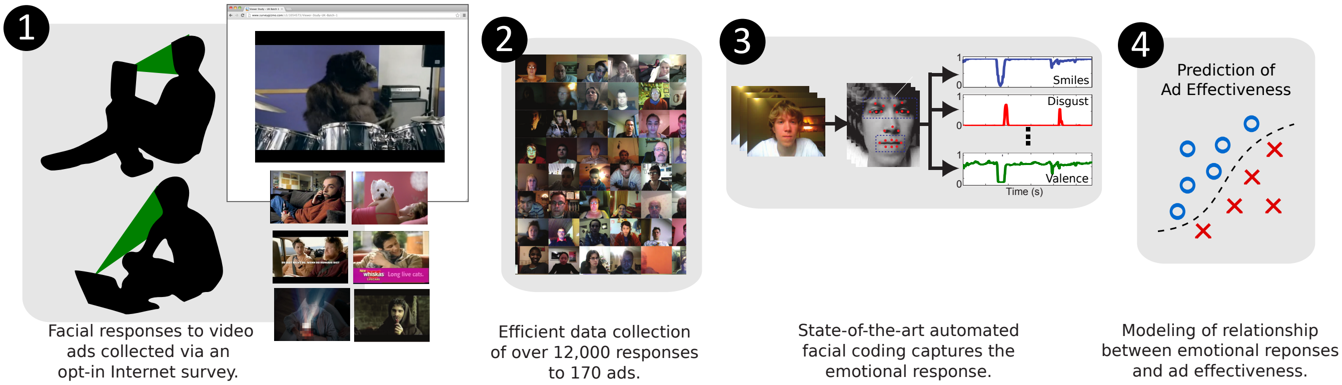
R. Kaliouby is with Affectiva, Inc., Waltham, USA. (e-mail: kaliouby@affectiva.com).

J. Cohn is with the University of Pittsburgh and Carnegie Mellon University, USA. (e-mail: jeffcohn@cs.cmu.edu).

R. W. Picard is a professor at the Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. (phone: 617-253-0611; e-mail: picard@media.mit.edu).

[1] http://www.comscore.com

[2] http://pewinternet.org/



Facial responses to video ads collected via an opt-in Internet survey.

Efficient data collection of over 12,000 responses to 170 ads.

State-of-the-art automated facial coding captures the emotional response.

Modeling of relationship between emotional reponses and ad effectiveness.

Fig. 2. Overview of the framework used in this paper. 1) Spontaneous and naturalistic facial responses to video ads are collected via software embedded into a web survey. 2) The data collection makes use of the connectivity of the Internet and ubiquitous nature of webcams to allow the efficient collection of over 12,000 responses to 170 ads. 3) State of the art automated facial coding was used to capture the expression responses of the viewers. 4) We modeled the relationship between facial responses and ad effectiveness measures, building a completely automated prediction of ad liking and change in brand purchase intent resulting from the ad. Such a system could be used in the copy-testing of ads.

sponses to content and measures of advertising effectiveness has been limited by traditional research methods, such as surveys, that are laborious, time consuming and often do not capture the temporal nature of emotional changes during an ad. In addition, in some cases questioning viewers about their opinions on ads is impractical to capture (e.g. when people are occupied by another task such as surfing the web or watching TV) and automated methods of prediction would be beneficial.

However, recently the connectivity of the Internet and the ubiquity of webcams has enabled large-scale collection of (opt-in) facial responses via online frameworks [4]. This approach is efficient as it allows large numbers of facial responses to be collected from viewers in a wide geographical area. It also helps avoid some of the pitfalls that arise from data collection in traditional market research settings. First, the viewers are in a natural viewing context rather than a lab, and second, the measurement is all remote and does not require electrodes to be placed on the body. Third, it is possible to collect a large amount of data at a fraction of the cost: compensation for each participant who watched 10 ads and completed a survey lasting 30 minutes cost less than $10.

The challenges of measurement of facial responses outside a controlled context are that there can be many sources of noise (lighting, position of the camera, social factors affecting the viewing experience) that are hard to control. However, careful experimental design can limit the impact of these factors. We have executed a number of experimental iterations collecting facial videos over the Internet [4], [7] and these have informed this study.

A key measure of advertising effectiveness is advertisement likability [8], [9]. Our preliminary work, considering over three thousand facial response videos, has shown evidence that facial expressions can predict ad liking [10]. However, McDuff *et al.* [10] only presented results for three ads and it was not clear how generalizable the system would be for a wider variety of ads. Other work has provided evidence that facial expressions can predict variables related to advertising success such as recall [11] and ad "zapping" [12]. A metric of high interest to advertisers is a viewer's purchase intent (PI) towards products from the advertised brand. Teixeira *et al.* [13] explored the relationship between facial expressions and viewers' PI. We will present a large-scale analysis of facial responses to ads aired over the past 12 years and evaluate what we believe is the first model that can automatically predict an ad's likelihood of driving purchase intent from facial responses. It would not have been possible to collect and analyze such a large dataset using traditional methods: recruiting thousands of individuals to come to a lab from around the globe and hand labeling expressions in over 3.7 million frames would have been prohibitively expensive and time consuming. Figure 2 shows the framework we use to collect the facial responses, automatically code the expression metrics and model the relationship with ad effectiveness measures. Prediction of the ability of an ad to be likable and to increase purchase intent is valuable in copy-testing of advertising content and potentially in the targeting of video content on online TV and video sharing sites.

The contributions of this paper are; 1) to present the largest dataset of facial responses to ads ever collected, 2) to model the relationship between facial responses and ad liking and changes in purchase intent, 3) to identify features of aggregate emotional responses that make an ad effective.

II. Related Work

A. Facial Expression Recognition

Over the past 15 years automatic analysis of facial expressions has received much attention from the computer vision, psychology and affective computing communities. Initially, much of the work focused on posed and acted facial behavior. However, recent work has focused increasingly on naturalistic and spontaneous behavior [4], [14], [15] and subtle expressions [16].

Most facial expression recognition systems follow a similar structure. Facial registration is performed to align the face and limit the impact of pose and position of the face. State-of-the-art face registration methods include Active Appearance Models (AAM) [17] and Constrained Local Models (CLM) [18]. Shape and/or appearance features are then extracted from a region of interest (ROI) and used as input to a computational model that maps features to expression or action unit labels. Commonly used features are local binary patterns (LBP) and histograms of oriented gradients (HOG). The most commonly used class of model is Support Vector Machines (SVM). A comprehensive review of methods for facial expression recognition can be found in [3].

Smile analysis is one of the most commonly used and robust forms of facial expression recognition. Whitehill *et al.* [15] present a state of the art smile detector trained on images found on the Internet. Their approach was an efficient way to source training data for the classifier. Our previous work has demonstrated accurate smile detection in uncontrolled settings over the Internet [10]. Facial behavior can be classified using FACs, discrete category labels (e.g. six emotional states) or using continuous measures of emotion such as valence and arousal/activation. A number of approaches for dimensional measurement of emotions from facial behavior (such as valence) have been presented recently [19]. For this work we use both custom classifiers for detecting AUs (e.g. AU02) and discrete emotion labels (e.g. disgust).

B. Media and Emotions

Kassam's [20] analysis of facial expressions demonstrates that both facial expressions and self-report responses have significant variance: results show that expression analysis provides unique insight into emotional experiences, different from information obtained via self-report questioning. Predicting emotional valence during exposure to media has been demonstrated [21], as has the prediction of media preferences from automatically measured facial responses [10], [22].

Joho *et al.* [23] show the possibility of detecting viewer's personal highlights from automatically analyzed facial activity. Zhao *et al.* [24] designed a video indexing and recommendation system based on automatically detected expressions of six emotions (amusement, sadness, disgust, anger, surprise and fear).

Timing information is important in mapping measured facial responses to preferences [25]. Facial measurements of emotion allow us to capture precise temporal information about a person's emotional experience. This work provides support that the conclusions in [25] extend to uncontrolled naturalistic settings and are true on a large-scale.

C. Market Research

Micu and Plummer [26] measured zygomatic major (AU12) activity using facial electromyography (EMG) whilst people watched TV ads. The results provided evidence that physiological measurements capture different information from self-reported feelings. This evidence aligns with [20]. Our real-world data support these findings.

Hazlett and Hazlett [11] measured facial EMG whilst viewers watched advertisements. They found that facial muscle movements during ads provided a more sensitive discriminator for recall than self-report measures and that peaks in facial EMG were related to emotion-congruent events in the ads.

TABLE I

NUMBER OF VIDEOS TESTED FROM EACH PRODUCT CATEGORY AND EMOTION CATEGORY (CATEGORIZED USING MTURK LABELERS). THE LARGEST PROPORTION OF ADS WERE INTENTIONALLY AMUSING. TOTAL NUMBER OF ADS: 170. TOTAL NUMBER OF AMUSING ADS: 75.

		Product Category				
		Petcare	Confec.	Food	Other	Total
Emotion Category	Amusement	14	46	7	8	75
	Heart-warming	7	2	0	4	13
	Cute	11	1	2	0	14
	Exciting	3	5	2	3	13
	Inspiring	2	3	2	2	9
	Sentimental	5	1	3	0	9
	No Majority	11	17	3	6	37
	Total	53	75	19	23	170

Teixeira *et al.* [12] showed that including affect is important in reducing "zapping" (skipping the advertisement) of online advertising. Berger and Milkman [27] found that positive affect inducing content was more likely to be shared than negative affective inducing content and that virality was also driven by highly arousing content. Recall is a commonly used measure of advertising effectiveness and emotions influence recall [28]. Ambler and Burne [29] found that ads with more intense emotion were more memorable and that β-blockers that suppress affect reduced the ability to recall ads.

However, in all these examples the data was collected in a laboratory setting and not a natural context. In addition, many of these examples only consider between 10 and 20 ads. Scaling the analysis is important to have more confidence that the findings will generalize. Our previous work [10] was the first showing automatically measured facial responses to online ads could predict measures of advertising effectiveness (ad liking and desire to watch again). Teixeira *et al.* [13] showed that entertainment (measured from smile activity) associated with the brand was more likely to increase purchase intent than entertainment not associated with the brand - more smiles do not always make an ad more effective. However, in McDuff *et al.* [10] and Teixeira *et al.* [13] only viewer smile responses were measured, not a larger set of expressions.

III. DATA AND DATA COLLECTION

A. Video Ads

We test 170 video ads from four countries (30 from France, 50 from Germany, 60 from the UK and 30 from the US). The videos were all originally aired between 2001 and 2012. The mean length of the video content was 32s (std = 14s). The ads were chosen as they represented a broad range of successful and unsuccessful content (as judged by the brands advertised) within the product categories.

Product Categories A majority of the video ads tested were for products in one of the three following categories: pet care, confectionery (chocolate, gum and candy) and food (instant rice and pasta products). Of the 170 ads 23 were from other product categories. Importantly, these were all products that might be bought frequently by category users and do not represent a long-term purchasing decision (such as a new car might). Table I shows the number of videos from each product category.

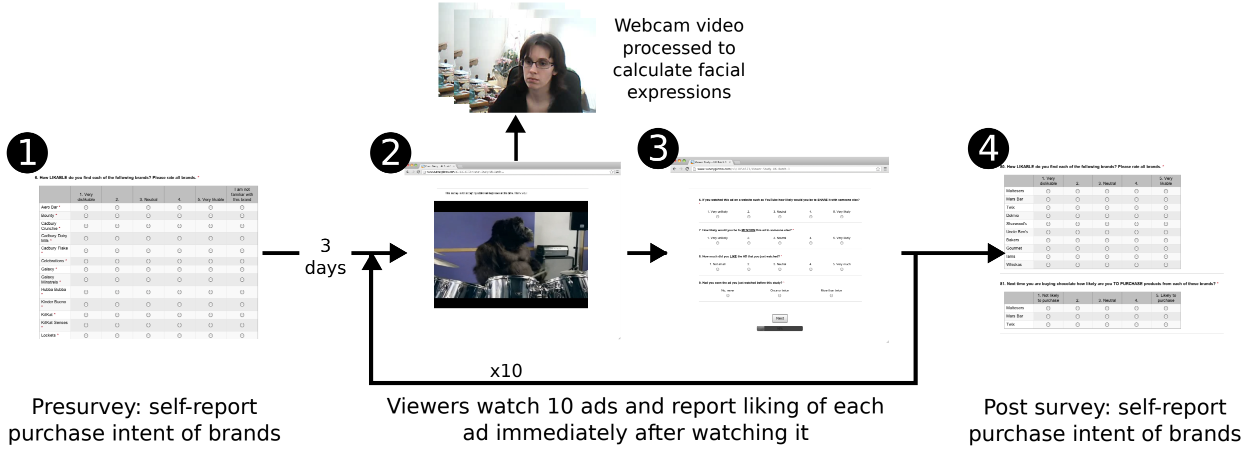

Webcam video processed to calculate facial expressions
1
2
3
4
3 days
x10
Presurvey: self-report purchase intent of brands
Viewers watch 10 ads and report liking of each ad immediately after watching it
Post survey: self-report purchase intent of brands

Fig. 3. Structure of the data collection survey. 1) A presurvey obtained baseline self-report of purchase intent for each brand. Consent was also requested to record videos. 2) After 3 days the participant was recontacted, video ads were watched and webcam video recorded simultaneously. 4) Self-reported liking questions follow each ad. Ten ads were viewed by each participant. 4) Post survey obtained self-report of purchase intent for each brand.

Emotion Categories The ads were not all designed to elicit the same emotions or to communicate the same messages. Two different ads could be designed to create very different feelings within the viewers. One ad may be designed to amuse whereas another may be designed to be sentimental. The affective goal of the ad is important contextual information when considering viewers' facial responses. We used Amazon's Mechanical Turk (MTurk) platform to crowdsource emotion category labels for the videos. At least three coders were recruited to watch each video and answer the following question: "*CHOOSE the words that best describe the type of FEELINGS that you think this video was designed to induce.*" Labelers were able to select one of more answers from the following list: Sentimental, Inspiring, Exciting, Romantic, Heart-warming, Amusing, Cute. The majority label was taken for the videos. Table I shows the number of videos from each emotion category. The initial list of seven possible labels was derived by the first author who watched all the ads and selected seven labels felt to best describe the emotional content.

B. Participants

Participants were recruited from four countries (France, Germany, the UK, the US) to view the ads and complete a survey. Recruitment was such that age groups, gender and economic status (annual salary) of the participants was as balanced as possible and also helped mitigate the effects of a possible self-selection bias. In addition, in all cases at least 70% of the viewers who watched each ad were a category user of the product being advertised. Figure 4 shows the distribution of viewers across gender, age and economic status.

Not all the participants we contacted had a functioning webcam or were willing to let their responses be recorded. In neither of these cases were they allowed to continue with the survey. Of the participants that started the survey, 48% reported having a working webcam. Of these 48% of participants, 49% stated they were happy to have their facial responses recorded (thus 23.5% could take part). These statistics show that quite a large number of people need to be contacted in order to collect a dataset using this method. However, contacts are inexpensive and this should not prevent large numbers of people participating. Perhaps a greater issue is the self-selection bias that is a result of only those with webcams and who are willing to be recorded being able to take part. In order to combat this we try to ensure an even demographic split as described above (see Figure 4). Interesting future work could consider quantifying the impact of the self-selection effects.

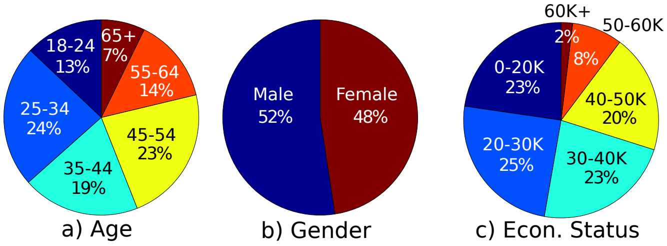


Fig. 4. Demographic breakdown of the 1,223 viewers in our study: a) age, b) gender, c) economic status (approximate annual salary in thousands of US dollars).

In total 1,223 people successfully completed the survey. Each participant watched 10 ads giving a total of 12,230 facial responses. Each ad was watched by an average of 72 viewers. Once a participant had taken the survey they were excluded from taking the survey again, even with a different set of ads.

C. Survey

The video content and facial expression capture software were integrated into an online survey. Each participant viewed 10 videos from their country. The survey structure is shown in Figure 3. Before taking part the participants were asked for permission to stream videos captured from their webcam to the server. Figure 5 shows a screenshot of the consent question.

Presurvey: People were contacted initially with a presurvey to capture baseline measures of purchase intent. They were asked the following **purchase intent** question about a number of brands.

Q. Next time you are buying [product category] how likely are you TO PURCHASE products from each of these brands?

Not likely		Neutral		Very likely
1.	2.	3.	4.	5.

They were contacted again after three days to complete the main part of the survey.

Main Survey: After being invited to complete the main survey the viewers watched 10 videos in a random order (to minimize ordering effects), prior to watching the videos viewers were briefly shown their webcam stream so that they could align the camera and ensure reasonable lighting, this greatly increases the quality of the resulting facial expression metrics that can be extracted. They were also asked to remove hats and to not chew gum or eat during the experiment.

Following each video viewers were asked the following **liking** question:

Q. How much did you **LIKE the AD** that you just watched?

Not at all		Neutral		Very much
1.	2.	3.	4.	5.

Following all the ads participants were once again asked the **purchase intent** question:

Q. Next time you are buying [product category] how likely are you TO PURCHASE products from each of these brands?

Not likely		Neutral		Very likely
1.	2.	3.	4.	5.

Thus we capture responses about individual liking of the ads viewed and pre- and post-measures of purchase intent for the brands advertised.

At the end of the survey participants were asked:

Q. *How COMFORTABLE did you feel during the study?*
Of the viewers 88% reported "very comfortable" to "neutral", 3% reported "very uncomfortable".
They were then asked:

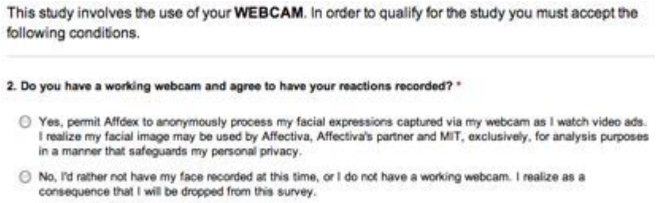
Q. *Did you behave differently than you would have if you were watching these ads NOT as part of a study?*
Of the viewers 71% reported "no differently", 25% reported "a little differently" and 4% reported "very differently".

These statistics along with observation of the recorded videos suggest that the responses of the viewers were in general natural. However, we should be aware that because the viewers were required to give consent to be recorded their responses may be slightly influenced by this. We believe that these data are more naturalistic than they would be if collected in a lab-based setting which is perhaps even more likely to cause unnatural behavior.

Participants were compensated approximately $8.10 for taking part in the survey (compensation was given in the local currency). The average time to complete the survey was 36 minutes.

IV. Automated Facial Analysis

The facial videos were analyzed using Affectiva's facial action classifiers. Figure 6 shows the facial expression analysis pipeline.



This study involves the use of your **WEBCAM**. In order to qualify for the study you must accept the following conditions.

2. Do you have a working webcam and agree to have your reactions recorded? *

○ Yes, permit Affdex to anonymously process my facial expressions captured via my webcam as I watch video ads. I realize my facial image may be used by Affectiva, Affectiva's partner and MIT, exclusively, for analysis purposes in a manner that safeguards my personal privacy.

○ No, I'd rather not have my face recorded at this time, or I do not have a working webcam. I realize as a consequence that I will be dropped from this survey.

Fig. 5. The consent forms that the viewers were presented with before watching the commercials and before the webcam stream began. Viewers also had to accept the standard flash webcam access permission before their camera was turned on.

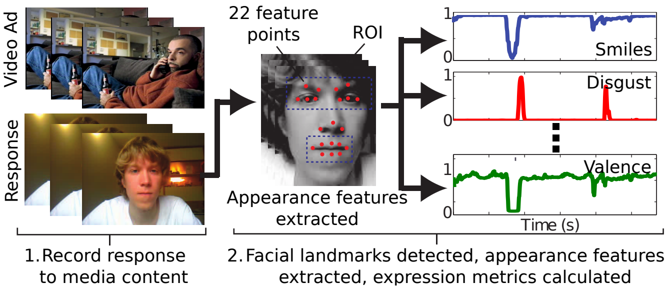

Video Ad
Response
22 feature points
ROI
Appearance features extracted
1
0
Smiles
1
0
Disgust
1
0
Valence
Time (s)
1. Record response to media content
2. Facial landmarks detected, appearance features extracted, expression metrics calculated

Fig. 6. Flow diagram of the facial expression analysis pipeline. 1) Facial videos were recorded as the media content was played back. 2) The Nevenvision facial feature tracker was used to detected facial landmarks in the frames of the recorded videos. Histogram of oriented gradient (HOG) features extracted from the region of interest with the frames were used to calculate the expression metrics.

A. Face Detection

The Nevenvision facial feature tracker[3] was used to automatically detect the face and track 22 facial feature points within each frame of the videos. The location of the facial landmarks is shown in Figure 6. The 12,230 facial videos amounted to a total of 4,767,802 frames of video. In 3,714,156 (77.9%) of the frames a face could be detected. For frames in which a face could not be detected the classifiers did not return a value.

B. Expression Detectors

To compute the expression probabilities we used custom algorithms developed by Affectiva. We use classifiers for eyebrow raises, smiles, disgust expressions and positive and negative valence expressions (Figure 7 shows examples - all with the original frame next to a cropped image of the facial region). We selected these facial expressions as they were deemed highly relevant to the context of advertising and viewer responses, and have been measured in previous work [11]. Support vector machines (SVM) with radial basis function kernels were used for classification in all cases. The signed distance of the sample from the classifier hyperplane was taken and normalized using a monotonic function that in the training phase rescaled points between [0, 1]. The classifier outputs were probabilistic and continuous moment-by-moment measures that were computed for each frame of the facial video, yielding one-dimensional metrics for each

[3] Licensed from Google, Inc.

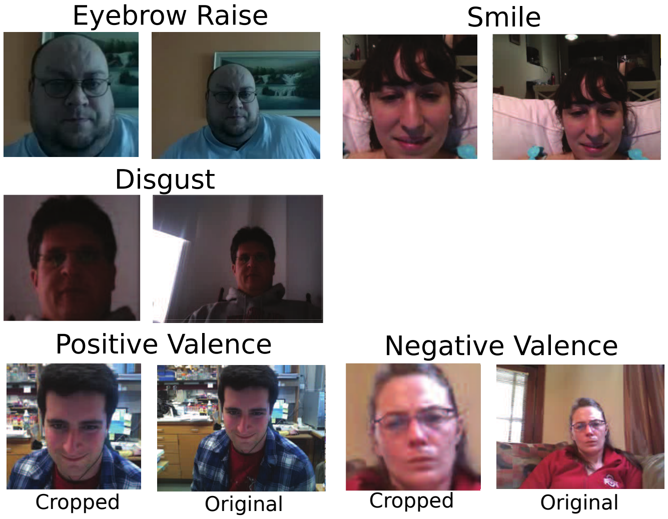


Fig. 7. Examples of eyebrow raise, smile, disgust and positive and negative valence expressions. Original video frames and cropped versions of the frames are shown.

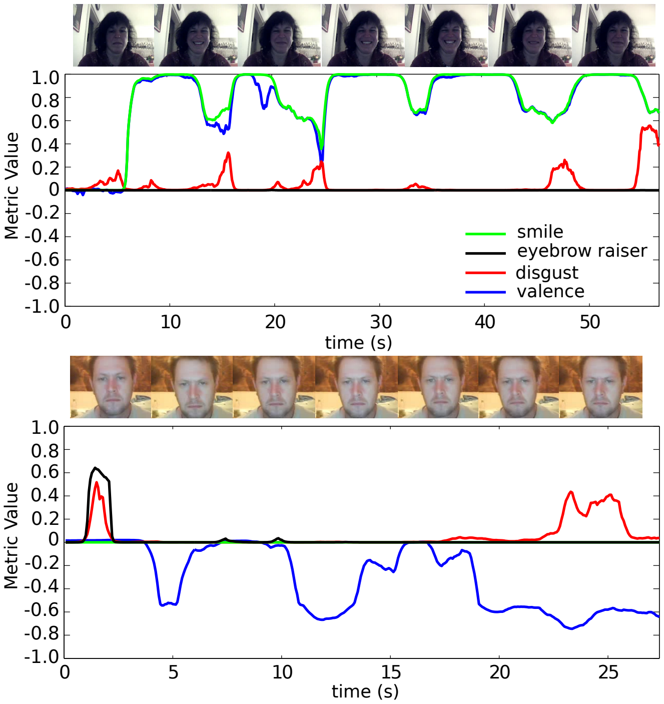


Fig. 8. Expression tracks: top) an example of a response with strong smiling and positive valence, bottom) an example of a response with strong disgust and negative valence. Frames of the corresponding video are shown above - the frames have been cropped to make the face larger.

video. Figure 8 shows example tracks with screenshots of the responses for two individuals.

Eyebrow Raise (E): The detector uses Histogram of Oriented Gradient (HOG) [30] features extracted from the whole face region of interest (ROI) as input to the SVM. The output is a continuous probability measure of an eyebrow raise action ranging from 0 to 1. Training examples were labeled as 1 with the presence of AU01 or AU02 and 0 otherwise.

Smile (S): The detector uses HOG features extracted from the whole face ROI as input to the SVM. The output is a continuous probability measure of a smile expression ranging from 0 to 1. This is a smile detector rather than an AU12 detector, training examples were labeled as 1 with the presence of a smile and 0 otherwise.

Disgust (D): The detector uses HOG features extracted from the whole face ROI as input to the SVM. The output is a continuous probability measure of a disgust expression ranging from 0 to 1. For the disgust classifier training examples were labeled as 1 with the presence of a disgust expression and 0 otherwise. In a separate experiment, videos with a high probability of containing disgust expressions were collected by showing people disgust inducing content.

Valence (V): The detector uses HOG features extracted from the whole face ROI. The output is a continuous value between -1 and 1 where -1 is a negatively valenced facial expression and 1 is a positively valence facial expression. For valence, a three class labeling system was adopted using the following criteria:

```
if (smile present) {valence = +1}
else if (AU04 or AU09 or AU15 present) {valence = -1}
else {valence = 0}
```

C. Training and Testing of Expression Detectors

All classifiers were trained and tested with more than 5,000 spontaneous images labeled by FACS trained human coders. Some of the images were taken from the same expression sequence. However, we tried to select as diverse a set of examples of each action as possible. These images were taken from webcam videos recorded in 20 separate studies across Asia, Europe and America. These videos are similar but different to the webcam videos collected in this study. The images from these datasets were labeled for the presence of an expression by human coders. For each video, three FACS trained human labelers coded for the presence or absence of: AU01, AU02, AU4, AU9, AU15, disgust and smile. The majority class label was taken. Table II shows the area under the receiver operating characteristic (ROC) curves for the smile, disgust and valence classifiers (as defined above). For the three label valence classification we report results for each combination: positive vs. negative, positive vs. neutral and neutral vs. negative examples.

TABLE II
AREA UNDER THE RECEIVER OPERATING CHARACTERISTIC CURVES FOR THE EYEBROW RAISE, SMILE, DISGUST AND VALENCE CLASSIFIERS.

	Classifier					
				Valence		
	Eye. R.	Smile	Disgust	+ve/-ve	+ve/neut.	neut./-ve
AUC	77.0	96.9	86.7	97.3	92.2	71.5

V. Facial Activity Characteristics

A. Expressiveness of Viewers

To characterize the expressiveness of viewers we analyzed the metrics across all videos. Frames for which the expression classifier output is smaller than 0.1 are classed as no expression present. In 82.8% of the 3,714,156 frames in which a face was detected there was no detected eyebrow raise, smile, disgust expression or non-neutral valence expression.

Figure 9 shows histograms of the number of frames with each expression (smiles, disgust and positive and negative valence) probability. Examples of frames from select buckets are shown. A vast majority of the frames did not have an eyebrow raise, smile, expression of disgust or positive or negative valence detected as present. Table III shows the percentage of frames which feature expressions of each metric value: only 6% of frames had an eyebrow raise > 0.1, 7.9% of frames had a smile > 0.1 and 5.5% of frames an expression of disgust > 0.1.

TABLE III
PERCENTAGE OF THE 3,714,156 FRAMES WITH EXPRESSIONS METRICS WITHIN 10 EVENLY SPACED CLASSIFIER OUTPUT BINS CENTERED ON THE VALUES SHOWN.

Bin	|Eyebrow R.|	|Smile|	|Disgust|	|Valence|
0.05	94.0	92.1	94.5	81.9
0.15	1.04	1.67	1.63	4.22
0.25	0.59	0.89	0.80	2.29
0.35	0.43	0.61	0.53	1.58
0.45	0.35	0.50	0.40	1.37
0.55	2.04	1.11	0.84	3.44
0.65	0.82	0.79	0.54	2.19
0.75	0.33	0.68	0.38	1.15
0.85	0.12	0.66	0.24	0.88
0.95	0.03	0.98	0.14	0.87

In only 54.5% of face videos were there any detected expressions greater than 0.1 at any point. In 36.9% of the face videos were there any detected expressions greater than 0.5 at any point. However, with an average of over 70 viewers for each ad we found detectable responses - greater than 0.5 - in at least one viewer for all the ads. In addition, we note that there are much larger numbers of detected positive valence expressions than detected negative valence expressions. Considering that most ads probably aim to induce positive affect this is to be expected.

B. Aggregate Characteristics

Figure 10 shows the mean valence metrics for the different ads tested (ordered by increasing mean positive valence). Interestingly, a large number of ads had negative valence mean expression metrics. These results, and those above, show that although responses are sparse, different people respond to the ads differently and the ads elicited a range of expressions (from strong positive valence to negative valence). A few ads elicited no aggregate positive or negative valence.

In the remaining part of the paper we focus on the prediction of aggregate level results (how effective is an ad across all viewers) rather than individual level results (how effective is an ad for a specific individual). Examples of individual level prediction can be found in [10].

VI. CLASSIFICATION

To test the predictive performance of the facial responses we build and test classifiers for predicting ad effectiveness performance (effectiveness based on the self-reported liking and PI responses) directly from the measured facial response metrics and contextual information (product category and country). Below we explain how we calculate the features, the labels and how we validate, train and test the models. Figure 11 shows a flow diagram of the aggregate metric calculation, feature extraction and classification.

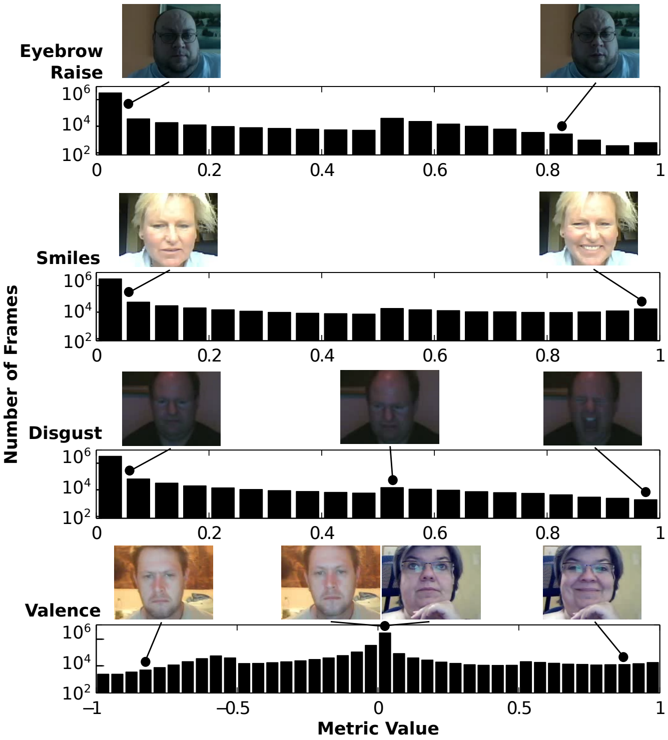


Fig. 9. Histograms of the number of frames with each expression probability. a) Smile, b) Disgust, c) Valence. In 82.8% of frames was there no detectable eyebrow raise, smile, disgust or positive/negative valence expression with magnitude above 0.1. Responses to ads in naturalistic settings are sparse but for all the ads there were expressive responses within the 70+ viewers. Note: the y-axis is logarithmic.

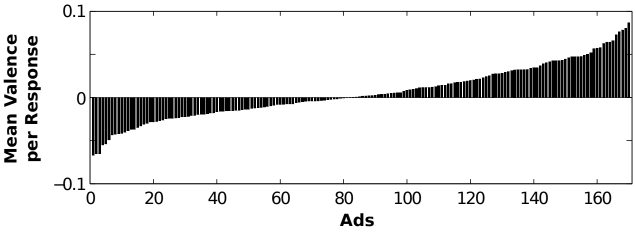


Fig. 10. Mean expression valence metrics for the 170 ads sorted by ascending mean valence.

A. Calculating Aggregate Metrics

We calculate aggregate expression metrics for each ad from the individual facial responses to that ad - Figure 11 (step 1). These were calculated as the mean metric intensity across all viewers to the ad (ignoring frames in which no face was detected). We compute mean tracks for the eyebrow raise, smile, disgust and valence classifiers.

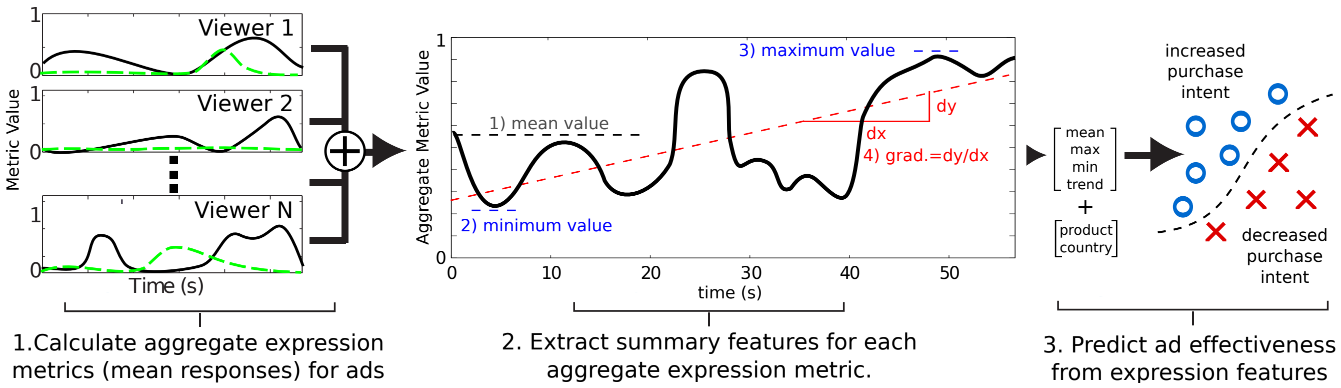

Viewer 1
Viewer 2
Viewer N
1
0
Metric Value
Time (s)
1.Calculate aggregate expression metrics (mean responses) for ads
Aggregate Metric Value
3) maximum value
1) mean value
2) minimum value
dy
dx
4) grad.=dy/dx
0
10
20
30
40
50
time (s)
2. Extract summary features for each aggregate expression metric.
mean
max
min
trend
+
product
country
increased purchase intent
decreased purchase intent
3. Predict ad effectiveness from expression features

Fig. 11. 1) Aggregate metric tracks calculated from all viewers who watched the ad. 2) Features extracted from each of the aggregate metric tracks: a) mean value, b) minimum value, c) maximum value, d) the gradient of the linear trend. 3) Summary features extracted from the facial expression metrics used to predict ad effectiveness.

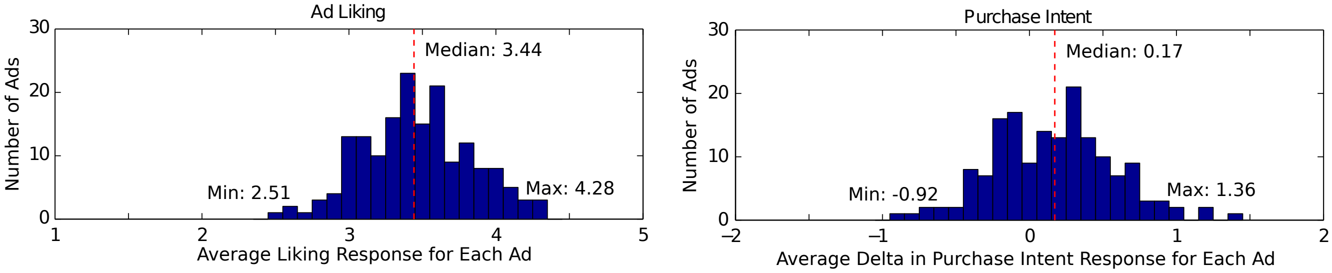


Fig. 12. Distribution of average: left) ad liking response and right) delta in purchase intent response for all ads. The median, minimum and maximum values are all shown. The report of liking was significantly greater than neutral (p<0.001).

B. *Extracting Summary Features*

Facial Metric Features: We extract summary features from the aggregate facial expression metrics. The summary features extracted from each summary metric were: mean, maximum, minimum and gradient. Figure 11 (step 2) shows how the features were extracted from an aggregate metric track. These four features for the four facial metrics classifiers led to a feature vector of length 16 for each ad.

Contextual Features: We use the product category and the country the ad is from as contextual features. These are coded as a binary matrix with columns that correspond to each of the five categories and columns that correspond to each of the four countries.

C. *Computing Labels*

The labels we use are taken from the viewers' self-report responses to the questions (in Section III-C) answered during the survey. We posed the problem as a two-class classification task due to the challenging nature of predicting ad performance from spontaneous facial responses. In this case discrimination between ads that were liked more or liked less than average, or increase purchase intent more or less than average, was still a very interesting task.

Liking Score: To compute the liking score for each commercial we calculate the average ad liking reported by each of the viewers, in response to the question *"How much did you LIKE the AD that you just watched?"*. We divide the ads into two categories - those with average liking greater than the median score and those with average liking equal to, or lower than, the median. Since we separate the ads using the median value the classes are inherently balanced in size. Five of the ads did not have complete labels therefore there are 165 liking examples.

Purchase Intent Score: To compute the purchase intent score for each commercial we calculate the mean delta in purchase intent reported by each of the viewers, in response to the questions *"Next time you are buying [product category] how likely are you TO PURCHASE products from each of these brands?"* which was asked in the pre-survey and at the end of the main survey. We divide the ads into two categories - those with average purchase intent delta greater than the median and those with average purchase intent delta equal, or lower than, the median. Seven of the ads did not have complete labels, therefore there are 163 purchase intent examples.

Figure 12 shows the distribution of labels for the liking score and purchase intent score. The average reported liking was significantly (p<0.001) greater than "neutral". As explained above, in both cases we normalize by the median average rating and split the data into two classes. The correlation between the liking and purchase intent scores for the ads was low (ρ=0.0691) and not significant. This suggests that indeed the metrics were capturing different types of the responses and that purchase intent is not driven entirely by ad liking.

D. *Model*

For this analysis we test the performance of an SVM model. A Radial Basis Function (RBF) kernel was used. During

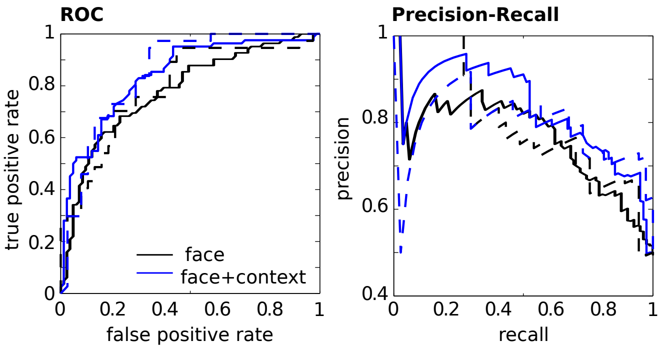


Fig. 13. Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves for the ad liking models varying the SVM decision threshold. Black) the performance using face features, blue) the performance using face and context features combined. Unbroken lines) results for all ads, broken lines) results for only the amusing ads.

validation the penalty parameter, C, and the RBF kernel parameter, γ, were each varied from 10^k with k=-3, -2,..., 3. The SVMs were implemented using libSVM [31]. The choice of parameters during the validation state was made by maximizing the geometric mean of the area under the receiver operating characteristic (ROC) and precision-recall (PR) curves (when varying the SVM decision threshold).

In order to test a model that can generalize we use a leave-one-ad-out training and testing scheme. As such, data for one ad is taken out of the dataset and the remaining data is used for validation and training (cross validation performed on the set and the median parameters selected). This process is repeated N times, where N is the number of ads.

TABLE IV
AREA UNDER THE ROC AND PR CURVES FOR THE AD LIKING CLASSIFIER: TOP) ALL ADS (N=165), BOTTOM) ONLY AMUSING ADS (N=75). COHEN'S κ BETWEEN THE PREDICTED AND SELF-REPORT LABELS FOR THE OPTIMAL CLASSIFIERS ARE SHOWN.

Ads	**Features**	ROC AUC	PR AUC	Cohen's κ
All	Naive	0.5	0.5	0.5
	Face	0.779	0.762	0.72
	Face & Context	0.840	0.828	0.76
Amusing	Naive	0.5	0.5	0.5
	Face	0.790	0.798	0.73
	Face & Context	0.850	0.797	0.76

TABLE V
CONFUSION MATRICES FOR THE OPTIMAL LIKING CLASSIFIER: TOP) ALL ADS (N=165), BOTTOM) ONLY AMUSING ADS (N=75). BASED ON THRESHOLD OF POINT CLOSEST TO (0,1) ON THE ROC CURVE.

Ads		Actual +ve	Actual −ve
		(High Liking)	(Low Liking)
All	Predict +ve	66	24
	Predict -ve	16	59
Amusing	Predict +ve	26	7
	Predict -ve	11	31

VII. RESULTS AND DISCUSSION

A. Ad Liking Prediction

Figure 13 shows the receiver operating characteristic (ROC) and precision-recall (PR) curves for the model for predicting ad liking score varying the SVM decision threshold in both cases. Table IV shows the area under the curve (AUC) for the receiver operating characteristic (ROC) and precision-recall (PR) curves for the ad liking score prediction model. We compare the performance using just the face features and a combination of the face and contextual features. Table V shows the confusion matrix for the SVM classifier (with the optimal decision threshold - closest point to (1,0) on the ROC curve) with face and context features for the ad liking score prediction. The Cohen's κ for the optimal classifier using each of the feature combinations is also shown in Table IV. During the validation process the median parameters C and γ were 10 and 0.1 respectively.

It is reasonable to think that ads with different aims (i.e. comical ads that aim to amuse vs. charity cause related ads that aim to evoke sympathy) would result in a different relationship between viewer's expressed responses and ad effectiveness. As a result we also performed the same analysis for just the ads labeled as intentionally amusing by the MTurk labelers. The AUC for the ROC and PR curves are shown in Table IV and confusion matrix in Table V. We see that the amusing ad models performs slightly better, with greater ROC AUC and PR AUC in three of four cases. For the amusing ad model only 18 of 75 ads are misclassified (76% accuracy).

Figure 14 shows examples of the true positives, true negative, false positives and false negatives from the best performing classifier. The emotion profiles of ads that generate high ad liking is a strong gradient and high peak in positive expressions (valence and smiles). Emotion profiles of ads that do not generate high ad liking are either low across all metrics (i.e. very few people show the expressions we detect) or feature a greater amount of negative (disgust) expressions than positive expressions (smiles). These results are congruent with previous work [32], [25] showing that peak and final emotions experienced are disproportionately weighted when people recall their feelings during the experience.

There are some cases that break these trends. For example the responses to one ad (shown in Figure 14(h)) showed large amounts of smiling and very low levels of disgust but the average liking score was below the median response. This ad had a liking score of 3.36 which is very close to the class boundary of 3.44 which explains why it would easily be misclassified. In other cases, examples in the positive class achieved similarly low facial responses (e.g. Figure 14(i and l)) and this explains why they would be misclassified.

A leave-one-ad-out training and testing scheme is not participant independent as each participant watched 10 ads during the experiment. Therefore, we repeated the ad liking analysis using a leave-ten-ads-out training, validation and testing scheme. The ROC AUC for the participant independent case was 0.821 and PR AUC was 0.752.

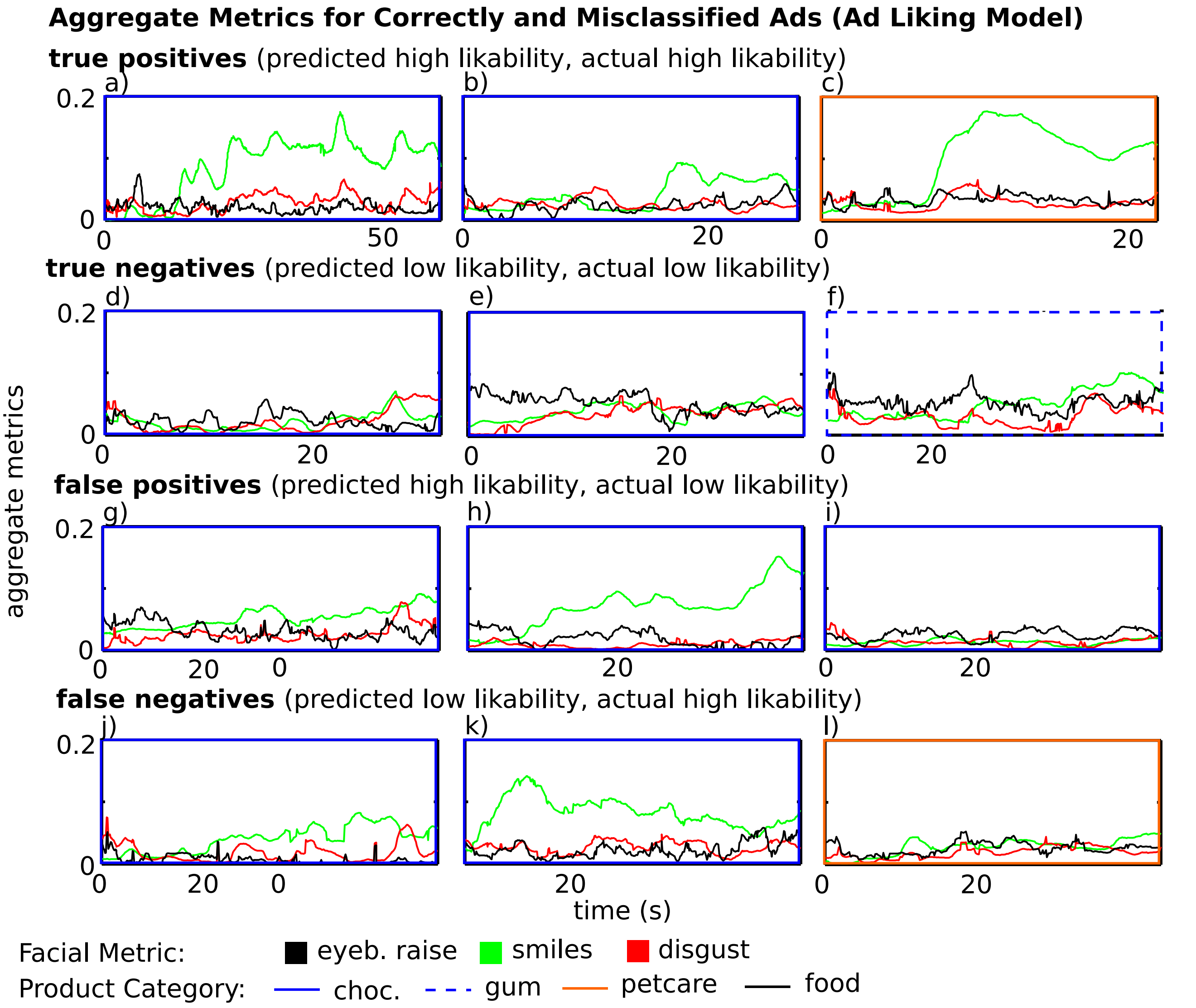


Fig. 14. Aggregate ad response metrics correctly and incorrectly classified by the ad likability model. True positives, true negatives, false positives and false negatives are shown. Aggregate: eyebrow raise - black, smiles - green, disgust - red. High peak levels of positive expressions, high expressiveness and strong increasing trend in positive expressions predict high ad likability. Low expressiveness predicts low ad likability. Individual plot outlines indicate the product category for the advertised product.

B. Purchase Intent Prediction

Figure 15 shows the ROC and PR curves for the purchase intent score prediction model. Again, we compare the performance using just the face features and a combination of the face and contextual features. In addition, we show results with all ads and just with the amusing ads. The Cohen's κ for the optimal classifier using each of the feature combinations is also shown in Table IV. Table VII shows the confusion matrix for the best performing SVM classifier for the purchase intent score prediction. Only 18 of the 74 ads (accuracy = 76%, F1-score = 0.757) were misclassified when considering just the amusing ads. During the validation process the median SVM parameters C and γ were 1.0 and 1.0 respectively.

Figure 16 shows examples of the true positives, true negative, false positives and false negatives from the best performing PI model. We also plot the timing of the appearances of the brand within each ad (broken grey line). The prediction performance was lower for the PI model than for the liking model suggesting that the relationship between facial responses and changes in PI is more complex, as expected.

TABLE VI
AREA UNDER THE ROC AND PR CURVES FOR THE PURCHASE INTENT CLASSIFIER: TOP) ALL ADS (N=163), BOTTOM) ONLY AMUSING ADS (N=74). COHEN'S κ BETWEEN THE PREDICTED AND SELF-REPORT LABELS FOR THE OPTIMAL CLASSIFIERS ARE SHOWN.

Ads	Features	ROC AUC	PR AUC	Cohen's κ
All	Naive	0.5	0.5	0.5
	Face	0.755	0.804	0.74
	Face & Context	0.739	0.741	0.71
Amusing	Naive	0.5	0.5	0.5
	Face	0.647	0.696	0.69
	Face & Context	0.781	0.811	0.76

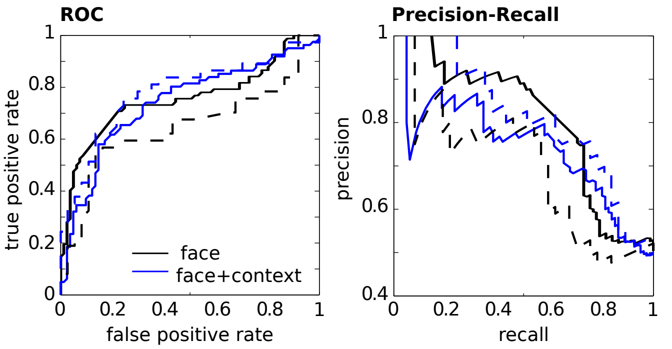


Fig. 15. Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves for the purchase intent models varying the SVM decision threshold. Black) the performance using face features, blue) the performance using face and context features combined. Unbroken lines) results for all ads, broken lines) results for only the amusing ads.

TABLE VII
CONFUSION MATRICES FOR THE BEST PERFORMING PURCHASE INTENT CLASSIFIER: TOP) ALL ADS (N=170), BOTTOM) ONLY AMUSING ADS (N=75). BASED ON THRESHOLD OF POINT CLOSEST TO (0,1) ON THE ROC CURVE.

Ads		Actual +ve (High Liking)	Actual −ve (Low Liking)
All	Predict +ve	52	19
	Predict -ve	29	63
Amusing	Predict +ve	28	9
	Predict -ve	9	28

Purely eliciting more and more positive expressions is not as successful at driving purchase intent as at driving ad liking. However, notice that for all the true positives and false negatives in Figure 16 the peak in aggregate smiling is preceded by a brand appearance, whereas this is not the case for any of the true negatives. These results support the work of Teixeira *et al.* [13] showing that emotion elicited by ads is more effective if associated with a brand. Our results suggest that brand appearances immediately prior to the peak positive emotion is a driver for increasing purchase intent. Furthermore, Figure 16 (g) shows a false positive that appears to exhibit the features of a good response (i.e. a brand appears preceding the peak positive response) but we also see that the peak in disgust responses is also shortly after a brand appearance. This suggests that negative emotions may get associated with the brand and outweigh the effects of the positive responses. This is something that would not have been identified had we only considered smile responses as was the case in [13].

Once again we repeated the PI analysis using a leave-ten-ads-out training, validation and testing scheme in order to test participant independent performance. The ROC AUC for the participant independent case was lower at 0.680. This is a challenging classification scheme as we leave out a lot of data in the training and validation process.

VIII. Conclusion and Future Work

We present the largest ever analysis of facial responses to online ads. Using an online framework and state-of-the-art facial expression analysis we capture and code 12,230 facial responses to 170 ads from four countries (France, Germany, UK, US). In total over three million frames were analyzed. This analysis would not have been practical with traditional laboratory data collection methods and manual coding of the frames of facial video.

We measured eyebrow raises, smiles, disgust and positive and negative valence expressions of the viewers on a frame-by-frame basis and mapped this to two key measures of advertising effectiveness, ad liking and changes in brand purchase intent. We note that facial responses to the ads (viewed in natural settings) were sparse. In only 17.2% of the frames was there a detectable eyebrow raise, smile, disgust or positive or negative valence expression. Almost 50% of the facial response videos had no detectable behavior. However, aggregate metrics reveal that there were detectable responses from subsets of the viewers to all the ads and this yields rich temporal affective information.

We built and tested a model for predicting ad liking based on the emotional responses of the viewers. The model performs accurately (ROC AUC = 0.850). A strong positive trend in expressed valence and high peak positive expressions suggest an ad will have high liking score. This supports previous work looking at individual responses. We built and tested a model for predicting changes in purchase intent based on the automatically coded facial responses (ROC AUC = 0.781). Performance in predicting both effectiveness measures was good. In addition we can gain insight into the structure of effective ads, such as where the brand should appear. Our results suggest that brand appearances immediately prior to the peak positive emotion is a driver for increasing purchase intent.

It is important to consider that we may not be detecting all of the facial activities that can occur. In this work we are only considering a combination of action units including AU02, AU4, AU09, AU10, AU12 and AU15. Due to the challenging nature of detecting naturalistic action units from low resolution videos we do not at this time focus on more actions. We selected the action units above to focus on because we felt they were most relevant for media measurement. However, future work should consider extending the findings to more AUs or combinations of AUs. Some preliminary work has shown the utility of predicting short-term sales impact of ads from automatically measured facial responses [33]. Future work will look at this relationship in more depth.

There are a number of extensions of this work that would be interesting to address in the future. We have only tested short video content (30-60s ads). There remain questions as to how this approach might generalize to longer content. Fleureau *et al.* [34] measured audience physiological responses to two-hour long movie content and revealed significant variations in arousal throughout the media. It is possible that facial responses to longer content might differ in frequency and duration compared to the content studied here. In this work we have only tested content from four countries. It would be very useful to extend these experiments to content in other markets (such as China or India) in order to assess the impact of cultural background on the role of emotions in advertising. Finally, the advertisements tested in this work were for

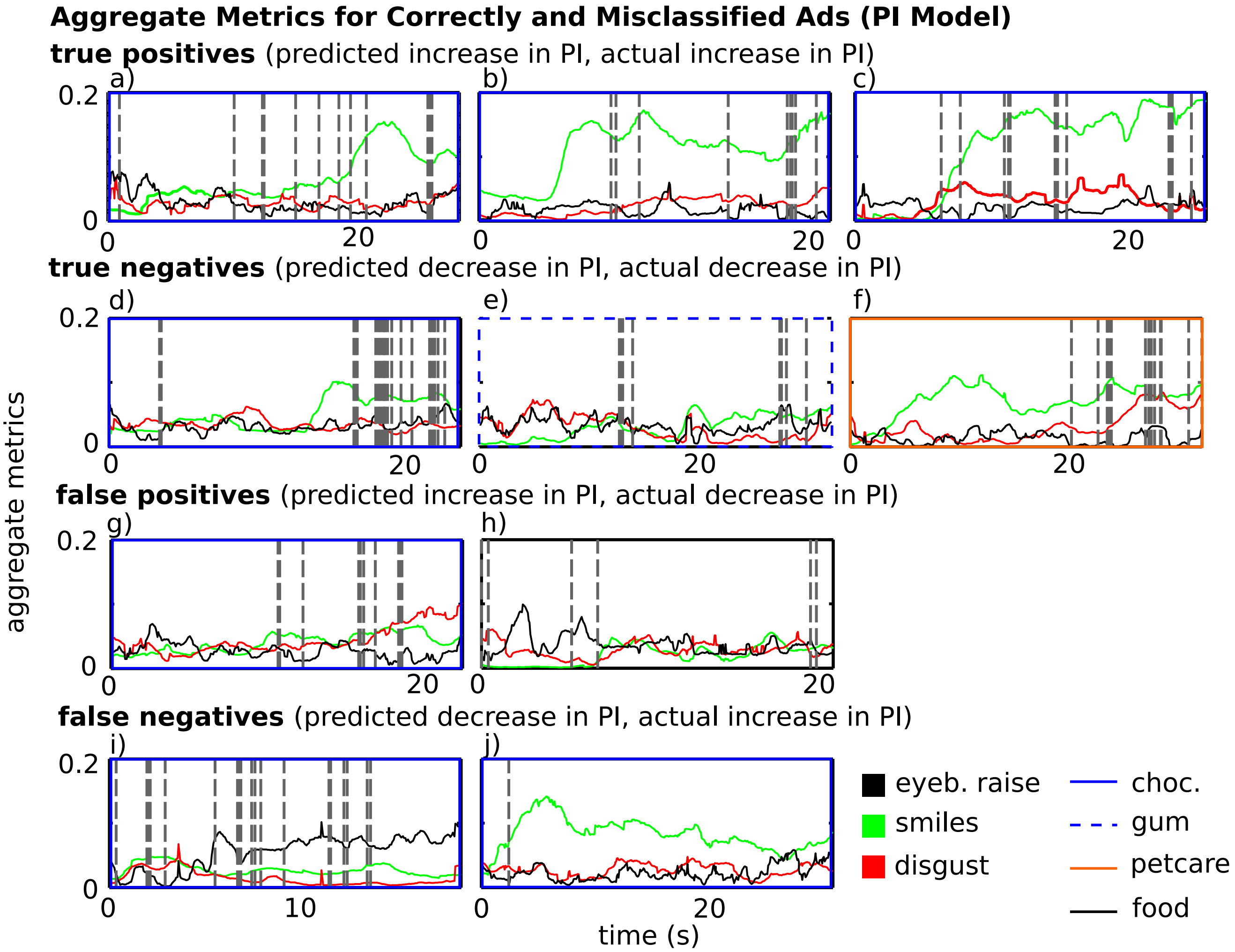


Fig. 16. Aggregate ad response metrics correctly and incorrectly classified by the purchase intent model. True positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are shown. Brand appearances within the ads are indicated by the vertical dashed lines. Notice how the peak in smile activity is preceded by a brand appearance in the TPs and not in the TNs. Aggregate: eyebrow raise - black, smiles - green, disgust - red. Individual plot outlines indicate the product category for the advertised product.

products which represented a short-term purchasing decision (e.g. a chocolate bar). However, there are many products that represent a much longer-term purchasing decision (e.g. cars), the nuances of these differences should be characterized. The approach presented here could be used in conjunction with content analysis of the audio and visual content of the ads. For instance understanding the link between emotional responses and scene changes, background music, brand appearances or other components.

Acknowledgment

Affectiva generously allowed access to the cloud-based expression analysis platform. This work was funded by the MIT Media Lab Members Consortium and MARS. Daniel McDuff was supported by an NEC fellowship. The reviewers provided very useful comments to improve this manuscript.

References

[1] P. Bolls, A. Lang, and R. Potter, “The effects of message valence and listener arousal on attention, memory, and facial muscular responses to radio advertisements,” *Communication Research*, vol. 28, no. 5, p. 627, 2001.

[2] P. Ekman and W. Friesen, “Facial action coding system,” 1977.

[3] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 1, pp. 39–58, 2009.

[4] D. McDuff, R. El Kaliouby, and R. Picard, “Crowdsourcing facial responses to online videos,” *IEEE Transactions on Affective Computing*, vol. 3, no. 4, pp. 456–468, 2012.

[5] J. Hernandez, M. E. Hoque, W. Drevo, and R. W. Picard, “Mood meter: counting smiles in the wild,” in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 301–310.

[6] J. P. Johnson, “Targeted advertising and advertising avoidance,” *The RAND Journal of Economics*, vol. 44, no. 1, pp. 128–144, 2013.

[7] D. McDuff, R. El Kaliouby, E. Kodra, and R. Picard, “Measuring voter’s candidate preference based on affective responses to election debates,” in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, 2013, pp. 369–374.

[8] R. Haley, “The arf copy research validity project: Final report,” in *Transcript Proceedings of the Seventh Annual ARF Copy Research Workshop*, 1990.

[9] E. Smit, L. Van Meurs, and P. Neijens, “Effects of advertising likeability: A 10-year perspective,” *Journal of Advertising Research*, vol. 46, no. 1, p. 73, 2006.

[10] D. McDuff, R. El Kaliouby, T. Senechal, D. Demirdjian, and R. Pi-

card, “Automatic measurement of ad preferences from facial responses gathered over the internet,” *Image and Vision Computing*, 2014.

[11] R. Hazlett and S. Hazlett, “Emotional response to television commercials: Facial emg vs. self-report,” *Journal of Advertising Research*, vol. 39, pp. 7–24, 1999.

[12] T. Teixeira, M. Wedel, and R. Pieters, “Emotion-induced engagement in internet video ads,” *Journal of Marketing Research*, 2010.

[13] T. Teixeira, R. W. Picard, and R. Kaliouby, “Why, when and how much to entertain consumers in advertisements? A web-based facial tracking field study,” *Marketing Science*, 2014.

[14] M. Pantic, “Machine analysis of facial behaviour: Naturalistic and dynamic behaviour,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3505–3513, 2009.

[15] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, “Toward practical smile detection,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 11, pp. 2106–2111, 2009.

[16] T. Senechal, J. Turcot, and R. El Kaliouby, “Smile or smirk? automatic detection of spontaneous asymmetric smiles to understand viewer experience,” in *Automatic Face and Gesture Recognition, 2013. Proceedings. Tenth IEEE International Conference on*, 2013.

[17] T. Cootes, G. Edwards, and C. Taylor, “Active appearance models,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 6, pp. 681–685, 2001.

[18] D. Cristinacce and T. Cootes, “Feature detection and tracking with constrained local models,” in *BMVC*, vol. 3. BMVA, 2006, pp. 929–938.

[19] H. Gunes and M. Pantic, “Automatic, dimensional and continuous emotion recognition,” *International Journal of Synthetic Emotions (IJSE)*, vol. 1, no. 1, pp. 68–99, 2010.

[20] K. S. Kassam, “Assessment of emotional experience through facial expression,” Ph.D. dissertation, Harvard University, 2010.

[21] D. McDuff, R. El Kaliouby, K. Kassam, and R. Picard, “Affect valence inference from facial action unit spectrograms,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, pp. 17–24.

[22] E. Kodra, T. Senechal, D. McDuff, and R. Kaliouby, “From dials to facial coding: Automated detection of spontaneous facial expressions for media research,” in *Automatic Face & Gesture Recognition and Workshops (FG 2013), 2013 IEEE International Conference on*. IEEE, 2013.

[23] H. Joho, J. Staiano, N. Sebe, and J. Jose, “Looking at the viewer: analysing facial activity to detect personal highlights of multimedia contents,” *Multimedia Tools and Applications*, pp. 1–19, 2011.

[24] S. Zhao, H. Yao, and X. Sun, “Video classification and recommendation based on affective analysis of viewers,” *Neurocomputing*, 2013.

[25] B. L. Fredrickson and D. Kahneman, “Duration neglect in retrospective evaluations of affective episodes.” *Journal of personality and social psychology*, vol. 65, no. 1, p. 45, 1993.

[26] A. C. Micu and J. T. Plummer, “Measurable emotions: How television ads really work,” *Journal of Advertising Research*, vol. 50, no. 2, pp. 137–153, 2010.

[27] J. Berger and K. Milkman, “What makes online content viral?” *Unpublished manuscript, University of Pennsylvania, Philadelphia*, 2011.

[28] A. Mehta and S. Purvis, “Reconsidering recall and emotion in advertising,” *Journal of Advertising Research*, vol. 46, no. 1, p. 49, 2006.

[29] T. Ambler and T. Burne, “The impact of affect on memory of advertising,” *Journal of Advertising Research*, vol. 39, pp. 25–34, 1999.

[30] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[31] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[32] C. Varey and D. Kahneman, “Experiences extended across time: Evaluation of moments and episodes,” *Journal of Behavioral Decision Making*, vol. 5, no. 3, pp. 169–185, 1992.

[33] D. McDuff, R. El Kaliouby, E. Kodra, and L. Larguinet, “Do emotions in advertising drive sales? use of facial coding to understand the relationship between ads and sales effectiveness,” in *European Society for Opinion and Marketing Research (ESOMAR) Congress*, 2013.

[34] J. Fleureau, P. Guillotel, and I. Orlac, “Affective benchmarking of movies based on the physiological responses of a real audience,” in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, 2013, pp. 73–78.

Daniel McDuff (S’09) received the bachelor’s degree, with first-class honors and master’s degree in engineering from Cambridge University. He received his PhD from the MIT Media Lab in 2014, working in the Affective Computing group. Prior to joining the Media Lab, he worked for the Defense Science and Technology Laboratory (DSTL) in the United Kingdom. He is interested in computer vision and machine learning to enable the automated recognition of affect. He is also interested in technology for remote measurement of physiology.

Rana El Kaliouby received the BSc and MSc degrees in computer science from the American University in Cairo and the PhD degree in computer science from the Computer Laboratory, University of Cambridge. She is co-founder, and chief technology officer at Affectiva, Inc. She is a member of the IEEE.

Jeffrey F. Cohn Jeffrey Cohn is a professor of psychology and psychiatry at the University of Pittsburgh and Adjunct Professor at the Robotics Institute, Carnegie Mellon University. He received his PhD in psychology from the University of Massachusetts at Amherst. He has led interdisciplinary and inter-institutional efforts to develop advanced methods of automatic analysis of facial expression and prosody and applied those tools to research in human emotion, interpersonal processes, social development, and psychopathology. He co-developed influential databases, including Cohn-Kanade and MultiPIE, and has cochaired the IEEE International Conference on Automatic Face and Gesture Recognition and ACM International Conference on Multimodal Interaction.

Rosalind W. Picard (M’81 - SM’00 - F’05) received the ScD degree in electrical engineering and computer science from MIT. She is a professor of Media Arts and Sciences at the MIT Media Lab, founder and director of the Affective Computing Group at the MIT Media Lab. She is also a cofounder of Affectiva, Inc. and Empatica, Inc. Her current research interests focus on the development of affective technologies for health and wellbeing. She is a fellow of the IEEE and member of the IEEE Computer Society.