



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Just the Faces: Exploring the Effects of Facial Features in Print Advertising

Li Xiao, Min Ding

To cite this article:

Li Xiao, Min Ding (2014) Just the Faces: Exploring the Effects of Facial Features in Print Advertising. Marketing Science 33(3):338-352. <https://doi.org/10.1287/mksc.2013.0837>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2014, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Just the Faces: Exploring the Effects of Facial Features in Print Advertising

Li Xiao

School of Management, Fudan University, Shanghai, People's Republic of China 200433, lixiao@fudan.edu.cn

Min Ding

Smeal College of Business, Pennsylvania State University, University Park, Pennsylvania 16802; and
School of Management, Fudan University, Shanghai, People's Republic of China 200433, minding@psu.edu

Human faces are used extensively in print advertisements. In prior literature, researchers have studied spokespersons in general, but few have studied faces explicitly. This paper aims to answer three questions that are important to both researchers and practitioners: (1) Do faces affect how a viewer reacts to an advertisement on the metrics that advertisers care about? (2) If faces do have an effect, is it large enough to warrant careful selection of faces when constructing print advertisements? (3) If faces do have an effect and the effect is large, what facial features elicit such differential reactions on these metrics, and are such reactions different across individuals and/or product categories? Relying on the eigenface method, a holistic approach widely used in the computer science field for face recognition, we conducted an empirical study to answer these three questions. The results show that different faces do have an effect on people's attitude toward the advertisement, attitude toward the brand, and purchase intention and that the effect is nontrivial. Multiple segments were identified and substantial differences were found among people's reactions to the faces in the ads across those segments. We also found that the effect of faces interacts with product categories and is mediated by various facial traits such as attractiveness, trustworthiness, and competence. Implications and directions for future research are discussed.

Keywords: face; facial features; advertising effectiveness; eigenface

History: Received: July 9, 2012; accepted: November 12, 2013; Preyas Desai served as the editor-in-chief and Michel Wedel served as associate editor for this article. Published online in *Articles in Advance* February 10, 2014.

Introduction

Print media plays an important role in advertising practice. According to Nielsen's (2013) report, global print advertising spending reached more than \$150 billion in 2012, accounting for more than a quarter of total media advertising spending in the world. In print advertising, it is common practice to have a human endorser present the brand or product and advocate its adoption. We collected more than 600 print ads that appeared in the United States from 2001 to 2012 from the AdForum database¹ and found that over 50% of print ads feature faces. A recent GfK MRI (2011) report indicated that only a small portion (less than 10%) of print ads use celebrity endorsers, a finding consistent with our search of the AdForum database. For many ads that use non-celebrity endorsers, consumers know nothing about these endorsers other than what they can infer from the endorsers' faces.

Although faces are widely used in print ads, the selection process in practice is ad hoc. We interviewed an advertising agency on how print ads are created for fashion brands. The normal procedures are as follows: (a) the advertising agency receives a request from a client company, (b) it sends a model request to multiple modeling agencies, (c) the modeling agencies send back pictures of recommended models (100+), (d) the advertising agency selects some models (about 10 to 20, based the fashion brand and its request) and photographs them wearing the same clothes, (e) the advertising agency gives the pictures to the client, and (f) the client decides which models to use (often in conjunction with the advertising agency), at which point copy testing may or may not be involved.

Although there is a large body of literature on the effects of endorsers/spokespersons on advertising effectiveness from various aspects—for example, gender (e.g., Gilly 1988), ethnicity (e.g., Wheatley 1971), age (e.g., Freiden 1984), profession (e.g., Ohanian 1991), fame (e.g., Heath et al. 1994), and overall attractiveness (e.g., Bower and Landreth 2001)—few studies deal directly with faces. Therefore, the effect of

¹ AdForum.com chooses what it deems the five best ads every week for inclusion in its database. Ads may be print, video, or online.

different faces on advertising effectiveness remains largely unknown. In this paper, we focus on faces of noncelebrity endorsers about whom viewers know little but their faces.

Borrowing face recognition techniques from the field of computer science, we empirically test how different faces affect advertising effectiveness in the print advertising context. Specifically, we aim to address three questions that are important to both researchers and practitioners:

1. Do faces affect how a viewer reacts to an advertisement (abbreviated as an “ad”) on the metrics that advertisers care about? While keeping certain aspects of the spokesperson constant (gender, ethnicity, age, etc.), will different faces elicit different responses from viewers? Here, we focus on three key metrics of advertising effectiveness that advertisers care most about, including attitude toward the ad (Aad), attitude toward the brand (Abrand), and purchase intention (PI) (Miniard et al. 1990, Goldsmith et al. 2000).

2. If faces do have an effect, is it large enough to warrant a careful selection of faces when constructing a print advertisement? If the effect of face exists but is trivial, clearly advertisers should not spend much time or money on face selection. Only if the effect exists and is substantial should they be careful when selecting the face to appear in an advertisement.

3. If faces do have an effect and the effect is large, what specific facial features elicit such differential reactions on the metrics, and are such reactions different across individuals and different product categories? This question deals with the heterogeneity of the effect of faces on viewers’ responses on key ad metrics. A critical step toward answering this question involves the decomposition of faces. As such, we relied on the eigenface method, which is a widely used approach for face recognition in the computer science field. The eigenface method is “the first really successful demonstration of machine recognition of faces” (Zhao et al. 2003, p. 412). Many face recognition algorithms have been derived from this approach. The eigenface method will be explained in further detail in the next section.

The rest of the paper is organized as follows: We first review the literature on face research. We then describe the empirical study and discuss the data analysis and results. Finally, we present a discussion and conclusions.

Relevant Literature

We first provide a review of previous research on the effects of faces and facial features. Then, we introduce the eigenface method and how it works to extract facial features from different faces. Finally, we review literature indicating that substantial heterogeneity exists in the way people make inferences from faces.

Existing Literature on Face Research

The effect of faces on people’s perceptions and behavior has been studied in multiple disciplines (see Zebrowitz 2006 for a review). In psychology, Todorov et al. (2005) showed that inferences of competence based solely on facial appearance could predict the outcomes of U.S. congressional elections to some extent. Dumas and Testé (2006) found that perceived facial maturity has an influence on juridic judgments. Collins and Zebrowitz (1995) found that appearance is significantly related to the types of jobs people hold and partly determines job status. Pincott (2010) found that women use facial masculinity to infer the attractiveness of males and form mating preferences. In marketing, Gorn et al. (2008) found that during public relations crises, chief executive officers with baby faces are perceived to be more trustworthy and creditable than those with mature faces. Solomon et al. (1992) showed that congruency between the facial attractiveness of a spokesperson and product image can elicit favorable communication responses. In computer science, Koda and Maes (1996) showed that faces of virtual agents have an impact on human–computer interaction. And Brahnham (2002) focused on techniques to customize virtual agents’ faces to match the personalities of users.

Despite the large body of literature, to our knowledge, the effect of face on the effectiveness of print ads has not been empirically demonstrated. The most relevant literature in this domain is Bower and Landreth’s (2001) study on how the overall physical attractiveness of models (but not specifically their faces) affects product evaluations. More important, the magnitude of a face effect, heterogeneity in reactions to a face, and the practical relevance to practitioners have never been addressed.

An important stream of face research is to identify what specific facial features drive perceptions or behavior. In psychology, researchers have used a physiognomic approach to extract facial features from different faces. Physiognomic method represents a face with a set of facial distances. For example, Berry and McArthur (1985) decomposed a male face into 11 facial distances and found that a face with bigger eyes and a wider chin is judged to be more baby-faced and thus more trustworthy. Cunningham et al. (1990) decomposed a male face into 26 facial distances and found that a male face with bigger eyes and a longer chin is perceived as more attractive in a female’s eyes. Other researchers focused on one feature, facial width-to-height ratio, and found a significant effect on perceived propensity for aggression (Carré et al. 2009) and cooperative behavior in a trust game (Stirrat and Perrett 2010).

As the above studies illustrate, the choice of physiognomic measurement is somewhat arbitrary and

only captures a subset of information from face. For example, Berry and McArthur (1985) chose 11 physiognomic features, and Cunningham (1986) chose 26, but it is unknown exactly how many features should be studied. Furthermore, because of model identification issues and multicollinearity concerns, both studies included some interaction terms in their models, such as eye area, eye roundness, and chin area, but they did not include all possible interactions. Many interactions that might have an effect are missing from the models—e.g., the ratio of eye distance to face length, the angle of the eyes to the nostril tip, etc. This method especially is not useful if the faces studied are similar on these typical physiognomic features, such as in the case of print advertisement where the actors used have similar measurements on most of these typical physiognomic features. People who do not fit the stereotypical look of a model are unlikely to be selected. Thus very limited variations exist among these typical physiognomic features in this context. Based on these limitations, to decompose faces we used a holistic method, the eigenface method, which can capture all subtle differences across faces.

The Eigenface Method

Eigenface is the gold standard method in computer science for face recognition and was first proposed by Turk and Pentland (1991). It uses principal component analysis (PCA) to project training face images onto a set of dominant eigenvectors. Because each dominant eigenvector looks roughly like a face, these eigenvectors are also called eigenfaces. This method is easy to implement, is computationally efficient, achieves good accuracy in face recognition tasks, and thus is widely used in the computer science field (Zhao et al. 2003). Its implementation is briefly described as follows (for more details, see Turk and Pentland 1991):

Step 1. Acquire a training set of face images. Ideally, these images are all the same size, in a fixed position, and facing forward under constant lighting. The faces show neutral emotion and do not wear accessories (e.g., glasses). Assume that there are M gray-scaled face images in the training set and the size of each image is $H \times W$ pixels.

Step 2. Stack each face image into a $HW \times 1$ vector, denoted as X_i . The training set of images is represented by a $HW \times M$ matrix, denoted as X .

Step 3. Calculate the mean $\bar{X} = (1/M) \sum X_i$ and subtract the mean from X . The mean vector \bar{X} could be unstacked back to an $H \times W$ matrix and displayed, which looks roughly like a face with blurry contour. This is called the *average face*.

Step 4. Use PCA to calculate the eigenvectors and eigenvalues of the covariance matrix. At this step, we obtain M eigenvectors and the corresponding M eigenvalues.

Step 5. Keep the K most important/dominant eigenvectors and project each image in the training set onto these K eigenvectors, stacked together and denoted as $EV(HW \times K)$. Thus, each image in the training set (X_i) can be approximated and represented by a unique loading vector, $L_i(K \times 1)$. Each image can be reconstructed as $X_i \approx EV \times L_i + \bar{X}$. The choice of K depends on how many eigenfaces are needed to get a good reconstruction of training images (Turk and Pentland 1991).

Once the average face and K dominant eigenvectors are stored, given a face image, either an image from the training set or a new image, it is easy to calculate the loading vector of this face image. The dissimilarity between two face images can be well represented by the difference between two corresponding loading vectors. An illustration of the eigenface decomposition is shown in Figure 1.

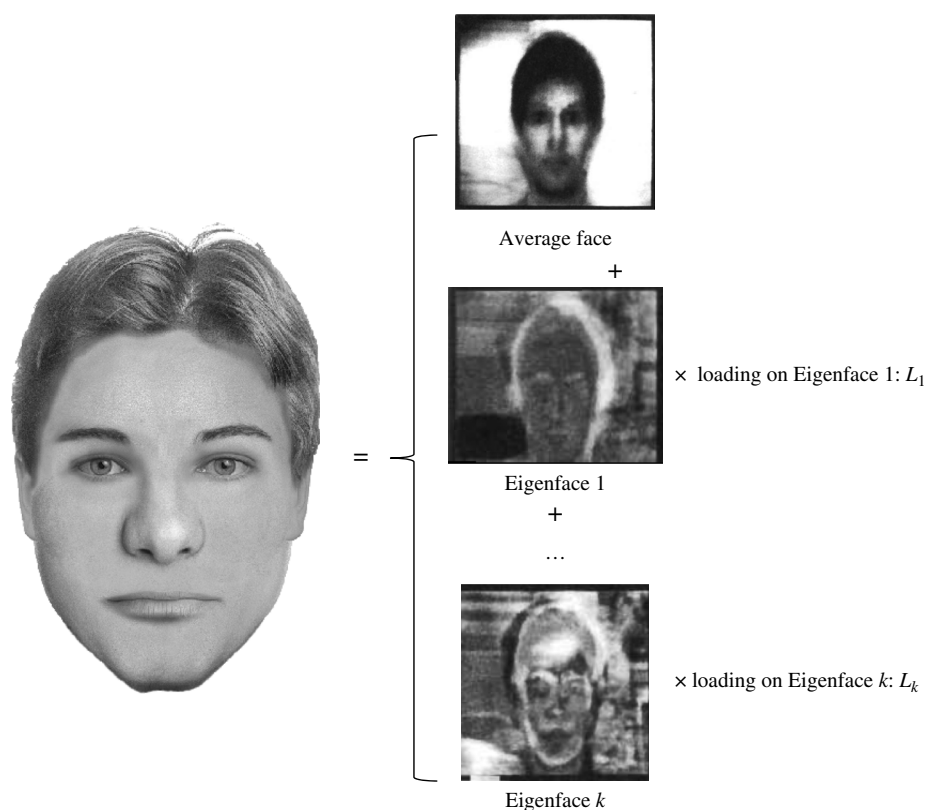
Despite its wide use in face recognition, the eigenface method has rarely been used to study the effect of facial features on viewers' responses. Only Branham (2002) made the first attempt to apply the eigenface method to her virtual agent study. However, the faces used in her study were synthetic faces generated by software, not real human faces. To date, no study has applied the eigenface method to real faces to study people's perceptions or reactions to print advertisements.

Unlike the physiognomic method, which requires a predefined set of facial distances, no prior information is needed for the eigenface method. The extracted eigenvectors contain the major differences among training images; thus it is a holistic measurement.

Heterogeneity in Viewers' Responses to Faces

All of the previous studies that link facial features to people's reactions were conducted at the aggregate level, not the individual level, and attempted to use one model to explain how people make inferences from faces. People were treated as homogeneous in these studies; individual differences were ignored. One underlying assumption is that people share in common what facial features influence their impressions and how important these facial features are in forming their perceptions. For example, based on Berry and McArthur's (1985) findings, a trustworthy face must have big eyes and a wide chin. The major theory underlying this assumption is that the perception of facial features has adaptive value and that trait impressions are based on those facial qualities that demand the greatest attention for the survival of the species—namely, physical fitness, age, and emotional state (Zebrowitz 1998). Faces with features that are indicative of these qualities are believed to have pronounced overgeneralization effects. Thus,

Figure 1 Eigenface Decomposition



Note. With the average face and k eigenfaces stored in the database, any face can be compressed and reconstructed by a $k \times 1$ loading vector (L_1, \dots, L_k) (Turk and Pentland 1991).

there is considerable consistency and universal consensus in the way people make inferences from faces (Cunningham et al. 1995).

However, this assumption has been greatly challenged. Ample evidence shows that the heterogeneity among people is not trivial and that it influences their way of making inferences from faces. A comparison between the results from the studies by Cunningham (1986) and Cunningham et al. (1990) reveals a gender difference in inferring attractiveness. Solomon et al. (1992) proposed six distinct types of attractive looks. Pincott (2010) reported that women from areas with good healthcare conditions (e.g., Sweden) prefer feminine-looking men, whereas women from areas without good healthcare systems (e.g., Russia) prefer masculine-looking men. Ekman and Friesen (2003) reported that culture and individual heterogeneity has an effect on both the exhibited face side and the recipient side. Based on these arguments, people should be studied in separate segments defined by factors such as gender, ethnicity, etc. Hence, multiple segments of people should be identified based on the way they make inferences from faces. Within each segment, people share similar preferences toward faces, whereas across segments, people show substantially different preferences toward faces.

Resolving this controversy for marketing practitioners is important, because quite different marketing strategies might need to be considered to optimize the effects of advertisements on consumer perceptions and behavior. If our results reveal that people are somehow similar and that the one best face exists, then an advertising agency's job is to find a spokesperson with a face that closely resembles the best face for the advertisement. However, if we find that people are heterogeneous in their facial preferences and multiple best faces exist with each appealing to a different segment of consumers, which may or may not depend on product categories, then an advertising agency would want to identify and define specific customer segments and create multiple advertisements, each with a distinct face that closely matches the best face for that specific segment of customers.

Experimental Design

Stimulus Advertisements

To mimic advertising practice, we used real models' faces as stimuli in this study. As a result, whatever effect we identify will not be artificial as a result of the poor choice of faces (e.g., using faces that will

never be selected by practitioners). We followed a rigorous face selection process in this study. First, we collected thousands of faces of male models from fashion websites such as Macy's, Brooks Brothers, and MyHabit.com; modeling agency websites such as Elite and Ford; and other graphic websites such as Google Images, Flickr, and Getty Images. Then, we narrowed the image pool down to a face database with 60 distinctive faces that satisfied the following criteria: (a) Caucasian male, between 25 and 35 years old; (b) no accessories, no obvious moustache, moderate hairstyle; (c) mild smile, positioned right-side up, facing forward; (d) not a celebrity (and no resemblance to a celebrity); and (e) reasonable resolution. Finally, we randomly selected 12 faces from the face database and used them as stimulus faces.

The product categories were selected based on the following criteria: (a) the category often features human faces in print ads, (b) the general public has some knowledge of the product category, and (c) it is appropriate to feature a man's face in an advertisement in the category. Of the categories that fit these criteria, we selected 12 that also have substantial variation in hedonism so that we could test whether people's reaction to faces are related to the hedonism characteristic of a product category. Hedonism is a variable often used in the literature to represent different product categories (e.g., Raghunathan and Irwin 2001, Carroll and Ahuvia 2006, Inman et al. 2009). It is indirectly indicated in the literature that hedonism might affect the relationship between facial features and people's responses. For example, Todorov et al. (2005) found that competent faces are more likely to be elected for political positions (less hedonic context); Englis et al. (1994) found that attractive faces are frequently featured in ads for fashion products such as clothing and perfume (more hedonic context). However, to our knowledge, the interaction between hedonism and facial features has never been studied in the literature. The 12 categories selected for this study are beer, restaurant, job search agent, men's cologne, coffee, computer, hotel, jeans, sports utility vehicle (SUV), sports shoes, camcorder, and car dealer.

Based on the face selection results (12 stimulus faces) and category selection results (12 categories), we created $12 \times 12 = 144$ stimulus ads. To create a more natural advertisement viewing experience for participants, the stimulus ads were created from real ads where the original face was replaced with one of the stimulus faces while holding all other elements constant with the following exception: the original brand names were replaced with some generic names and logos were removed. We hired professional graphic designers to create the stimulus ads to make sure that they looked reasonably natural at a

quick glance. At the end of the empirical study, we asked the participants whether they observed anything out of place or abnormal in the ads. Only a few participants indicated that they sensed that some of the faces in the ads might have been photoshopped.

For four categories (restaurant, job search agent, computer, and sports shoes), we included the original faces in the stimulus ads as a baseline. The other eight original faces were not used because they wore accessories such as glasses, had beards, or were celebrities. Altogether, we created 148 stimulus ads.

Extraction of Facial Features

Using the eigenface method, we first extracted the facial image from the original pictures and removed the noise related to lighting and angle. Second, we normalized each face and placed it into a 180×180 pixel-sized plane. This allowed us to construct the training set of 60 images. Third, we used the PCA method to obtain the eigenvectors and eigenvalues. We retained 12 eigenvectors that explain almost 75% of the variance among training images and can reconstruct the faces reasonably well. The root-mean-square pixel-by-pixel error in representing cropped versions of training face images, i.e., reconstruction errors, was less than 10%. The average face and 12 eigenfaces extracted from the 60 training images are shown in Figure 2.

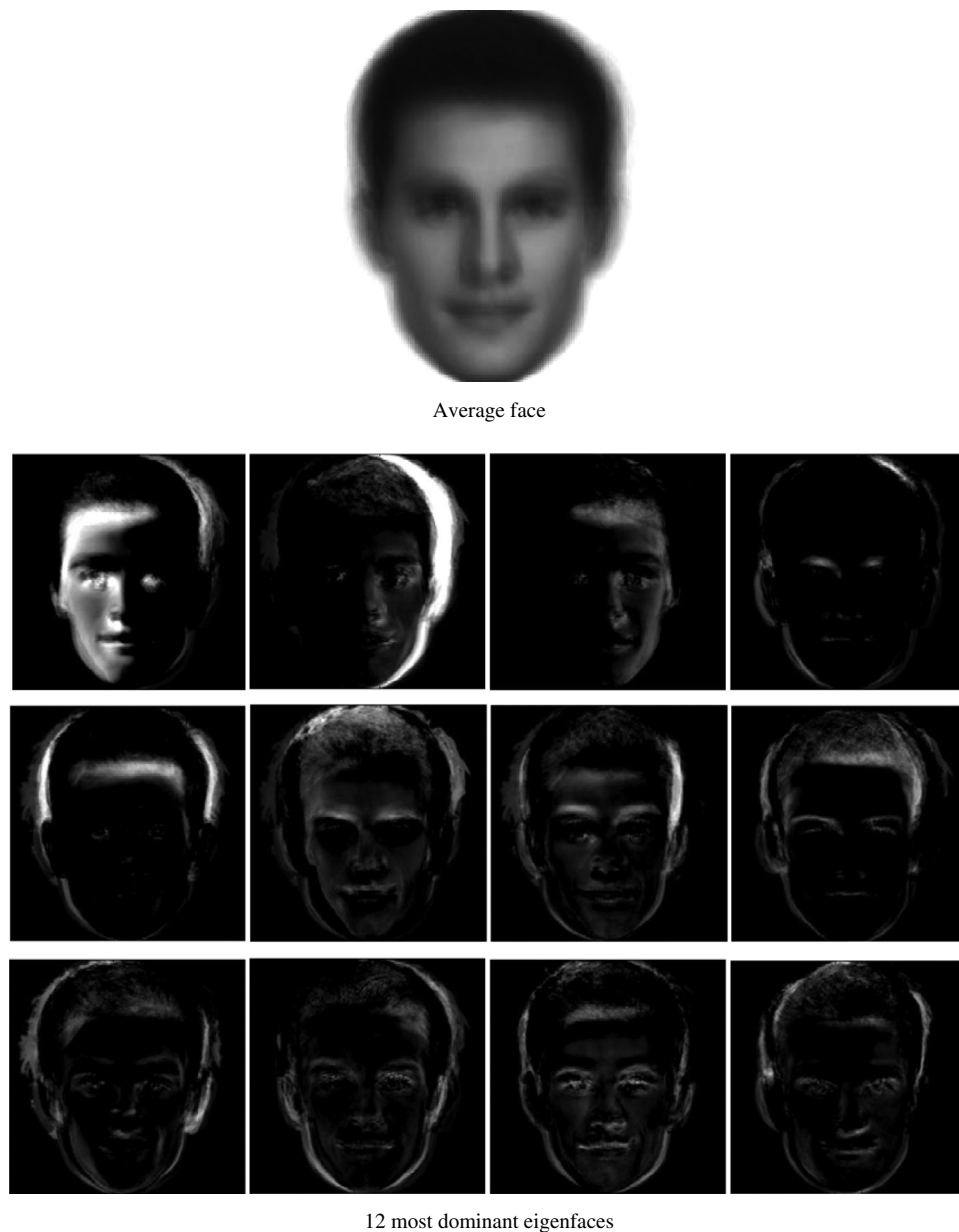
Finally, the loading vector (12×1) of each stimulus face was calculated. Eigenface loadings ranged from $-4,130.69$ to $5,467.63$ and thus were resized by dividing by 1,000. The resized eigenface loadings were the facial features we extracted from the stimulus faces. These loadings were used as independent variables in our subsequent data analysis.

Experimental Procedures

A total of 989 participants participated in this study. Of those, 676 undergraduate students at a major U.S. university completed the study in an on-campus computer laboratory for course credit. The remaining 313 participants were recruited from the Amazon Mechanical Turk subject pool; each of them received \$2 for their participation. In terms of gender, 472 were male (48%) and 517 (52%) were female. In terms of ethnicity, 755 participants were Caucasian (77%), 25 were African American (3%), 42 were Hispanic (4%), 125 were Asian (12%), 22 were multiracial (2%), and 20 were categorized as other (2%). Ages ranged from 18 to 69 years old, with a mean of 24.36 and a standard deviation of 10.27 years.

We applied a between-subject design. Each participant was presented with 12 print advertisements, one for each product category. Within each product category, one stimulus advertisement out of 12 (or 13 if the original face was used in the category) was randomly assigned to each participant. The sequence of

Figure 2 The Average Face and 12 Most Dominant Eigenfaces



Notes. As shown above, the average face and the 12 most dominant eigenfaces were extracted from 60 training faces that were used in the present study, including the 12 stimulus faces used in the empirical study. The contribution of each eigenface to explain variances among 60 faces decreases from left to right and from the upper row to the lower row.

12 product categories was also randomized for each participant. The average time spent on each advertisement evaluation was 11.4 seconds. After participants viewed all 12 print advertisements, they began evaluating them one by one on attitude toward the ad, attitude toward the brand, and purchase intention. After evaluating all 12 ads, participants were asked to evaluate the hedonism levels of the 12 product categories, followed by a survey about demographics, cognitive style, and value system. At the end of the empirical study, they were asked to evaluate the faces they saw on six trait dimensions that have been identified in the

literature as facial characteristics that affect people's responses: baby-facedness (Berry and McArthur 1985, Dumas and Testé 2006, Gorn et al. 2008), masculinity (Pincott 2010), attractiveness (Solomon et al. 1992), trustworthiness (Berry and McArthur 1985, Gorn et al. 2008, Stirrat and Perrett 2010), aggressiveness (Carré et al. 2009), and competence (Todorov et al. 2005).

We used the scale from Miniard et al. (1990) to measure three key ad metrics. Three items (bad/good, uninteresting/interesting, dislike/like) were used to measure attitudes toward the advertisement on a seven-point scale. Attitudes toward the brand

were measured using a three-item seven-point scale (unfavorable/favorable, negative/positive, dislike/like). Purchase intention was measured using a two-item seven-point scale (unlikely/likely, improbable/probable). Reliabilities for the three constructs are 0.88, 0.89, and 0.91, respectively.

In our study, hedonism was measured using a four-item seven-point scale (unpleasant/pleasant, awful/nice, sad/happy, disagreeable/agreeable) (Crowley et al. 1992). Reliability for hedonism is 0.89. A 36-item seven-point scale was used to measure cognitive style (Hauser et al. 2009), and a 30-item nine-point scale was used to measure value style (Schwartz 1992). Each of the six trait dimensions was measured on a seven-point bipolar scale (i.e., baby-faced/mature, feminine/masculine, unattractive/attractive, untrustworthy/trustworthy, unaggressive/aggressive, and incompetent/competent).

Estimation and Results

Effect of Faces on Key Ad Metrics at the Aggregate Level

As indicated in the experimental design section, the only difference in the stimulus ads for each product category was the face appearing on the ad; therefore, any differences in key ad metrics can be attributed to the face used in the ad. For each category, we calculated the means for each face for each key ad metric (see Table 1).

The results show that faces do have an effect on advertising effectiveness. For example, by comparing means, Face 6 in the beer ad significantly outperformed Face 12 on all three key ad metrics. For each category and each key ad metric, the face with highest mean significantly beats the face with lowest mean at the $p < 0.05$ level.

A comparison between the results for the 12 stimulus faces and the original faces justifies careful face selection. For example, for sports shoes, by changing from the original face to Face 7, the print ad achieved about a 0.5-point increase on evaluations of all three key ad metrics. As our face selection procedures show, these 12 stimulus faces were randomly selected from among 60 male models' faces. It is very possible that by applying our method to a large sample of faces, practitioners would achieve even greater improvement.

It is worth noting that the preferences for faces were different across categories, which indicates a strong interaction between face effect and category. For example, for attitude toward ad, Face 11 achieved the highest mean for the jeans ad but the lowest for the car dealer ad.

To help answer the second question on whether practitioners should care about such a difference in

effect, we calculated the percentages of high-rating participants (evaluation of key ad metric > 6) for each category and each key ad metric. For example, in business practice, practitioners care most about the customers who give high ratings to their products or services (Urban and Hauser 1993). We report both the percentage for each face for each ad (in Table 2) as well as absolute maximum improvement between the best and worst face for each product category (in Figure 3).

The effect is quite striking. For example, Table 2 shows that for the cologne ad, by merely changing the face from Face 2 to Face 1, advertisers could achieve a more than 10% increase in attitude toward ad (from 8% to 25.8%), attitude toward brand (from 5.3% to 21.6%), and purchase intention (from 9.3% to 20.6%). In other words, such a change would double or even triple the number of high rating participants. We plotted the maximum absolute percentage improvement (the percentage of high-rating participants for the best face of high rating minus the percentage of high-rating participants for the worst face of high rating) in Figure 3.² Figure 3 shows that by merely changing from the worst face of high rating to the best face of high rating, each ad could achieve an increase in the number of high-rating people ranging from 4% to 21% for each key ad metric. The interaction of face effect with product categories can also be seen in a comparison of high-rating people. For example, Figure 3 shows that for the cologne ad, the face effect on all three key ad metrics is huge, whereas for the computer ad, the face effect is much smaller. For the restaurant ad, the face effect for Aad and Abrand is substantial, but it is minor for PI.

These results indicate that different faces do have an effect on advertising effectiveness, and that the effect is nontrivial. Hence, we conclude that advertisers should carefully select faces/spokespersons to appear in print advertisements.

Segmentation Analysis Using Facial Features

We used 12 eigenface loadings and 12 interaction terms between eigenface loadings and hedonism as independent variables and the three key ad metrics as dependent variables simultaneously in the segmentation analysis. To explore possible heterogeneity in how people respond to facial features, we used the latent class multivariate regression model proposed by Ramaswamy et al. (1993). Specifically,

$$Y_i \sim G_i(Y_i; \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\Delta}) = \sum_k w_k f(Y_i | X_i, \boldsymbol{\beta}_k, \boldsymbol{\Delta}_k), \quad (1)$$

² We did not plot the maximum relative percentage improvement (the percentage of high-rating participants for the best face of high rating divided by the percentage of high-rating participants for the worst face of high rating) because, for several categories, the percentage of high rating participants for worst face of high rating was zero.

Table 1 Mean of Key Ad Metrics by Product Category and Face

Category	Face 1	Face 2	Face 3	Face 4	Face 5	Face 6	Face 7	Face 8	Face 9	Face 10	Face 11	Face 12	Original
Beer													
Aad	3.26	3.56	2.97	2.81	3.04	3.66	3.05	3.03	3.21	3.12	2.77	2.63	N/A
Abrand	3.37	3.60	2.99	3.07	3.21	3.73	3.35	3.15	3.43	3.37	3.28	2.90	
PI	3.30	3.65	2.85	2.86	2.92	3.44	3.23	2.92	3.07	3.12	2.96	2.70	
Restaurant													
Aad	4.75	4.41	4.42	4.54	4.63	4.46	4.41	4.83	4.15	4.39	4.32	4.59	4.75
Abrand	4.88	4.78	4.69	4.82	4.90	4.58	4.78	5.05	4.42	4.69	4.57	4.95	5.00
PI	4.81	4.97	4.77	4.87	4.96	4.68	4.80	5.06	4.57	4.76	4.53	4.89	4.95
Job search agent													
Aad	3.25	3.52	3.22	3.24	3.25	3.71	3.06	3.56	3.04	3.35	3.65	3.21	3.37
Abrand	3.64	3.81	3.48	3.73	3.56	3.90	3.29	3.75	3.28	3.64	3.91	3.62	3.66
PI	3.24	3.34	2.83	3.03	2.95	3.41	2.69	3.15	2.76	3.03	3.44	3.09	3.09
Cologne													
Aad	4.93	3.91	4.69	4.36	4.46	4.65	4.56	4.14	4.12	4.65	4.55	4.02	N/A
Abrand	4.80	3.90	4.57	4.39	4.38	4.63	4.34	4.21	4.10	4.62	4.49	3.99	
PI	4.21	3.25	3.95	3.84	3.81	4.17	3.77	3.64	3.45	4.01	3.99	3.23	
Coffee													
Aad	3.74	3.10	3.62	2.83	3.30	3.80	3.70	3.01	2.93	3.42	3.12	2.88	N/A
Abrand	3.81	3.30	3.71	3.07	3.41	3.84	3.78	3.31	3.34	3.45	3.28	3.29	
PI	3.16	2.85	3.28	2.67	3.19	3.50	3.39	2.84	3.15	3.18	2.99	2.84	
Computer													
Aad	3.08	3.34	3.17	3.04	2.54	3.19	3.39	2.79	3.01	3.10	3.32	3.12	3.30
Abrand	3.33	3.54	3.18	3.36	2.82	3.22	3.53	3.07	3.09	3.23	3.41	3.38	3.20
PI	2.68	2.88	2.73	2.73	2.19	2.62	2.85	2.31	2.39	2.85	2.88	2.58	2.44
Hotel													
Aad	3.37	3.51	3.58	3.18	3.02	3.61	3.97	3.53	3.08	3.24	3.42	3.20	N/A
Abrand	3.52	3.94	3.89	3.50	3.51	3.88	4.23	3.81	3.44	3.75	3.65	3.65	
PI	3.29	3.72	3.64	3.19	3.24	3.45	4.00	3.48	3.14	3.42	3.13	3.37	
Jeans													
Aad	3.70	3.29	3.66	3.06	3.03	3.81	3.56	3.59	3.54	3.18	3.94	3.30	N/A
Abrand	3.68	3.19	3.67	2.97	2.99	3.70	3.51	3.71	3.39	3.25	3.80	3.19	
PI	3.26	2.71	3.19	2.61	2.66	3.42	3.00	3.13	2.98	2.76	3.35	2.83	
SUV													
Aad	4.32	3.92	4.36	4.19	3.95	3.69	4.17	3.93	3.63	4.13	4.26	3.85	N/A
Abrand	4.20	3.92	4.17	4.05	3.81	3.78	4.07	3.90	3.72	3.99	4.10	3.95	
PI	3.09	2.81	2.92	3.23	2.93	2.83	3.13	2.91	2.73	2.89	2.99	3.22	
Sports shoes													
Aad	4.75	4.26	4.59	4.52	4.69	4.58	4.99	4.29	4.42	4.35	4.58	4.38	4.46
Abrand	4.76	4.38	4.51	4.57	4.69	4.53	4.84	4.30	4.28	4.33	4.57	4.30	4.37
PI	4.14	3.98	3.99	3.92	4.19	4.13	4.33	3.78	3.66	3.64	3.92	3.73	3.84
Camcorder													
Aad	3.61	3.39	3.83	3.21	3.45	3.90	3.48	3.56	3.46	3.34	3.52	3.40	N/A
Abrand	3.75	3.60	4.05	3.50	3.63	4.07	3.59	3.78	3.62	3.52	3.54	3.56	
PI	2.99	2.90	3.02	2.80	3.02	3.07	2.90	2.89	2.97	2.96	2.68	2.90	
Car dealer													
Aad	2.49	2.36	2.25	2.23	2.04	2.36	2.08	2.35	1.89	2.24	1.73	2.44	N/A
Abrand	3.00	2.85	2.67	2.69	2.52	2.58	2.53	2.83	2.40	2.76	2.20	2.76	
PI	2.42	2.42	2.30	2.18	2.09	2.20	2.05	2.44	2.09	2.48	1.74	2.51	

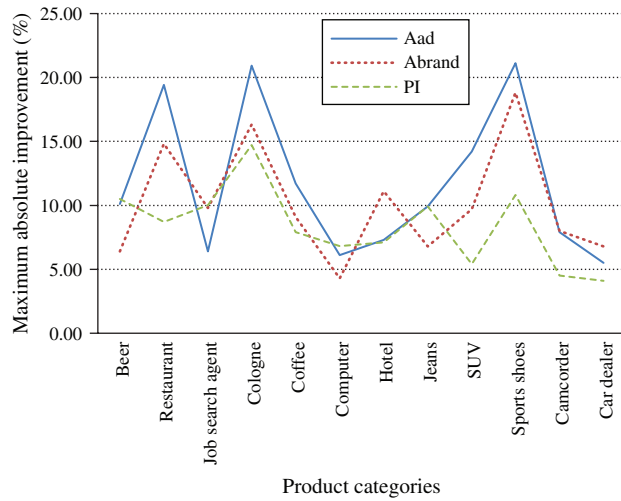
Notes. For each category and each key ad metric, the face that achieves the highest mean is deemed the best face of mean and marked in bold, whereas the face with lowest mean is deemed the worst face of mean and marked in bold and italic. For each category and each key ad metric, evaluations from participants who viewed the ad with best face of mean are significantly higher than evaluations from those who viewed the same ad with worst face of mean at the $p < 0.05$ level. In four categories, four original faces were also tested in the experiment and thus compared here. For each category, the original face refers to the face that was featured in the real ad. So even though they are all called original face, they are unique for each category and different from each other.

Table 2 Percentage of High-Rating Participants (Key Ad Metric Evaluated Above 6) for Each Key Ad Metric by Product Category and Face

Category	%												Original
	Face 1	Face 2	Face 3	Face 4	Face 5	Face 6	Face 7	Face 8	Face 9	Face 10	Face 11	Face 12	
Beer													
Aad	11.9	9.3	6.2	9.2	4.7	12.7	6.3	2.5	8.3	7.8	4.4	6.3	N/A
Abrand	8.3	4.0	2.1	5.7	4.7	8.5	7.5	3.8	8.3	7.8	2.2	2.5	
PI	10.7	14.7	4.1	10.3	8.2	12.7	10.0	8.9	8.3	7.8	5.6	7.5	
Restaurant													
Aad	28.8	18.8	16.0	22.4	17.4	22.2	18.4	30.2	10.7	19.5	16.9	25.7	18.1
Abrand	28.8	27.5	21.3	30.3	21.7	23.5	21.8	31.7	19.0	21.8	16.9	22.9	30.6
PI	35.0	31.3	29.3	31.6	33.3	32.1	32.2	34.9	27.4	31.0	27.7	30.0	36.1
Job search agent													
Aad	3.8	3.8	2.5	6.8	3.5	6.6	2.6	7.4	2.8	9.0	8.6	7.6	7.0
Abrand	8.8	11.3	2.5	9.5	7.1	3.9	2.6	7.4	1.4	7.5	11.1	8.9	7.0
PI	10.0	6.3	2.5	9.5	4.7	9.2	0.0	4.4	2.8	4.5	7.4	7.6	7.0
Cologne													
Aad	25.8	8.0	28.9	20.5	20.5	17.5	22.5	16.0	12.9	21.2	27.7	9.5	N/A
Abrand	21.6	5.3	17.1	13.3	11.0	15.0	21.1	16.0	8.9	16.5	20.5	11.9	
PI	20.6	9.3	19.7	13.3	12.3	12.5	12.7	16.0	7.9	18.8	18.1	6.0	
Coffee													
Aad	7.1	3.3	12.8	3.8	5.1	10.3	5.7	9.4	4.9	11.4	8.3	1.1	N/A
Abrand	5.7	4.4	4.7	1.3	5.1	10.3	4.3	8.2	4.9	5.7	3.6	2.3	
PI	8.6	5.5	5.8	3.8	6.4	8.0	11.4	7.1	6.1	10.2	3.6	5.7	
Computer													
Aad	3.6	6.5	5.4	2.4	3.4	8.5	5.3	4.2	2.8	3.8	4.5	5.1	7.5
Abrand	2.4	3.9	5.4	4.7	1.1	2.8	3.9	4.2	1.4	2.5	4.5	3.8	4.5
PI	3.6	1.3	6.8	3.5	1.1	0.0	2.6	0.0	1.4	2.5	3.0	2.5	4.5
Hotel													
Aad	7.1	9.9	6.7	3.5	4.9	8.4	8.3	4.9	5.4	2.6	5.7	4.3	N/A
Abrand	7.1	9.9	5.6	2.4	4.9	8.4	8.3	13.4	7.6	6.5	5.7	7.4	
PI	5.7	7.4	6.7	7.1	6.1	2.4	9.5	7.3	5.4	5.2	4.3	7.4	
Jeans													
Aad	11.0	10.0	8.9	6.0	2.2	12.0	7.8	9.0	3.8	5.6	9.7	8.3	N/A
Abrand	8.8	10.0	8.9	8.3	3.2	8.4	3.9	5.1	6.3	5.6	8.3	8.3	
PI	11.0	10.0	8.9	3.6	1.1	9.6	3.9	6.4	6.3	3.3	6.9	6.9	
SUV													
Aad	17.4	14.1	19.8	19.0	15.4	11.0	16.8	14.6	8.5	22.7	15.7	12.3	N/A
Abrand	13.0	8.5	10.5	8.3	6.6	11.0	11.9	3.4	7.0	12.0	11.2	9.9	
PI	4.3	1.4	3.5	6.0	3.3	3.7	5.9	2.2	4.2	6.7	2.2	1.2	
Sports shoes													
Aad	22.5	12.3	18.4	16.9	28.1	24.4	30.8	9.6	12.5	15.0	21.9	14.3	22.5
Abrand	23.8	11.1	21.1	18.2	25.0	14.1	24.4	8.4	6.3	12.5	17.8	16.7	19.7
PI	13.8	9.9	13.2	18.2	17.2	17.9	19.2	8.4	9.4	10.0	12.3	13.1	16.9
Camcorder													
Aad	3.4	2.7	9.1	6.3	4.7	10.6	2.9	5.1	5.6	4.2	7.5	3.8	N/A
Abrand	2.3	4.0	9.1	5.1	3.5	9.4	2.9	3.8	6.9	1.4	3.8	6.4	
PI	2.3	4.0	1.8	5.1	2.4	5.9	1.4	3.8	2.8	1.4	2.8	2.6	
Car dealer													
Aad	5.5	0.0	0.0	0.0	1.4	3.6	1.1	2.0	1.2	1.2	1.2	0.0	N/A
Abrand	6.8	1.2	1.1	1.5	1.4	2.4	0.0	3.9	1.2	1.2	1.2	0.0	
PI	4.1	0.0	1.1	1.5	1.4	1.2	0.0	3.9	1.2	1.2	1.2	1.4	

Notes. For each category and each key ad metric, the face that achieves the highest percentage of high-rating participants is deemed the best face of high rating and is marked in bold, whereas the face with the lowest percentage of high-rating people is deemed the worst face of high rating and is marked in bold and italic. For each category and each key ad metric, the percentage of high-rating participants for the best face of high rating is significantly higher than the percentage for the worst face at the $p < 0.05$ level. Note that for most categories and key ad metrics, the best and worst faces of high rating are the same with best and worst faces in terms of mean (see Table 1) but are different for some cases.

Figure 3 Maximum Absolute Percentage Improvement for High-Rating Participants



Notes. For each category and each key ad metric, the maximum absolute percentage improvement equals the percentage of high-rating participants for the best face of high rating minus the percentage of high-rating participants for the worst face of high rating. For each category and each key ad metric, the best face of high rating refers to the face that achieves the highest percentage of high-rating participants, whereas the worst face of high rating refers to the face that achieves the lowest percentage of high-rating participants.

where

$$f(\mathbf{Y}_i | \mathbf{X}_i, \boldsymbol{\beta}_k, \Delta_k) = (2\pi)^{-(T/2)} |\Delta_k|^{-(1/2)} \exp\left\{-\frac{1}{2}(\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_k)' \Delta_k^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_k)\right\} \quad (2)$$

and

$$\pi_k \geq 0, \quad \sum_{k=1}^K \pi_k = 1. \quad (3)$$

Here, $\mathbf{Y}_i = [\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{iT}]'$ is a $T \times 3$ matrix with $i = 1, \dots, I$ as an indicator of participant and T as the total number of replications, and 3 refers to the three key ad metrics, i.e., attitude toward the ad, attitude toward the brand, and purchase intention in our case; \mathbf{X}_i is a $T \times J$ matrix of independent variables for participant i , where J refers to the number of independent variables; π_k is the prior probability of segment k ; K is the predefined number of segments, an assumption to be imposed on the finite mixture model; $\boldsymbol{\beta}_k$ is the $J \times 1$ parameter vector for segment k ; and $\Delta = (\Delta_1, \dots, \Delta_K)'$, where Δ_k is the $T \times T$ variance-covariance matrix for segment k .

By introducing an indicator variable

$$z_{ik} = \begin{cases} 1 & \text{iff participant } i \text{ belongs to segment } k, \\ 0 & \text{otherwise,} \end{cases}$$

the likelihood can be presented as

$$L = \prod_i \prod_k [w_k f(\mathbf{Y}_i | \mathbf{X}_i, \boldsymbol{\beta}_k, \Delta_k)]^{z_{ik}}. \quad (4)$$

We used the Latent GOLD Syntax Module for implementation,³ which is mainly based on the expectation-maximization algorithm for model estimation (Vermunt and Magidson 2005). Considering that we dealt with only 12 faces, although the model including all 24 independent variables was identified, it was very unstable and the convergence was bad. Thus, we added a variable selection procedure to the segmentation model. Given K , we used a stepwise greedy search for variable selection (Bishop 1996). We used the Bayesian information criterion (BIC) for variable selection (Andrews and Currim 2003). In this case, a variable will enter into the model if its coefficient is significant in at least one segment, and a variable will exit out of the model if its coefficient is insignificant in all segments. The heuristic is briefly described below.

Step 1. Select a K .

Step 2. Run a K -segment finite mixture model with empty set $P_0 = \{\emptyset\}$, i.e., only intercepts, no x 's.

Step 3. Select the next best variable $x^+ = \arg \min_{x \notin P_k} [BIC(P_k + x)]$ to enter into the K -segment finite mixture model.

Step 4. Update $P_{k+1} = P_k + x^+$; $k = k + 1$.

Step 5. Remove the worst variable $x^- = \arg \min_{x \in P_k} [BIC(P_k - x)]$ from the K -segment finite mixture model if removing it will help decrease the BIC value of the model.

Step 6. Update $P_{k+1} = P_k - x^-$; $k = k + 1$.

Step 7. Repeat Steps 5 and 6 until removing any variable in the model will not decrease the BIC value.

Step 8. Repeat Steps 3–7 until adding any variable into the model or removing any variable from the model will not help decrease the BIC value, or all 24 variables are included in the model.

We tested for $K = 1, 2, 3, 4, 5$, and 6, and the five segment result is the best model based on the BIC criterion. The variables selected were $hedonism \times L1$, $hedonism \times L4$, $hedonism \times L5$, $hedonism \times L6$, $hedonism \times L7$, $hedonism \times L10$, $hedonism \times L11$, and $hedonism \times L12$.⁴ The five segments were composed of 40, 82, 386, 344, and 137 participants, respectively. The estimation for the best model is shown in Table 3. Substantial heterogeneity is shown across segments. For example, in terms of attitude toward the ad, participants in Segments 1 and 2 do not favor faces that load heavily on Eigenface 1, especially in ads for hedonic products; participants in Segments 4 and 5, by contrast, do favor faces that load heavily on Eigenface 1, especially in ads for hedonic products; and participants in Segment 3 do not care about face loadings on Eigenface 1

³ We thank Wayne DeSarbo for recommending the model and software.

⁴ To test the robustness of the result, we also ran the mixture model with the eight selected variables for $K = 1, \dots, 8$. The model with $K = 5$ achieves the smallest BIC value as well.

Table 3 Preference of Facial Structure of Each Segment, by Different Key Ad Metrics

Variable	Segment 1 (40)			Segment 2 (82)			Segment 3 (386)			Segment 4 (344)			Segment 5 (137)		
	Aad	Abrand	PI	Aad	Abrand	PI	Aad	Abrand	PI	Aad	Abrand	PI	Aad	Abrand	PI
<i>Hedonism * L1</i>	−0.26**	−0.01	−0.11*	−0.21**	−0.18**	−0.18**	−0.01	0.01	0.05**	0.19**	0.19**	0.24**	0.14**	0.12**	−0.08**
<i>Hedonism * L4</i>	0.42**	0.00	0.41**	0.31**	0.28**	0.21**	0.00	−0.03	−0.09**	−0.20**	−0.16**	−0.23**	−0.12**	−0.08	0.07
<i>Hedonism * L5</i>	0.48**	0.11	0.52**	0.31**	0.32**	0.19*	−0.08	−0.08	−0.12**	−0.41**	−0.34**	−0.44**	−0.29**	−0.14*	0.31**
<i>Hedonism * L6</i>	1.01**	−0.12	0.70**	1.46**	1.31**	1.42**	−0.03	−0.11	−0.24**	−1.41**	−1.27**	−1.61**	−0.63**	−0.56**	0.68**
<i>Hedonism * L7</i>	−0.40	−0.06	−0.45	−1.00**	−0.84**	−1.05**	0.14	0.17*	0.34**	1.37**	1.26**	1.51**	0.95**	0.84**	−0.32**
<i>Hedonism * L10</i>	−0.41**	0.00	−0.58**	−0.64**	−0.63**	−0.59**	0.13**	0.12**	0.18**	0.74**	0.71**	0.89**	0.61**	0.49**	−0.28**
<i>Hedonism * L11</i>	−0.76**	0.19	−0.95**	−0.98**	−0.90**	−0.87	0.11	0.18**	0.18**	0.92**	0.83**	1.04**	0.42**	0.27**	−0.41**
<i>Hedonism * L12</i>	1.07**	−0.12	0.54**	1.61**	1.44**	1.44**	−0.31**	−0.32**	−0.45**	−1.48**	−1.33**	−1.63**	−0.87**	−0.65**	0.59**

*Significant at $p < 0.10$; **significant at $p < 0.05$.

at all, no matter how hedonic the featured product would be. It is worth noting that all variables entering into the best model are interaction terms, which gives more evidence supporting the conjecture that face effect interacts with product categories. Because the sizes of Segments 3 and 4 are substantially bigger than those of Segments 1, 2, and 5, we focus our following discussion mainly on Segments 3 and 4.

To further explore the heterogeneity across segments, we ran a classification and regression tree algorithm (a supervised learning method; see Bishop 2006 for more details) of the segment labels on the demographic information as well as the cognitive style and value system evaluation. Three demographic vari-

ables (age, gender, and ethnicity), three cognitive items (prefer the whole versus parts and details, visually driven, and analytical), and two values (authority and choosing own goals) emerged as important variables to classify the participants into five identified segments. A pairwise comparison on these eight variables across five segments was run afterwards to explain segment heterogeneity. We also conducted a set of mediation tests (Baron and Kenny 1986) with the six trait variables we measured. Results are shown in Table 4.

It is shown that Segment 3 mainly comprises young, visually driven people who value choosing their own goals. The effect of eigenface features on

Table 4 Segment Characteristics

Trait	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
Size	40	82	386	344	137
Demographics ^a	More female	More female	Younger Less Caucasian	More male Mainly Caucasian	Older
Cognitive style ^{a,b}	Prefer parts and details More visually driven Less analytical	Less visually driven	More visually driven	Prefer the whole Less visually driven More analytical	
Value system ^{a,c}	Value authority less	Value choosing own goals more	Value choosing own goals more	Value authority more Value choosing own goals less	Value authority less
Baby-facedness				Aad, PI	Aad, Abrand
Masculinity		Aad, Abrand	Aad, Abrand, PI	Aad, Abrand	Aad, Abrand, PI
Attractiveness	Aad	Aad, Abrand, PI	Aad, Abrand, PI	Aad, Abrand, PI	Aad, Abrand, PI
Trustworthiness	Aad, PI	Aad, Abrand, PI	Aad, Abrand, PI	Aad, Abrand, PI	Aad, Abrand, PI
Aggressiveness			PI	Aad, Abrand, PI	PI
Competence	Aad	Aad, Abrand, PI	Aad, Abrand, PI	Aad, Abrand, PI	Aad, Abrand, PI

^aIf the feature is not included in the cell, it means that this feature for this segment is similar to the entire sample. For example, gender information is described for Segments 1, 2, and 4, but not for Segments 3 and 5. This means the gender distribution in Segments 3 and 5 is similar to the entire sample.

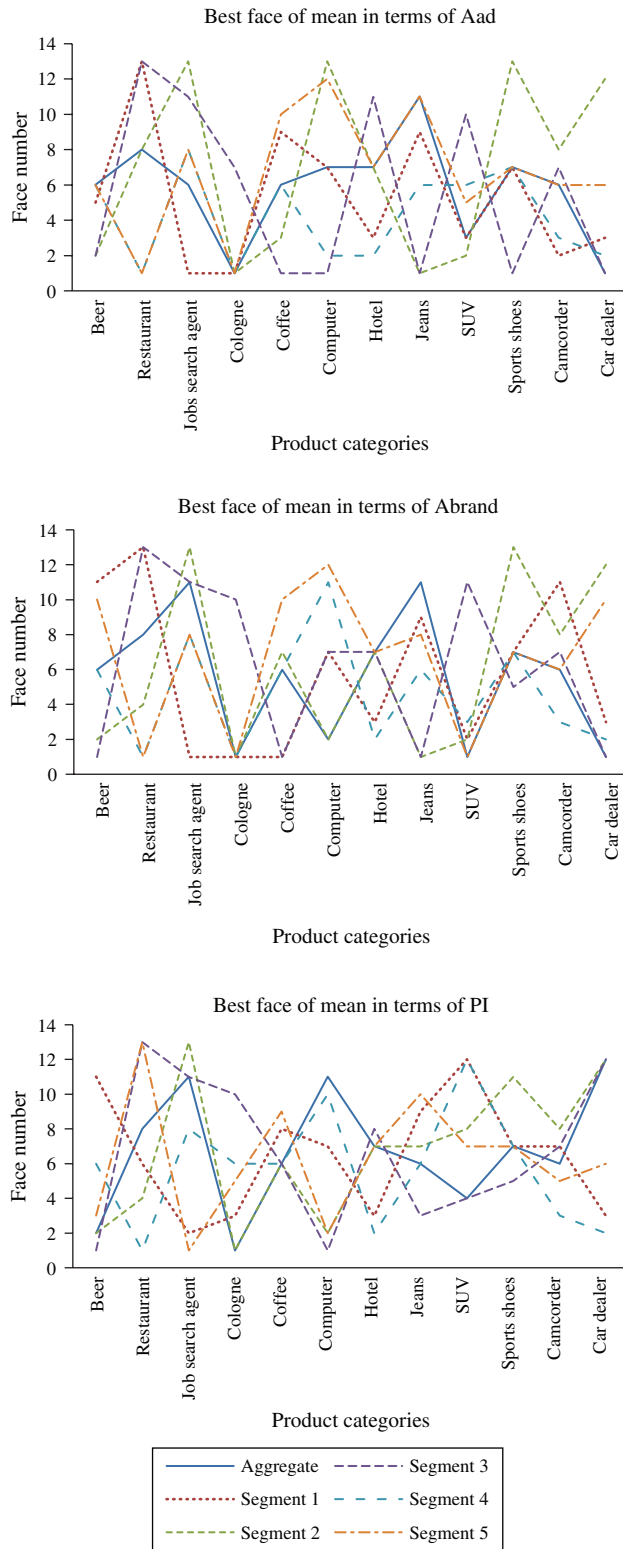
^bThe original scales for these three items are as follows: "Please indicate how much you agree or disagree with the following statements from 1 (strongly disagree) to 7 (strongly agree)."

- I'm usually more interested in the whole than in parts and details.
- I will read an explanation of a graph/chart before I try to understand the graph/chart on my own.
- I find that to adopt a careful, analytical approach to making decisions takes too long."

^cThe original scales for these two items are as follows: "Please rate the values below very carefully and then rate each value on an importance scale 'AS A GUIDING PRINCIPLE IN MY LIFE,' from 7 (of supreme importance) to 6 (very important), to 3 (important), to 0 (not important), to −1 (opposed to my values).

- Authority.
- Choosing own goals."

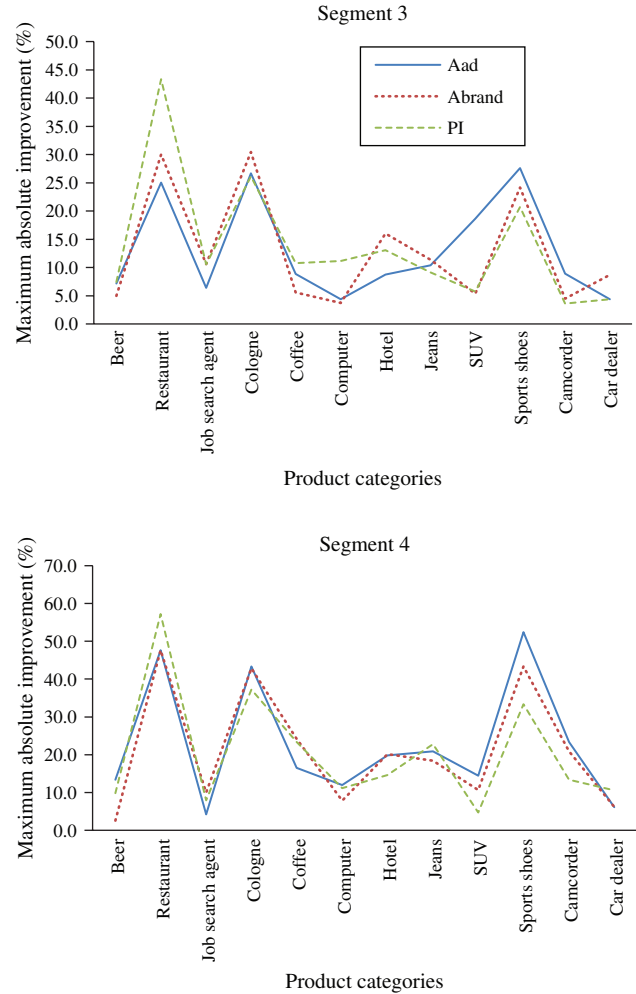
Figure 4 Best Face of Mean for Each Category and Segment



Note. The original face is denoted as Face 13 if applied.

all three key ad metrics appears to be mediated by masculinity, attractiveness, trustworthiness, and competence; the effect of eigenface features on purchase intention is partially mediated by aggressive-

Figure 5 Maximum Absolute Percentage Improvement for High-Rating Participants, by Segment



Note. Because Segments 1, 2, and 5 are of small size, here we only present the segment-level results for Segments 3 and 4.

ness, but aggressiveness does not play a role for attitude toward the ad or attitude toward the brand. Compared with Segment 3, Segment 4 is more male and less white, and it comprises people with holistic views who are not visually driven, are analytical, value authority, and do not care much about choosing their own goals. In Segment 4, the effect of eigenface features on ad metrics is mediated by attractiveness, trustworthiness, aggressiveness, and competence. Baby-facedness plays a role in attitude toward the ad and purchase intention, but not in attitude toward the brand; masculinity plays a role in attitude toward the ad and brand, but not in purchase intention. It is worth noting that even though attractiveness, trustworthiness, and competence are important mediating variables for both Segments 3 and 4, the mediation effect is different across the two segments. For example, for Segment 3, competence partially mediates the effects of L6 and L7

on attitude toward the ad, whereas for Segment 4, competence partially mediates the effect of L1, L4, L5, L6, L7, L10, and L11 on attitude toward the ad. Therefore, although competence is important for both Segments 3 and 4, the faces that are regarded as competent are different for these two segments. The competence order of the 12 stimulus faces for Segment 3 is totally different from the order for Segment 4. Face 9 is regarded as the most competent and Face 3 the least competent for Segment 3; Face 5 is regarded as the most competent and Face 1 the least competent for Segment 4.

Based on these results, we conclude that people make reasonably consistent inferences from facial features, and substantial heterogeneity exists in the way people make inferences.

Effect of Faces on Key Ad Metrics at the Segment Level

How do the segmentation results help advertisers understand people's heterogeneity in making inferences from facial features? To address this question, we compared the best faces of mean at the segment level (see Figure 4).

We found substantial heterogeneity from segment to segment. All three panels in Figure 4 show that different faces are preferred for different segments and diverge from the best face of mean at the aggregate level. For example, for the computer ad, participants at the aggregate level favored Face 7 the most and Face 5 the least in terms of evaluation of attitude toward the ad, but at the segment level, participants in Segment 3 favored Face 1 the most and Face 4 the least, whereas those in Segment 4 favored Face 2 the most and Face 1 the least.

Substantial heterogeneity can also be easily seen by comparing high-rating people at the segment level, as shown by the maximum absolute improvement in percentage of high-rating people at the segment level (see Figure 5). Different faces are preferred, and the magnitude of face effect also differs across segments. Interestingly, in contrast with the diverging trend for the three key ad metrics at the aggregate level (see Figure 3), the three key ad metrics move in a similar direction at the segment level. This provides evidence that segmentation with three dependent variables works.

Hence, ad agencies should not only be careful when selecting faces for their ads, but they also should take heterogeneity among target audiences into consideration. Ideally, multiple advertisements, each with a different face, should be created to appeal to different segments of the target audience.

Conclusions and Discussion

In this paper, we empirically demonstrated the effect of different faces on advertising effectiveness for various product categories. To answer the three questions that were raised at the beginning of this paper, we conclude that (1) faces do affect how a viewer reacts to an ad on the metrics that advertisers care about, (2) the effect is substantial, and (3) people show reasonably consistent preferences toward faces, and substantial heterogeneity among individuals and product categories exists in how viewers react to advertisements. Moreover, eigenface features are practical for segmenting people based on their preferences toward different faces.

The present study contributes theoretically to the existing literature in several ways. First, we focused specifically on faces in print advertisements and empirically demonstrated the effect of different faces on advertising effectiveness by using real ads and real faces. Second, we introduced a quantitative method, the eigenface method, into marketing research, which will hopefully encourage future face studies in marketing. Third, we provided evidence for the interaction of face effects with product categories, and we empirically demonstrated that hedonism is a useful category feature that differentiates product categories and interacts with face effects. Finally, we resolved the controversy over people's heterogeneity in face preferences while contributing to the face literature in general.

In practice, this study has several implications for advertisers and ad agencies. The substantial effect of different faces on advertising effectiveness indicates that ad agencies should be careful when selecting faces to appear in print advertisements. Ad agencies should also pay attention to possible heterogeneity in the preferences of the target audience and use different faces to target different customer segments. In addition, the methods used in the present study provide a new approach for professionals interested in conducting a quantitative study to assist in the screening and selection of print media spokespersons. It is worth noting that for the categories that included the original faces, the original faces were very rarely selected as the best face of either mean or high rating. Thus, the ad hoc face selection process in practice does not generate the best result, showing the value of our quantitative method.

We propose using the eigenface method for face screening purposes. From an ad agency perspective, such face screening can be done with the following steps. First, ask modeling agencies to submit the faces of the models they represent and pool all of the faces into one database (containing perhaps 1,000 or more faces of professional models). Second, decompose these faces into loadings on a set of eigenfaces.

As a result, every face in the face database will be represented by a vector of loadings on these eigenfaces. Third, measure the hedonism of a product category and identify the segment(s) that best match the segments of the target customers. Fourth, identify the top 5 or 10 faces from the face database using the results from this paper. These top faces could be used for copy testing later. The ad agency need only do the first and second steps once. The third step is specific for each client and requires some very reasonable effort. The fourth step is also specific for each client but can be done automatically. Note that an ad agency can conduct a study similar to what we did in this paper for a specific client and a specific product category using the client's target segment as the sample. The results will be used for the third step above. This may yield a more precise choice of face to represent the client, and it obviously requires substantially more effort (but is still feasible).

The eigenface method is good for most practices that use faces extensively, such as print ads that feature a real human face or a virtual agent that features a synthetic face. Unlike a real human face whose eigenface loadings are fixed corresponding to a specific face, a virtual agent's face is synthetic so that its eigenface loadings can be more easily changed so as to morph the face. For a virtual agent, the steps described above can be used not only for face screening but also for face morphing to construct the optimal virtual agent face for a certain condition.

This paper is the first attempt to apply a quantitative approach to studying faces in the marketing field. We hope the introduction of this quantitative approach (specifically, the eigenface method) will encourage fruitful future research on faces. Here, we list several promising directions for future research. First, the eigenface method suffers from the limitation that it is hard to interpret and quite nonintuitive. By looking at the eigenfaces, it is very hard to tell exactly how each eigenface is different from the others. Future research should focus on creating an intuitive way to interpret the eigenface method and its result. Second, this paper focuses on static faces with constant expressions; future research could go one step further by studying the effect of different facial expressions. Third, in the current study, we controlled all elements except for the faces and product categories. It might be interesting to study the interaction of face effect with other elements (body shape, hairstyle, costume, gender, age, etc.) in the future. Fourth, we used one category feature, hedonism, to represent product categories. Other category features might also interact with face effect. In addition to category features, other variables in the ad context, such as brand personality (Solomon et al. 1992), might

interact with face effect as well. It might be worthwhile exploring other variables for interactions with face effect. Fifth, we used greedy search for variable selection. Greedy search is a generic method that is sometimes unable to generate the global optimal model. It might be worthwhile exploring other variable selection methods (e.g., the Metropolis-Hastings method) in future research. Finally, this study focuses on print ads. Future researchers might consider studying the effects of faces in video ads, given the dominant role of TV advertising in total media ad spending (Nielsen 2013).

Acknowledgments

The authors thank Gary Lilien, Wayne DeSarbo, Jun Liu, Xinwei Deng, and participants from the 2011 Marketing Science Conference in Houston and the Annual Ph.D. Student Presentations at Pennsylvania State University for their helpful comments. This research is supported by a Smeal Small Research Grant and a National Natural Science Foundation of China fund [Grant 71232008].

References

- Andrews R, Currim I (2003) A comparison of segment retention criteria for finite mixture logit models. *J. Marketing Res.* 40(2):235–243.
- Baron R, Kenny D (1986) The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Personality Soc. Psych.* 51(6):1173–1182.
- Berry DS, McArthur LZ (1985) Some components and consequences of a babyface. *J. Personality Soc. Psych.* 48(2):312–323.
- Bishop CM (1996) *Neural Networks for Pattern Recognition*, 1st ed. (Oxford University Press, Oxford, UK).
- Bishop CM (2006) *Pattern Recognition and Machine Learning* Springer Science + Business Media, New York).
- Bower A, Landreth S (2001) Is beauty best? Highly versus normally attractive models in advertising. *J. Advertising* 30(1):1–12.
- Brahnam S (2002) Modeling physical personalities for virtual agents by modeling trait impressions of the face: A neural network analysis. Doctoral dissertation, Department of Computer Science, City University of New York, New York.
- Carré JM, McCormick CM, Mondloch CJ (2009) Facial structure is a reliable cue of aggressive behavior. *Psych. Sci.* 20(10):1194–1198.
- Carroll B, Ahuvia A (2006) Some antecedents and outcomes of brand love. *Marketing Lett.* 17(2):79–89.
- Collins M, Zebrowitz L (1995) The contributions of appearance to occupational outcomes in civilian and military settings. *J. Appl. Soc. Psych.* 25(2):129–163.
- Crowley A, Spangenberg E, Hughes K (1992) Measuring the hedonic and utilitarian dimensions of attitudes toward product categories. *Marketing Lett.* 3(3):239–249.
- Cunningham MR (1986) Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *J. Personality Soc. Psych.* 50(5):925–935.
- Cunningham MR, Barbee AP, Pike CL (1990) What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness. *J. Personality Soc. Psych.* 59(1):61–72.

- Cunningham MR, Roberts AR, Barbee AP, Druen PB, Wu CH (1995) Their ideas of beauty are, on the whole, the same as ours: Consistency and variability in the cross-cultural perception of female physical attractiveness. *J. Personality Soc. Psych.* 68(2):261–279.
- Dumas R, Testé B (2006) The influence of criminal facial stereotypes on juridic judgments. *Swiss J. Psych.* 65(4):237–244.
- Englis BG, Solomon MR, Ashmore RD (1994) Beauty before the eyes of beholders: The cultural encoding of beauty types in magazine advertising and music television. *J. Advertising* 23(2):49–64.
- Ekman P, Friesen WV (2003) *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions* (Malor Books, Cambridge, MA).
- Freiden J (1984) Advertising spokesperson effects: An examination of endorser type and gender on two audiences. *J. Advertising Res.* 24(5):33–41.
- GfK MRI (2011) Despite risks, celebrity endorsers raise print advertising awareness. Press release (February 23). Retrieved January 2014, http://www.gfkmri.com/assets/PR/GfKMRI_022311PR_CelebrityEndorsers.htm.
- Gilly M (1988) Sex roles in advertising: A comparison of television advertisements in Australia, Mexico, and the United States. *J. Marketing* 52(2):75–85.
- Goldsmith R, Lafferty B, Newell S (2000) The impact of corporate credibility and celebrity credibility on consumer reaction to advertisements and brands. *J. Advertising* 29(3):43–54.
- Gorn G, Jiang Y, Johar G (2008) Babyfaces, trait inferences, and company evaluations in a public relations crisis. *J. Consumer Res.* 35(1):36–49.
- Hauser JR, Urban GL, Liberali G, Braun M (2009) Website morphing. *Marketing Sci.* 28(2):202–223.
- Heath T, McCarthy M, Mothersbaugh D (1994) Spokesperson fame and vividness effects in the context of issue-relevant thinking: The moderating role of competitive setting. *J. Consumer Res.* 20(4):520–534.
- Inman JJ, Winer R, Ferraro RS (2009) The interplay among category characteristics, customer characteristics, and customer activities on in-store decision making. *J. Marketing* 73(5):19–29.
- Koda T, Maes P (1996) Agents with faces: The effect of personalization. *Proc. 5th IEEE Internat. Workshop Robot Human Comm.*, 189–194.
- Miniard P, Bhatla S, Rose R (1990) On the formation and relationship of ad and brand attitudes: An experimental and causal analysis. *J. Marketing Res.* 27(3):290–303.
- Nielsen (2013) Global AdView Pulse Lite Quarter 4, 2012. Report (April 22). Retrieved January 2014, <http://nielsen.com/us/en/reports/2013/global-adview-pulse-lite--q4-2012.html>.
- Ohanian R (1991) The impact of celebrity spokespersons' perceived image on consumers' intention to purchase. *J. Advertising Res.* 31(1):46–54.
- Pincott J (2010) Why women don't want macho men. *Wall Street Journal* (March 27). Retrieved January 2014, <http://online.wsj.com/news/articles/SB10001424052748704100604575145810050665030>.
- Ragunathan R, Irwin J (2001) Walking the hedonic product treadmill: Default contrast and mood-based assimilation in judgments of predicted happiness with a target product. *J. Consumer Res.* 28(3):355–368.
- Ramaswamy V, DeSarbo WS, Reibstein DJ, Robinson WT (1993) An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Sci.* 12(1):103–124.
- Schwartz SH (1992) Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Adv. Experiment. Soc. Psych.* 25:1–65.
- Solomon M, Ashmore R, Longo L (1992) The beauty match-up hypothesis: Congruence between types of beauty and product images in advertising. *J. Advertising* 21(4):23–34.
- Stirrat M, Perrett D (2010) Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psych. Sci.* 21(3):349–354.
- Todorov A, Mandisodza A, Goren A, Hall C (2005) Inferences of competence from faces predict election outcomes. *Science* 308(5728):1623–1626.
- Turk M, Pentland A (1991) Face recognition using eigenface. *J. Cognitive Neuroscience* 3(1):71–86.
- Urban G, Hauser J (1993) *Design and Marketing of New Products*, 2nd ed. (Prentice Hall, Englewood Cliffs, NJ).
- Vermunt J, Magidson J (2005) *Technical Guide for Latent GOLD 4.0: Basic and Advanced* (Statistical Innovations, Belmont, MA).
- Wheatley J (1971) The use of black models in advertising. *J. Marketing Res.* 8(3):390–392.
- Zebrowitz L (1998) *Reading Faces: Window to the Soul? New Directions in Social Psychology* (Westview Press, Boulder, CO).
- Zebrowitz L (2006) Finally, faces find favor. *Soc. Cognition* 24(5):657–701.
- Zhao W, Chellappa R, Rosenfeld A, Phillips P (2003) Face recognition: A literature survey. *ACM Comput. Surveys* 35(4):399–458.