Facial Expressions as Predictors of Online Buying Intention

Sun Joo Ahn, Maria Jabon, Jeremy Bailenson
Stanford University

Abstract

This study attempts to correlate the movement of facial features to consumer purchasing behavior and ultimately detect patterns in facial expressions that will predict behaviors. The facial expression of forty-two undergraduate and graduate students were recorded while they participated in an online shopping experiment where they rated twenty-four products pretested for humor and involvement levels and chose five final items to purchase. Datasets were then created by postprocessing the videos to extract the students' facial features and the movements of these features were calculated using computer software. Finally, collected datasets were analyzed using two learning algorithm classifiers. The results were promising in that classifiers were able to predict buyer intent substantially above chance. Some of the best predictions were made regarding purchase likelihood of subjects, more specifically on humorous and high-involvement product datasets. Theoretical implications and possibilities for practical applications using facial expressions as predictors of online buying intention are discussed based on these results.

* Acknowledgement: Authors would like to thank and acknowledge Dai Nippon Printing Co., Ltd. for their support and cooperation in this research.

Understanding the relationship between nonverbal behaviors and their corresponding mental states has received much attention in the field of communication (see Burgoon, Dillman, & Stern, 1993, for a review). In particular, the face has been an area of concentrated research interests as a nonverbal indicator of emotion and now has quite an established history of studies that connect specific facial expressions to specific types of emotions (Dillard & Wilson, 1993). The bulk of prior research focuses on corresponding facial expression to emotions and affective states but not many have turned to facial expressions as predictors of actual behavior.

This paper is an attempt to use computer algorithms to detect patterns in facial gestures and link them to behaviors that immediately follow the expression. In particular, the rewards for success in such short-term forecasting will be greatest in online environments. In dealing with the outpour of information on the internet, many users are forced to rely on snap judgments and first impressions. Thus, speed is imperative to any online activity and using computers to read facial expressions will enable the prediction of online behavior in real-time.

Applications of such studies will allow computers to 'read' user faces, predict the next sequence of behaviors, and act preemptively. This sort of progress in new media will allow technology to interact with, instead of simply react to, users. If computers can automatically 'read' what users want, customization will be possible at a more personal level according to each individual user needs. Customized content on the web is known to positively affect attitudes and behavior of users (Kalyanaraman & Sundar, 2006) and observing facial gestures as predictors of imminent behavior may serve as a reliable method for personalization on the web.

Our hypothesis does not involve causal relationships where we predict certain facial expressions will lead to certain behaviors. On the contrary, this is a bottom-up approach in which we track all facial expression data and look for a correlational pattern which results in expected behavior. This requires the analyses of vast amounts of facial tracking data, which is enabled by computer algorithms based on computational models, or computers that can 'read' human faces.

Interdisciplinary methods are gaining greater validity and popularity across all fields of science, and computational modeling has provided technical support with communication theories in previous work. For example, Bradley (2007) adopted computational modeling to reconfirm Gerbner's Cultivation Theory using neural network simulations and derived results that closely resemble data gathered from human subjects. Methodologies similar to the one used in the current study have previously been verified by Bailenson and colleagues in a facial-tracking study which successfully demonstrated real-time classification of emotions using an automated machine learning algorithms (Bailenson et al., 2007). Using computer learning algorithms complements the shortcomings of subjective measures and human coders to in assessing facial expressions. Not only will this ensure reliability among the analyses between all subjects but the computers will also be able to detect minute movements that escape the human eye.

Online shopping was chosen as the main task of this study to induce facial expressions because we felt that shopping would keep subjects engaged in the experiment, yielding animate facial expressions. Also, online shopping is an activity that involves a combination of attitudes, decisions and behavior that may be observed in a single shopping instance. This paper explores the possibility of predicting whether or not a consumer will purchase a certain product by reading facial expressions moments prior to the purchase.

Online Consumer Behavior

Whereas offline consumers make a single discrete choice for a single brand in a static choice content, Internet shoppers display dynamic choice behavior. Internet choice behavior may be described as an evolving series of interrelated choices, where both consumer and market play roles in shaping the context of subsequent choice events depending upon the outcome of earlier encounters (Bucklin et al., 2002). Perhaps the greatest difference between shopping in brick and mortar shops and online shopping takes place during the pre-purchase search. Compared to traditional shopping mediums the ease, access and convenience with which consumers can search online is vastly superior (Bakos, 2001). The myriads of flash banner ads that lead to a complex web of page links, search engines, and e-mail recommendations from automated agents flood online consumers with new types of informational opportunities that cannot be found in supermarkets or offline retail stores. As a result, Bucklin et al. found substantial (and substantive) differences in the choice behavior between Internet choice and offline choice.

Furthermore, using clickstream data recorded in server log files, Bucklin and Sismeiro (2003) were able to find that the browsing behavior of website visitors changed both within and across site visits. Clickstream is a term which “denotes the path a visitor takes through one or more websites” (Bucklin et al., 2002) which helps researchers understand and predict consumer behaviors online. Two aspects of browsing behavior – the visitor’s decision to continue browsing or exit the site, and the length of time spent in each page –predicted future purchase.

Until recently, clickstream has been the most widely used measure of online consumer behavior. However, in more recent years, clickstream data has been criticized. Fox and Spencer (2006) point out that although clickstream data can point out the ‘when’ and ‘what’ of web visits, it fails to answer questions about the ‘how’ and ‘why’ of consumer site use. Also, its rich volume of data is rather deceptive because clickstream data is unable to cover all site traffic in complete detail, leaving gaps due to technological shortcomings and may overestimate the actual use of the site (Phippen, Sheppard & Furnell, 2004). Finally, when the most important information about a consumer could be personal attributes such as demographics, traits, or beliefs, clickstream data is only able to save a very limited aspect of the consumer profile.

Attention is another important factor to take into consideration because clickstream does not indicate whether consumers are actually looking at the website during the full amount of time shown in the data. It could be that they engage in another task while the webpage is merely open. Alternative methods that use eye-tracking devices to monitor pupil center/corneal reflection (PCCR) or pupil dilation can directly measure the level of attention paid (Dreze & Hussherr, 2003; Hembrooke, Pan & Joachims, 2005). Yet, these devices are costly and far more intrusive compared to the common high-resolution web camera we use in this study.

In this light, observing facial gestures using a web camera is a relatively feasible and accessible alternative to eye-tracking, and supplementing the facial data with click-through data will provide greater accuracy in predicting online consumer behavior. Online behavior involves countless dispositional (individual) and environmental (situation) variables and it is almost impossible for face-tracking to become a standalone solution for predicting purchase intention. Yet, its cost-effectiveness and ease of use will make it an appealing method of data collection for online behavior prediction in the future.

Facial Gestures as Expressions of Affective and Cognitive Processes

Scholars have been trying to figure out ways to ‘read’ the face for many centuries. In addition to reading emotions and affective states from facial expressions, people even go as far as to infer personality traits from facial appearance (for instance, Willis & Todorov, 2006). The face is undoubtedly a very important source of nonverbal communication. Before we go on to our attempt at connecting facial expressions to behaviors, we should first briefly examine what we know about the face and its expressions.

To begin with, the relationship between facial expressions and emotion is so strong that making a facial expression will often trigger an emotional response from the person making the expression (Ekman, 1982). It seems that the more fundamental forms of emotional expressions such as sadness and happiness retain a great deal of physical similarity across cultures (Ekman & Friesen, 1971). Ekman and Friesen specified distinct muscular movements for each of the distinct emotional types, assuming that for each emotion there exists a hardwired program connecting subjective state, autonomic response, processes within the brain, and resulting behavior (Ekman & Friesen, 1976). They created the Facial Action Coding System in 1978 which became a major tool they used to study facial expression, emotion, and in particular, deception in depth by observing muscular movements of the face. While people are able to inhibit emotions to a certain degree, with individual differences in abilities, it is often difficult for people to control emotions expressed in their faces (Ekman et al., 1991).

Researchers also demonstrate individual differences in the relationship between emotion and facial expressions. Some people are naturally more adept at controlling the emotions expressed in their faces (i.e., have good 'poker-faces'.) Berntson et al. (1992) postulated that there are stable individual differences among subsets of populations in terms of the operation of sympathetic nervous systems (Berntson et al., 1992).

Their results indicated that: (a) Only a small subset of the population can be classified reliably as internalizers or as externalizers; (b) whether individuals are classified as internalizers or externalizers, both sympathetic and facial expressive responses intensify as the intensity of the emotional stimulus increases; (c) sympathetic and facial expressive responses to strong emotional stimuli are negatively correlated; and (d) individuals classified as generalizers exhibit somewhat variable responses, perhaps because they represent heterogeneous subsets of individuals.

Despite their classifications, their general conclusion is that given a certain degree of emotional stimuli that is not extreme, both internalizers and externalizers alike will show positive correlations between facial expressions and emotional stimuli. This is an important point to note since it implies that regardless of inherent personal traits, facial expressions involuntarily reflect internal processes, unless the expressor has an overwhelming reason to hide his emotions (e.g. when someone must lie to hide their crime).

Facial expressions, however, are not just restricted to showing emotions. Ekman (1997) questions whether the use of the words 'expression' and 'communication' is appropriate when referring to facial gestures. Although emotions are inevitably expressed through the face, facial gestures are not mere representations of internal states. Emotions affect the brain, the heart, the nervous systems, and many other physiological and psychological changes in complex processes, and facial gestures indicate that such changes are occurring. This indicates that facial expressions tell us more than simple emotions. In sum, involuntary facial gestures are a combination of emotions, cognition, and physiological changes in our body that are expressed with the movement of specific facial muscles. This is what the current study would like to focus on; gaining enough information within facial expressions to predict future behavior.

Using Computers to Detect Human Responses

Many researchers have built computational systems that can detect human affective responses. Rosalind Picard is one of such pioneers in the field of affective computing, and with her colleagues at the MIT Media lab, she has experimented with computer algorithms and affective agents which work with humans and read their affective states to help them learn and make affective-cognitive decisions (Ahn & Picard, 2005) and allow self-monitoring. To give computers affective perceptual abilities of the face in particular, researchers have used algorithms that recognize patterns from facial expressions recorded on video or from stress patterns from thermal imagery of the face (see Picard & Daily, 2005 for a more detailed review). Picard discusses the use of real-time, fully automated tracking of facial features for computer vision that tracks eyes and eyebrows (Kapoor & Picard, 2002). Picard suggests that capturing such physiological data from participants is an accurate representation of how and what they feel, and a far better option than self-report questionnaires that interrupt the participants' affective-cognitive processes (see Reynolds & Picard, 2005 and Picard & Daily, 2005 for further details).

Alex Pentland and Tanzeem Choudhury from the MIT Media Lab also examine face recognition as an alternative to using a keyboard or a mouse in smart environments and as a safer means of personal identification system (Pentland & Choudhury, 1999). Pentland continues to search for a dependable means for behavioral prediction via wearable sensors. Since advanced modern technology has minimized the bulk and the weight of cameras and audio recorders, he proposed that these wearable technologies can recognize common daily human activities in real time and facilitate social interactions (e.g. negotiations, face-to-face communication, organizational dynamics) by predicting the participants' next actions (Choudhury & Pentland, 2002; Choudhury & Pentland, 2003; Pentland et al., 2004).

The results of efforts to model the dynamics of facial expressions have led to quite a few experiments where scholars attempt to apply more realistic faces to virtual agents (see Deng, Bailenson, Lewis & Neumann, 2006 for a detailed review of these works). Such endeavors hold great promises for future work in the field of HCI since it means that objective readings of facial-point movement will yield predictions of subjective affect and attitudes. For instance, when automated observations of facial gestures by computers lead to accurate predictions of purchasing behavior, we will see revolutionary changes in advertising and marketing strategies, as well as in the structure and dynamics of online commercial activities.

Research Design of the Current Study

This study examines an online shopping situation, using facial tracking software and learning algorithms to learn to classify facial expressions of the person sitting in front of the monitor. Very few, if any at all, studies to date have attempted to predict online shopping behavior with facial expressions. Consequently, this is largely an exploratory experiment to search for a pattern of facial gestures that correlate with intention to purchase products.

The experiment itself required two online shopping tasks where subjects were presented with a group of products to rate and choose. Products were pretested and categorized for differing degrees of involvement and humor.

Involvement is a concept often used vaguely in business or marketing literature that generally indicates the level of engagement people feel when they consume a certain product or content. Shopping is an activity that can be rational, emotional, or both depending on the reasons, motives and goals of the consumer. This is because shopping is a dynamic process of decision-making that involves cognitive and affective investments. For the purposes of this study, involvement will refer to the degree of personal relevance or “connection” triggered by products. Our prediction is that different levels of personal involvement with the presented products will trigger different cognitive and affective internal processes which may be manifested differentially on the face. High involvement products may produce a more predictive facial expression than low involvement products.

We also felt that humor should be included as a factor that may influence product choice since humor appeals are so popular among marketing practitioners. Of course, researchers have expressed difficulty in even firmly grasping the concept of humor. Robinson and Smith-Lovin (2001) comment that like beauty, humor is a word that we all know well and use commonly but cannot define with ease.

This paper does not attempt to measure how humorous a certain product is or how effective humor is in persuading consumers to purchase a product. For the purposes of this project, humor is operationalized simply as an aspect of the product which induces a certain kind of distraction or orienting-response which will elicit richer facial expressions. In their correlational study, Deckers and Hricik (1984) suggest that humor response may be paralleled to orienting-response, which is an instinctive reaction to a novel and unexpected stimulus. We expect products with humorous content to trigger an orienting-response which will hopefully be distracting enough for internalizers to actively express their cognitive or affective states through facial gestures. Then these facial reactions will not be measured as humor ratings but be correlated to likeability and purchase intentions. Not only have studies such as the one carried out by Deckers, Kuhlhorst, and Freeland (1987) concluded that facial expressions do not necessarily correlate with humor perception but we use humor as a distraction stimulus and have no need to measure the felt level of humor. Because humor may directly manifest itself in facial expressions, we predicted the predictive models may perform best when people were evaluating humorous products.

Stimuli

A pretest of 73 products was conducted among 21 graduate and undergraduate Stanford students in an online survey format. The test assessed the level of involvement and humor to select the final products to use in the actual experiment.

The products for this pretest were selected from www.amazon.com and www.t-shirthumor.com which are general websites with high accessibility. 44 products were chosen as high involvement/low humor (10 phones, 10 laptops, 5 printers, 10 watches, 5 games, 4 jukebox), 25 products were chosen as low involvement/low humor (10 pen, 10 soap, 5 tissue) and 4 products were chosen as medium involvement/high humor (4 shirts).

A computer randomized the products and presented a total of 24 products to each survey respondent. The survey was carried out online and four rating questions were attached to each product. The questions asked: 1) how much does this product tell other people something about you? 2) how important is this product to you? 3) how interesting is this product to you? 4) how funny is this product to you?

The first three questions were related to product involvement and were adapted from the work of Bauer et al. on product involvement and consumer decision-making styles (Bauer et al., 2006). In their study, Bauer and colleagues attributed three factors to product involvement: sign value, importance, and pleasure. In the past, risk was included as an additional factor, but they omitted the risk factor in accordance to more recent research trends. The first three questions used in our online survey replicate these three factors using the same five-point Likert scale. The final question gauges the level of amusement people feel when looking at the product, also measured with a five-point Likert scale.

Respondents were advised not to use to the back function in the web browsers and to finish rating all 24 products in one sitting.

After the pretest, a total of 24 products were selected as the final stimuli from the pool of 73 products for the main experiment, with 6 products each in the low involvement/low humor, medium involvement/low humor, high involvement/low humor, and medium involvement/high humor categories.

Scores from the first three rating questions related to involvement were averaged to derive the 6 products in three different involvement categories with low humor. Products were selected to represent product variety in each category. In the high involvement category, 3 phones and 3 laptops with the highest mean scores that range from 3.3 to 5.0 were selected. In the low involvement category, 2 pens, 2 soaps and 2 tissues were selected for the lowest mean scores ranging from 1.16 to 1.67. In the medium involvement category, 3 soaps, 2 laptops, and 1 phone were selected within the mean score range of 3.0 to 3.42, showing a good mixture of some products from the lower tail of the high involvement category and some from the higher tail of the low involvement category.

Scores from the humor rating question were arranged in ascending order to select 6 products with the highest humor scores, ranging from 3.25 to 3.75. These products were 3 shirts with funny pictures, 1 soap, 1 pen, and 1 game.

Histograms from both the involvement categories and the humor category showed normal distributions.

Humorous	Low Involvement	Med. Involvement	High Involvement
Soap 2 (3.75)	Soap 10 (1.67)	Soap 9 (3.0)	Laptop 5 (5.0)
Shirt 1 (2.75)	Tissue 4 (1.58)	Laptop 9 (3.17)	Laptop 7 (5.0)
Shirt 2 (3.0)	Pen 7 (1.58)	Soap 5 (3.0)	Phone 5 (4.3)
Pen 6 (3.0)	Soap 1 (1.5)	Phone 1 (3.08)	Laptop 6 (4.3)
Game 2 (3.5)	Tissue 5 (1.5)	Soap 4 (3.33)	Phone 6 (3.7)
Shirt 3 (3.25)	Pen 8 (1.16)	Laptop 10 (3.42)	Phone 4 (3.3)

Table 1. Final set of 24 product stimuli and mean rating scores from the pretest

Experiment Procedure

A total of 42 graduate and undergraduate students were recruited for participation in this experiment. These participants were called in to complete a two-part online task. The subjects participated in the study in a separate room while the researcher observed in an adjoining room, entering the subject’s room only to set up the computer for the next task. A web camera was set up on top of the monitor to record the facial gestures of all the participants. Although all the participants were aware of the video recording, they stopped paying attention to the camera once they became immersed in their shopping task. Instructions were read out aloud to the participants before beginning each part of the task. Subjects had to remove glasses and wear contact lenses as well as remove caps and hats for the collection of data that would be legible by the computer vision program.

Task design

The first part of the task was to answer two rating questions regarding the 24 products selected from the pretest. The first question asks, ‘What is your overall reaction to this product?’ and measures the participant’s preference of the product based on how much he/she likes the product. The second question asks, ‘How likely is it that you would purchase this product?’ and measures how likely the participant is to purchase the product when he/she is given the opportunity to buy it. These two questions were adapted from a product evaluation study carried out by Lim and colleagues (Lim, Olshavsky & Kim, 1988), and then adjusted to a five-point Likert scale in place of its original 10-point scale. In particular, the 5 point intention scale was borrowed from Jamieson and Bass’s study (1989).In the first part of the task, subjects looked at the picture of the product for 10 seconds and the two questions discussed above showed up on the right-hand side of the page after 10 seconds. This was designed to give participants a good amount of time to take a close look at the product itself before losing their attention to the questions. When they finished answering both questions, they clicked ‘continue’ on the web page and went on to rate the next product.

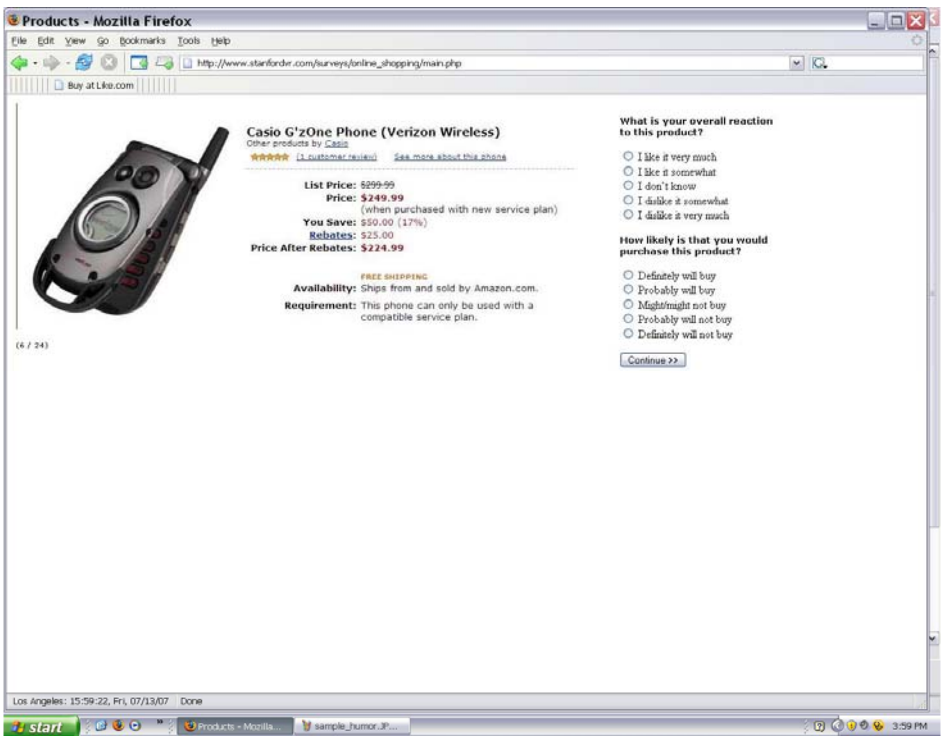


Figure 1. Sample screenshot of a high-involvement product (cellular phone)

The second part of the task showed a complete list of the 24 products rated by the participant in thumbnail pictures and required the participant to choose the final set of 5 products that he/she would like to purchase. This was to relate the purchase intention measured in the first task to actual purchase behavior measured by this second task.

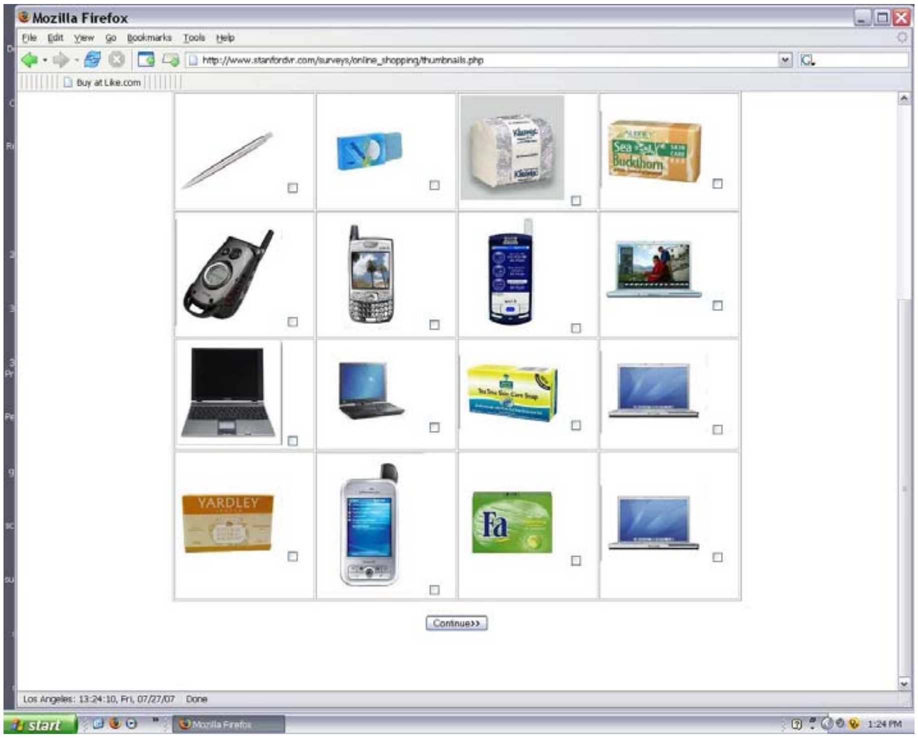


Figure 2. Sample screenshot of second task

Dataset Construction

Of the data from 42 subjects that participated, due to various technological difficulties in recording video unobstructed by hands, hair and lighting which impeded the facial tracking software, we were able to use only 24 for our final dataset. Sixteen of the subjects used for the final dataset were female and eight were male subjects. The dataset construction first required the extraction of facial features from the web camera videos taken during the study and synchronizing that data with the user responses to the survey. Statistics such as averages, velocities, minimums, maximums, standard deviations, and ranges were then computed on the facial points along with metafeatures such as PERCLOS rate (percentage of the time the eyes are closed) and mouth openness level. Wavelet transforms were then applied to the facial features and finally all the data was accumulated into a rich dataset for our computational models used to predict buyer intent. Figure 3 summarizes the phases of analysis performed.

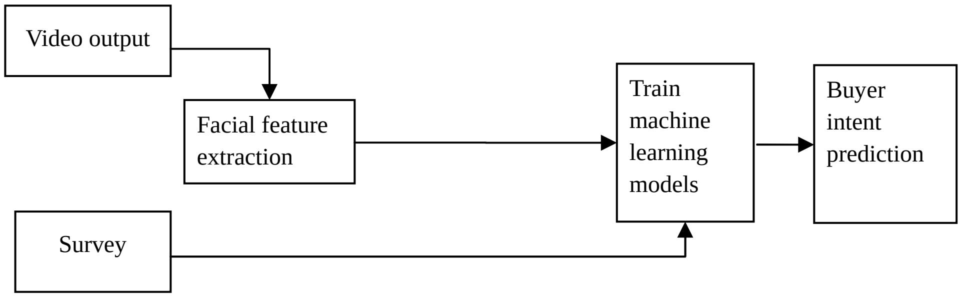


Figure 3. Data analysis procedure

Extracting Facial Features

Videos of subject faces captured during the task were post-processed using the OKAO vision library, which tracks 14 points of the face along with eye and mouth openness level and the head movements such as pitch, yaw, and roll at a rate of 20 frames per second. In Figure 4 we present a screenshot depicting the OKAO face tracking points and in Table 2 we describe the comprehensive list of OKAO face tracking features.

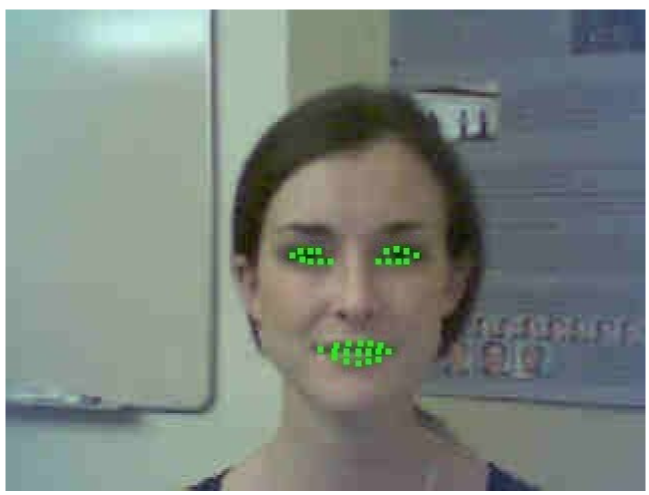


Figure 4. OKAO computer vision algorithm tracking points on subject's face

Confidence	Right Pupil X	Right Upper Lip
Average X	Left Pupil	Left Upper Lip
Average Y	Upper Lip Center	Right Lower Lip
Standard Deviation between all points and the average point	Lower Lip Center	Left Lower Lip
Right Eye Ratio	Right Mouth Corner	Gaze Tilt
Left Eye Ratio	Left Mouth Corner	Gaze Pan
Mouth Ratio	Left Outer Eye Corner	Left Eye Open Level
Pitch	Left Inner Eye Corner	Right Eye Open Level
Yaw	Right Outer Eye Corner	Mouth Open Level
Roll	Right Inner Eye Corner	

Table 2. The exhaustive list of the OKAO library output points of subject's face

Synchronizing Facial Data and Survey Output

We synchronized the facial data retrieved from OKAO with the user responses collected from the rating questions presented during the first task of the experiment by calculating the offset into the video of each response from their respective timestamps. We then collected the 200 frames of facial data (10 seconds) prior to each response for each subject and tagged each frame with labels such as subject ID, gender, item number, involvement, humor ranking, like ranking, buy ranking, and *buy tag* (whether or not this item was one of the five bought in the last part of the task or not).

Creating Final Datasets for Classification

The result of the calculations described above was a comprehensive dataset including 576 entries, each having 576 attributes comprised of all the facial data statistics described above along with the buy rating (1-5), like rating (1-5), five item rating (1 if they bought the item and 0 otherwise), gender (male or female), involvement level (low, medium, or high), and humor level (humorous or not humorous).

In order to fully utilize this rich dataset, we split it into 59 different datasets, each of which would emphasize different aspects of the data. We split first by whether we were looking at a buy, like, or five item rating. This allowed us to predict buy intention and likability separately. Within the buy and like subcategories we also split between gender, involvement, and humor categories which allowed us to explore the possibility of greater predictive power in each of these categories. Finally, since we were only concerned with a binary prediction, intent to buy ("yes") or no intent to buy ("no"), we further divided each like and buy sub-dataset by categorizing in one of four ways: either with a rating of one or two being considered a "no" and a rating of four or five being considered a "yes", with a rating of one, two, or three being considered a "no" and only a rating of five being considered a "yes", with a rating of one, two, or three being considered a "no" and a rating of four or five being considered a "yes", or with only a rating of one being considered a "no" and only a rating of five being considered a "yes". All of these divisions yielded a total of 59 distinct datasets (Table 3).

	All Subjects	Male	Female	Humorous	Low Involvement	Medium Involvement	High Involvement
Buy	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes
	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes
	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes
	1 = no 5 = yes	1 = no 5 = yes	1 = no 5 = yes	1 = no 5 = yes	1 = no 5 = yes	1 = no 5 = yes	1 = no 5 = yes
Like	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes	1,2 = no 4,5 = yes
	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes	1,2,3 = no 4,5 = yes
	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes	1,2,3 = no 5 = yes
	1 = no 5 = yes	1 = no 5 = yes	1 = no 5 = yes	1 = no 5 = yes	1 = no 5 = yes	1 = no 5 = yes	1 = no 5 = yes
5 Items	0 = no 1 = yes	0 = no 1 = yes	0 = no 1 = yes	X	X	X	X

Table 3. Dataset subdivisions

Predicting Consumer Ratings

In order to predict the consumer's likeability and purchase likelihood of items based upon the data we collected, we experimented with two state-of-the-art classifiers—Support Vector Machines and Logitboost with a decision stump weak classifier (Freund & Schapire 1996, Friedman, Hastie & Tibshirani 2000). We built the classifiers using the publicly available tool, WEKA. We trained a Support Vector Machine classifier with a linear kernel and a Logitboost using 40 iterations. In all of our experiments, we validated our models using 5-fold cross validation. Both classifiers were then run on each of the 59 datasets, resulting in a total of 118 analyses.

Evaluating Performance

To evaluate the performance of each classifier we established three operational definitions derived by calculations from the classifier outputs: Yes Performance, No Performance, and Chance Precision.

Yes Performance (“yes hits”), is the rate of correctly classified instances in the “yes” category. That is, Yes Performance would be high when a classifier is able to successfully ‘guess’ “yes” instances that turn out to be “yes.” Yes Performance would be low when a classifier guesses “yes” instances that turn out to be “no” instances.

Similarly, *No Performance (“no hits”)* is the rate of correctly classified instances in the “no” category. In the same way as Yes Performance, No Performance would be high when a classifier is able to successfully ‘guess’ “no” instances that turn out to be “no.”

Chance Precision is the overall ratio of correctly classified instances in the “yes” and “no” categories respectively, divided by the total number of instances. There would be a separate Chance Precision for Yes Performance and No Performance. These definitions are summarized in Figure 5.

Yes Performance =
Correctly classified “yes” instances / Total number of “yes” instances
Chance Precision for Yes Performance =
Total number of “yes” instances / Total number of all instances
No Performance =
Correctly classified “no” instances / Total number of “no” instances
Chance Precision for No Performance =
Total number of “no” instances / Total number of all instances

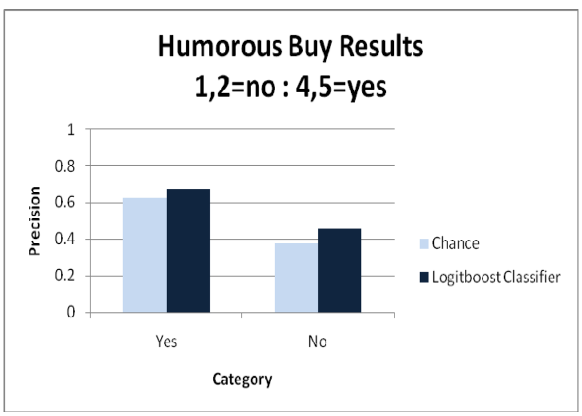
Figure 5: Performance Definitions

One must note that there is sometimes a tradeoff between “yes” and “no” performance in the classified categories. For example, in a dataset comprised of 75% “yes” instances, the classifier could simply always predict “yes” and get Yes Performance of 1. However, this high accuracy would be at the cost of a very low No Performance (0), since the classifier would missing every single “no” instance. Therefore when evaluating our classifiers we looked for not only high accuracy, but also high performance rates in both the “yes” and “no” categories.

For the purpose of brevity we do not discuss the results of all of our 59 datasets, and opt instead to present only results from datasets that are simultaneously at least 5% above Chance Precision in both Yes and No Performance.

Results

Of the 59 possible datasets, we found the split with a rating of 1, 2 = “no” and a rating of 4, 5 = “yes” and the split that uses 1 = “no” and 5 = “yes” to provide more consistent results. Therefore our analysis focused upon the four best results of Logitboost and SMO classifiers. These results indicate how well these classifiers are able to predict consumer behavior based on face-tracking data, moments prior to the actual behavior.



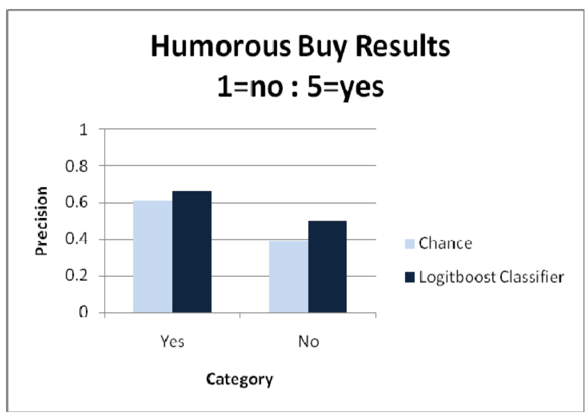
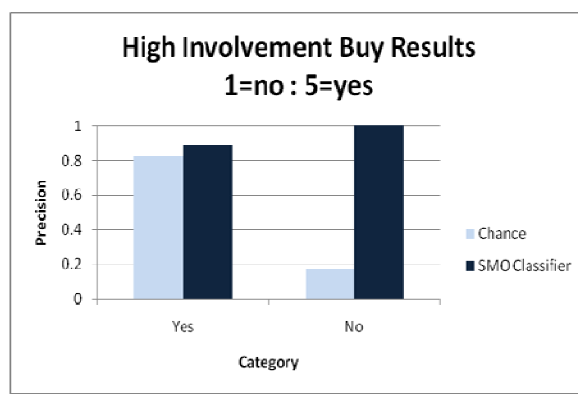


Figure 6. Best results from Logitboost and SMO Classifiers



The most significant results were obtained from the buy dataset, with the humorous products in particular. The dataset using the Logitboost Classifier with the 1, 2 = “no” and 4, 5 = “yes” split shows a Yes Performance of 0.676 which is 5.1% above Chance Precision and a No Performance of 0.457 which is 8.2% above Chance Precision. This means that the Logitboost Classifier was able to predict consumers’ purchasing intentions to buy humorous products 5.1% better than chance. The same dataset enabled the classifier to predict consumers' purchasing intentions *not* to buy humorous products 8.2% better than chance.

A 1 = "no" and 5 = "yes" split using the Logitboost Classifier records Yes Performance as 0.667 which is 5.6% above chance and No Performance as 0.5 which is 11.1% above chance. The same split using a SMO Classifier fares even better. The Yes Performance is 0.7 which is 8.9% above chance and No Performance is 0.5 which is 11.1% above chance.

The best performance of all was seen in the high involvement products (also buy dataset), using the SMO Classifier, with 1 = "no" and 5 = "yes" split. In this dataset, the Yes Performance was 0.889 which is 6.1% above chance and No Performance was close to 1.0 which is 82.7% above chance.

These results suggest a strong and realistic possibility of successfully predicting buyer intent from facial expressions made some seconds prior to purchase. There were also some distinct trends in the performance of the classifiers. In general, the classifiers performed better on buy data, rather than like data in all categories. The classifiers also performed very well in the humorous buy dataset where both the "yes" hit rate and "no" hit rate were relatively high. In the involvement category the classifiers performed significantly better on the high involvement items than for the low involvement items.

Discussion

This paper has explored the possibility of detecting correlations between specific patterns of facial expressions and behaviors and has successfully demonstrated that facial expressions can forecast purchasing behavior. If completely automated and computerized vision algorithms are able to predict the behavior of consumers, it will open up revolutionary horizons for businesses and industries. Admittedly, the effect sizes of this first attempt were not huge but the results are significant and offer exciting prospects for future research.

The fact that classifier performance was significantly higher for the buy dataset than for the like dataset implies that consumers incorporate different information processing tactics when they evaluate products (rating the likeability) or make purchasing decisions (rating the purchase likelihood). There has been previous work by Sood and Forehand (2005) where experimental tasks were differentiated between choice and judgment, proving that memory tracing for choice task is higher than for judgment task. Choice tasks can be thought in parallel to our buy ratings in this study and judgment tasks to our like ratings. The authors agree that choice tasks encourage consumers to rely on heuristics, seek less information, and base decisions on dimensional comparisons. Then the buy rating may be based more on heuristics than the like ratings. Since heuristic decision-making takes into account peripheral cues such as emotions, and increase in emotional stimuli leads to an increase in facial expressions, subjects are likely to show richer facial expressions for buy ratings than for like ratings, thus increasing the probability of better predictions.

Products that invoked amusement in people showed markedly higher prediction rates than for non-humorous items. As discussed previously, this could be accounted for by the fact that humor triggers an orienting response where individuals display a certain level of arousal and increased attention to process the novel stimulus. This could lead to a moment of ‘leakage’ in facial expressions where the individual temporarily loses cognitive control of the face and true expressions ‘leak’ to the surface, thus allowing the face to display his/her genuine thoughts or feelings. Even when the computer software accurately tracks all movements of data points on the face, we cannot gain predictive power unless these facial expressions reflect genuine cognitive and/or affective changes. Unexpected exposure to humor could interfere with conscious efforts to inhibit facial expressions as people focus their attention to the novel stimulus and lead to greater expressivity.

Another possibility is that people may show a certain pattern of facial movements when positive affective states are invoked. Advertisements that appeal to humor may be an interesting case for future research, as the prediction of purchase likelihood after seeing a funny ad may yield better results than non-humor ads. If this is true, advertisers could present humorous ads of various products until they obtain enough facial data to detect the ‘yes’ expression (faces that indicate purchase likelihood) and then concentrate their marketing efforts for that select product. Then successful facial tracking could lead to genuine one-to-one customization realized through automated learning algorithms.

What is particularly interesting is that among the humor products, predictive power was higher for buy datasets than for like datasets. Whereas most classifier results tended to be biased toward either ‘yes’ or ‘no’ in prediction, humorous products displayed equally high predictive powers for both ‘yes’ and ‘no’ items. Reflecting on the arguments presented above, it should be noted that both humor products and buy ratings trigger decision-making processes that rely on heuristic and dimensional comparisons. This is closely related to affect-based decisions and peripheral cues, which implies that consumer behavior can be best predicted through change in facial expressions when consumers are encouraged into affective states or to make ‘hedonic’ consumption decisions. Practical implications of this finding can lead to diverse and creative applications, such as playing music on online shopping sites or incorporating humor to the site interface in order to better predict consumer behavior.

Another result that adds strength to this explanation is the fact that high involvement products generally had higher predictive power than the medium or the low involvement products. Intuitively, high involvement products encourage consumers to become personally engaged in the decision-making process and such personal connections are bound to bring in heuristic cues heavily loaded with emotions and personal memories. Then, high involvement products may result in higher predictive power because heuristic-based decision-making processes yield richer facial expressions. In particular, we obtained better results with buy ratings (which also incorporate heuristic choices) than like ratings in the high involvement product category as well. In future studies, placing greater emphasis on the manipulation of affect and personal involvement may result in more distinctive results.

Consumer psychology and behavior are intricate combinations of high- and low-level decision making processes, shopping environments and individual personalities. Gauging this comprehensive activity based on the movements of 14 points on the face was an ambitious task to begin with. In particular, with no prior experiments of this kind to depend on, there were severe limitations on building theoretical backgrounds for this study. Despite such difficulties, we found results that are encouraging enough to further pursue the idea of predicting human behavior from modeling the dynamics of the face. Hopefully, there will soon come a day when our eyes literally become ‘the window to the soul.’

References

Ahn, H., & Picard, R.W. (2005). Affective cognitive learning and decision-making: A motivational reward framework for affective agents. The 1st International Conference on Affective Computing and Intelligent Interaction, October 22-24, 2005, Beijing, China.

Bailenson, J.N., Pontikakis, E.D., Mauss, I.B., Gross, J.J., Jabon, M.E., Hutcherson, C.A., Nass, C., & John, O. (2007, in press). Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human Machine Studies.*

Bakos, Y. (2001). The emerging landscape of retail e-commerce. *Journal of Economic Perspectives*, *15*, 69-80.

Bauer, H., Sauer, N., & Becker, C. (2006). Investigating the relationship between product involvement and consumer decision-making styles. *Journal of Consumer Behavior, July-August*, 342-354.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*, 1293-1295.

Berntson, G.G., Cacioppo, J.T., Crites, S.L., Uchino, B.N., Smith, G., & Snydersmith, M.A. (1992). *Journal of Personality and Social Psychology, 62(1),* 110-128.

Bradley, S. D. (2007). Neural network simulations support heuristic processing model of cultivation effects. *Media Psychology, 10,* 449-469.

Bucklin, R.E., Lattin, J.M., Ansari, A., Gupta, S., Bell, D., & Coupey, E., et al. (2002). Choice and the internet: From clickstream to research stream. *Marketing Letters (August),* 243-256.

Bucklin, R.E. & Sismeiro, C. (2003). A model of website browsing behavior estimated on clickstream data. *Journal of Marketing Research, 40(August),* 249-267.

Bucklin, R.E. & Sismeiro, C. (2004). Modeling purchase behavior at an e-commerce website: a task-completion approach. *Journal of Marketing Research, 41(August),* 306-323.

Burgoon, J. K., Dillman, L., & Stern, L. A. (1993). Adaptation in dyadic interaction: Defining and operationalizing patterns of reciprocity and compensation. *Communication Theory, 3,* 196-215.

Choudhury, T., & Pentland, A. (2002). The sociometer: A wearable device for understanding human networks. CSCW 2002.

Choudhury, T., & Pentland, A. (2003). Sensing and modeling human networks using the sociometer. Proceedings of ISWC 2003.

Deng, Z., Bailenson, J., Lewis, J. P., Neumann, U. (2006). Perceiving Visual Emotions with Speech. Proceeding of the 6th International Conference on Intelligent Virtual Agents. California, USA. 21-33, August.

Dillard, J.P., & Wilson, B.J. (1993). Communication and affect: Thoughts, feelings and issues for the future. *Communication Research, 20,* 637-647.

Dreze, X., & Hussherr, F. (2003). Internet advertising: is anybody watching? *Journal of Interactive Marketing, 17(4),* 8-23.

Ekman, P., & Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*, 124-129.

Ekman, P., & Friesen, W.V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior, 1(1),* 56-75.

Ekman, P. (1982). *Emotion in the human face*. Cambridge University Press.

Ekman, P. (1997). Should we call it expression or communication? *Innovations in Social Science Research, 10(4),* 333-344.

Ekman, P., O'Sullivan, M., Friesen, W.V., & Scherer, K.R. (1991). Face, voice and body in detecting deception. *Journal of Nonverbal Behavior, 15*, 125-135.

Fox, M., & Spencer, B. (Eds.) (2006). *Proceedings of the 8th International Conference on Electronic Commerce: The new e-commerce – Innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet,* 2006, Fredericton, New Brunswick, Canada, August 13-16, 2006. ACM 2006. 463-470.

Freedman, J. (1964). Involvement, discrepancy, and change. *Journal of Abnormal and Social Psychology, 69 (3),* 290-295.

Freund, Y. & Schapire R. (1996). *Experiments with a new boosting algorithm. Machine Learning Proceedings to the Thirteenth International Conference.* Morgan Kaufman, San Francisco, 148-156.

Gasper, K. & Clore, G.L. (1998). The persistent use of negative affect by anxious individuals to estimate risk. *Journal of Personality and Social Psychology, 74,* 1350-1363.

Hembrooke, H., Pan, B., & Joachims, T. (2005). In Google we trust: users' decisions on rank, position and relevancy. *Journal of Computer-Mediated Communication, Special Issue on Search Engines.*

Holbrook, M.B., & Gardner, M.P. (1993). An approach to investigating the emotional determinants of consumption durations: Why do people consume what they consume for as long as they consume it? *Journal of Consumer Psychology, 2,* 123-142.

Jamieson, L., & Bass, F. (1989). Adjusting stated intention measures to predict trial purchase of new products: A comparison of models and methods. *Journal of Marketing Research, 26(3),* 336-345.

Kalyanaraman, S., & Sundar, S. S. (2006). The psychological appeal of personalized content in web portals: Does customization affect attitudes and behavior? *Journal of Communication, 56,* 110-132.

Kapoor, A., & Picard, R.W. (2002). Real-time, fully automatic upper facial feature tracking. Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition, May 20-21, 2002, Washington D.C.

Lim, J., Olshavsky, R., & Kim, J. (1988). The impact of inferences on product evaluations: Replication and extension. *Journal of Marketing Research, 25(August),* 308-316.

Pentland, A., & Choudhury, T. (1999). Personalizing smart environments: Face recognition for human interaction. IEEE Computer, February 2000. Special issue on Biometrics.

Pentland, A., Curhan, J., Khilnani, R., Martin, M., Eagle, N., & Caneel, R., et al.. (2004). A Negotiation Analyzer. Appears UIST 2004, Santa Fe, NM, October 25-27.

Phippen, A., Sheppard, L., & Furnell, S. (2004). A practical evaluation of web analytics. *Internet Research, 14(4),* 284-293.

Picard, R.W., & Daily, S.B. (2005). Evaluating affective interactions: Alternatives to asking what people feel. CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches, April 2005, Portland, OR.

Reynolds, C., & Picard, R.W. (2005). Evaluation of affective computing systems from a dimensional metaethical position. 1st Augmented Cognition Conference, in conjunction with the 11th International Conference on Human-Computer Interaction, July 22-27, 2005, Las Vegas, NV.

Schwarz, N., & Clore, G.L. (1983). Mood, misattribution, and judgments of well-being: Informative and direct vs. functions of affective states. *Journal of Personality and Social Psychology, 45*, 513-523.

Schwarz, N., & Clore, G.L. (1988). How do I feel about it? Informative functions of affective states. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition, and social behavior* (pp.44-62). Toronto: Hogrefe International.

Sood, S., Forehand, M. (2005). On self-referencing differences in judgment and choice. *Organizational Behavior and Human Decision Processes, 98*, 144-154.

Willis, J., Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17 (7),* 592-298.