NES
New
Economic
School

VISUAL ELICITATION OF BRAND PERCEPTION

Daria Dzyabura
Renana Peres

Working Paper
No 260

NES Working
Paper series

December
2019

Visual Elicitation of Brand Perception

Daria Dzyabura

New Economic School, Moscow, Russia
ddzyabura@nes.ru
Skolkovskoe shosse, 45, Moscow, Russia
+7(977)5607344

Renana Peres

Hebrew University of Jerusalem, Israel
renana.peres@mail.huji.ac.il
Mount Scopus Campus, Jerusalem, Israel 91905
+972 (054)7725723

December 2019

Visual Elicitation of Brand Perception

Abstract

Understanding how consumers perceive brands is at the core of effective brand management. In this paper, we present the Brand Visual Elicitation Platform (B-VEP), an electronic tool we developed that allows consumers to create online collages of images that represent how they view a brand. Respondents select images for the collage from a searchable repository of tens of thousands of images. We implement an unsupervised machine-learning approach to analyze the collages and elicit the associations they describe. We demonstrate the platform's operation by collecting large, unaided, directly elicited data for 303 large US brands from 1,851 respondents. Using machine learning and image-processing approaches to extract from these images systematic content associations, we obtain a rich set of associations for each brand. We combine the collage-making task with well-established brand-perception measures such as brand personality and brand equity, and suggest various applications for brand management.

KEYWORDS: *Image processing, machine learning, branding, brand associations, brand collages, Latent Dirichlet Allocation.*

Statement of Intended Contribution

Understanding how consumers perceive brands is at the core of brand management. We propose and implement a direct brand perception elicitation method called Brand Visual Elicitation Platform (B-VEP). We then apply the method to several important problems in brand management.

Method. The method consists of the respondent task and the analysis. In the elicitation task, participants create online collages of photographs to represent their relationships with a brand on an online platform that we developed. We then use image processing and unsupervised machine learning methods to identify patterns in the resulting collages and extract associations. The primary goal of the method is to measure consumer brand associations in a direct, unaided, scalable way.

Applications. We apply the method to several important problems in brand management. Specifically, we:

- Create a prototypical collage for each brand, by finding a set of photos that represent the distribution of associations for the focal brand. This set can be used as a mood board to help graphic designers generate visual brand content, and to visually communicate the brand associations.
- Measure the uniqueness of associations: what do consumers associate with the brand significantly more/less than with other brands in its category?
- For each of the nine product categories in the data, relate the associations to brand favorability to identify desirable and undesirable associations in each category.
- Relate the associations to well-established brand metrics: characteristics of brand personality and brand equity.
- Find brands in different categories with similar associations, to suggest possible strategic alliances.

With our insights, brand managers can evaluate the retrieved associations against their desired positioning goals, aim their marketing mix elements to be consistent with these associations, to enhance the associations that fit to this positioning, and repress the undesirable associations. Creative advertising teams will be able to use our method to select images from photo repositories to fit with the current brand associations.

Visual Elicitation of Brand Perception

Introduction

Understanding how consumers perceive brands is at the core of brand management. It helps managers develop and position new products, understand the competitive landscape, and create effective marketing communications. As a result, measuring brand perception is also a central topic for marketing academics.

Brand perception is often conceptualized as an associative network, where concepts related to the brand attributes, benefits, and attitudes are represented as memory nodes whose connectivity represents the brand. Keller (1993) argues that these associations are diverse – they can relate to product related attributes, to attributes related to the marketing mix (price, packaging), user and usage. Associations can relate to the functional benefits of the products, to experiences and symbols, as well as to attitudes. The favorability of these associations, their strength, and their uniqueness determine the brand relative position to other brand, its competitive advantage, and, resulting from that, its brand equity. Under this framework, a brand manager's task is to manage what consumers associate with her brand: strengthen desired associations and weaken undesired ones. Because brand associations can be so diverse, eliciting and measuring them in an interpretable way, across brands and individuals, is challenging.

The proliferation of online social media platforms on which users contribute brand-relevant content has made possible scalable, unaided brand tracking by mining this user generated content (UGC). Extracting brand insights from UGC has recently received a lot of attention, in both industry and academia. The abundance of brand-related content that consumers post makes monitoring these brand conversations important for brand managers. These data have the advantage that they are unaided, and consumers can freely discuss any topic related to the brand. Researchers have used text data, such as reviews (Lee and Bradlow 2011), blogs (Gelper, Peres, and Eliashberg 2018), microblogs (Culotta and Cutler 2016), and discussion forums (Netzer et al. 2012), and recently visual data (Jalali and Papatla 2016; Zhang et al. 2017; Liu, Dzyabura, and Mizik 2019; Pavlov and Mizik 2019) to identify topics frequently discussed with brands.

However, for the purpose of understanding consumer brand perceptions, UGC suffers from some shortcomings. First, it is available for only certain categories: whereas the brand Nike generates a lot of social media commentary, finding social media posts on the brand Colgate, for instance, is difficult (Lovett, Peres, and Shachar 2013). Second, it is difficult to control the characteristics of the content contributors. For example, users who have a stronger relationship with the brand (Labrecque 2014), or who hold a particularly strong positive or negative opinion, may contribute more (Lovett et al. 2013). Finally, even a given consumer who contributes brand content may not contribute her true opinion of the brand: Consumers may post strategically to signal about themselves to the public (Han, Nunes, and Drèze 2010; Lovett et al. 2013) and serve their self-presentation needs (Seidman 2013).

To circumvent these challenges of UGC, we create a brand perception elicitation platform, inspired by qualitative elicitation approaches used in psychology and marketing. We developed an online Brand Visual Elicitation Platform (B-VEP, hereafter) for eliciting brand perceptions by asking consumers to create an online collage of images. The basic premise is that although the exact representation of brand associations in the human brain is not known, thoughts occur, in many cases, as images and visual metaphors, and therefore visual research methods are considered to better reflect the emotions, the cultural experiences and the attitudes that constitute the associations, as opposed to verbal methods that focus more on the discourse of these experiences (Reavey 2012). Also, use of images has been demonstrated to successfully act to disrupt well-rehearsed narratives of people (Reavey 2012), and hence might be effective in revealing hidden, often unarticulated associations and ideas.

Our methodology shares some elements with Zaltman's Metaphor Elicitation Technique (ZMET), a collage-based interviewing technique (Zaltman and Coulter 1995; Zaltman and Zaltman 2008). In ZMET, participants are asked to create a collage of pictures to represent how they view a brand. The method, which has been widely used by practitioners (Catchings-Castello 2000), argues that consumers store a rich visual representation of the brand and their relationship to it, and creating collages is an efficient method to elicit these metaphors (Zaltman and Coulter 1995).

Using images to reveal brand associations is supported by the extensive use of visual stimuli by firms to build brand identity, convey messages, and shape consumers' attitudes and preferences (Wedel and Pieters 2008). The human ability to process and relate to pictures and images (Kress and Van Leeuwen 1996; Palmer 1999), and to associate them with feelings and emotions (Cho, Schwartz and Song 2008) makes visual elements a key factor in customer-brand communication (Wedel and Pieters 2008; McQuarrie 2008). Visual elements such as product packaging (Greenleaf and Raghubir 2008), store design (Meyers-Levy and Zhu 2008), graphic design of ads (Pieters, Rosbergen and Wedel, 1999; Wedel and Pieters 2000; Rayner, Miller and Rotello 2008), and the visual context on which the brand is displayed (Cho, Schwartz and Song 2008) have shown to have a considerable impact on consumers' responses to brands. Like other qualitative direct-elicitation approaches (see Steenkamp and Van Trijp, 1997, for a review), a collage making task results in data that are less directed by consumers' strategic goals when posting on social media, can be applied for any brand, and can be used to gather responses from a controlled sample of consumers. It also has the advantage of being fully unaided and free form, allowing consumers to express their views in terms of a wide range of concepts. However, because it requires the presence of an interviewer, it is expensive to conduct at scale.

The method we propose builds on the principles of the collage-making procedure of ZMET and automates it both in the collage making step, as well the analysis. The collages are created online and can be collected from any desired sample of respondents. Participants can choose photos from their collages from a broad repository of hundreds of thousands of photos, using free browsing as well as keyword search, to accurately depict their relationship with the brand through the collage.

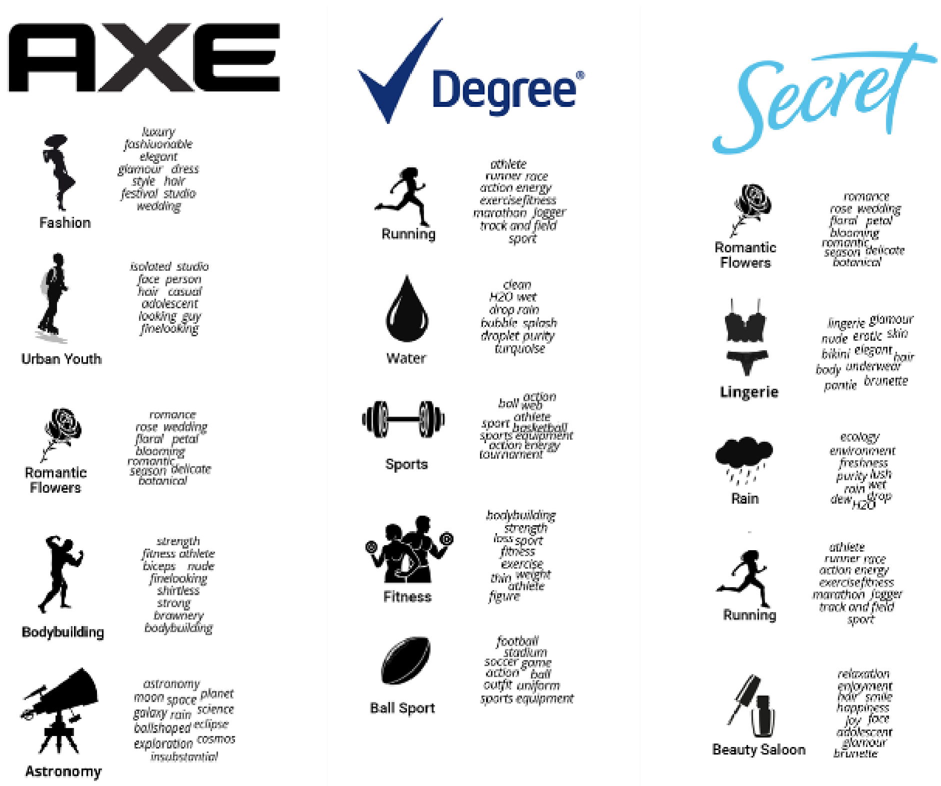
We analyze the collages using a machine-learning back end to derive quantitative insights from the collages. This step of content extraction combines several machine learning algorithms: image tagging, word embedding, and LDA topic mining. By using unaided elicitation and unsupervised learning algorithms, we do not limit the dimensions on which the brand perceptions are measured. The scalability of our approach makes surveying a large consumer population about any brands feasible.

To implement the method, we developed a designated software platform, which automates the collage creation task. We conduct a proof of concept for the insights it can provide by gathering collage representations for 303 major national US brands, from 1,851 respondents. By pooling the collected responses from all the brands, we generate, using unsupervised machine learning, a single, unified space of 150 brand associations, relating to objects, actions, adjectives, characters, places, sceneries, concepts, and metaphors, on which all these brands are mapped, to form the equivalent of a very high-dimensional perceptual map. Figure 1 presents three example brands from our data—*AXE*, *Degree*, and *Secret*—and their most frequently occuring associations. Note that the set contains associations relating to attributes, benefits, and attitudes (Keller 1993), that go beyond the standard dimensions of brand personality and brand equity. Although the three brands describe functionally similar deodorants, each brand has a distinctive set of associations: *AXE*'s strongest associations are of fashion, *urban youth*, *flowers*, *astronomy*, *bodybuilding*, and *band*. *Degree*, on the other hand has more athletic associations, such as *running*, *water*, *sports* and *fitness*, and *Secret*'s associations are more romantic and delicate including *flowers*, *lingerie*, *rain*, and *beauty salon*.

To complete the analysis, we present several potential applications of the method for brand managers: first, we show how to index repositories of photos according to their fit to the associations of a brand and use it to create prototypical collages, which can serve as mood boards, or inspiration boards for the brand visual image. Second, we show how to measure brand uniqueness relative to its category. Third, we relate the associations to brand favorability measure. Forth, we relate the associations to well established perceptual dimensions, by combining the task with a brand perception survey, including dimensions related to brand personality (Aaker 1997) and brand equity (Lovett et al. 2014, Mizik and Jacobson 2008). Fifth, we show how to use the similarity and distance in the association space, to detect potentially valuable communalities between brands, for example for potential collaborations.

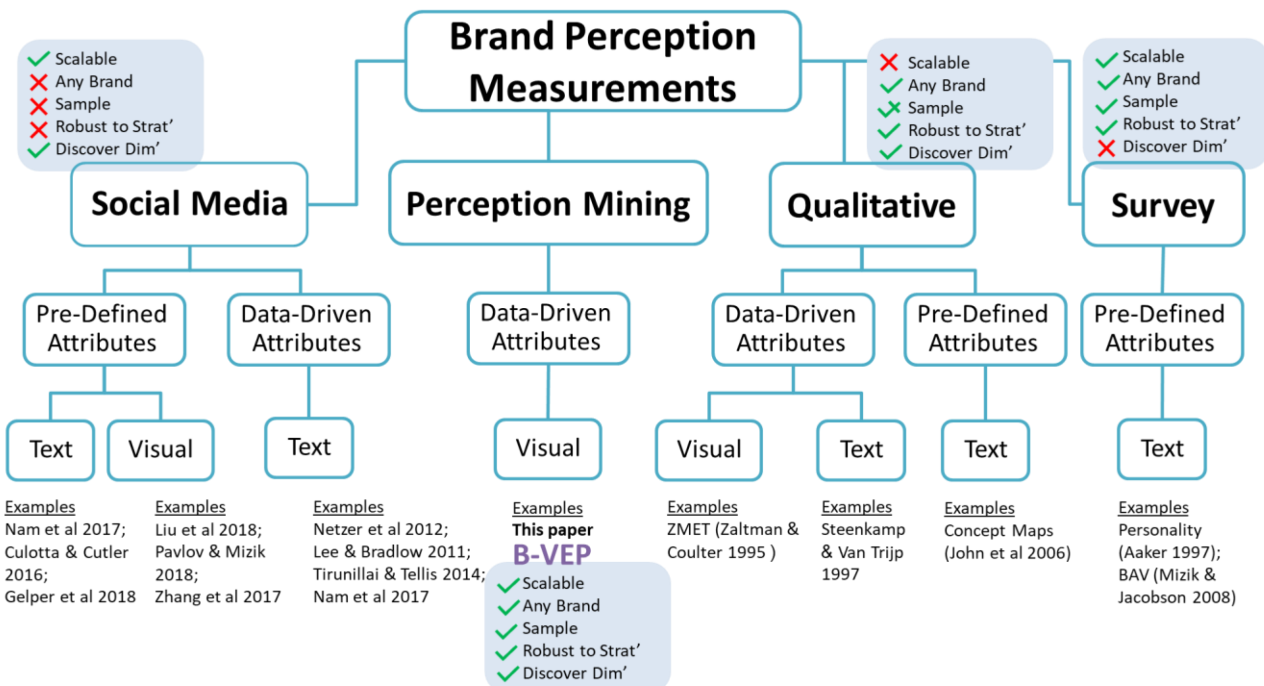
Our contribution is that we *propose* and *implement* a brand perception elicitation method, and then *apply* the method to several important problems in brand management. We develop a brand association direct-elicitation method, which is unaided, scalable, rich, and not sensitive to biases introduced by using UGC. Figure 2 provides a diagram summarizing extant brand perception elicitation methods and how B-VEP contributes to this literature. We categorize the existing methods by the type of data they are based on: survey, qualitative, or social media/UGC. We indicate, for each of these categories, whether or not they are scalable for many brands and respondents, can be done for any brand, the researcher can control the sample, robust to consumer strategic responses, and whether the dimensions along which brands are measured are pre-defined by the researcher or discovered from the data. Note that while we label qualitative data as being able to capture a sample according to the researcher's needs, these methods tend to be costly and time consuming therefore there are usually executed on small number of respondents and are not concerned with generalizability. A directly elicited measure of visual brand perception provides an important benchmark for UGC based metrics.

Figure 1: An illustration of the strongest association topics (with the associations constructing them), in decreasing order for *AXE*, *Degree*, and *Secret*.



The other methodological contribution we suggest is the unsupervised machine-learning approach to extract the brand perceptions portrayed in the collages. This analysis results in a rich set of brand associations (see Figure 1 and Table 1) that contain many types of the associations suggested in Keller's conceptual model (Keller 1993), as well as their strength.

Figure 2. Mapping the literature on brand perception measurement methods. For each method category we indicate whether or not it is scalable for many brands and respondents, can be done for any brand, the sample can be controlled by the researcher, robust to consumer strategic responses, and whether it allows discovering new dimensions that were not pre-defined.



We also contribute by applying the method to several important problems in brand management. Specifically, we:

- Create a prototypical collage for each brand, by finding a set of photos that represent the distribution of associations for the focal brand. This set can be used as a mood board to help graphic designers generate visual brand content, and to visually communicate the brand associations (see Figure 5).

- Measure the uniqueness of associations: what do consumers associate with the brand significantly more/less than with other brands in its category? (Table 4 and Web Appendix D[1])
- For each of the nine product categories in the data, relate the associations to brand favorability to identify desirable and undesirable associations in each category (Web Appendix E).
- Relate the associations to well-established brand metrics: characteristics of brand personality and brand equity (see Tables 5 and 6 and Web Appendix F).
- Find brands in different categories with similar associations, to suggest possible strategic alliances (see Web Appendix G).

Our work helps in retrieving brand associations that are far richer and deeper than what can be obtained using elicitation with typically a small number of aided dimensions, or by a simple search and verbal descriptions. With our insights, brand managers can evaluate the retrieved associations against their desired positioning goals, aim their marketing mix elements to be consistent with these associations, to enhance the associations that fit to this positioning, and repress the undesirable associations. Creative advertising teams will be able to use our method to select images from photo repositories to fit with the current brand associations.

Theoretically, our work suggests a way to quantify many aspects of the brand image (Keller 1993) as part of a single elicitation task. – It results in a rich space of various types of brand associations on which a large number of brands can be mapped. Our method enables assessing the favorability of associations, their strength, their uniqueness, and their connection to brand equity.

[1] The Web Appendices are found at https://www.dropbox.com/s/raden5wlg1wzkxo/Appendix%20A%20B%20C%20D%20E%20F%20G%20-%20To%20link%20to%20paper.xlsx?dl=0

Brand Visual Elicitation Platform (B-VEP)

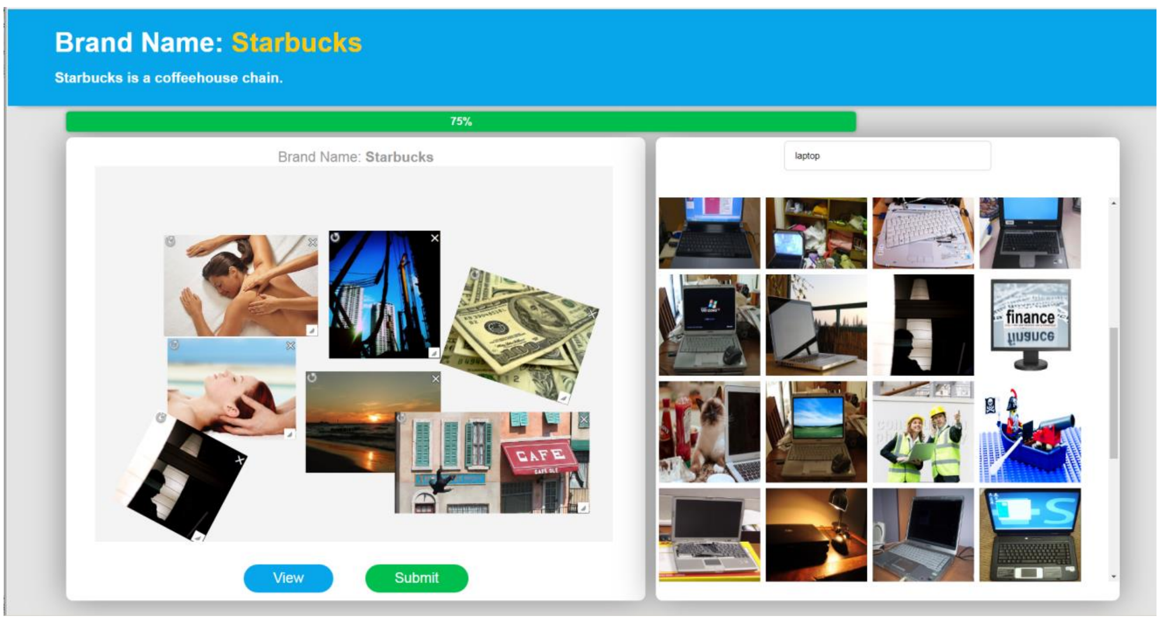
Our main data-collection tool is a software platform we developed on which consumers can create collages for brands. Collage creation is an expressive technique that has been used in research in psychology (Koll, Von Wallpach, and Kreuzer, 2010) and marketing (Zaltman and Coulter 1995; Zaltman and Zaltman 2008). Collages are known to support creative and metaphorical thinking by asking consumers to elaborate on their opinions and thoughts about their experience with a brand, and to merge pictures in various forms to one assembled composition (Davis and Butler-Kisber, 1999). Collage-making is an unaided visual elicitation technique that helps uncover hidden associations and emotions that could have remained undetected through other techniques (Koll, Von Wallpach, and Kreuzer, 2010) and therefore is appropriate for eliciting visual brand representation. Although traditionally, collage-making is a qualitative research method, we used its principles to develop an online collage-creating platform that can be carried out for a large number of brands and people, and quantitatively analyzed.

The collage-making procedure in this study is as follows. A respondent was assigned to a brand and was asked to think carefully about the brand, namely, "What are your emotions, associations, and expectations with respect to the brand? What does the brand mean to you? Recall your experiences with the brand. What are the colors associated with the brand? What shapes? What objects? What images?" The respondents were then shown several instruction screens explaining how to create the collage. Next, they were taken to the collage-making screen. Figure 3 provides a screen shot of the screen on which the collages were created.[2] The screen is

[2] The collage task, as well as the other parts of the questionnaire, can be found at http://bvep.ResearchSoftwareHosting.org

divided into two sections: The left-hand side is the "canvas" on which the respondent creates the collage (in Figure 3 the brand is *Starbucks*), and the right-hand side contains a large repository of photos for the respondent to choose from. Respondents could drag photos from the right- to the left-hand side to create the collage on the canvas. They could move, resize, and rotate the images once they had dropped them on the canvas. The right-hand screen contains the photo repository. Respondents were able to either scroll through photos randomly, or search for keywords, and retrieve photos relevant to that keyword. For example, in the screen shot in Figure 3, the user searched for the keyword "laptop."

Figure 3. The collage canvas. The photo repository is on the right side of the screen, and the canvas is on the left.



The photo repository is a key element in the platform. First, it should be large and diverse enough for respondents to not feel constrained by the images and to be able to accurately convey their perception of the brand with the available images. Second, the images should push the respondents to think about the entire spectrum of associations, including functional, experiential, and symbolic, user related and usage related (Keller 1993) associations of the brand beyond the product related attributes or the obvious brand elements. For example, a collage for Levi's should not simply be a collection of photographs of jeans, or Levi's logo.

With the above goals in mind, we created the photo repository and designed the right hand side of the screen. We began by downloading a large set of photographs from Flickr, a photo-sharing website. Flickr allows users to label the photographs they upload or view. To make the repository as rich as possible, and to help ensure participants could find photographs that represented what they were trying to communicate, we downloaded photos by querying Flickr's API for the top 4,000 nouns, verbs, and adjectives in the English language, and downloaded the first 50 photo results for each. Overall, the image database consists of 100,000 photographs (because many photographs are labeled with multiple labels).

We also implemented a search engine on the platform, so participants could more easily find images for their collage. The search engine returns photos that on Flickr have the labels of the queried word, in randomized order. For example, in the screen shot in Figure 3, the user searched for the term "laptop." The ability to retrieve photos by search words helped respondents tailor the collage to achieve a more accurate representation of their brand perception. Since each search word retrieved many photos (e.g. there are 46 photos in our repository labeled with "laptop", 318 labeled with "family", 3621 labeled with "nature" etc.), the search option did not limit the users, but was rather used as an initial aid in browsing through the repository. We wanted to ensure the collage represented the respondent's perception of the brand beyond simply the product category and the company's own marketing efforts. We also wanted to encourage respondents to retrieve personal and meaningful associations. To that end, we constrained the words respondents could search for. The system does not allow them to search for the brand itself, the category, or the type of product. If they did, they saw an error message, saying the word was not allowed as a search term for this brand. For example, when creating a collage for *Levi's*, the user would not have been able to search for "Levi's," "clothing," "apparel," or "jeans." Research assistants manually generated the list of these "banned" keywords for each brand.

Each respondent was assigned elicitation tasks for three brands sequentially. To ensure respondents only created collages for brands they were familiar with, respondents first had to rate their familiarity with 10 brands on a 5-point scale (5 = *very familiar*, 1 = *not at all familiar*). Three focal brands were selected randomly from those the respondent rated 4 or 5. If a respondent was not familiar with any of the brands, another set of 10 brands was presented, and if, after three sets of 10 brands, no brand was scored with 4 or 5 on familiarity, the survey terminated for that respondent.

Respondents were encouraged to spend as much time as needed to create a thoughtful collage. If a respondent submitted a collage after less than 2 minutes had elapsed, or if the collage contained less than six photographs, a pop-up screen appeared asking if she was sure she wanted to submit. After submitting the collage, respondents were asked to score the task's level of difficulty on a 5-point scale, with 5 being very difficult. As a sanity check, and to make sure respondents understood the task, respondents were also asked to briefly verbally describe the collage and explain their choice of images. Finally, research assistants checked each collage manually and removed the data if the participant did not appear to have invested sufficient effort in the collage. The criteria for deletion were to delete collages that took less than 1 minute to make, that used only 1-2 photos, and that the responses for the brand characteristics (see below) were identical for all items (e.g. respondent chose to rank the brand only 1 or only 5 or 3 on all the 49 items). In total, 17% of the collages were removed.

Designing the software platform was a major undertaking. Its user friendliness and clarity were essential for engaging respondents and obtaining high-quality collages. The user interface was designed following design best practices (Johnson 2013) using professional web designers. All screen, instruction, and error messages were extensively tested for clarity and understandability by an internal team of 10 users, as well as an external beta test team of 50 Amazon Mechanical Turk (MTurk) users.

Data

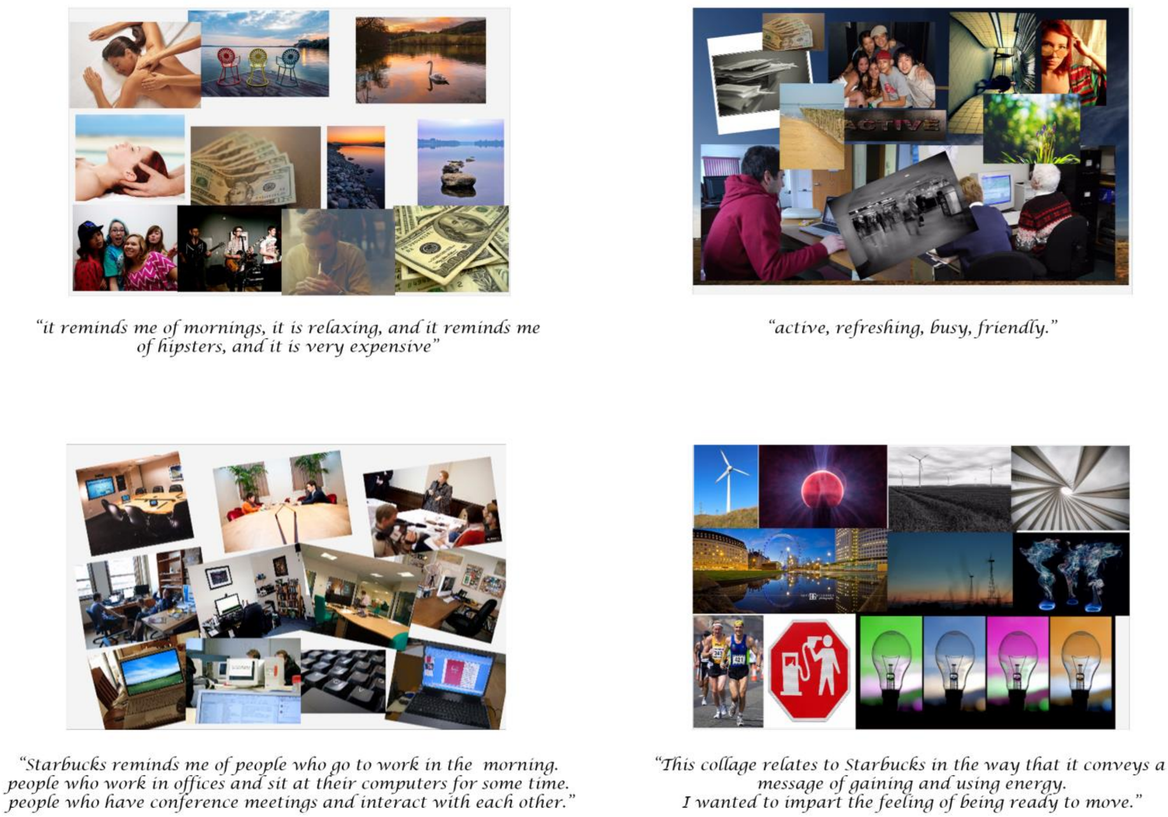
Respondents for the task were recruited on MTurk and received $2.50 for completing the entire task. Although our sample was not created to be demographically representative, it is quite balanced, skewed toward younger males. In total, all our respondents were US residents, 43.5% were males (and 56.5% females), 26% were 18-29 years old, 41% were 30-39 (the age group 18-39 forms 36.5% of the US population), and 33% were 40-69 (this age group forms 54% of the US population). Note that we used MTurk as a proof of concept and a means to recruit a large number of subscribers from the general population. If needed, a firm could use a more representative sample of respondents. Each respondent completed the task for up to three brands, or up to 30 minutes, whichever came first. That is, if the respondent was only finished with her first or second brand after 30 minutes, she was taken to the final screen, which thanked her for participating, and terminated the study. The time limit helped us avoid fatigued respondents.

Brand Collages. We collected 4,743 collages from 1,851 respondents (3,937 were approved by the research assistants). The data include an average of 15.6 collages per brand, for 303 national US brands from 9 categories: beauty (40 brands), beverages (65 brands), cars (29 brands), department stores (17 brands), food and dining (84 brands), home design and decoration (16 brands), household cleaning products (19 brands), apparel (23 brands), and over-the-counter medications (10 brands). The list of brands is an updated version of Lovett et al. (2013), excluding TV shows, video games and movies, and since-discontinued brands. The full list of brands is presented in Web Appendix A. On average, each collage took 8 minutes, and included 11.45 pictures. The average reported level of difficulty of creating the collages was 2.5 on a 5-point scale, with 5 being the most difficult.

Mostly, respondents used mixed methods of browsing through the photo repository and searching for specific search options. The search option was not used frequently, and 690 collages (17.5% of the approved collages) did not use search at all. The median number of search words used in a collage is 5, and the average is 6.4. Also, respondents did not make many attempts to use the “banned” words - out of the 25,262 search words used, only 1,111 (less than 5%) attempts were made to use the "banned" words. To further verify that the search function did not restrict or bias the collages, we compared, for each brand, the associations derived from the brand’s collages who used above median number of search words, to collages in which the number of search words used was below median. We found that this specific split was not significantly different from any random split of collages (see Appendix 1 for details).

Figure 4 presents a sample of four collages for the brand *Starbucks*, from four different respondents, along with the verbal description. We see that the collages contain rich, meaningful information about the brand associations. They are not simply showing people drinking coffee, or images related to Starbucks’ brand elements. At first glance, these collages appear to be very different from each other, and do not seem to demonstrate any communality. However, as we show next, these collages share specific visual elements that create a distinctive, unique set of associations, which is unique to *Starbucks* and differentiates it from the other brands in the sample, as well as from brands in its category. We will discuss how associations are extracted in the next section.

Figure 4: Examples of four collages for the brand *Starbucks*, made by four different respondents, with their verbal descriptions.



Brand Characteristics. In addition to the collage, we collected data on respondents' perceptions of the brands according to well-established brand metrics. After completing each collage, respondents were asked to rate the brand on each of 49 items, on a 5-point scale. The set of items is the union of Aaker's (1997) personality dimensions, and BAV brand equity items that constitute the four pillars of brand equity (Lovett et al. 2014). The items were presented in a randomized order. We consulted with a BAV team to operationalize the survey as closely as possible to the way they operationalize theirs. A major difference is that BAV's survey is done on a representative sample, whereas our sample, as explained above, is not truly representative. To show face validity, we calculated the correlation between the average brand score on each BAV item in our survey and the scores we received from BAV 2016-2017 data for these brands. The average correlation is 0.58 ($p<0.05$), lending face validity to our measurements.

Feature Extraction from the Collage Data

The collage-creation task generated a set of collages for each brand. Our goal is to extract and summarize interpretable associations from the raw brand collages and organize them into a single, unified space on which all brands can be mapped and analyzed. To do so, we use image-tagging methods to extract the visual elements of the collages, and identify patterns among tags in brand collages.

In many image-processing applications (e.g., Liu, Dzyabura, and Mizik 2019), the objective is to solve an image-classification problem. Therefore, the visual features extracted from images do not have to be interpretable, and typically include low-level features such as edges, corners, color histograms, shapes, line directions, and texture, or even more abstract deep-learned features. Our goal in this paper is different: we do not use the visual elements as an intermediate stage in solving a prediction problem. Rather, we are looking for what associations set one brand's collages apart from others' thus creating a mapping from brands and brand characteristics to visuals. Hence we are interested in extracting and summarizing *interpretable* features. For this reason, we turn to image tagging.

We used a commercially available image tagging tool called Clarifai (Rangel et al. 2016), which is pre-trained on a corpus of millions of photos and uses deep convolutional neural networks to classify the content of photos into over 11,000 semantic tags (labels) relating to the objects, scenery, actions, emotions, adjectives, and other visual elements (Howard 2013). Clarifai offers several options for pre-trained models, of which we used "general 1.3." Each photo is assigned the 20 tags with the highest confidence scores. For example, the photo at the bottom left of the bottom-right collage for *Starbucks* in Figure 4, showing men in a running competition, is tagged with athlete, competition, race, runner, marathon, track and field, jogger, running, athletics, fitness, action, energy, exercise, footrace, hurry, endurance, motion, effort, jog, man, and sport. The 4,743 collages in our dataset contain 91,856 photos, yielding 5,426 unique tags (the approved 3,937 collages had 4,601 unique tags).

In order to extract meaningful associations from this large set of tags, we need to group similar tags into association topics, and then measure the distribution of topic in the collage of each brand.

Extracting association topics from collage tags documents. The set of tags for each brand contains the extracted set of consumers' associations with the brand. However, the large number of total tags makes these results hard to interpret. Moreover, many of these tags are virtually synonyms, or higher-level labels of one another, for example, "girl" and "child". In order to identify which associations occur more often in one brand than others, these similar tags need to be grouped together as they represent one association.

We analyze the tags using a topic modeling approach called Guided LDA, a semi-supervised variation of the popular unsupervised topic modeling algorithm, Latent Dirichlet Allocation (Jagarlamudi, Daume III and Udupa, 2012). Each collage is treated as a document, and each tag as a word. Given enough training data, regular LDA would likely identify that "girl" and "child" are part of the same topic, if they indeed are a robust association occurring in many collages, and frequently occur together. However, given that our documents are relatively short for LDA, we found that we greatly enhance performance by initiating words with similar meanings to be in one topic. The guided LDA approach allows the researcher to specify an initialization, or seed, set of topics. For example, say we want the words "girl", "boy", and "child" to converge towards topic 1. In the initialization set, we can push these words to lie in this specific topic. How much extra boost is given to a term in a topic is controlled by a confidence parameter. The ability to initialize words to topics allows us to incorporate knowledge of word meanings, which regular LDA does not take into account.

***Initializing words to topics*.** We obtain the set of topics used for initializing Guided LDA by obtaining a word embedding for each tag, and clustering the tags in the embedded space. To obtain the word embedding, we used Stanford's Global Vectors tool, GloVe. GloVe is an unsupervised algorithm that is pre-trained on over 6 billion text tokens from Wikipedia and the linguistic data English GigaWord 5th Edition website. During the training phase, the algorithm uses global matrix factorization methods, in combination with local context window methods, and applies them on the training corpus to create a 300-vector dimensional space (Pennington, Socher, and Manning 2014). The algorithm takes into consideration factors such as word-to-word co-occurrence, context similarities, and word analogies. We used this 300-dimensional space provided by GloVe as input to our analysis, and represented each tag in our dataset as a point in this space. We then clustered the resulting vectors using a *k*-means clustering algorithm (Scikit-learn machine learning Python package). The clustering procedure results in 465 word clusters, which is about 10% of the original set of tags. We removed clusters that occur in fewer than 50 collages, and fewer than 6 tags, leaving us with 120 word clusters to be used as input into Guided LDA.

As is common in text mining, we remove the most- and least-frequently occurring tags. Specifically, we removed tags occurring fewer than 10 and more than 2000 times in the corpus, resulting in a total vocabulary of 2,596 unique tags (out of the original 4,601). The resulting corpus was used as input to guided LDA. The result is (1) a set of topics, each topic being a distribution over tags, and (2) the distribution over topics for each collage. The main parameter to be set by the researcher is the total number of topics. We experimented with different values of this parameter. We compared the resulting topic distributions in terms of the similarity of high-probability tags in one topic. We found that the results with 150 topics had the most face validity.

Finally, we named each topic manually, using three research assistants, majoring in English literature, based on the tags with the highest probability for the topic, to make sure topic names are meaningful. These 150 topics form a rich set of brand associations, including objects (*animals*, *food*, *people*, etc.), constructs (*abstract art*, *horror*, *contemporary*, *delicious*, *famous*, *fantasy*, *illness*), occupations (*musician*, *bodybuilding*, *baking*), nature (*beach*, *misty*, *snowscape*, *wildlife*), and institutions (*corporate*, *army*, *investment*, *school*). In Keller's (1993) terminology, these associations represent product related attributes (e.g. *alcoholic drinks*), non-product related attributes (e.g. *baby*, *holiday party*), functional, experiential and symbolic benefits (e.g. *fitness*, *cityscape*, *popstar*), and attitudes (e.g. *American flag*).

Web Appendix C contains the distribution of tags in the association topics, as well as the topic names. For example, the association *aeronautics* is composed of the tags air, flight, airplane, aircraft, flying, military, jet, with probabilities 7.7%, 7.4%, 6.3%, 6.3%, 5.1%, 4.8%, 3.9%, respectively. The topic *cityscape* is composed of downtown, cityscape, skyline, skyscraper, modern, office, tower, bridge with probabilities 9.4%, 9.4%, 7.9%, 7.4%, 7.4%, 5.3%, 5.0%, 4.4% respectively. The topic *running* is composed of athlete, runner, race, action, energy, exercise, fitness, marathon, jogger with probabilities 7.8%, 6.4%, 5.9%, 5.8%, 5.7%, 5.0%, 4.5% respectively.

These 150 association topics (termed associations hereafter) constitute the set of dimensions on which we will map the brands. Note that the dimensions may change for a different set of brands. Next, we analyze how the brands relate to these associations.

To validate the results of the association extraction, we run an additional study as follows. Participants are given a set of associations extracted from a collage, and two different collages to choose from, one of which is the correct collage (from which the presented associations were extracted). They are asked to indicate which of the two collages best matches the presented set of associations. Participants were recruited on MTurk, and paid $1. A total of 46 participants completed the study, each completing 20 tasks, giving us a total of 920 choices. Of these, 784, or 85.2% were correct, which validates the association extraction algorithm[3].

[3] The user interface of the validation experiment can be found in http://collages.researchsoftwarehosting.org/

Results - Brand Associations

For each brand, we average the association distribution extracted from the brand's collages. Guided LDA outputs the probability of each of the 150 topics occurring in each collage. Let θ_{ik} represent the probability with which association k occurs in collage i. We compute the average of the association distributions across the collages for a given brand, namely, $\frac{\sum_{i \in I_b} \theta_{ik}}{|I_b|}$, where I_b is the set of collages for brand b. Table 1 presents the top 10 highest weighted associations for all the brands in the beauty category. The results for all brands are presented in Web Appendix C.

Consider, for example, the makeup brands *CoverGirl* and *Maybelline*. Both are associated with *glamour* and *flowers*, however *Maybelline* has stronger associations for *lingerie* and *hairstyling*, while *CoverGirl* is more strongly associated with *Music festival* and *Holiday party*. Of the beer brands (Web Appendix C), Budweiser is associated with *ball sports*, *fire*, *water*, *auto racing*, and *youth*; and *Corona* with *beach*, *ocean*, *breakfast*, *lingerie*, and *pool*. Recall that each of these associations represents a large number (see Web Appendix B) of objects, concepts, emotions, and activities respondents chose to present in their collages. The associations relate to the brand's product attributes, usage, users, functional, symbolic, and experiential benefits, as well as attitudes towards the brand. The association weights enable measuring the strength of each association in a given brand. As we show later in section Applications for Brand Management , it also enables to calculate brand uniqueness, relative positioning, favorability, and relationships with other brand metrics.

Table 1: The top 10 most frequently occurring associations for Beauty Products brands (in decreasing order of probabilities)

Brand name	Most frequent associations (top 10)									
Always	Glamour	Therapy	Flowers_botanical	Beach	Water	Running	Flowers_romantic	Dance	Flowers_tropical	Hairstyling
AVEENO	Fashion	Streams	Flowers_botanical	Baby	Water	Frosty	Erotic	Delicate_fablic	Rain	Ocean
AVON	Glamour	Hand	Flowers_botanical	Produce	Frosty	Mountain	Fruits	Flowers_romantic	Finance	Beauty_salon
AXE	Fashion	Urban_youth	Flowers_romantic	Astronomy	Bodybuilding	Band	Ball_sports	Suit	Military	Fitness
Bath & Body Works	Flowers_romantic	Water	Flowers_botanical	Therapy	Fruits	Streams	Bedroom	Frosty	Bathroom	Juice
Caress	Flowers_romantic	Fruits	Water	Streams	Glamour	Cat	Resort	Therapy	Car	Fashion
Chanel beauty	Flowers_romantic	Lingerie	Geometric	Jewelry	Alcoholic_drinks	Old_town	Glamour	Resort	Baking	Wheat
Charmin	Water	Flowers_romantic	Bedroom	Cat	Family	Delicate_fablic	Geometric	Animals	Youth	Sailing
Clean & Clear	Beauty_salon	Bathroom	Flowers_botanical	Flowers_tropical	Water	Countryside	Streams	Water_birds	Furniture	Old_town
Clinique	Glamour	Hairstyling	Flowers_romantic	Eye	Painting	Flowers_botanical	Fashion	Ocean	Wedding	Insects
Colgate	Water	Family	Herbs	Glamour	Child	Frosty	Streams	Dental	Youth	Business_school
CoverGirl	Flowers_romantic	Glamour	Holiday_party	Water	Child	Music_festival	Fashion	Flowers_botanical	Cat	Carnival
Crest	Water	Power_energy	Bathroom	Rain	Frosty	Flowers_botanical	Youth	Steel	Mountain	Glamour
Degree	Running	Water	Sports	Fitness	Ball_sports	Bathroom	Steel	Flowers_tropical	Industry	Therapy
Dial Soap	Water	Rainstorm	Ocean	Bedroom	Bathroom	Flowers_tropical	Streams	Hairstyling	Rain	Countryside
Dove	Flowers_romantic	Streams	Water	Warm_fabrics	Erotic	Bathroom	Ocean	Flowers_botanical	Raindrop	Beauty_salon
Garnier Fructis	Fruits	Streams	Fashion	Flowers_romantic	Flowers_botanical	Hairstyling	Power_energy	Ocean	Rainstorm	Baking
Gillette	Wedding	Suit	Sailing	Water_sports	Modern_building	Ball_sports	Bodybuilding	Alcoholic_drinks	Furniture	Raindrop
Head & Shoulders	Hairstyling	Flowers_tropical	Water	Juice	Beach	Streams	Youth	Resort	Bathroom	Rain
Herbal Essence	Flowers_botanical	Rain	Hairstyling	Flowers_romantic	Juice	Metalwork	Flowers_tropical	Streams	Fruits	Water
Irish Spring	Streams	Water	Erotic	Mountain	Bathroom	Rain	Flowers_romantic	Countryside	Flowers_tropical	Running
Jergens	Flowers_romantic	Baby	American_flag	Prey_birds	Fruits	Animals	Sports	Hand	Ocean	Water
Kleenex	Rainstorm	Furniture	Child	Ocean	Baby	Water	Bedroom	Mountain	Flowers_romantic	Prey_birds
Kotex	Flowers_romantic	Fashion	Child	Water	Glamour	Cat	Delicate_fablic	Dining	Baby	Prey_birds
Loreal	Prey_birds	Hairstyling	Glamour	Fashion	Church	Flowers_romantic	Beauty_salon	Hand	Old_town	Delicate_fablic
Mary Kay	Glamour	Hairstyling	Fashion	Flowers_romantic	Lingerie	Water	Curved_lines	Coffee	American_flag	Warm_fabrics
Maybelline	Glamour	Eye	Hairstyling	Fruits	Lingerie	Arts_and_Crafts	Cat	Flowers_romantic	Abstract_art	Flowers_botanical
Neutrogena	Water	Flowers_romantic	Flowers_botanical	Hairstyling	Bathroom	Ocean	Baby	Streams	Rain	Glamour
Nivea	Glamour	Flowers_botanical	Flowers_romantic	Water	Lingerie	Delicate_fablic	Insects	Countryside	Streams	Hairstyling
Olay	Flowers_romantic	Flowers_botanical	Glamour	Hairstyling	Water	Breakfeast	Flowers_tropical	Streams	Prey_birds	Ocean
Old Spice	Bodybuilding	Bathroom	Heavy_vehicle	Cat	Running	Mountain	Streams	Water	Industry	Snowscape
Pantene	Hairstyling	Bathroom	Flowers_romantic	Rain	Beach	Ocean	Retail	Water	Lingerie	Autumn
ProActiv	Water	Hairstyling	Flowers_romantic	Beauty_salon	Produce	Eye	Flowers_botanical	Horror	Raindrop	Finance
Revlon	Glamour	Fashion	Flowers_botanical	Rainstorm	Modern_building	Raindrop	Countryside	Happy_Nature	Arts_and_Crafts	Street_art
Scott Tissue	Cat	Frosty	Prey_birds	Delicate_fablic	Child	Power_energy	Animals	Industry	Fowl	Equiade
Secret	Flowers_romantic	Lingerie	Rain	Running	Beauty_salon	Flowers_botanical	Glamour	Fashion	Cottage	Military
Sephora	Hairstyling	Flowers_romantic	Fruits	Glamour	Water	Abstract_art	Horror	Lingerie	Resort	Flowers_botanical
Suave	Water	Flowers_botanical	Rainstorm	Streams	Flowers_romantic	Foggy_landscape	Hairstyling	Autumn	Lingerie	Snowscape
Tampax	Beauty_salon	Running	Lingerie	Flowers_romantic	Fashion	Water	Sports	Youth	Hairstyling	Cat
Tresemme	Hairstyling	Flowers_romantic	Glamour	Fashion	Flowers_botanical	Beauty_salon	Suit	Prey_birds	Jewelry	Countryside

Table 2: The top 10 most frequently occurring associations for each category in their category averages (in decreasing order of probabilities)

Category	Most frequent associations (Top 10)									
Beauty products	Flowers_romantic 0.046	Water 0.045	Hairstyling 0.034	Flowers_botanical 0.033	Glamour 0.032	Streams 0.023	Fashion 0.019	Bathroom 0.018	Beauty_salon 0.016	Fruits 0.016
Beverages	Water 0.031	Streams 0.025	Ball_sports 0.023	Fruits 0.022	Ocean 0.021	Alcoholic_drinks 0.017	Band 0.017	Countryside 0.017	Water_sports 0.016	Juice 0.015
Cars	Traffic 0.049	Car 0.044	Cityscape 0.037	Finance 0.028	Steel 0.024	Modern_building 0.020	Countryside 0.019	Old_town 0.019	Aeronautics 0.017	Desert 0.017
Clothing products	Fashion 0.028	Sports 0.025	Clothing 0.021	Band 0.020	Street_art 0.018	Running 0.017	Lingerie 0.017	Cityscape 0.016	Hairstyling 0.015	Glamour 0.015
Department stores	Retail 0.050	Finance 0.032	Clothing 0.025	Business_school 0.023	School 0.022	Furniture 0.019	Fashion 0.019	Cityscape 0.019	Modern_building 0.018	Family 0.017
Food and dining	Dining 0.064	Family 0.030	Youth 0.027	Baking 0.021	Child 0.020	Produce 0.019	Farm 0.018	Fire 0.017	Retail 0.017	Furniture 0.016
Health products and services	Family 0.032	Hospital 0.030	Flowers_botanical 0.027	Business_school 0.026	Child 0.024	Baby 0.023	Cat 0.021	Water 0.020	Running 0.019	Flowers_romantic 0.019
Home design	Furniture 0.047	Steel 0.034	House 0.029	Modern_building 0.026	Water 0.022	Family 0.020	Construction 0.019	Dining 0.016	Geometric 0.016	Bedroom 0.016
Household products	Water 0.050	Flowers_romantic 0.030	Furniture 0.029	Flowers_botanical 0.029	Frosty 0.024	Bathroom 0.020	Bedroom 0.019	Family 0.019	Cat 0.019	Mountain 0.018

To validate the brand-association relationship, we run an additional validation study, which follows a similar format to the study used for collage validation. Participants are given a set of associations for a brand, and two different brands to choose from, one of which is the correct brand (for which the presented associations were extracted). They were asked to indicate which of the two brands best matches the presented set of associations. Participants were recruited on MTurk, and paid $1. A total of 91 participants completed the study, giving us 1707 choices. Of these, 1280, or 75% were correct, which validates the brand association relationship[4].

We note from Table 1 that certain associations, such as *flowers*, *water* and *hairstyling* are particularly prevalent in the Beauty Products category. Table 2 presents category averages: the top 10 most frequently-occurring associations in each category, and their average probability of occurring in a collage. The results have face validity in that most of the high probability associations are closely related to the category, e.g. traffic for cars, and furniture for home design and decoration.

***Comparing the content of collages to search keywords and verbal descriptions*.** Our methodology relies on previous research suggesting that visual research methods are powerful in reflecting emotions and attitudes, and therefore can be useful in eliciting deep associations and hidden metaphors (Zaltman and Coulter 1995; Wedel and Pieters 2008; Reavey 2012). To demonstrate the richness of visual associations, we compared the associations derived from the collages to two other means that we collected through the task. The first is the search words users could use while browsing for photos (see Appendix 1for detailed description). The second is the verbal descriptions respondents provided after making the collage, to provide us with a better idea as to what respondents had in mind when creating the collage. Specifically, they were asked to "Describe how your collage relates to brand X."

We assessed the relative richness of search words, verbal descriptions and visuals using several richness metrics. First, we measured the average number of search terms (for the search words), words

[4] The user interface of the validation experiment can be found in http://positiveness.researchsoftwarehosting.org/

(for the descriptions), and Clarifai tags (for the visuals) per collage, as well as their overall number. We also measured the language richness through the relative proportion of speech parts. Finally, we applied the association extraction method, and report how many associations were obtained after the LDA optimization, and how many association topics, on average, are significantly different for a brand relative to its category (this will be explained in more detail when analyzing brand uniqueness). These metrics are presented in Table 3.

Table 3: Comparison of the richness of search words, verbal descriptions and visuals

Richness dimension	Measure	Search words	Verbal descriptions	Visuals	Comment
Volume	Average number of terms/words/tags used per collage	6.41*	10.9	234	*17.5% of the collages used 0 search words.
Volume	Overall number of terms/words/tags over all collages	25,262	86,623	469,903	
Language richness	The relative proportion nouns- adjectives- adverbs- verbs- conjunctions in the terms/words/tags	79.7%-12.9%-1.1%-4.9%-1.3%	28.9%-9.9%-3.6%-17.4%-33.8%	87%-7.9%-0.8%-3%-0.8%	
Elicited Associations	No of associations (LDA topics)	20	30	150	*All verbal descriptions went through a pre-processing step of stemming (reducing inflected words to their word stem e.g. family and families will be counted as one unique word).
Distinctiveness of associations	Average number of significantly different associations per brand relative to the category	4.1	5.7	25.1	

The results indicate that visuals are richer than search words or verbal descriptions. They yield more features (469,963 tags vs. 86,623 words and 25,262 search terms), the analysis generates 150 associations compared with 30 for the verbal descriptions and 20 for the search words. These associations are more distinctive – the number of significant associations per brand relative to the category is 25.1 for visuals, relative to 5.7 for the verbal descriptions and 4.1 for the search words. Interestingly, compared with verbal descriptions, they contain considerably fewer conjunctions, which enhances even more the meaningfulness of the tag derived from the photos.

Applications for Brand Management

We now highlight several applications of B-VEP for brand managers. The first is using the associations to index an entire photo repository and compute, for each photo, how closely it resembles the association distribution for the brand. We demonstrate how to use this indexing to create a ***prototypical collage***, or a mood board for each brand: that is, a collection of photos that together capture the average distribution of associations and provide a visual representation of the brand. The second is measuring ***brand uniqueness***: for each brand, we test how its set of associations differs from those of other brands in its category. The third is the relationship between the associations and ***brand favorability*** measures, we find that the corresponding associations differ by category, e.g. favorable associations for cars are different from favorable associations for beverages. The forth is to relate the brand for ***commonly used brand metrics*** – brand personality (Aaker 1997) and brand equity (Mizik and Jacobson 2010). We find a set of associations that correspond to horizontal characteristics on which brands can differentiate (such as *cool, down-to-earth*, etc.), which transcend the product category. The fifth is how to use the similarity and ***distance in the association space***, to detect potentially valuable communalities between brands, which could be used for potential collaborations.

Photo indexing and a brand's prototypical collage. Our method can be used to provide a measure of the fit between a visual (e.g. a photo, a collage), and a brand. That is, we can use our method to index repositories of images according to their fit with the association set of a brand. Such indexing can have various applications. For example, it can help brand managers and graphic designers to search for brand images which reflect the current set of associations of a brand and to create mood boards, or prototypical collages for brands by visually displaying their associations.

To illustrate this, we used Guided LDA to calculate the distribution of associations of all the images in our photo repository. Then we used a greedy algorithm to choose for each brand the 10 photos (which were not included in any of the brand original collages), that together generate the highest similarity to the brand associations vector. We measured similarity as the cosine similarity between the normalized 150-dimensional topic distribution vectors of the photos and of the brand. Recall that cosine similarity ranges between 0 to 1, where 1 indicates maximum similarity. The average similarity between the brand association vector and the representative collage is 0.899, indicating that we managed to create a set of prototypical collages[5]. Figure 5 presents the prototypical collages for the three deodorant brands *AXE*, *Degree*, and *Secret*. Note that we could have chosen photos from other photo repositories, or created collages containing more or fewer than 10 photos.

Figure 5: Prototypical collages, or mood boards, for the deodorant brands *AXE*, *Degree* and *Secret*, based on cosine similarity between the brand association distribution and the photos in the photo repository.

[5] The representative collages can be found in https://www.dropbox.com/sh/t1gc61mkx2k5lyz/AACL6rXp0le-SisLK8jXuhX2a?dl=0.

Brand uniqueness. To determine how a brand stands out from others in its category, we test whether an association occurs with a significantly higher/lower probability in collages for the focal brand than for other brands in the same category. We choose to compare the brand to its category, rather than simply to all other brands in the set, in order to get rid of category-level averages - for example, beauty brands have on average more flowers, water, hairstyling, and glamour than do car brands. Specifically, we perform a Mann-Whitney test to compare $\{\theta_{ik}: i \in I_b\}$ to $\{\theta_{ik}: i \notin I_b\}$, for each association k. We report for each brand, the associations for which these two samples are statistically significantly different. We present the results for the brands in the Beauty Products category in Table 4. Results for the full set of brands are presented in Web Appendix D.

The left-hand set of columns (most associated with) contains the five associations which occur significantly more frequently in the collages for the brand whereas the right-hand set of columns (least associated with) contains the five associations which occur significantly less frequently for this brand than others in the category For example, we can see that, relative to the average beauty brand, the deodorant brand *Degree* is most associated with *running*, *sports*, *fitness*, *ball sports*, and *water*, meaning these associations appear in its collages significantly more frequently than for the average brand in the category. The flowers, which have a strong presence in the category, although existing in its associations (Table 2) do not differentiate *Degree* from the category. In the Cars category, most cars are associated (see Table 2) with *traffic*, *cityscape* and *steel*, but *Ferrari* (see Web Appendix D) has, relative to other car brands, strong associations also with *aeronautics*, *delicate fabrics*, and *lingerie*, and less strong associations with *industry*, *school* and *church* than the average car brand. *Jeep*, positioned as an outdoorsy brand, has significantly lower weights, in the association distribution vector of its collages, of the *cityscape* and *modern building* than the average car brand.

Table 4: The top 5 most and least frequent associations for Beauty Products brands relative to the category (ordered in each section from most to least unique)

Brand name	Most associated with (Top 5), relative to the category					Least associated with (Top 5), relative to the category				
Always	Therapy	Dance	Running	Glamour	Flowers_tropical	Hairstyling	Computer	Youth	Beach	Pool
AVEENO	Streams	Frosty	Foggy_landscape	Diving		Abstract_art	Water_birds	Misty		
AVON	Seats	Cutlery	Healthy_cooking	Dining		Water	American_flag	Car	Aeronautics	Erotic
AXE	Urban_youth	Bodybuilding	Band	Ball_sports	Suit	Water	Military	Family	American_flag	Prey_birds
Bath & Body Works	Therapy	Fruits	Aeronautics	Juice	Water	Cat	Running	Music_festival	Sports	Bodybuilding
Caress	Streams	Abstract_art	Water			Steel	Carnival	Desert	Flowers_romantic	Juice
Chanel beauty	Light	Mountainiring	Beauty_salon	Science		Countryside	Baby	Autumn	Bedroom	Dogs
Charmin	Bedroom	Water	Cat	Delicate_fablic	Snowscape	Horror	Mountain	Baby	Child	Therapy
Clean & Clear	Happy_Nature	American_flag				Theater	Misty	Cottage	Dining	Autumn
Clinique	Glamour	Hairstyling	Eye	Painting	Arts_and_Crafts	Horror	House			
Colgate	Water	Family	Herbs	Child	Dental	Flowers_romantic	Fruits	Water_sports	Running	Music_festival
CoverGirl	Holiday_party	Music_festival	Alcoholic_drinks	Wedding	Religion	Streams	Fashion	Glamour	Hairstyling	Military
Crest	Bathroom	Steel	Healthy_cooking	Dental	Abstract_art	Prey_birds	Candy	Construction	Sparkling	Industry
Degree	Running	Sports	Fitness	Ball_sports	Water	Hairstyling	Glamour	Prey_birds	Wedding	Abstract_art
Dial Soap	Water	Ocean	Streams	Popstar	Clothing	Glamour	Fashion	Wedding	Jewelry	Suit
Dove	Flowers_romantic	Streams	Erotic	Water	Bathroom	Fruits	Produce	Sports	Old_town	Business_school
Garnier Fructis	Streams	Diving	Religion	Aquarium		Cat	Family	Prey_birds	Furniture	Youth
Gillette	Suit	Sailing	Water_sports	Modern_building	Bodybuilding	Flowers_romantic	Fruits	Holiday_party	Rally	Ball_sports
Head & Shoulders	Juice	Herbs	Aeronautics	Snowscape	City_twilight	Rainstorm	Fruits	Wheat	Produce	Popstar
Herbal Essence	Water_birds	Autumn	Prey_birds	Fowl	Wildlife	Bathroom	Youth	Bedroom	Family	Child
Irish Spring	Streams	Water	Mountain	Countryside	Rain	Hairstyling	Glamour	Lingerie	Cat	Bedroom
Jergens						Streams	Youth	Pool	Misty	Happy_Nature
Kleenex	Rainstorm	Photography	Candy			Glamour	Flowers_romantic	Fashion	Lingerie	Flowers_botanical
Kotex	Fashion	Business_school	Baby	Warm_fabrics	Wheat	Streams	Alcoholic_drinks	Pool	Industry	Erotic
Loreal	Hairstyling	Glamour	Horror			Water	Countryside	Flowers_tropical	Rain	Flowers_romantic
Mary Kay	Glamour	Curved_lines	Fashion	Symbol	Sparkling	Fruits	Countryside	Mountain	Sailing	Misty
Maybelline	Glamour	Eye	Hairstyling	Curved_lines	Candy	Frosty	Countryside	Mountainiring		
Neutrogena	Baby	Ocean	Foggy_landscape	Hairstyling	Urban_youth	Cat	Steel	Herbs	Military	Religion
Nivea	Aeronautics	Street_art	Candle	Clock		Rainstorm	Geometric	Sailing	Abstract_art	Warm_fabrics
Olay	Flowers_romantic	Flowers_botanical	Flowers_tropical	Insects	Adverture_quest	Running	Erotic	Bodybuilding	Fitness	Construction
Old Spice	Bodybuilding	Heavy_vehicle	Running	Sports	Military	Flowers_romantic	Water	Fashion	Prey_birds	Delicate_fablic
Pantene	Beach	Autumn	Ocean	Lingerie	Kitchen	Therapy	Horror	Baby	Geometric	Train
ProActiv	Eye	Metalwork	Bathroom	Computer	Cartoon	Car	Theater	Pollution	Symbol	
Revlon	Fashion	Glamour	Flowers_botanical	Modern_building	Street_art	Streams	Rain	Mountain	Mountainiring	Aeronautics
Scott Tissue	Cat	Frosty	Prey_birds	Child	Animals	Water	Clothing	Glamour	Flowers_botanical	Jewelry
Secret	Flowers_romantic	Cottage	Baking	Golf	Healthy_cooking	Ocean	Finance	Rally	Photography	Dining
Sephora	Hairstyling	Fruits	Flowers_romantic	Eye	Lingerie	Power_energy	Juice	Herbs	Industry	American_flag
Suave	Band					Fashion	Bedroom	Furniture	Delicate_fablic	Geometric
Tampax	Beauty_salon	Modern_building				Streams	Holiday_party	Light	American_flag	Alcoholic_drinks
Tresemme	Hairstyling	Fashion	Glamour	Seats	Diving	Fruits	Rain	Streams	Foggy_landscape	Autumn

Brand favorability. Next, we identify desirable and undesirable associations in each category. Recall that after submitting each brand collage, the respondent was asked a series of questions about the brand. One of the survey items was to rate the brand on being "High Quality" – 1 being the lowest quality and 5 being the highest. We regress this rating on the associations extracted from the collage. One collage is a data point, and we run the regressions on collages for brands separately for each category. The results are presented in Web Appendix E.

For example, for cars, the associations *alcoholic drink*, *cityscapes, house, fashion* and *suit* have a positive and significant coefficient, that is, they occur more frequently in collages for which the respondent rates the brand is higher quality. The associations *music festival, healthy cooking, breakfast, rain, dance, and ruin* have negative coefficients. Interestingly, while certain associations, such as *ruin* have either a negative or non-significant coefficient for all categories, some associations have opposite signs in some categories. For example, while *breakfast* and *healthy cooking* is negative for Cars, both are positive in Food and Dining. Negative associations for Food and Dining include *pollution, traffic, industry, vehicle, finance, computer,* and *ruin.* The association *beach* is positive for Food and Dining but negative for Beverages. The association *house* is positive for Cars but negative for Beverages.

Because the High Quality characteristic is a vertical dimension, i.e. one on which all brands would prefer to be rated highly, we conducted this analysis at the category level. Indeed, one would expect positive and negative associations to be specific to a product category. Next, we look at more horizontal brand characteristics: those which some brands want to have and others don't. For example, while some brands would like to be perceived as sincere and down-to-earth, others may aim to be perceived as glamorous or sophisticated. We expect that such characteristics transcend the product category, and conduct the analysis on all brands from all categories.

Brand Associations and Brand Personality and Equity Characteristics. We explore the relationship of the brand associations extracted from the collages with the frequently used brand characteristics: Aaker's brand-personality characteristics (Aaker 1997) and Young and Rubicam's BAV equity characteristics (Lovett et al. 2014). Understanding the relationships between specific brand associations with brand dimensions of personality and equity can assist brand managers in cultivating and using the visual representation that will support the desired personality and equity characteristics for the brand. For example, what associations should a manager develop with her brand in order to make the brand more down-to-earth?

Recall that each respondent, after completing the collage for a brand, was asked to score the brand on the items of the brand-personality and brand-equity characteristics. Altogether, the respondents rated the brand on 49 characteristics, a unified set of the Aaker brand-personality traits and the BAV brand-equity pillars (some of the characteristics occur in both). In order to measure relationships between these characteristics and our identified brand associations, we regress these ratings on the corresponding collage's distribution of topics (associations).

Specifically, let I be the set of collages, K be the set of associations, and S be the set of brand characteristics, rated by each respondent on a 1-5 scale. There is a total of 49 characteristics in the survey, that is $|S| = 49$. Let y_{is} be the rating on characteristic s after collage k. We run the following regression for each characteristic:

$$y_{is} = \alpha_s + \beta_{sk}\theta_{ik} + \varepsilon_{is}\, , i = 1, \dots, |I|, k = 1, \dots, |K|.$$

We note that the resulting coefficients β_{sk} represent to what extent the topic k occurs more/less in collages in which the brand is rated higher on characteristic s. Recall that we have 3,937 observations and 150 regressors for each regression.

Tables 5 and 6 present the significant associations with the five highest positive coefficients and the five with the most negative coefficients, for each of the items used to construct the brand personality characteristics (Table 5) and the equity pillars (Table 6). Web Appendix F presents the full table. For example, the personality trait Glamorous (which is part of the Upper Class facet in the Sophistication personality factor of Aaker's scheme) is positively associated with associations such as *wedding*, *eye*, *fashion*, and *glamour*, and least associated with *heavy vehicles*, *construction* and *patriotism* (see Web Appendix B for the complete cluster associations) meaning that brands that are ranked high on Glamorous have fewer collages containing these associations. Recall that all with the Guided LDA, all tags appear in all topics. What determines the nature of the topic is the probability of the tag to belong to the topic. Therefore, the tag "face" can be part of more than one association.

The personality trait Rugged (which is part of the Ruggedness factor in Aaker’s scheme) is positively associated with the associations of heavy vehicle, military, bicycle, industry and dessert, and highly negatively associated with *therapy*, *church*, *candy*, *arts and crafts*, and *sparkling* (see Web Appendix B for the complete tags related to each association). The equity characteristic Innovative, which is part of the Differentiation BAV equity pillars, is correlated with high frequency of associations such as *hand*, *religion*, *painting*, *cityscape*, and *light*, and negatively correlated with *patriotism*, *chest*, *ruin*, *symbol* and *cowboy*, meaning that brands that score high on innovative will contain fewer visuals of these associations in their collages.

The results presented in Tables 5 and 6 and Web Appendix F demonstrate that brand personality and equity traits systematically relate to particular associations. Mapping brands in this very rich, unstructured space of visual content reveals that the meaning of certain visual content is systematically related to established brand measures.

Table 5: The associations with strongest positive and negative coefficients associated with each personality characteristic. The positive are arranged in decreasing order, and the negative are arranged in increasing order (from the most negative to the least negative). *N*=3937.

	Most associated with					**Least associated with**				
Sincerity										
Cheerful	Happy nature***	Pool***	Beach***	Curved lines***	Hand***	Ruin***	Suit***	Photography***	Metalwork***	Pollution***
Down to earth	Happy nature***	Farm***	Dental***	Herbs**	Heavy vehicle***	Suit***	Modern building***	Cityscape***	Finance***	Aquarium*
Family oriented	Child***	Chest***	Happy nature***	Joy***	Seats***	Aquarium***	Suit***	Music festival***	Horror***	Auto racing***
Friendly	Happy nature***	Painting***	Joy***	Herbs**	Child***	Suit***	Aquarium***	Metalwork***	Photography***	Pollution***
Honest	Herbs**	Hand***	Happy nature**	Religion***	Baby***	Aquarium**	Theater**	Pollution***	Suit***	Rally***
Original	Hand***	Herbs*	Painting**	Vehicle*	Light***	Patriotism**	Metalwork***	Seats**	Urban youth**	Business school***
Real	Herbs**	Hand***	Happy nature*	Religion***	Snows cape**	Suit***	Fitness**	Pollution**	Candy**	Finance***
Sentimental	Hand***	Winter**	Child***	Coffee***	Baby***	Casino***	Suit***	Fitness***	Adventure quest**	Modern building***
Sincere	Winter**	Hand***	Herbs*	Happy nature**	Painting**	Aquarium***	Adventure quest***	Suit***	Pollution***	Symbol***
Small town	Farm***	Heavy vehicle***	Chest***	Bicycle***	Wheat***	Eyewear***	Church**	Abstract art***	Suit***	Modern building***
Wholesome	Herbs***	Baby***	Winter**	Happy nature**	Insects***	Suit***	Music festival***	Symbol***	Pollution***	Modern building***
Excitement										
Contemporary	Painting***	Casino**	Eye**	Hand**	Therapy*	Ruin***	Seats***	Heavy vehicle***	Metalwork***	House***
Cool	Music festival***	Adventure quest**	Street art***	Curved lines***	Beach***	Ruin***	Seats***	Chest***	Symbol***	Heavy vehicle***
Daring	Auto racing***	Street art***	Wildlife***	Painting***	Religion***	Seats***	Baby***	Ruin***	Child***	Chest**
Exciting	Music festival***	Hand***	Curved lines***	Light***	Beach***	Seats***	Ruin***	Patriotism***	Metalwork***	Hospital***
Imaginative	Painting***	Curved lines***	Hand***	Raindrop***	Light***	Seats***	Patriotism***	Ruin***	Suit***	Heavy vehicle***
Independent	Religion***	Old town***	Curved lines**	Hand**	Light***	Animals***	Symbol***	Candy***	Pollution**	Seats**
Spirited	Painting***	Curved lines***	Music festival***	Pool***	Hand***	Seats***	Theater**	Ruin***	Suit***	Computer***
Trendy	Painting***	Raindrop***	Music festival***	Curved lines***	Eye***	Seats***	Heavy vehicle***	Farm***	Patriotism**	Ruin***
Unique	Religion***	Hand***	Light***	Eye**	Painting*	Patriotism***	Seats***	Symbol***	Metalwork***	Joy***
Upper class	Wedding***	Hand***	Cityscape***	Therapy***	Old town***	Seats***	Patriotism***	Joy***	Symbol***	Heavy vehicle***
young	Dance***	Beauty salon***	Beach***	Street art***	Happy Nature***	Seats***	Theater***	Cowboy***	Patriotism***	House***

Table 5: Continued.

	Most associated with					Least associated with				
Competence										
Confident	Painting***	Golf*	Herbs**	Church*	Hand**	Symbol***	Rally***	Ruin***	School***	Seats*
Corporate	Science***	Adventure quest***	Vehicle***	Winter**	Pollution***	Happy nature***	Flowers tropical***	Water sports***	Carnival***	Joy***
Hard Working	Industry***	Herbs***	Heavy vehicle***	Military***	Bicycle**	Aquarium*	Wedding***	Music festival***	Candy***	Casino*
Intelligent	Hand***	Painting**	Raindrop**	Religion***	Astronomy***	Pollution***	Symbol***	Joy***	Bicycle**	Arts and crafts***
Leader	Hand***	Military***	Light***	Religion**	Wildlife**	Theater*	Symbol***	Ruin***	Bicycle*	Pollution**
Reliable	Happy nature***	Winter*	House***	Baby***	Dogs***	Aquarium***	Theater**	Pollution***	Rally***	Symbol***
Secure	Herbs**	Religion***	Winter*	Military***	Hand**	Aquarium**	Rally***	Candy***	Pollution**	Fowl***
Successful	Painting**	Herbs*	Happy nature*	Sparkling**	Religion**	Pollution***	Ruin***	Heavy vehicle***	Seats**	Symbol**
Technical	Vehicle***	Aquarium*	Wildlife***	Train***	Steel***	Seats***	Happy nature***	Pool***	Youth***	Delicate fabric***
Sophistication										
Charming	Hand***	Wedding***	Herbs*	Painting**	Eye**	Heavy vehicle***	Pollution***	Industry***	Metalwork***	Seats***
Feminine	Wedding***	Therapy***	Hairstyling***	Glamour***	Flowers romantic***	Patriotism***	Auto racing***	Industry***	Heavy vehicle***	Bicycle***
Glamorous	Wedding***	Eye***	Fashion***	Glamour***	Old town***	Joy***	Seats***	Heavy vehicle***	Construction***	Patriotism**
Good looking	Hand***	Wedding***	Therapy***	Painting***	Raindrop***	Seats***	Casino***	Heavy vehicle***	Symbol***	Ruin***
Smooth	Winter***	Therapy***	Wedding***	Raindrop***	Delicate fabric***	Pollution***	Heavy vehicle***	Seats***	Rally***	Symbol***
Up to date	Painting***	Curved lines***	Hand***	Casino*	Religion***	Seats***	Theater**	Ruin***	Bicycle**	Cowboy***
Ruggedness										
Masculine	Bicycle***	Military***	Heavy vehicle***	Auto racing***	Photography***	Happy nature***	Therapy***	Seats***	Delicate fabric***	Arts and crafts***
Outdoorsy	Bicycle***	Water sports***	Desert***	Autumn***	Farm***	Church***	Candle***	Symbol***	Horror***	Suit***
Rugged	Heavy vehicle***	Military***	Bicycle***	Industry***	Desert***	Therapy***	Church**	Candy***	Arts and Crafts***	Sparkling***
Tough	Military***	Industry***	Heavy vehicle***	Equiade***	Photography***	Sparkling***	Holiday party***	Youth***	Arts and crafts***	Candy***
Western	Patriotism***	Equiade***	American flag***	Farm***	Heavy vehicle***	Flowers tropical***	Church***	Candle***	Adventure quest**	Abstract art***

Table 6: The associations with the strongest positive and negative coefficients associated with each brand equity pillar. The positive are arranged in decreasing order, and the negative are arranged in increasing order (from the most negative to the least negative). *N*=3937.

	Most associated with					**Least associated with**				
	BAV Brand Equity									
Esteem										
High Quality	Therapy***	Happy nature***	Hand***	Wedding**	Religion**	Symbol***	Pollution***	Casino***	Seats***	Patriotism**
Regard	Herbs***	Happy nature***	Winter*	Therapy**	Coffee***	Pollution***	Theater***	Ruin***	Symbol***	Adventure quest**
Differentiation										
Different	Hand***	Auto racing***	Curved lines***	Light***	Vehicle**	Seats***	Patriotism**	Symbol***	Metalwork***	Joy***
Distinctive	Light***	Hand***	Sparkling**	Power energy**	Curved lines**	Metalwork***	Patriotism**	Seats***	Ruin***	Construction***
Dynamic	Painting***	Golf**	Curved lines***	Hand***	Light***	Patriotism***	Seats***	Ruin***	Metalwork***	Heavy vehicle***
Innovative	Hand***	Religion***	Painting**	Cityscape***	Light***	Patriotism***	Chest***	Ruin***	Symbol***	Cowboy***
Relevance										
Relevance	Herbs**	Hand***	Flowers botanical***	Baby***	Therapy*	Theater***	Suit***	Adventure quest**	Pollution***	Symbol**
Knowledge										
Familiarity	Herbs**	Hand***	Flowers botanical***	Baby***	Therapy*	Theater***	Suit***	Adventure quest**	Pollution***	Symbol**

Similarities between the associations of brands. Our association elicitation method enables measuring the similarity of the association between brands. We calculated the cosine similarity between the normalized (sum of squares is equal to 1) association distribution vectors of all the pairs of brand in our sample. Recall that cosine similarity is a way to compare two vectors, by calculating the angle they create. The number ranges from 0 to 1, where 1 indicates identical vectors. Web Appendix G describes the similarity matrix. This is a symmetric 303X303 matrix, with the diagonal equals to 1, whose value indicate how brands are perceived similar (or different) in the association space. For example, we see that the brand *Cheesecake Factory* is highly similar in associations to the baking appliances brand *Kitchen Aid* (cosine similarity of 0.84); The family dining chain *Golden Corral* is very similar in its associations to the supermarket cooked food brand *Hormel* (cosine similarity of 0.91); *Barnes and Noble* has similar associations to the pain drug brand *Aleve* (cosine similarity of 0.7); and, nicely, *Febreze* had a 0.7 cosine similarity to *Ashley Furniture*. This similarity can be an indication for the potential of brand alliance, cross category perceptual maps, and positioning inquiries.

Discussion

In this paper, we propose and implement a novel brand-association-elicitation tool (which we term B-VEP). The elicitation task allows participants to portray their relationship with the brand through a collage of photographs. Visual images have the advantage of being better reflecting the emotions, cultural experiences and attitudes that constitute the consumer associations, as opposed to verbal methods that focus more on the discourse of these experiences (Reavey 2012). Use of images has been demonstrated to successfully act to disrupt well-rehearsed narratives of people revealing hidden, unarticulated ideas. The analysis uses unsupervised machine learning methods to avoid "strangling" the data: rather than looking for specific pre-defined associations, we let the data speak and identify associations using topic modeling. The resulting set of associations is rich, and spans a variety of objects, occupations, nature, constructs, and institutions, just to name a few.

Using this tool, we gathered a large set of consumer brand perceptions on 303 brands. We applied it to explore several important questions for brand management: creating mood boards for each brand, consisting of a collection of photographs that capture the distribution of consumers' associations with the brand; finding unique associations, on which the brand differs from others in its product category; identifying favorable and unfavorable associations for each category; testing which associations are related to commonly used brand metrics, such as brand personality and brand equity; and, finally, measuring association-based similarities between brands from different categories, which may identify potential for brand alliances or strategic partnerships.

We see these applications as just scratching the surface of the potential of using visual elicitation. We hope that future research will build on this work in other directions. For example, one future direction might be identifying brand extension strategies. *Starbucks*' stronger association is *baking* (see Web Appendix C), and *dining* is also one of the top 5. While *Starbucks* does offer food and baked goods, this association might imply a need for more dining choices that can be further explored (interestingly, *Dunkin Donuts*, which by definition offers baked goods, has much weaker associations of *baking* and *dining*). In the Beauty Products category (Table 1), *Clinique's* strongest association is Hairstyling (hair, face, glamour, studio, nude, lips, skin). However, their product line hardly includes hair products. These insights can be a starting point for exploring further directions by brand managers.

Another potential avenue for future work is to identify systematic relationships between perceptual dimensions and elements of visual design, such as shapes, colors, texture, etc. While modern visual design provides many guidelines on how these elements can be used in a composition to create a certain perception, few of these are empirically tested on brand-related imagery.

An interesting theoretical question related to underlying psychological mechanisms is the evolvement of brand associations and their relationships with brand characteristics (Torres and Bijmolt 2009). On one hand, one could argue that consumers think about brands in terms of characteristics such as personality and equity, and then create in their minds images to represent these characteristics (e.g., they perceive the brand as innovative, and the concept of innovativeness evokes metaphors such as transistors, and therefore they associate the brand with visuals containing transistors). On the other hand, one could think of the brand as evoking sets of metaphors, and the characteristics of these metaphors reflect, in turn, the way consumers perceive the brand (e.g., the brand evokes the association of a transistor, transistors are perceived as innovative, which forms, among other things, the innovative perception of the brand). B-VEP can help to address this question through tasks such as collage building of synthetic brands with predefined controlled characteristics, or creating collages describing characteristics (e.g. innovative) and test their similarity to associations of brands.

Our tool can be useful to explore heterogeneity among consumers' brand perceptions. By collecting a large number of collages per brand, we can learn how individual differences in personality, values, lifestyle, and other variables of interest, influence the brand perception. Insights from such studies can be useful for segmentation, optimizing marketing communications, and creating a better fit between brands and their consumers.

In sum, modern software and image-processing tools open many new opportunities for marketing researchers. B-VEP allows researchers and firms to gather and harvest visual brand-related data directly from consumers, which complements existing brand metrics, as well as the rapidly growing field of visual social media monitoring.

References

Aaker, Jennifer L. (1997), "Dimensions of brand personality," *Journal of Marketing Research*, 34 (3), 347-356.

Catchings-Castello, Gwendolyn (2000), "The ZMET alternative," *Marketing Research* 12 (2) 6.

Cho, Hyejeung, Norbert Schwarz, and Hyunjin Song (2008), "Images and Preferences A Feelings-As-Information Analysis," in *Visual Marketing*, Lawrence Erlbaum Associates, New York.

Culotta, Aron, and Jennifer Cutler (2016), "Mining brand perceptions from twitter social networks," *Marketing Science*, 35 (3), 343-362.

Davis, D., and Butler-Kisber, L. (1999), "Arts-based representation in qualitative research: Collage as a contextualizing analytic strategy," Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada).

Gelper, Sarah, Renana Peres, and Jehoshua Eliashberg (2018), "Talk Bursts: The Role of Spikes in Pre-release Word-of-Mouth Dynamics," *Journal of Marketing Research*, Forthcoming.

Greenleaf, Eric, and Priya Raghubir (2008), "Geometry in the Marketplace," in *Visual Marketing*, Lawrence Erlbaum Associates, New York.

Han, Young Jee, Joseph C. Nunes, and Xavier Drèze (2010), "Signaling status with luxury goods: The role of brand prominence," *Journal of marketing*, 74 (4), 15-30.

Howard, Andrew G (2013), "Some improvements on deep convolutional neural network based image classification," arXiv preprint arXiv:1312.5402 (2013).

Jalali, Nima Y., and Purushottam Papatla (2016), "The palette that stands out: Color compositions of online curated visual UGC that attracts higher consumer interaction," *Quantitative Marketing and Economics*, 14 (4), 353-384.

Jagarlamudi, Jagadeesh, Hal Daumé III, and Raghavendra Udupa (2012), "Incorporating lexical priors into topic models," In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 204-213.

John, Deborah Roedder, Barbara Loken, Kyeongheui Kim, and Alokparna Basu Monga (2006), "Brand concept maps: A methodology for identifying brand association networks," *Journal of Marketing Research* 43 (4), 549-563.

Johnson, Jeff (2013), *Designing With the MindiIn Mind: Simple Guide to Understanding User Interface Design Guidelines*. Elsevier, 2013, New York.

Keller, Kevin Lane (1993), "Conceptualizing, measuring, and managing customer-based brand equity," *Journal of marketing*, 57 (1), 1-22.

Koll, Oliver, Sylvia Von Wallpach, and Maria Kreuzer (2010), "Multi-method research on consumer–brand associations: Comparing free associations, storytelling, and collages," *Psychology and Marketing*, 27 (6), 584-602.

Kress, Gunther R., and Theo Van Leeuwen (1996), *Reading images: The grammar of visual design*, Routledge, New York.

Labrecque, Lauren I (2014), "Fostering consumer–brand relationships in social media environments: The role of parasocial interaction," *Journal of Interactive Marketing*, 28 (2), 134-148.

Lee, Thomas Y., and Eric T. Bradlow (2011), "Automated marketing research using online customer reviews," *Journal of Marketing Research*, 48 (5), 881-894.

Liu Liu, Daria Dzyabura, Natalie V. Mizik (2019), "Visual Listening In: Extracting Brand Image Portrayed on Social Media," forthcoming, *Marketing Science*.

Lovett, Mitchell, Renana Peres, and Ron Shachar (2013), "On brands and word-of-mouth," *Journal of Marketing Research*, 50 (4), 427-444.

Lovett, Mitchell, Peres, Renana, and Shachar. Ron (2014), "A dataset of brands and their characteristics," Marketing Science, 33 (4), 609-617.

McQuarrie, Edward F. (2008), "Differentiating the Pictorial Element in Advertising – A Rhetorical Perspective," in Visual Marketing, Lawrence Erlbaum Associates, New York.

Meyers-Levy Joan, and Rui Zhu (2008), " Perhaps the store made you purchase it: toward an understanding of structural aspects of indoor shopping environments, in in Visual Marketing, Lawrence Erlbaum Associates, New York.

Mizik, Natalie, and Jacobson, Robert (2008), "The financial value impact of perceptual brand attributes," *Journal of Marketing Research*, 45 (1), 15-32.

Nam, Hyoryung, Yogesh V. Joshi, and P. K. Kannan (2017), "Harvesting brand information from social tags," *Journal of Marketing*, 81 (4), 88-108.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), "Mine your own business: Market-structure surveillance through text mining," *Marketing Science*, 31 (3), 521-543.

Palmer, Stephen E. (1999), *Vision science: Photons to phenomenology*, Cambridge, MA: The MIT Press.

Pavlov, Eugene and Natalie Mizik (2018), "Increasing consumer engagement with firm-generated social media content: The role of images and words," Working Paper.

Pennington, Jeffrey, Richard Socher, and Christopher, D Manning. (2014), "Glove: Global vectors for word representation," *Proceedings of the Empirical Methods in Natural Language Processing* (EMNLP 2014) 12.

Pieters, Rik, Edward Rosbergen, and Michel Wedel (1999), "Visual attention to repeated print advertising: A test of scanpath theory," *Journal of marketing research*, 36 (4), 424-438.

Rangel, José Carlos, Miguel Cazorla, Ismael García-Varea, Jesus Martínez-Gómez, Élisa Fromont, and Marc Sebban (2015), "Scene classification based on semantic labeling," *Advanced Robotics*, 30 (11-12), 758-769.

Rayner, Keith, Brett Miller, and Caren M. Rotello (2008), "Eye movements when looking at print advertisements: The goal of the viewer matters," *Applied Cognitive Psychology*, 22 (5), 697-707.

Reavey, Paula (2012), *Visual methods in psychology: Using and interpreting images in qualitative research*, Chapter 1. Reavey, Paula, ed. Routledge, 2012.

Seidman, Gwendolyn (2013), "Self-presentation and belonging on Facebook: How personality influences social media use and motivations," *Personality and Individual Differences*, 54 (3), 402-407.

Steenkamp, Jan-Benedict, and Hans Van Trijp (1997), "Attribute elicitation in marketing research: a comparison of three procedures," *Marketing Letters*, 8 (2), 153-165.

Tirunillai, Seshadri, and Gerard J. Tellis (2014),"Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation." *Journal of* Marketing Research, 51 (4), 463-479.

Torres, Anna and Tammo H. Bijmolt (2009). Assessing brand image through communalities and asymmetries in brand-to-attribute and attribute-to-brand associations. European Journal of Operational Research, 195(2), 628-640.

Wedel, Michel, and Rik Pieters (2000), "Eye fixations on advertisements and memory for brands: A model and findings," *Marketing Science*, 19 (4), 297–312.

Wedel, Michel, and Rik Pieters (2008), "Introduction to visual marketing," in *Visual Marketing*, Lawrence Erlbaum Associates, New York.

Zaltman, Gerald, and Robin Higie Coulter (1995), "Seeing the voice of the customer: Metaphor-based advertising research," *Journal of Advertising Research*, 35 (4), 35-51.

Zaltman, Gerald and Zaltman, Lindsay H. (2008), *Marketing Metaphoria, What Deep Metaphors Reveal about the Minds*. Harvard Business Press, Boston Massachusetts.

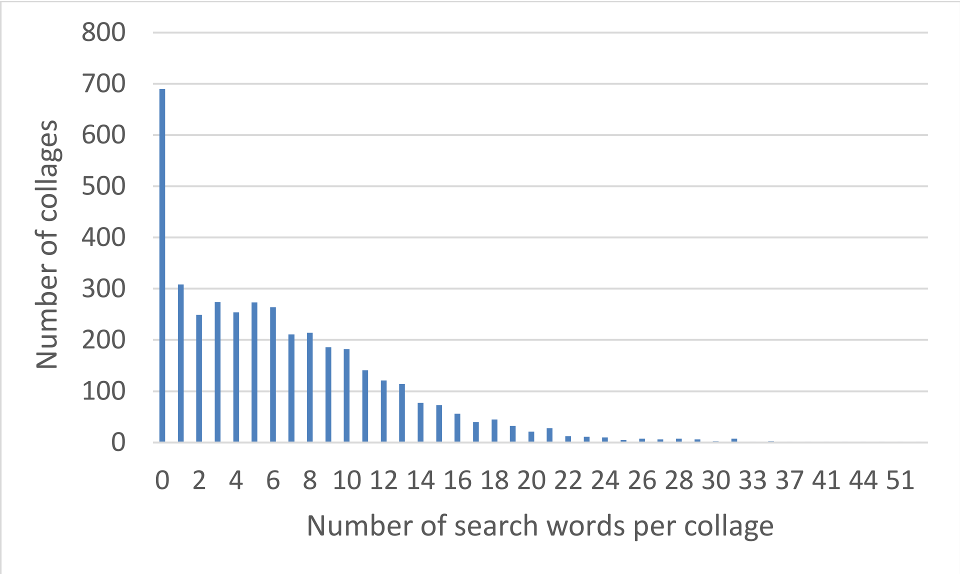
Zhang, Shunyuan, Dokyun Lee, Param Vir Singh, Kannan Srinivasan (2017), " How Much Is an Image Worth? Airbnb Property Demand Estimation Leveraging Large Scale Image Analytics," Working Paper.

Appendix 1 – The usage of search words

Participants select the images for their collages from a large photo repository (the right hand of Figure 3 in the paper). To help them browse through the repository, we implemented the option to search for keywords. Participants were not requested to use the search option, and it was implemented as an aid to help users navigate through the very large image repository. However, since the search is using words, one may wonder whether this usage of words undermined B-VEP's main focus as a visual elicitation tool. To verify that this is not the case, we have conducted the following tests and measurements:

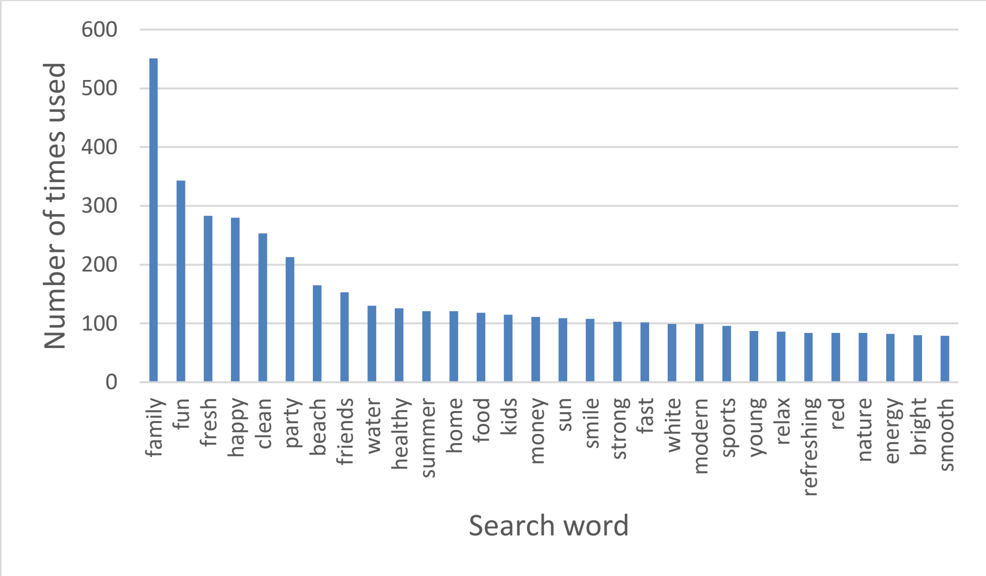
1. ***Usage of the search keywords is infrequent***– Figure A1.1 below displays the distribution of search words per collage - out of our 3,937 approved collages, 690 (17.52%) did not uses any search word. The median number of search terms used in a collage is 5 indicating that half the collages used 5 search terms or less. The average number of search terms per collage is 6.41.

Figure A1.1: The distribution of search terms used in a collage across collages



2. ***Search words are repetitive*** – The search words that used by participants are limited and are often repetitive. Out of the 25,262 search terms (consisting of 28,505 separate words) used by respondents, only 6,475 were unique (21.3%). The top 30 words (0.5% out of the number of unique words) are responsible for 4,465 searches (17% out of the total number of searches). Figure A1.2 shows a histogram of the common search terms. Using a small number of search terms over all the brands indicates that the search words, at least as used for this data collection, are not brand specific and might have a limited power in providing a unique brand association.

Figure A1.2: The top 30 search terms over all the collages



3. ***Each Search retrieves a large number of photos*** –The photo repository has ~100,000 photos, each of them is labeled by Flickr users with dozens of labels. Therefore, each search word retrieves multiple photos, from which the user needs to keep scrolling to choose the most appropriate one. For example, the search word "family" retrieves 318 photos, "nature" retrieves 3,621 photos, "child" retrieves 629 photos, and "happy" retrieves 210 photos. In Figure 3 in the paper, the participant searched for the word "laptop" and retrieved 46 Flickr photos labeled with "laptop." The participant could have chosen a laptop with people sitting next to it, children playing a laptop in a restaurant, a laptop in an office, a school etc. The chosen picture contains many additional visual items which, we believe, reflect additional feelings, attitudes and associations, the user had for the brand which are not part, and might not be even related to the original search word "laptop." Therefore, the search can be viewed as an aid in the browsing, but not one that limits or constraints it.

4. ***Users rarely use the* "*banned*" *words*** – In order to encourage users to elicit rich associations and avoid obvious collages, users were generally directed to "not choose pictures that show the brand logo (or a logo of any other brand), type of product, or product category." If they did during the search they received an error message. Out of the 25,262 search terms used, only 1,111 (less than 5%) attempts were made to use the "banned" words. Interestingly, despite this restriction, the collages of users still managed to capture the functionality of the brand. In Table 2 in the paper, we present the common associations per category, and indeed, the category information is still present and significant at the collages. For example – Beauty Products are associated with *flowers*, *water*, *hairstyling* and *glamour*, and Cars with *traffic*, *car*, *cityscape*, *finance* and *steel*. Hence, it seems that our restriction is not a significant factor, which interferes with the flow of the collage making. It rather manages to help respondents to create rich collages, while still keeping the essence of the product.

5. ***Search word usage does not impact the collage –*** Although, as we showed above, respondents do not use the search words much, we wanted to verify whether collages that used more search words generate different associations than collages that used them less. To do so, we carried out the following procedure:
   a. ***Split*** **-** the collages of each brand into two equal groups – the collages which used a below median number of search terms (1,968 collages, average of 2.9 search terms per

collage), and collages that used above median number of search terms (1,968 collages, average of 11.6 search words per collage).

b. ***Elicit Associations* -** We applied our association elicitation method (the Guided LDA topic extraction) on each of these two groups and extracted associations.

c. ***Test Similarity* –** For each brand, we calculated the cosine similarity between the normalized (sum of squares is equal to 1) association distribution vectors of the above median group and the below median group. Recall that cosine similarity is a way to compare two vectors, by calculating the size of angle they create. The number ranges from 0 to 1, where 1 indicates identical vectors.

d. ***Estimate similarity relative to any random partition* –** We compared the cosine similarity of the association vectors of the above-below media groups, to the cosine similarity values obtained by 100 other random equal partitions of the brand collages. That is, if a brand has *n* collages, they form $\binom{n}{n/2}$ partitions of size n/2. We sampled 100 of these partitions for each brand, calculated the cosine similarity of their association vectors, and checked what percentile does the above-below similarity falls into. If, indeed, it is equivalent to any of the other partitions, the percentile should fall in the range 0-1 in a uniform distribution. Figure A1.3 below presents the percentiles for our 303 brands. Indeed, the distribution is no significantly different from Uniform (Chi-square p value of 0.12).

Figure A1.3: The ordered percentiles, for all brands, of the cosine similarity LDA topic distribution vector of the above-below median partition within 100 other equal size random partitions.