Working Paper
Last update: June 19, 2019

Letting Logos Speak:

Leveraging Multiview Representation Learning for Data-Driven Logo Design

Ryan Dew*
The Wharton School
University of Pennsylvania

Asim Ansari
Columbia Business School
Columbia University

Olivier Toubia
Columbia Business School
Columbia University

June 19, 2019

Logos serve a fundamental role in branding as the visual figurehead of the brand. Yet, due to the difficulty of using unstructured image data, prior research on logo design has been largely limited to non-quantitative studies. In this work, we explore logo design from a data-driven perspective. In particular, we aim to answer several key questions: first, to what degree can logos represent a brand's personality? Second, what are the key visual elements in logos that elicit brand and firm relevant associations, such as brand personality traits? Finally, given text describing a firm's brand or function, can we suggest features of a logo that elicit the firm's desired image? To answer these questions, we develop a novel logo feature extraction algorithm, that uses modern image processing tools to decompose unstructured pixel-level image data into meaningful visual features. We then analyze the links between firm identity and the features of logos through a deep, multiview generative model, which links visual features of logos with textual descriptions of firms and consumer ratings of brand personality by learning representations of brand identity. We apply our modeling framework on a dataset of hundreds of logos, textual descriptions from firms websites, third party descriptions of firms, and consumer evaluations of brand personality to explore these questions.

Keywords: logos, branding, machine learning, multiview learning, deep generative modeling, image processing

*Corresponding author. Email: ryandew@wharton.upenn.edu

1. Introduction

Logos are the most distinct marks of brands, adorning everything from packaging to advertising. Designers create logos to represent the essence of brands, and firms motivate brand and logo redesigns with an intention to convey a new idea. Yet, despite the overwhelming significance of logos, and the substantial costs of logo redesigns, marketing scholars have paid relatively little attention to the logo design process. In this work, we show that there is a science to the logo design process that can be captured by models, and that such models can serve as a basis for understanding the meaning conveyed by logos, as well as aid brands in the design of logos consistent with their brand identities. In particular, we synthesize novel image processing techniques with a deep, multiview representation learning framework, to capture the links between a brand's function and identity, its logo features, and how consumers perceive the brand's personality.

Our data-driven treatment of logos, together with a multiview representation learning framework, allows us to understand the branding and design process from three related perspectives:

1. **The researcher's perspective.** What types of logo features are associated with specific brand identity and brand personality traits? By considering the logo features as inputs to the model, and text and brand personality as the outputs, we can understand how different logo features contribute to consumer perceptions.

2. **The designer's perspective.** Given a description of a brand, or a desired consumer-level perception, which logo features are most commonly used to achieve that identity? This mirrors the design process, where a company-supplied brief is used as the basis for designing a logo, and relies on being able to use text and/or brand personality as an input to predict logo features.

3. **The brand manager's perspective.** Given a new logo, how will consumers perceive that logo? Or, given a set of candidate logos, which may vary on key design elements, which logo best matches a company's intended brand perception? This, again, requires being able to go from a logo as input, to a set of brand descriptors, like keywords or brand personality perceptions.

Our work makes several contributions. Foremost, it is the first paper to study logos from a holistic and quantitative perspective. This is important, first, because it adds a level of objectivity to the design process: while our model cannot replace the creative touch of designers, it can offer both designers and firms guidance in crafting their brand identities, in an objective fashion. When weighing competing designs and opinions, an objective prediction of the reactions of consumers to a logo design can allow managers to make a data-driven decision, in what has historically been viewed as a creative domain. A model-based approach lets us simultaneously assess the many facets of logo design, and work with unstructured, natural data like text. Moreover, because a model-based approach allows us to make design recommendations, it can be used even by budget-strapped firms to thoughtfully design their logos in a data-driven fashion. By basing our framework on representation learning, where unstructured data is converted to dense, numeric representations of brands in a latent space, we gain an additional design-related benefit: by interpolating between brands in the latent space, we can come up with novel combinations of existing brand identities, thus facilitating the creativity process.

From a methodological perspective, ours is among the first papers in marketing to directly use image data, without relying on human coders. Specifically, our work presents a novel approach to working with unstructured, visual data, through a theory-driven image processing approach. Our feature extraction algorithm decomposes logos into meaningful features, many of which are driven by prior theory about how logos convey meaning. The set of these features forms a "visual dictionary" which we can use to describe logos in a way that is meaningful to designers, and that is also amenable to probabilistic modeling. Working directly with image data is important for wide and general applicability of our framework, as well as for scalability: for brand managers or designers to use our model in practice, it cannot be based on the inputs of human coders.

Our work is also among the first in marketing to synthesize both unstructured text and image data. The model we develop for that purpose is called a multimodal variational autoencoder. Variational autoencoders (VAE) are popular machine learning tools for learning representations of complex data. In this work, we develop a multimodal variational autoencoder, which learns representations of brands across all of the ways in which brand is manifest: text, logo, and brand personality. The task of learning unified representations across domains is an instance of multiview or transfer learning. As we operationalize transfer learning via learned representations that are shared across domains, it is also an instance of representation learning (Li et al., 2016). For inference, we draw on the weakly supervised product-of-experts inference network approach, introduced by Wu and Goodman (2018), which learns a set of functions that can infer a brand's latent representation, given any of the modalities, and which can then be used to predict any of the other domains. For example, given a textual description of the brand, we can predict which features we expect to find in that brand's logo, and how we expect consumers to perceive the brand's personality.

The rest of the paper is organized as follows: in Section 2, we review the existing literature on logo design and aesthetics in marketing. In Section 3, we describe the unique dataset we have compiled to calibrate our model. In Section 4, we briefly describe how images are stored at the data-level, then describe our logo feature extraction algorithm. In Section 5, we present descriptive and "model-free" predictive evidence of the links between design, brand personality, and firm function. In Section 6 , we develop a multi-view learning model of brands and their logos, and in Section 7, we show the results of applying that model to our data, including examples of the learned representations, logo recommendations, and links to brand personality. Finally, we conclude with a summary of on-going research and directions for further study.

2. Literature

There is a sizable literature, especially in consumer behavior, on how consumers react to aesthetics, both in logos and in other aspects of marketing. Much of this literature describes how specific logo features lead to different consumer reactions and impressions. Other papers discuss how these reactions vary cross-culturally, or the mechanisms governing consumers' reactions to various visual stimuli. In this section, we review those findings, which we then draw on in Section 4 as the basis for our logo feature extraction algorithm.

2.1. Logos

A limited amount of research in marketing has been done specifically on firm logos, starting with Henderson and Cote (1998), where they use factor analysis on a set of logo traits, coded by experts, to come up with a set of constructs that describe logos generally: natural, harmonious, elaborate, parallel, repetition, proportion, and roundness. Of their factors, only natural, harmonious, and elaborate (from now denoted NHE) seem predictive of outcome measures generally. In Henderson et al. (2003), they test whether these constructs hold cross-culturally, finding little difference of the predictive power of NHE in Asia versus the United States. This cross-cultural work is then expanded by van der Lans et al. (2009), again using NHE, together with three "objective design measures"—repetition, proportion, and parallelism, all determined by expert coders from disparate geographies. They find the NHE dimensions are universally good descriptors of design, even cross culturally. Together, these studies support the idea that NHE provide a good proxy for design elements of logos.

Other work has looked at specific aspects of logos. Klink (2003), for example, studies the link between the brand name and the traits of the logo, finding ties between the phonetic structure of the name and the traits used in the logo, such as color and angularity. Walsh et al. (2010) find that moving from an angular logo to a round logo produces generally mixed responses in consumers, dependent on their level of commitment to the brand. The idea of circular versus angular logos is also explored in Jiang et al. (2015), where they find that the mere circularity or angularity of the logo affects perceptions of the product and the company, through perceived hardness or softness, which in turn influences attribute judgments. Other studies look at the orientation of the logo, including Cian et al. (2014), who find that different logos can evoke the idea of movement, often through the positioning of the logo elements or the horizontal orientation of the logo, which in turn affects consumers' engagement and attitudes. Even more recently, Schlosser et al. (2016) find that upward diagonals convey greater activity than downward diagonals, leading to more favorable product evaluations, greater efficacy beliefs, and greater post-consumption satisfaction. Together, these studies imply that among the objective design measures employed in a design model should be traits like color, angularity, and orientation.

Finally, there has been a significant amount of work done on typeface and font. Doyle and Bottomley (2006) provide an excellent overview and study of fonts in logos, describing both the background of typeface research, and studying specifically the appropriateness of a given typeface for describing a particular product or brand. They define appropriateness in terms of abstract connotations,[1] where abstract connotation is captured by Osgood's evaluation, potency, and activation dimensions (EPA), a set of factors that has been shown across contexts (including typeface) to capture abstract connotations. They find that congruence in EPA between the font and the product leads to more frequent choice of the product. In another study, Hagtvedt (2011) shows that incomplete typeface can lead to both perceptions of untrustworthiness and increased innovativeness. Hence, an understanding of the role of font is important.

2.2. Aesthetics

While academic work specifically on logos has been relatively limited, there is a large body of work on aesthetics and perception, some of it in marketing, especially in the domain of consumer response to advertising.

Color In marketing, Deng et al. (2010) study consumers' preferences for color combinations in product design. They have three main findings. First, of the three common dimensions of color—hue, saturation, and lightness—they find people tend to de-emphasize lightness, relative to the other two. Second, in product design, people prefer generally similar colors, but with a single contrast color, where the contrasting color is often used to highlight a single distinctive element. Finally, they find that people generally prefer a small number of colors. Kareklas et al. (2014) also explore color in marketing. They find that people exhibit an automatic preference for white over black in product choice and advertising, similar to the implicit bias observed in other studies in psychology. Relatedly, Semin and Palma (2014) find that white is perceived as more feminine, whereas black is perceived as more masculine. In psychology, more work has been done on color. For example, Valdez and Mehrabian (1994) study the effect of color on emotions, finding that of the three key color dimensions, saturation and lightness drive emotional responses along the pleasure, arousal, and dominance dimensions. They also find shades of blue, green, and purple to be the most pleasant, and shades of yellow to be the least pleasant.

Font Besides logos, font and typeface have also been explored both in the domain of advertising, and in impression management generally. Childers and Jass (2002) explore the influence of typeface on perceptions, finding that the semantic connotations of typeface can influence consumers' ratings of products. Henderson et al. (2004) take a different approach and analyze many extant fonts in an effort to summarize their impressions and design features. They come up with a set of four factors—pleasing, engaging, reassuring, and prominent—that describe typeface impressions, and six factors—elaborate, harmony, natural, flourish, weight, and compressed—that describe typeface design, based on the

[1] Abstract connotations differ from direct connotations, like, for example, a font with "snowcaps" being associated with something cold.

typology literature and ratings of experts, and conclude that there may be universal design elements that can help managers in impression management.

Orientation In an early study on advertising, Meyers-Levy and Peracchio (1992) show that the camera angle of an ad showing a product can influence consumers' judgments of the product, moderated by processing motivation. Specifically, they find that when processing motivation is low, looking up at the product yields more favorable judgments; alternatively, when processing motivation is moderate, looking at an eye-level product is best. More recently, Chae and Hoegg (2013) find that in cultures where reading is done from left to right, products are viewed more favorably when positioned congruently with this spatial orientation (and vice versa). Deng and Kahn (2016) find that the location of the product image on its packaging (top/left or bottom/right) influences the item's perceived weight (lighter or heavier respectively).

Other A host of other papers discuss other aspects of aesthetics that might be relevant for logo design. For example, Navon (1977) finds that global features are processed more readily and fully than local ones, a trait we might expect to operate also in logos. More recently, Pieters et al. (2010) use eye-tracking to study the visual complexity of advertisements. They come up with two distinct aspects of visual complexity: feature complexity and design complexity. Feature complexity simply refers to variation in basic features like color and edges, and is measured by variance at the pixel level, while design complexity refers to variation in the elaborateness of the design, and is measured by six general principles: quantity of objects, irregularity of objects (shape), dissimilarity of objects, detail of objects, asymmetry of object arrangement, and irregularity of object arrangement.

Relevant to relating brand constructs to visual elements, Orth and Malkewitz (2008) decompose package design into five distinct "types"—massive, contrasting, natural, delicate, and nondescript—and relate those types prescriptively to brand personalities. In a review article, Spence (2012) discusses cross-modal effects, including visual perceptions associated with tastes and textures (e.g. the angularity of carbonation or bitterness), which could be relevant determinants of logo design. Spence argues that firms can use these principles to set up an appropriate cross-modal expectation for a consumption experience, thereby enhancing it. This, in turn, is based off earlier work that discusses consumers preferences for congruity in the consumption experience (e.g. a fancy logo matching a fancy experience; see Patrick and Hagtvedt (2011) for an example of this kind of effect).

3. Data

In this section, we describe the dataset we have compiled of brands and their logos. Our goal is to understand both what brand-relevant concepts a given logo conveys, and how a firm can design a logo consistent with those concepts. To that end, our dataset consists of four components: logos, textual descriptions of firms from the firms' websites, industry labels, and brand personality ratings from consumers reacting to both the logo and textual description.

Our insights derive from learning the links between existing logos and these other variables; hence, for our insights to meaningfully capture good design practices, we must ensure that the firms we gather data for have given some thought to the design of their logos. We thus chose firms that were either rated as having a strong brand identity by brand specialists, or were highly profitable and recognizable, with the rationale that these firms have likely invested in their brand identity as part of their success. Specifically, we looked at all firms that were either listed in the Interbrand brand consultancy's list of Top 100 Global Brands of 2016, listed as among the top 500 most valuable American brands of 2016 by the brand valuation consultancy Brand Finance, or listed in the Forbes 500 in 2016. There was a large degree of overlap between the lists, leaving us with a final sample of 715 firms.

Logos Firms typically employ a variety of logos for different purposes. Broadly speaking, a logo may be comprised of three key features: marks, logotype, and subtext. Marks are the non-textual parts of the logo (e.g. the Apple apple, or the Nike swoosh); the logotype is the primary textual identifier, usually displaying the brand name; and the subtext is other text, often a brief descriptor of the brand. A logo always has either a mark or a logotype, while some logos have both, and some include a subtext. Some firms employ variants of their logo for different purposes, which may consist of either just the mark, or just the logotype, or the mark and logotype omitting the subtext, or a logo where the colors are inverted (e.g. blue lettering on a white background becomes white lettering on a blue background). Determining which logo to use thus requires some amount of judgment on the part of the researcher. As a rule, we selected logos with white backgrounds, if such a logo is in use. Similarly, we selected the logo with both logotype and mark, if it is in use by the firm. For other aspects of the logo, including subtext and the orientation of the mark relative to the logotype, we used the version that appeared most commonly on the firm's online marketing materials.

Text To understand which firms use which design features, we collected *web descriptions* from the firms' websites, consisting of both functional and brand-relevant text taken directly from firms' websites. We collected this data in two batches: in one, we asked Amazon mechanical turk users to find text on the firm's website that describes how the firm views its brand, and that does *not* merely describe what the firm does. We guided workers toward the About Us, Mission Statement, Corporate Values, or Investor Relations pages of firms' sites. In a second batch, we asked workers to find text that describes what the firm does, and is not identical to the text already supplied. In both cases, we gave incentives for workers to provide long descriptions.

Industry Labels In addition to the full textual descriptions, as a simpler measure for capturing what firms do, we also collected *industry labels* from the database Crunchbase. Crunchbase is commonly used by investors to learn about firms. One feature that Crunchbase captures is a set of standard tags describing what the firms do. For example, the ride-sharing company Uber has the labels Customer Service, Mobile Apps, Public Transportation, Ride Sharing, and Transportation. In total, there were 615 labels across our 715 companies. These labels are further organized into category groups, which reflect similar activities. For example, Public Transportation, Ride Sharing, and Transportation are all be categorized under the group Transportation.

Brand Personality Finally, we also collected *brand personality ratings* from consumers, following the framework of Aaker (1997), as a simple way of understanding brand impressions in the minds of consumers. Specifically, we used Amazon Mechanical Turk to elicit brand personality perceptions from U.S.-based consumers, by showing participants both the logo and the text describing the firm, and then we asking them to rate the extent to which they thought each of a set of traits describes the focal firm, based on the logo and text provided. We used the original set of 42 personality traits from Aaker (1997), as well as three reverse-coded attention check traits.[2] We gathered 20 responses per brand.

4. Logo Feature Extraction

The primary barrier to using visual data in models is the difficulty of working with unstructured image data. Many methods have been developed for incorporating images in models, with much of the literature coming from the computer vision and machine learning communities. Broadly, there are two approaches to using images in models: the first uses raw pixel-level data as the input to a probability or machine learning model. This is common, for example, in models of image recognition or image captioning, where the model is typically based on a neural network, and the task is a supervised prediction task. A second approach first processes the image, then uses the outputs from this processing as an input to the model. A common approach here is to create an image "dictionary" of representative image features. In our work, we follow the second approach: to incorporate the logos into our model of design, we first process the logo image into logo features, through a novel logo feature extraction algorithm based on modern image processing methods. Our algorithm is rooted in the literature on logo design and consumers responses to aesthetics, and distills a logo into components that are meaningful for consumers and designers. This approach facilitates understanding of our final model, as each of the inputs is human interpretable.

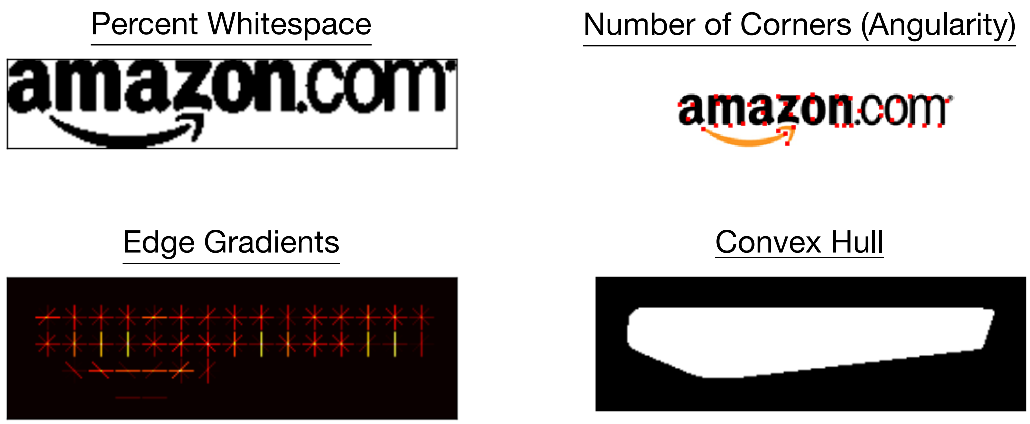


Figure 1: Examples of global features, using Amazon's logo as an example. Percent whitespace captures the percentage of pixels that are white (background), within the convex hull of the logo. The number of corners is a measure of angularity computed via the Harris corner detector. Edge gradients capture directionality of edges in the logo, and are computed by computing numerical gradients sliding over a binarized (black and white) version of the logo. The convex hull is the smallest convex polygon containing all of the non-background pixels.

4.1. Algorithm Overview

Our algorithm has three general stages: in the first stage, which we term summarization, we compute a variety of features from the logo as a whole, which we refer to as global summary features. Examples of these features are given in Figure 1, using Amazon's logo as an example. One such computation is a density-based color quantization, where we learn how many distinct colors are in each logo. In the second stage of the algorithm, which we term segmentation, we assign each pixel in the logo to one of these colors, then segment the logo into regions that are separated either by color or by background (i.e. the color white). For each of these segments, we separate them into characters and marks. This process is illustrated in Figure 2, again using Amazon's logo as an example. In the final stage, which we term tokenization, we cluster several of the features across logos, including the color, hull shape, and mark shape, to form a dictionary of logo features.

4.2. Visual Features

A comprehensive listing of all of our visual features can be found in the table in Appendix A, including descriptions on each feature, and how each feature connects to the literature outlined in Section 2. In this section, we briefly describe each logo feature, grouping them into color, format and shape, font, and other features. Note, however, that these groups are just for expositional convenience; in our analysis, we treat each feature independently.

[2]The reverse-coded traits were honest/dishonest, exciting/boring, and good-looking/ugly. Any participant who answered that both traits are descriptive of the firm was automatically removed.

Figure 2: Examples of the segmentation process, using Amazon's logo as an example. The original logo is at top. Beneath that is the segmented logo, where black identifies the background, and distinct regions are marked by different color regions. We then apply a template matching and filtering algorithm to identify which of these regions are characters (bottom-right), and assume the remainder are the marks (bottom-left).

Color The full dictionary of colors is given in Figure 3. This is computed by clustering colors across all of the logos in the dataset. Besides for just computing which colors are present in a logo, we also compute which color is the dominant color (one per logo), which colors are accent colors (all colors except the dominant color), and how much whitespace there is within the convex hull of all logo pixels. We also compute other summary statistics about color in the hue-saturation-value (HSV) color space, including the mean and standard deviation of the saturation and lightness channels.

Format and Shape These variables include features like whether or not the logo has a mark, how many marks there are, and what the aspect ratio of the logo is. We also compute the convex hull of the logo, which is the smallest convex polygon that contains all of the non-background pixels. We then cluster these hulls across logos to form a dictionary of logo shapes, which is shown in Figure 4. We also do something similar for the shape of the marks: for each mark, we standardize its shape and convert it to greyscale, then cluster across marks into 14 representative mark types. This maintains more details than the convex hull approach, allowing us to see, for example, the difference between solid and hollow circles, but is also typically more noisy. We give examples of these classes in Figure 5.

Font Font is a crucial feature of logos. We therefore have developed an elaborate procedure to identify and describe characters and their fonts. Specifically, for each segment of the

Name	R	G	B	Color	Name	R	G	B	Color
White	253	253	253		Dark Blue	30	42	124	
Black	20	18	18		Light Gray	165	164	167	
Red	226	33	41		Light Blue	54	153	204	
Blue	25	89	152		Light Green	99	178	67	
Dark Green	34	120	77		Yellow	245	202	36	
Orange	239	131	40		Tan	186	164	103	
Dark Gray	116	111	111		Dark Red	174	39	63	

Figure 3: The color dictionary: This table shows the RGB color channel values of the cluster centers for the representative set of colors, along with the actual color encoded by those values. These were obtained by clustering in the LAB color space across logos, which is meant to capture differences in human color perception.

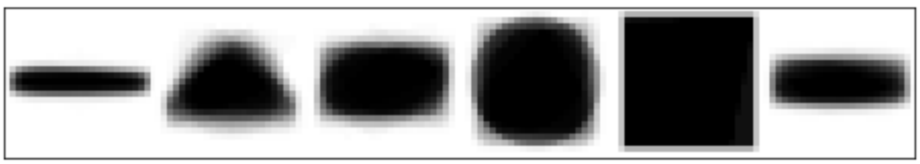


Figure 4: The hull classes: This table shows the six typical shapes of logos, as characterized by their convex hulls. Each logo in our dataset is assigned to one of these classes.

logo, we apply a template matching procedure, to try match the segment to an extensive collection of fonts, which we curated to capture the intricacies of font design as exhaustively as possible. This font dictionary captures a range of font families, forms, and stylings, including examples of fonts from all Vox-ATypI font classes, a standard font classification scheme used by font experts.[3] We illustrate our complete font typology in Figure 6.

Other There are several other features which we found in the literature review to be important aspects of design: complexity, symmetry, repetition, and orientation. For each of these, we include direct or indirect measures aimed at capturing that feature, without the need for a human coder. For complexity, we include a number of measures, including the number of distinct colors, the number of segments, the perimetric complexity (the ratio of edge pixels to interior area), and the greyscale entropy (the average variance of pixel intensities across sliding windows). We also include measures of both horizontal and vertical symmetry, computed by looking at the correlation between halves of the image. For repetition, we look at the different subregions of the logo, and compute correlations between size and complexity across them, as a proxy for repetitive structure. For orientation, we compute both measures of position of the mark relative to the text, and also edge-based metrics. Several of these features are illustrated in Figure 1.

[3] https://en.wikipedia.org/wiki/Vox-ATypI_classification

Cluster	Sample of marks
6	
7	
9	

Figure 5: The mark classes: This table shows three examples of our mark classes, with 10 randomly sampled examples of each. Each mark is assigned to a single class.

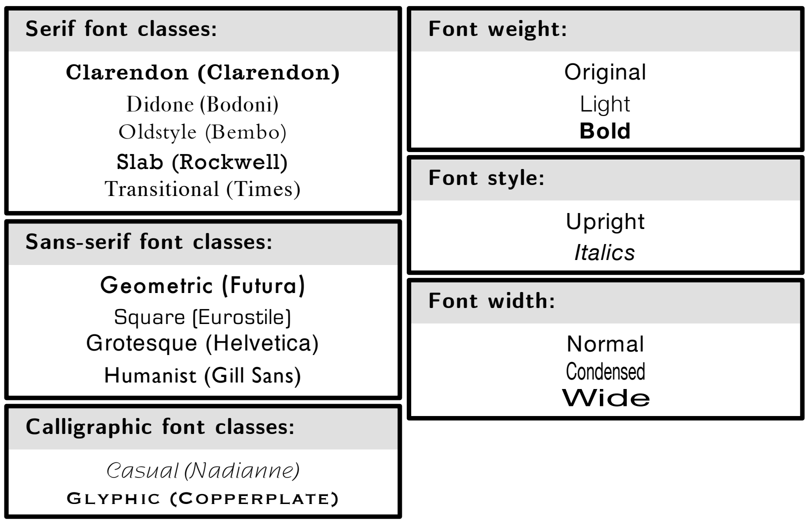


Figure 6: Font classification system employed by the algorithm: fonts were matched to a font family, weight, style, and width.

5. Exploring the Data

Before describing our multiview learning framework, we first provide some simple descriptive evidence illustrating the interplay between logo features, firm function, as operationalized by industry labels, and brand personality perceptions. There are two primary goals of this section: first, to build familiarity with the logo data, and second, to motivate the full model, by illustrating the complex interplay between each of our domains (logo, firm website text, and brand personality).

To capture the interplay between these variables in an intuitive and interactive fashion, we rely on visualizations commonly called forest plots, which show how one focal variable, the outcome, varies as a function of another (binary) variable.[4] In the case of a true binary variable, like color (e.g. does a logo have blue in it?), the plot shows the difference in the outcome for firms that have the variable (e.g. the logo has blue), compared to firms that do not have the variable (e.g. the logo has no blue in it). For real-valued variables,

[4]We include OLS-based explorations in Appendix C.

we consider a median split: comparing firms that are in the top 50% for the variable to firms that are in the bottom 50% for that variable. In the remainder of this section, we highlight a few of these plots. However, we also provide a web app that allows the reader to explore the full set of possible forest plots, which can be accessed at `https://rdew.shinyapps.io/explore_logo_data`.

5.1. Brand Personality

In our data, brand personality provides an especially insightful portrait as to how consumers perceive the firm. In Figure 7, we present a series of plots that look at how brand personality perceptions vary as a function of logo features, that are both intuitive, and corroborate some of the findings from the literature on logos and aesthetics, adding validity to our data and logo processing algorithm.

The first of these plots compares consumer's brand personality perceptions across the three most common dominant logo colors: black, blue, and red. We can see, for instance, that black logos tend to score low on down-to-earth, but high on dimensions like daring, spirited, and imaginative. Interestingly, they also seem to score high on upper class and charming, but also on outdoorsy and tough. This result, in isolation, seems surprising, as upper class and charming seem quite different than outdoorsy and tough. This unintuitive result highlights the need for understanding the whole combination of logo features, jointly: black, alone, may be used to convey a multitude of brand identities. Logo design must thus rely on many facets, simultaneously, to build a personality-consistent logo.

One crucial feature beyond color is font, and in the second plot in Figure 7, we explore different font features. In many cases, these features also match intuitions: serif fonts are perceived as more sophisticated, less rugged. Condensed lettering is more down-to-earth but less intelligent, while wide lettering is tough. Bold lettering and light lettering move in opposite directions, with bold letters being perceived as more down-to-earth and tough, while light letters are daring and sophisticated. Italics are tough and down-to-earth, but not upper class. We can thus begin to see that the *combination* of color and font can bring out unique identities, when one, in isolation, cannot. It is this combination of features that the model of logo design which we propose in the next section aims to capture.

In the final plot of Figure 7, we see some of the global features of logos. These features are less intuitive than color and font, but have been emphasized more in the literature. For instance, we see that horizontally symmetric logos (feature `h_sym`) tend to be perceived better along almost all dimensions, except intelligent, perhaps reflecting the role of harmony in positive affect discussed in Henderson and Cote (1998). We find horizontal orientation (feature `hor`) is related to tough and outdoorsy brands, whereas upward-diagonal orientation (feature `up_diag`) appears positively related with cheerful, spirited firms. This latter point lends some support for the findings of Schlosser et al. (2016), who found that upward diagonals convey activity. Angularity, as captured by the number of corners (feature `ncorners`), is positively associated with down-to-earth and tough logos, and negatively related to the

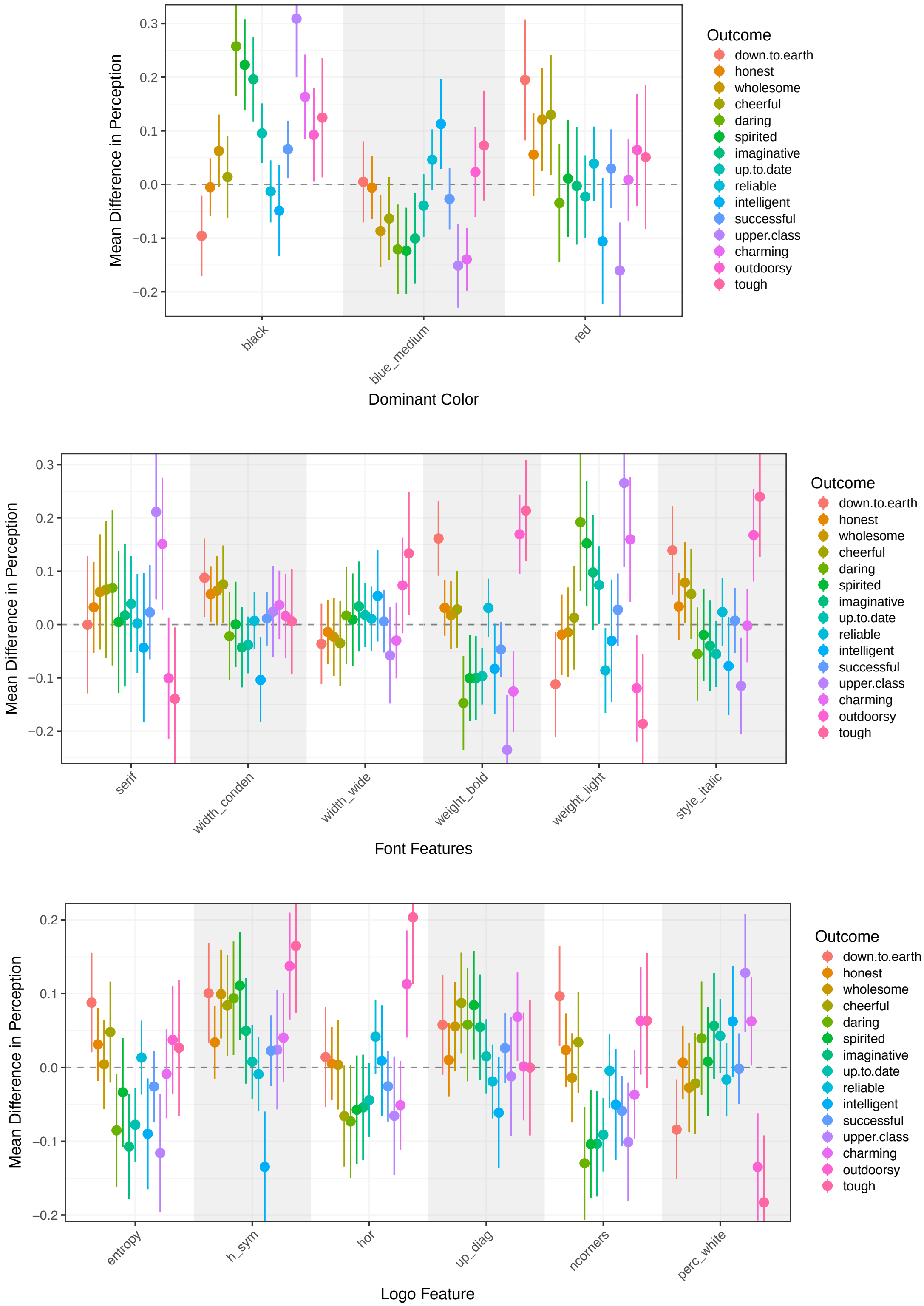


Figure 7: Each color in the plot represents a different brand personality factor, denoted in the legend. On the x-axis are features of the logo. On the y-axis is the difference in brand personality perception for firms that have a certain feature (or fall in the top half of firms for that variable), versus firms that do not have that feature (or fall in the bottom half).

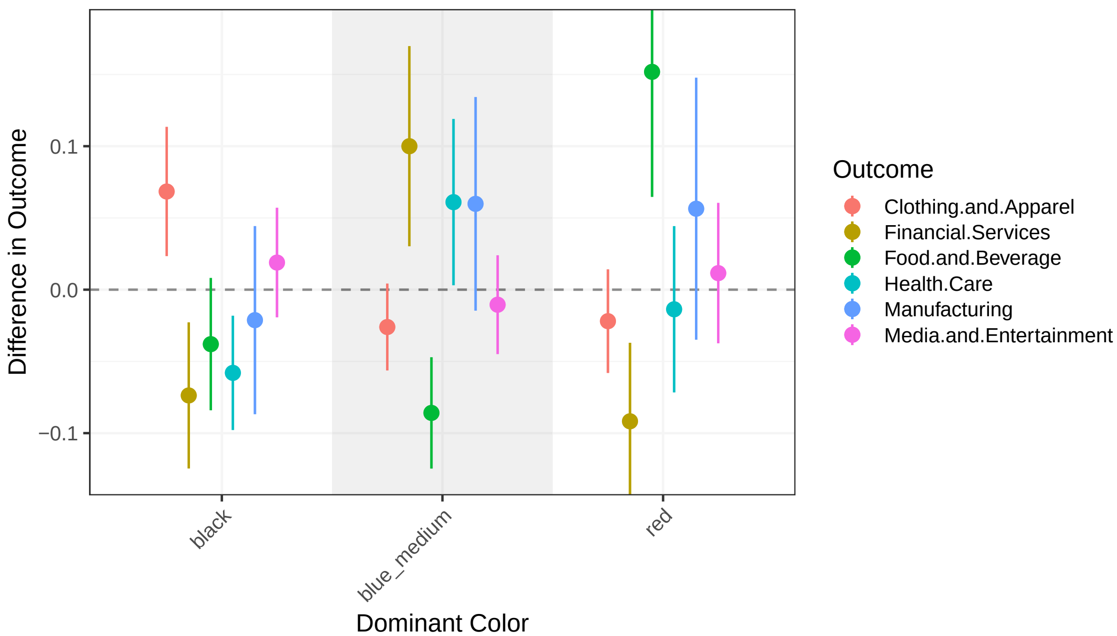
others. This appears to support the findings of Jiang et al. (2015), where angularity is found to be associated with durability. Percentage whitespace (variable `perc_white`) has a positive association with upper class and charming, and not with outdoorsy and tough, which is reminiscent of the findings of Semin and Palma (2014) about the femininity of white.

5.2. Industry Labels

As a simple measure of what a firm does, we rely on industry labels from the database Crunchbase. Apart from conveying brand image, logos may also simply convey what it is that a firm does. Firms may rely on logos as a signal, such that consumers can identify the firm as fitting their expectations for what kind of product or service they will receive. As such. In Figure 8, we show a set of forest plots, similar to those for brand personality, this time exploring variation in terms of industry labels, focusing on the dominant color of the logo. At the top, we show: given a logo has a specific feature, is that company more or less likely to be in a certain industry? At the bottom, we consider the other direction: given a firm is in a particular industry, is it more or less likely to have a particular feature in its logo?

We can see some of these relationships are quite strong and intuitive. For instance, blue is associated with the financial services industry, but not food and beverage, and the reverse is true for red. Black is associated with clothing and apparel companies, which is also consistent with the brand personality link of black with upper class and charming, as many clothing and apparel companies are also luxury brands. However, we also see, again, that the story is complicated. For example, while we saw in the brand personality analysis that black logos are perceived as rugged, it is not necessarily the case that companies in rugged industries, like manufacturing, are using black logos.

These visual analyses study effects in isolation. They thus raise the question: what is the right *combination* of logo features a firm should employ to be perceived a certain way? We see, for instance, that red is positively associated with food and beverage companies, but negatively with an upper class brand personality perception. What combination of logo features might convey the idea of an upper class fast food company? In addition, the industry label is a very simple way of operationalizing what a firm does, just as brand personality is a simple way of operationalizing what a brand is. To answer questions regarding combinations of features, and to facilitate the use of unstructured, textual data that may more accurately reflect the nuances of a company, we need a model that is able to make sense of this type of data, and simultaneously weigh all of these distinct aspects of brand identity.



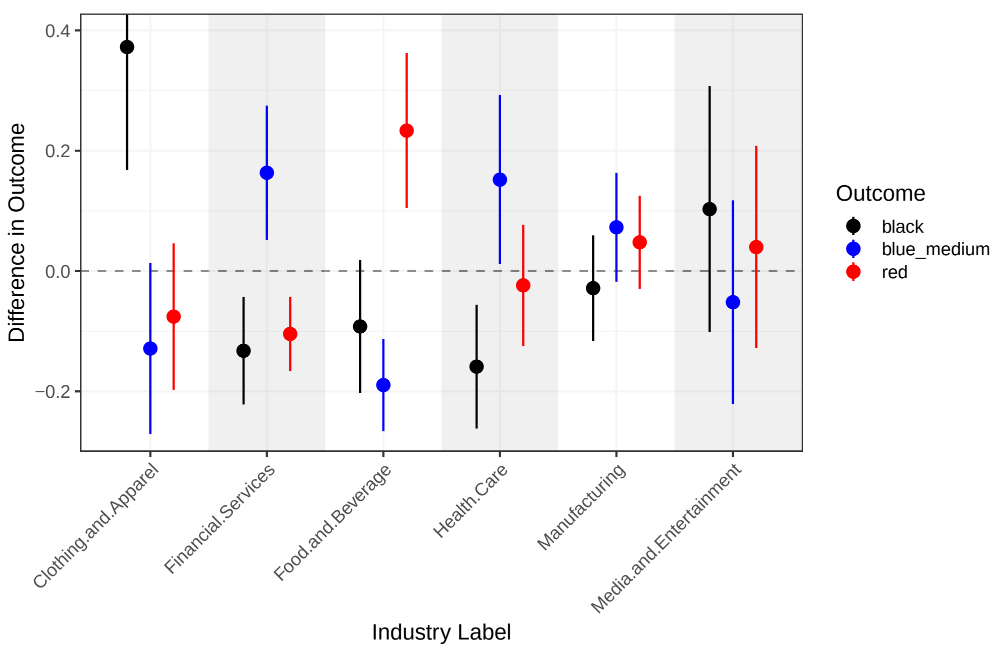


Figure 8: We show the relationship between industry label and logo color in two different ways: at the top, we show, given a logo has a certain dominant color, whether the logo is more or less likely to be labeled with a certain industry label. At the bottom, we show, given a logo has a certain industry tag, is it more or less likely to have each of the dominant colors.

6. Model of Logo Design

In this section, we describe our model of logo design. Specifically, our model draws on methods from deep generative modeling (Rezende et al., 2014; Kingma and Welling, 2013; Ranganath et al., 2014) and multiview learning (Li et al., 2016; Wu and Goodman, 2018) to learn joint representations of brands that can then be used to predict each of our domains of interest. Specifically, our model synthesizes our three main data sources: the *web descriptions*, which are how the firm describes itself in text, the *logo*, comprised of the features we described in Section 4, and finally consumers' perceptions, operationalized by their brand personality evaluations. By leveraging multiview representation learning, we are able to understand the links across these domains, in a flexible fashion, without specifying a priori which feature or modality is the dependent variable, and which are the independent variables. In turn, this enables us to answer questions from all three perspectives described in the intro: the researcher's perspective, the designer's perspective, and the brand manager's perspective.

6.1. Multimodal VAE

Our modeling framework is based on the variational autoencoder (Rezende et al., 2014; Kingma and Welling, 2013), a deep learning model designed to learn generative models of data. A variational autoencoder contains two components: an inference network, and a decoder network. The inference network is a deep neural network takes data as input, and outputs the parameters of an approximate posterior distribution for those data's latent parameters. The decoder network is a separate deep neural network that takes the vector representations of the data, and outputs a probability distribution over the original data. In this way, the two together provide dimensionality reduction: the inference network doing the inference from data, and the decoder making predictions.

Our specific implementation is a multimodal variational autoencoder (MVA), similar to that used in Wu and Goodman (2018), which learns a representation of the joint distribution across many domains of interest. In generality, suppose we have D domains of interest, indexed by $d = 1, \ldots, D$. In our application, these will be text, logo, and brand personality. For each of the brands in our data, $b = 1, \ldots, B$, denote the data of brand b in domain d as x_b^d. In each domain, there are different features (words for text, logo features for logos, personality traits for brand personality). We will index these features $j = 1, \ldots, V_d$. Note that, in our work, we observe data in all of the domains for each brand, but that the framework also allows for missingness. For each one of the domains, we posit a probability model that captures the features of that domain.

An overview of the basic components of the model is the following:

- Each brand has its own *latent representation*, denoted z, that captures the "brand identity" of the brand in dense, numeric, vector form.

- Given the representation, z, the *decoder network*, parametrized by a set of hyperparameters θ, transforms z into the parameters, μ, of feature-specific *probability models*, $p(x|\mu = f(z;\theta))$, which in turn predict the features we expect that brand to have.

- Given data, x, from some or all of the domains, the *inference network* is a function, parametrized by a set of hyperparameters ϕ, computes an approximation of the posterior distribution of z, $p(z|x;\phi)$.

These three components are deeply interconnected, such that describing each, in isolation, is challenging. We thus first, in Section 6.2, describe the probability models, which capture the traditional "likelihood" of the data. Then, we describe the decoder network used to link the representations to these models in Section 6.3. Finally, we describe the inference network, which can be used to learn the latent representations from the data, in Section 6.4.

6.2. Domain Probability Models

Conditional on the joint representation z_b, each brand's features are modeled using domain-specific probability models, the parameters of which, μ_b, are inferred from the decoder network described in the following section, $\mu_b = f(z_b;\theta)$. The specific models used for our data are:

- *Text:* For determining which words to include, we stemmed and tokenized the full vocabulary, removed standard stopwords, then filtered out words that occurred in less than twenty different brand descriptions. For modeling this textual data, we then use a simple binary model, capturing whether or not a given word is present in the textual description. That is, for each brand b, for each word w, we model:

$$P(x_{bw}^{\text{Text}} > 0) = \frac{1}{1 + \exp(-\mu_{bw}^{\text{Text}})} \tag{1}$$

  This simple coding captures the idea that firms choose to use a set of words, and that we are interested in whether or not a firm chooses to label itself a certain way (e.g. as "innovative"). Although the number of times a given word is repeated may contain information, it may also merely reflect how much text was present on the firm's website, or any number of unrelated factors. Hence, we only model whether or not a given word is present.

- *Logo features:* Many of the logo features exhibit very different statistical properties. In the appendix, we describe all of the logo features, together with their data types. In our model, conditional on the latent representation z_b, each of these features is drawn independently. For each one of these features, we then use an exponential family distribution that has support on that data type. Specifically, for real-valued data, like entropy, we use a normal distribution (or a lognormal distribution for continuous values with only positive support), such that for a real-valued feature indexed j, we have:[5]

$$x_{bj}^{\text{Logo}} \sim \mathcal{N}(\mu_{bj1}^{\text{Logo}}, \sigma_{bj}^{\text{Logo}}),\ \sigma_{bj}^{\text{Logo}} = \log(e^{\mu_{bj2}^{\text{Logo}}} - 1)) \tag{2}$$

Note that, for two parameter families, like the normal, we learn both the mean and the variance. For binary data, like whether the logo has a mark, we use a bernoulli distribution, equivalent to the model for text described above. For choice data, like the dominant color, where we have one of $m = 1, \ldots, M_j$ possible options, we use a categorical distribution, such that:

$$x_{bj}^{\text{Logo}} \sim \text{Categorical}(\text{Softmax}(\boldsymbol{\mu}_{bj})), \tag{3}$$

$$\boldsymbol{\mu}_{bj} = (\mu_{bj1}, \ldots, \mu_{bjM_j}) \tag{4}$$

- *Brand personality:* Similar to the real-valued logo features, brand personality in our data is also real-valued: it is the average of all respondents ratings, measured between 0-4. We approximate this using a normal model, again with the mean and variance learned from the latent representation.

6.3. Latent Representations and the Decoder Network

To learn the parameters μ_{bj}^d, we assume that each brand has a K-dimensional latent vector representation, which we denote $z_b = (z_{b1}, \ldots, z_{bK})$, that is shared across the domains. For each component of this representation, z_{bk}, we assume a unit normal prior:

$$z_{bk} \sim \mathcal{N}(0, 1).$$

Given this representation, the parameters μ_{bj}^d are computed from a deep neural network. The structure of this neural network may depend on the domain. Typically, we will use dense layers with rectified linear activation units (ReLU) and skip connections, which means the following sequence of computations:

$$\mathbf{h}_{b1}^d = \max(0, \mathbf{a}_{d0} + W_{d0} z_b) \tag{5}$$

$$\ldots \tag{6}$$

$$\mathbf{h}_{bL_d}^d = \max(0, \mathbf{a}_{d(L_d-1)} + W_{d(L_d-1)}^h \mathbf{h}_{b(L_d-1)}^d + W_{d(L_d-1)}^z z_b) \tag{7}$$

$$\mu_{bj}^d = a_{dL_dj} + (\mathbf{w}_{dL_dj}^h)'(\mathbf{h}_{bL_d}^d) + (\mathbf{w}_{dL_dj}^z)' z_b \tag{8}$$

Intuitively, we are applying the same operation (the ReLU) in sequence. At each layer of the model, we compute a new representation of the brand which we call the hidden units at layer ℓ, denoted by $\mathbf{h}_{b\ell}^d$, using both bias (intercept) parameters $\mathbf{a}_{d,\ell}$ and kernel (coefficient) parameters $W_{d,\ell}^h$, $W_{d,\ell}^z$. We combine these hidden units with the original representation z_b,

[5] The $\log(e^y - 1)$ structure in Equation 2 is the inverse of the so-called softplus function, $y = \log(1+e^x)$, which is commonly used to enforce positivity, as a more numerically stable alternative to a simple exponentiation.

in what is known as a skip connection, to learn the hidden units of the next layer.[6] This operation is repeated L_d times for the number of layers in the network for domain d. At each layer, the number of hidden units (meaning the dimension of $\mathbf{h}$) may change, which allows the network to learn different levels of abstraction of the data. Moreover, as the operations are nonlinear, this network theoretically corresponds to learning an arbitrary nonlinear relationship between the data and the representation. In effect, this means we can capture quite complex joint distributions. The more hidden units, and the more layers, the more expressive the model.

We denote the whole set of parameters,

$$\theta_{dj} = (a_{dL_dj}, w_{dL_dj}, \{a_{d\ell}, W_{d\ell}\}_{\ell=1,\ldots,L_d-1}),$$

and this whole operation as:

$$\mu_{bj}^d = \mathrm{DNet}_d(z_b; \theta_{dj}),$$

where $\mathrm{DNet}(\cdot)$ stands for "decoder network." Note first that, across j, many of the components of θ_{dj} will be shared within a domain. We may also use θ_d to refer to all of the network parameters within domain d across all j. Also, note again that the exact nature of this network can differ across domains: the above conveys the general structure. We describe the specifics of each domain's network in a later section.

6.4. Multiview Inference Networks

The key task in using the MVA framework is learning the representations z_b. Once we know z_b, we can use z_b to make predictions across modalities via the probabilistic decoder. Important to our framework, we would like to be able to learn z_b given information on only a subset of the domains. Then, we can use the representation z_b and the decoder to make predictions for the unseen modalities. In practice, this means we could use the MVA to generate a logo template, given a textual description, to generate words describing a specific set of logo features, or to predict brand personality assessments given either visual or textual information.

The goal of the inference network is to go from data x_b to an approximate posterior distribution for the latent representation z_b. In most models, learning latent parameters is accomplished by model training, using either maximum likelihood, MCMC, or variational inference. Inference networks transform the problem of inference of latent parameters into a problem of learning a function, parametrized by a (deep) neural network, such that given any data, we can obtain an approximate posterior distribution for the latent variables of interest, simply by evaluating the function. Using similar notation as above, a generic inference network can be written as:

$$\xi_{bk} = \mathrm{INet}(x_b; \phi),$$

[6] We include skip connections to avoid a phenomenon called latent variable collapse, in which models like ours get stuck in uninformative local optima (Dieng et al., 2018).

where this condensed notation stands for a neural network given by:

$$\mathbf{h}_{b1}^{\mathrm{Inf}} = \max(0, \mathbf{c}_{d0} + V_0 x_b) \tag{9}$$

$$\dots \tag{10}$$

$$\mathbf{h}_{bL}^{\mathrm{Inf}} = \max(0, \mathbf{c}_{L-1} + V_{L-1}\mathbf{h}_{b(L-1)}^{\mathrm{Inf}} \tag{11}$$

$$\xi_{bk} = \mathbf{c}_{Lk} + (\mathbf{v}_{Lk})'(\mathbf{h}_{bL}^{\mathrm{Inf}}), \tag{12}$$

and where ξ_{bk} is the vector of parameters of a (mean field) approximation to the true posterior, $q(z_{bk};\xi_{bk}) \approx p(z_{bk}|x_b)$.[7] In the case of a VAE, this approximation is assumed to be normally distributed, such that:

$$q(z_{bk};\xi_{bk}) = \mathcal{N}(z_{bk}; \mu = \xi_{bk1}, \sigma = \xi_{bk2}). \tag{13}$$

In our model, the goal is transfer learning via multiview representation learning: we want to be able to go from data in one domain, to the joint representation, and then to make predictions in all domains. To facilitate that, we implement a training procedure for multimodal variational autoencoders, similar to that of Wu and Goodman (2018). Specifically, we implement not a single inference network, but a set of modality-specific inference networks, which take as inputs data from a single domain, and outputs the posterior of the full representation. That is, we learn D distinct inference networks,

$$\xi_{bk} = \mathrm{INet}_d(x_{bd}; \phi_d),$$

corresponding to the model's "best guess" at the posterior distribution, given data from only one domain. This forces the model to learn multimodal representations, and avoids the case wherein some of the latent dimensions specialize in predicting only one domain. Although we defer our description of the full inference algorithm to a later section, the way this works intuitively is that, during training, at each iteration of the training algorithm, the data is randomly split into D bins. Then, for brands in bin d, only the data from domain d, x_b^d, are used to learn the parameters z_b. The result is, after training is done, we have D functions which we can use to go from one domain to infer the full posterior of z_b, which in turn, lets us make predictions about the other domains, via z_b.

6.5. Network Structures

As described previously in Equation 5, we use a skip structure in our decoder network (Dieng et al., 2018), where each layer of the network also contains the latent representation z, in addition to the hidden units h. This structure helps avoid a phenomenon called latent variable collapse, in which VAE-type models learn uninformative representations of the data very close to the prior. The remaining structure of the data is then specifying

[7]Note that, while the inference networks and decoder networks are all functions modeled with deep neural networks, these neural networks are modeled as a priori independent; that is, there is no imposed dependency between the two.

the dimensionality of the latent representation z_b, as well as how many layers, and how many hidden units are in each layer, for each domain of the data, for both the decoder and inference networks.

In this work, all of our components are small, relative to much of the deep learning literature. This is because we are using relatively small data, with already somewhat structured and pre-processed inputs, that are already represented at higher levels of abstraction.[8] Specifically, we assume there are 10 latent variables ($K = 10$). In the decoder networks, we assume that each domain's network has two layers, with 20 hidden units in the top layer for each. For text, we use 60 hidden units in the second (bottom) layer, reflecting its higher dimensionality than other domains. For the rest, we use a bottom layer of 40 units.

Under our multiview inference framework, we assume a four part structure: one inference network for each of the domains, and a fourth inference network which is given access to all of the domains. For these, we assume the topmost layer has 20 hidden units, but again assume a different number of hidden units for the bottom layer, reflecting our assumptions about how information-rich each domain is: for the full information inference network, we assume the second (bottom) layer has 80 hidden units; for the text, logo, and BP inference networks, we assume the second layer has 40 units.

6.6. Inference

We perform inference on the model via a form of Variational EM (VEM), implemented in the Edward probabilistic programming language. This follows the standard inference procedure for VAEs, as introduced in Rezende et al. (2014) and Kingma and Welling (2013), with only a slight twist to allow for our multiple decoder and inference networks. Specifically, the goal is to infer the model parameters for the decoder networks and the inference networks. In the classical VAE, the following loss function is minimized:

$$\ell(\theta, \phi) = \sum_{b=1}^{B} -E_{z\sim q(z;\xi_b=\mathrm{INet}(x_b;\phi))}[\log p_\theta(x_b \mid z)] + \\ \mathrm{KL}(q(z; \xi_b = \mathrm{INet}(x_b; \phi)) \mid\mid p(z)). \tag{14}$$

This is equivalent to the standard evidence lower bound (ELBO) for doing variational inference on the latent parameters, z, but where the variational approximation is given by the inference network (Blei et al., 2017). Another interpretation is that the first term encourages a good reconstruction of the data, while the second term regularizes estimates toward the prior. This is referred to as variational EM, as the variational procedure approximates the distribution of the latent variables z, but the model parameters θ are optimized for the likelihood of the data.

[8] Many deep learning frameworks operate at the pixel level, which is the most raw, unprocessed format for an image. As we described in Section 4, we choose a structured approach for feature extraction because it makes the results of the model much more intuitive, as we will see in the Section 7.

Feature(s)	Full Data	Logos	Text	BP	Intercept Only
Binary Text	0.096	0.102	0.094	0.126	0.157
Binary Logo	0.122	0.135	0.126	0.182	0.212
Real Logo	0.472	0.504	0.487	0.686	0.753
BP Ratings	0.190	0.200	0.181	0.210	0.405

Table 1: The own- and cross-modal reconstruction error across all of the inference networks, relative to an intercept only model.

In our multiview inference framework, the $p_\theta(x_b \mid z)$ from Equation 14 decomposes into a product of the domain-specific decoder networks and feature-specific probability distributions. Moreover, we modify the above to form a stochastic batched inference procedure where, for each iteration of our optimization, we split the data into four equally sized bins, such that for the first bin, we use the full inference network; for the second bin, we use the text inference network; for the third bin, we use the logo inference network; and for the fourth bin, we use the BP inference network. Returning to Equation 14, this means that, in our optimization, at each iteration, the $q(z; \xi_b = \text{INet}(x_b; \phi))$ used for observation b depends on the bin that brand b is assigned to in that iteration. Such a procedure is similar, but not exactly equivalent to that suggested by Wu and Goodman (2018). We run this stochastic procedure until the crossmodal log probability of the data, $p_\theta(x_b \mid z)$, converges.

7. Model Results

7.1. Model Fit

To begin, we examine how the model fits the data, and to what degree the inference networks produce equivalent representations.

7.1.1. Reconstruction Error

The metric by which VAEs are often evaluated is what's called reconstruction error: how well does the model do at reconstructing the data it is meant to represent? In our case, for each inference network, the error can be decomposed into the part that is own-modal reconstruction error (the modality that was input to the network), and a part that is cross-modal reconstrucion (i.e. the heldout modalities). In Table 1 we compare absolute error rates across the inference networks for several components of the model, using the last batch of training as the input data for the inference networks. We compare this to an "intercept-only" benchmark, wherein the average value of each feature is used as the prediction for all inputs.

There are three interesting patterns to note: first, the model is able to reproduce the

	Logos	Text	BP
Full Data	0.9	0.966	0.575
Logos	.	0.899	0.568
Text	.	.	0.576

Table 2: Average over brands of the correlation between z_b as learned by different inference networks.

data significantly better than a naive intercept-only model. Second, we notice that the BP-based inference network does worse on all cross-modal reconstruction errors. This is not surprising: relative to the other modalities, brand personality is a very high-level, abstract input, with significantly fewer features. As such, it is unable to match the representations learned by the other inference networks. Finally, for all networks except brand personality, we find that the reconstruction error rates are roughly equivalent. This is because, in all cases, the decoder network is the same, regardless of the inference network, and moreover, at each iteration, the firms that are used in each inference network are randomly shuffled. Hence, the model is incentivized to learn coherent representations across the inference networks, which result in nearly equivalent hit rates.

7.1.2. Data Complementarity

Given these patterns in the reconstruction error, we are also interested in understanding to what degree the learned representations are coherent across inference networks. That is, to what degree is the posterior mean of z_b when inferred through, for example, the full inference network, correlated with, for example, the brand personality inference network? In Table 2, we show the average correlation between representations, averaging over all brands. From this, we see that there is by and large agreement. The full and text networks learn the closest representations, which makes sense, given the richness of those two data sources. By using just brand personality, we get correlated, but not equivalent representations. This explains why the errors in Table 1 vary the way they do: brand personality is not rich enough to achieve the same degree of precision as the other modalities. Moreover, the brand personality representation is strongly correlated with the full representation in several of the dimensions of the latent 10-dimensional representation, but weakly correlated in others. This supports the idea that these modalities are complementary: brand personality captures some aspect of the brand that's displayed in text and logo, but that other aspects of textual and visual identity are independent of the brand personality.

7.2. Exploring the Latent Space

Now that we understand how the model approximates the multimodal representation, we can start exploring what representations the multiview variational autoencoder, or MVA, learns.

7.2.1. Neighbors in z Space

In general, it is difficult to interpret the latent space generated by our MVA, as the links from the representation to the data through the decoder are highly nonlinear. One question we can ask is, given a focal brand, which brands are closest to it in the latent space? We show this analysis for four brands in Table 3.

Brands that are closeby in z_b space are predicted to have similar properties across the different modalities. In some cases, the results in Table 3 are very intuitive. For instance, McDonald's tends to be close to many mass market, affordable chain stores, with dense, simple logos, often operating in the food industry. Starbucks' closest neighbors share circular properties, as well as operating within the slightly upper scale food space. Nike's closest neighbor is Adidas, which is similar both in terms of aesthetics and function to Nike. Finally, Actavis, a pharmaceutical manufacturer, is close to other manufacturing and B2B firms, with again similar logos, especially in terms of font, color scheme, and mark complexity.

McDonalds and Supervalu To help build intuition about Table 3, we consider the very first example: McDonalds, and its nearest neighbor Supervalu. As just described, McDonalds and Supervalu have many superficial similarities: they both have red, bold logos, and operate in the discount food space. These similarities are also reflected in the data. If we consider how people perceive these brands, vis-a-vis brand personality, there are considerable similarities, as plotted in Figure 9. Moreover, the words that the two brands use to describe themselves are also similar: the correlation between McDonalds binary text vector and that of Supervalu is $r = 0.24$, nearly double the correlation of McDonalds with other firms on average ($r = 0.13$). These similarities across all modalities among these two brands is what the model is detecting, leading to their similar representations in z space.

Nike and Disney There are some less intuitive findings in Table 3 as well. Perhaps most interesting are Nike's neighbors besides Adidas: Disney, Polaris, and Lego. Let's consider the similarities between Nike and Disney. Aesthetically, their logos are, in fact similar, in terms of color and layout. Their brand personalities are also aligned, as we show in Figure 10. What's striking about this plot is on how many dimensions both brands score near the top of the scale, including on dimensions like successful, imaginative, and family-oriented. There are also some differences, especially related to the ruggedness of Nike. Finally, the words they use to describe themselves are also similar: the correlation between Nike's binary word vector and that of Disney is $r = 0.2$, compared to $r = 0.12$ on average across all brands. Hence, while perhaps surprising at first glance, there are deep connections between the brand identities of Nike and Disney, which the model detects, and then subsequently predicts.

Focal Brand	Neighbors in z_b space			
McDonalds Fast food	**Supervalu** Retailing and grocery	**Old Navy** Apparel	**Dollar General** Discount retailer	**Kroger** Grocery
Starbucks Coffee	**Chipotle** Fast casual restaurant	**Whole Foods** Organic grocer	**L'Oreal** Personal care	**Minute Maid** Juice and beverage
Nike Footwear and apparel	**Adidas** Footwear and apparel	**Disney** Media and entertainment	**Polaris** Snowmobiles and ATVs	**Lego** Toys
Actavis Pharmaceutical manufacturing	**Praxair** Industrial gases	**Autoliv** Automotive safety supplier	**Clorox** Consumer products	**Optum Health** Health services

Table 3: The 4 closest brands to each focal brand in z_b space, including their logo, name, and a brief description.

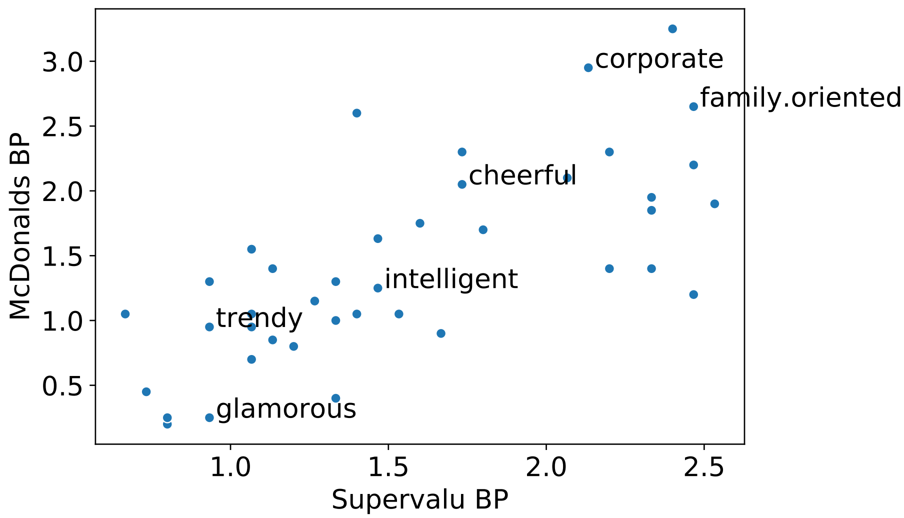


Figure 9: Brand personality ratings of McDonalds versus Supervalu, showing the high correlation between personality perceptions of the two brands. Each dot represents a different brand personality trait, with several suggestive traits labeled.

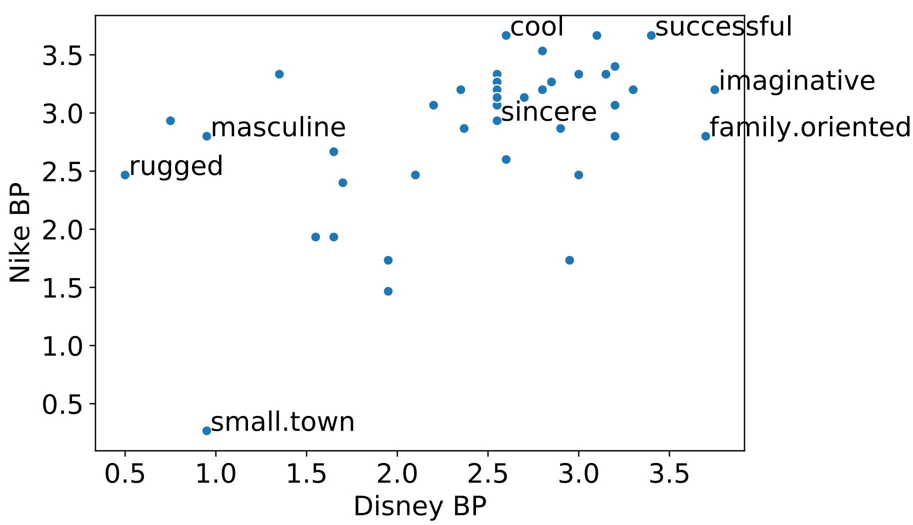


Figure 10: Brand personality ratings of Nike versus Disney, showing the high correlation between personality perceptions of the two brands. Each dot represents a different brand personality trait, with several suggestive traits labeled.

Rank (Closest)	p = 0.9	p = 0.7	p = 0.5	p = 0.3	p = 0.1
1	M	OLD NAVY	Cadbury	Cadbury	NIKE
2	OLD NAVY	M	OLD NAVY	NIKE	adidas
3	SUPERVALU	Kroger	GAP	Disney	Disney

Table 4: Linear interpolation between McDonalds and Nike, showing three brands, in order, whose z_b are closest to $z = pz_{\text{McDonalds}} + (1-p)z_{\text{Nike}}$.

7.2.2. Interpolating Between Brands

Another way to attempt to understand the latent 10-dimensional space learned by our MVA is to use it to interpolate between brands. Intuitively, the MVA converts a large set of features with very different statistical properties into compact, continuous vector representations. Continuous movement in this latent space thus allows for continuous movement among brand identities, slowly shifting the predictions of the model. We can use such movement in the latent space to interpolate between brand identities.

McDonalds and Nike Midpoint Analysis For instance, drawing on our previous analyses, we may ask the question: which brand identities emerge by interpolating between McDonalds and Nike? To answer this question, we consider new z values of the following form:

$$z = pz_{\text{McDonalds}} + (1-p)z_{\text{Nike}}.$$

We consider $p = 0.9, 0.7, 0.5, 0.3, 0.1$. We then find the actual z_b vectors that are closest to this interpolated value for each value of p. We show the results in Table 4. In general, we find a few transitions that happen between these identities: first, we see the apparel companies like Old Navy that were previously similar McDonalds emerge as the most similar to the interpolation. We also see the element of "value" fade away, as firms like Supervalu and Dollar General disappear, and firms like Gap appear. At the midpoint, we see Cadbury, a chocolate company, emerge as the midpoint. Finally, as we move toward Nike, we see Disney and Adidas again emerge, although Disney emerges sooner than Adidas.

It is interesting to consider why the model identifies Cadbury as the closest brand to the midpoint of McDonalds and Nike. First, it's worth noting that, while Cadbury is the closest in terms of z distance to the midpoint, it is not exactly at the midpoint. In fact, on several dimensions, the Cadbury z_b is actually quite far from the midpoint. Thus, in some sense, no brand exists at the exact midpoint between McDonalds and Nike. There are, however,

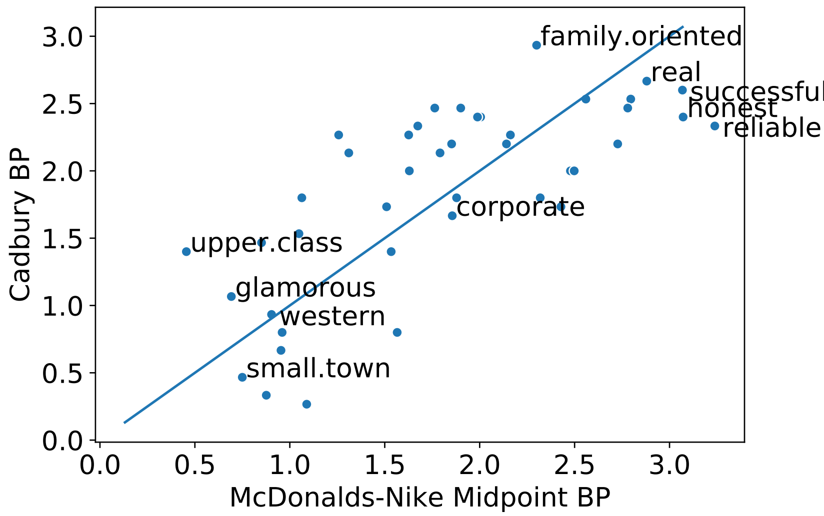


Figure 11: The actual brand personality of Cadbury is plotted against the predicted brand personality of the fictitious brand that lies at the midpoint between McDonalds and Nike in z space.

some clear similarities between this fictitious midpoint brand and Cadbury. For example, in Figure 11, we plot the brand personality of Cadbury against what the model predicts for the fictitious midpoint brand. We see a close, but not exact correlation: the model predicts a midpoint brand that scores high on reliability, honesty, and success, while low on upper class, glamorous, and small town. Cadbury, on the other hand, does not score as low on things like upper class and glamorous, but agrees with much of the rest of the profile. We can also use text to understand what kind of brand would exist at this intersection: the model expects words like meaning, compete, step, footprint, citizen, force, healthier, dollar, happen, creation, and breakthrough. Intuitively, these words do appear to be a midpoint between McDonalds and Nike, emphasizing dollars, creating, health, competition, and footprint.

Other Interesting Midpoints While by no means comprehensive, there are many other interesting findings like the above, which fall out of the model's ability to interpolate between brand identities. These include:

- Under Armour as the midpoint between Nike and Gucci: Under Armour is positioned to some degree as an upscale fitness clothing brand. It thus makes sense that the midpoint between a fairly mainstream athletic brand, Nike, and a luxury fashion brand, Gucci, would be a brand like Under Armour.
- Booking and Priceline are at the midpoint between Google and Hyatt, emphasizing again this clean interpolation between brand identities and firm functionalities, with

The Gucci of Nike

$$0.5z_{\text{Gucci}} + 0.5z_{\text{Nike}} \approx z_{\text{UnderArmour}}$$

The Google of Amazon

$$0.5z_{\text{Google}} + 0.5z_{\text{Amazon}} \approx z_{\text{eBay}}$$

The Mercedes of Old Navy

$$0.5z_{\text{Mercedes}} + 0.5z_{\text{OldNavy}} \approx z_{\text{RalphLauren}}$$

The Google of Hyatt

$$0.5z_{\text{Google}} + 0.5z_{\text{Hyatt}} \approx z_{\text{Booking}}$$

Figure 12: Interpolating between brands in the latent space. Here, we illustrate the visual similarities, in addition to the functional and brand-related similarities described in the text.

Booking and Priceline being search engines for hotels.

- eBay is at the midpoint between Amazon and Google, which is fascinating, given eBay's visual similarity to Google, but functional similarity to Amazon.

- Ralph Lauren is at the midpoint between Mercedes-Benz and Old Navy. Ralph Lauren is a more upscale and luxurious apparel brand, relative to Old Navy.

In addition to functional and brand-related similarities, there are also visual similarities, as illustrated in Figure 12.

Implications for Design In the sense that the model allows for interpolation and "arithmetic" between brand identities, it mirrors the logo design process. Logo designers often start with a survey of an industry, competitors, and audience, and determine the key elements of design that convey meaning in each of these spaces. In coming up with a final design for a focal brand, the task is then one of interpolating: for instance, how do we think of the Starbucks of Chinese cuisine? The Uber of healthcare? How can we infuse a little bit of the brand identity of Gucci into the fast food industry? By being able to formulate such questions mathematically, as vector operations in a latent space, we make this process of interpolation data-driven.

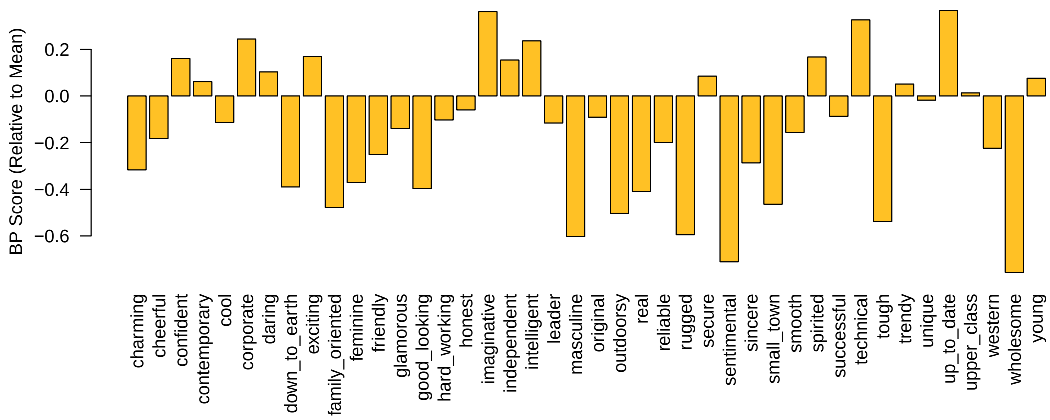


Figure 13: The predicted brand personality profile for our first randomly generated brand, the "cold, modern corporation." We show the predictions as differences from the average brand personality rating for each trait (i.e. positive = more than average, negative = less than average).

7.3. Generating Brand Identities

Our MVA is what is known as a deep generative model in machine learning. This term arises because the model can be used to generate data that mirrors the input data. Generation under the model simply involves sampling new zs from the prior, $z_k \sim \mathcal{N}(0, 1)$, then passing these new z vectors through the decoder networks. Such generation gives us several insights: first, it allows us to further explore the structure the model has learned, by seeing what brand identities it generates. Second, it can provide a mechanism for idea generation, as the simulated brands may be structured fusions of the input data. Finally, and perhaps most importantly, it gives us a way of validating the model: by randomly sampling brand identities from the model, we can see if the model has learned coherent patterns, thus lending additional credibility to the learned latent space.

To illustrate that the model has, indeed, learned coherent representation, we will consider a case study, with a randomly sampled z vector given by:

$$z = (0.61,\ 1.24,\ 0.96,\ 1.55,\ -0.26,\ -0.17,\ 2.96,\ -1.09,\ -0.48,\ -0.68).$$

Below, we describe the features of a brand with this representation. As we show below, this z corresonds to a brand identity which we label a "cold, modern corporation." The predicted brand personality profile corresponding to this z is displayed in Figure 13. From this, we high scores on up-to-date, imaginative, technical, and corporate, and low scores on wholesome, sentimental, tough, and family-oriented. Together, this paints a picture of a technical, modern corporation.

The words that the model predicts are most likely to appear on the brand's website

Color		Font		Layout	
Feature	Prob	Feature	Prob	Feature	Prob
Has: blue dark	0.671	Font: wide	0.823	Has mark	1.000
Has: blue medium	1.000	Font: bold	0.956	Mark pos: bottom	0.967
Has: yellow	0.992	Font: no italics	0.981	Mark pos: top	0.556
Accent: blue medium	0.995	Class: geometric square	0.744		
Accent: yellow	0.998	Class: clarendon	0.525		

Table 5: Binary logo features that the model predicted would occur with greater than 50% probability for the generated brand, together with the predicted probabilities.

Feature	Value	Feature	Value
# Characters	-0.46	Aspect Ratio	-1.31
# Colors	-0.30	Entropy	0.33
# Corners	-0.33	Perimetric Complexity	-0.71
# Marks	-0.27	Horizontal Symmetry	0.50
# Regions	-0.54	Vertical Symmetry	-0.90
% Whitespace	-0.92	Mean Lightness	-0.75
Vertical Edges	0.39	Mean Saturation	0.53
Down Diag Edges	-0.16	SD Lightness	0.44
Horizontal Edges	-0.28	SD Saturation	0.55
Up Diag Edges	-0.30		

Table 6: Real-valued logo features that the model predicted for the generated brand. These values are standardized values (z-scores), and hence can be interpreted as standard deviations different from the average value of the feature.

are displayed in a word cloud in Figure 14. In addition to the top predicted words, we also show the words that are relatively likely and relatively unlikely for the simulated brand. In general, there are certain words that many firms use, including product, business, customer, world, provide, and service, which may not be as relevant to understanding the focal brand. We see that these two word clouds support the identity conveyed by brand personality: among the relatively likely words, we find technical words like data, app, problem, and implement. In the relatively unlikely words, we find things like provide, family, culture, and life.

Finally, we can see the visual features we expect to find in this firm's logo by examining Tables 5, 6, and 7. An interpretation of this logo by the author is presented in Figure 15.[9] It is harder to objectively interpret these visual elements, but we claim that this logo template appears to share similar elements to other logos in, for instance, the technology space.

This simulation illustrates that the model learns, and can generate, coherent brand

[9]The authors are not designers, as may be obvious from the interpretation.

Most Likely

Relatively Likely

Relatively Unlikely

Figure 14: At top, a random sample of the words that the model predicts will occur with greater than 50% probability, drawn proportional to their probability. At bottom left, the words that the model predicts will occur significantly more than they occur on average. At bottom right, the words that the model predicts will occur significantly less than they occur on average.

Feature	First (Prob)	Second (Prob)	Third (Prob)
Dominant Color	Med. Blue (0.999)	Dark Blue (0.001)	Yellow (0.000)
Hull Class	Circle (0.652)	Triangle (0.293)	Med. Rect./Oval (0.028)
Mark Class	Wispy Horiz. (0.847)	Circular (0.113)	Square (0.023)
Font Serifs	Sans-Serif (0.671)	Serif (0.329)	No Chars (0.000)

Table 7: Predicted categorical logo features for the generated brand. For each feature, we list the top three most likely outcomes under the model, together with their probabilities. (Throughout, the abbreviation “Med.” stands for “Medium.”

Figure 15: A rendering by one of the authors of a logo matching the features described in Tables 5, 6, and 7.

identities. We include additional simulated identities in Appendix D. Next, we discuss the decision support implications for the model, with an eye to answering the questions we outlined in the introduction.

7.4. Crossmodal Inferences

Finally, from both a design and managerial perspective, the most critical component of our model is the ability to move across modalities. That is, to predict, for instance, a logo, from a textual brief. This is the designer's perspective we outlined in the introduction. Moving from text and brand personality to logo features allows us to inform the design process in a data-driven fashion, by automatically translating text and survey data into visual templates. The ability to go from a logo to text and personality is also important, insofar as it allows for both the evaluation of potential identities, and for "letting the logos speak," to gain a better understanding of common design patterns. These are the researcher's and manager's perspectives we outlined in the introduction. In this section, we illustrate two of these channels: going from brand personality to text and logo, and going from text to brand personality and logo.[10]

Before delving into those illustrations, we describe how the process works, mathematically. In all cases, crossmodal predictions work through the modality specific inference networks, combined with the full decoder network. Specifically, given data on domain d for a new firm, denoted x^d_{new}, we can learn the approximate posterior of that brand's representation, z_{new}, via the modality d inference network:

$$z_{\text{new}} \sim \mathcal{N}(\xi_{\text{new},1}, \xi_{\text{new},2}),\ \xi_{\text{new}} = \text{INet}_d(x_{\text{new}_\text{d}}; \phi_d).$$

We can then make predictions for any of the domains by passing the expectation for z_{new}, $E(z_{\text{new}}) = \xi_{\text{new},1}$ through the decoder network for any of the domains of interest, d^*:

$$p(x^{d^*}_{\text{new}} \,|\, z_{\text{new}}) = p(x^{d^*}_{\text{new}} \,|\mu^{d^*}_{bj} = \text{DNet}_{d^*}(z_{\text{new}}; \theta_{d^*})).$$

This reveals the practicality of this multiview inference network approach: evaluating a

[10]Going from logo to text and brand personality is a work in progress.

conditional posterior predictive is equivalent to evaluating two functions: the inference network of the given domain d to infer the posterior of the latent parameter z, and the decoder network of the domain of interest d^*, conditional on the inferred z.

7.4.1. Brand Personality to Textual and Visual Identity

Given a brand personality profile, our goal is to use the MVA framework to understand what words might describe a firm with that personality, and what features are likely in that firm's logo. As a case study, we will focus on a firm that is a rugged, masculine, reliable, and hard working firm, with brand personality profile (relative to the mean) displayed in Figure 16.

Using the brand personality encoder network, we can learn a posterior distribution over z for a brand with this personality. Plugging the mean of that posterior distribution for z into the text decoder network, we find the most likely words are those shown visually in the word cloud in Figure 17, and in Table 8. Visually, the model expects to find again a blue logo, similar to the randomly generated firm in the previous section. The accent colors it expects now are again yellow, but also light blue. The font it expects is distinct from the random profile: it expects that this firm will use bold condensed letters, as opposed to wide. In terms of convex hull, it gives the highest probability to a circular or wide ovular/rectangular logo. Finally, similar to the random logo, it expects this firm will have a dark logo with low whitespace, and with a lower than average aspect ratio, indicating that it is less wide and more tall than average. We again provide a non-professional rendering of a logo meeting many of these criteria in Figure 18.

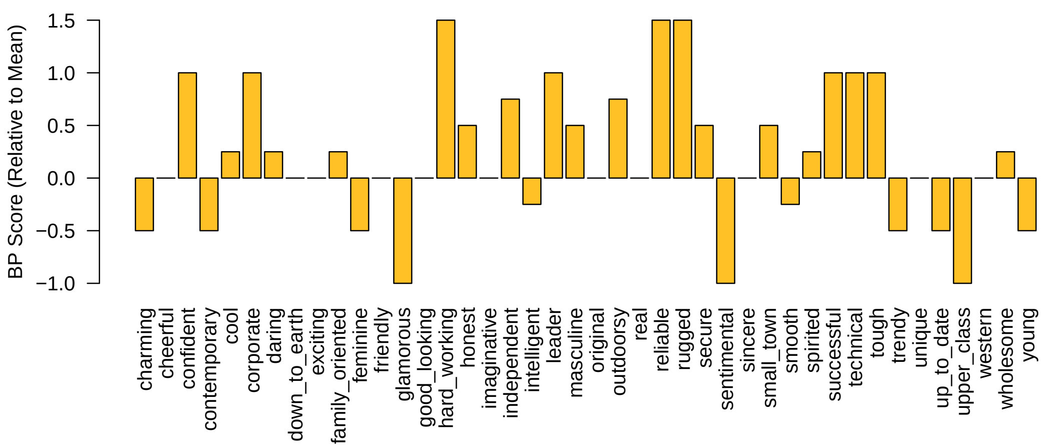


Figure 16: Brand personality of our focal firm for doing crossmodal inferences. Personality values are shown relative to the mean (i.e. differenced from the mean personality value across all firms).

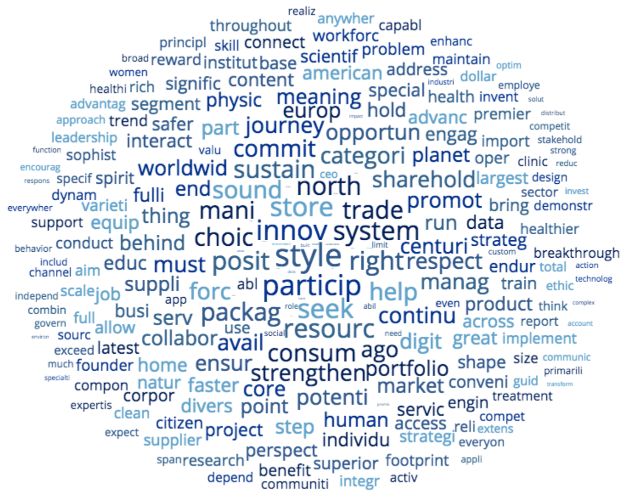


Figure 17: A word cloud reflecting the most likely words generated from the crossmodal inference procedure for the focal brand personality profile, corresponding to words that would likely be on the website of a firm with that brand personality.

	Top 20 Words
Most Likely	promis, regul, unit, whole, men, shop, accomplish, effici, specialti, women, account, environment, strong, solut, mobil, visit, exceed, divis, heritag, abil
Relatively Likely	regul, ceo, meaning, compet, scientif, whole, treatment, footprint, sector, trend, dollar, forc, implement, latest, faster, healthier, everywher, clinic, sophist, compon
Relatively Unlikely	improv, compani, time, experi, state, high, around, deliv, also, day, offer, countri, best, can, everi, creat, provid, us, new, work

Table 8: Likely and unlikely words generated from the crossmodal inference procedure for the focal brand personality profile, corresponding to words that would (not) be on the website of a firm with that brand personality.

7.4.2. Text to Logo and Brand Personality

Our final illustration of crossmodal inferences illustrates the direction that most approximates the design process: from a textual description to a logo and a prediction of brand personality perceptions. For this section, we will focus on a firm that was not included in our original dataset: Shake Shack. Shake Shack is a modern fast casual restaurant, serving burgers, hot dogs, milkshakes, and french fries, based out of New York City. We processed this text as we did the brands in our original sample, and present a summary of the text from their website in Figure 19. We then used the text inference network to infer Shake Shack's latent z_b. This was then passed to our logo and personality inference networks, to predict the feature's of Shake Shack's logo, and the way consumers will perceive their brand

Figure 18: Rendering of a logo containing many of the traits the model predicts given the focal brand personality.

personality.

In Figure 20, we present the brand personality predictions, which we assess to be relatively accurate: Shake Shack is a fairly trendy, contemporary take on fast food. It is generally perceived as (relatively) glamorous and exciting, especially in its association with New York City, and cheerful in both what it does, and how it portrays itself. In Tables 9, 10, and 11, we give the logo predictions, which are somewhat less accurate. Interestingly, the accent color of (light) green is accurately predicted, as is the square font, the high perimetric complexity, the vertical symmetry, and the higher variation in lightness. But many of the other predictions, including the dominant color of brown, the bold font, and the left placement of the mark are off.

We may then ask, why do the model's predictions differ from reality in the case of Shake Shack? One interpretation is that Shake Shack has intentionally deviated from the mold, to draw on certain poignant associations. For instance, an interesting element of the Shake Shack logo is its resemblance to an old neon sign, emblematic of an old school hot dog stand. While a standard fast food or fast casual restaurant may indeed feature bold font, Shake Shack differs from the model to emphasize its heritage. Moreover, the emphasis on blacks, instead of browns or blues, is characteristic of sophistication, charm, and luxury. Originating in New York, and marketing itself as a more upscale experience, relative to the standard burger and fries chains, perhaps this relative emphasis is thus strategic in its appeal to the New York demographic, and its key point of differentiation from the competition.

8. Conclusions

In this work, we have explored logo design and brand identity from a data-driven perspective. Our primary contributions are an approach to working with logos as data in a way that is both automatic and human interpretable, predictive results which can help identify specific features of interest and understand patterns in design, and finally a multiview learning model that mimics the design process, and in which we introduce a new approach

Figure 19: A word cloud representing the text from Shake Shack's website, processed in the same way as our original textual data.

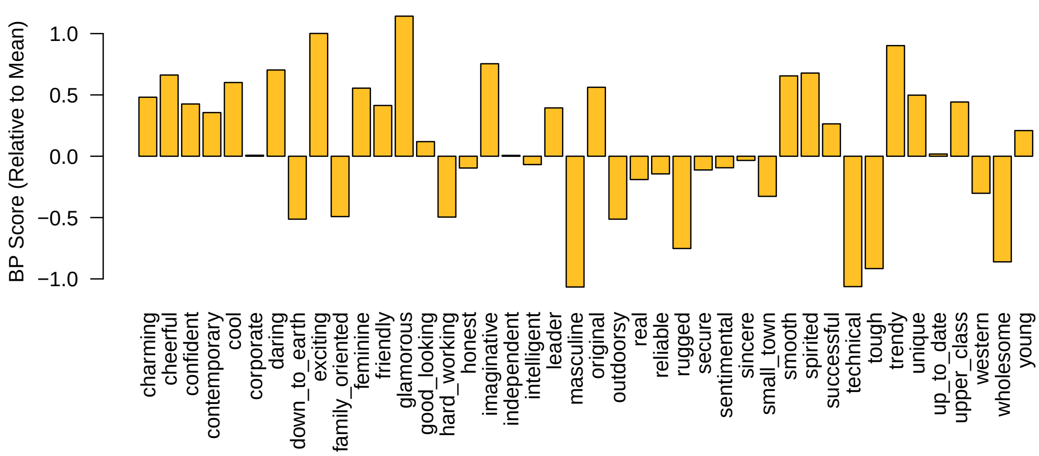


Figure 20: Brand personality predictions for Shake Shack, relative to the mean brand personalities in our sample, based on a crossmodal prediction from Shake Shack's text.

for using variational autoencoders for multiview learning. Our feature extraction algorithm makes the process of understanding logo design both objective, in the sense that it is done automatically through image processing, and useful for designers, in the sense that the

Color		Font		Layout	
Feature	Prob	Feature	Prob	Feature	Prob
Has: Dark Blue	0.729	Weight: Bold	0.746	Mark pos: Left	0.509
Has: Med Blue	0.826	Weight: Original	0.654	Has Mark	1.000
Has: Light Green	0.861	No Italics	1.000		
Accent: Dark Blue	0.820	Class: Geometric Square	1.000		
Accent: Light Blue	0.923				
Accent: Light Green	0.959				

Table 9: Binary logo features that the model predicted would occur with greater than 50% probability for Shake Shack, together with the predicted probabilities.

Feature	Value	Feature	Value
# Characters	0.22	Aspect Ratio	0.37
# Colors	0.59	Entropy	0.52
# Corners	0.05	Perimetric Complexity	0.60
# Marks	-0.26	Horizontal Symmetry	-0.47
# Regions	-0.01	Vertical Symmetry	0.87
% White	-0.37	Mean Lightness	-0.33
Vertical Edges	1.26	Mean Saturation	-0.28
Down Diag Edges	0.50	SD Lightness	0.15
Horizontal Edges	-0.75	SD Saturation	-0.56
Up Diag Edges	-1.14		

Table 10: Real-valued logo features that the model predicted for the Shake Shack. These values are standardized values (z-scores), and hence can be interpreted as standard deviations different from the average value of the feature.

Feature	First (Prob)	Second (Prob)	Third (Prob)
Dominant Color	Brown (0.847)	Med. Blue (0.141)	Dark Blue (0.006)
Hull Class	Med. Rect./Oval (0.665)	Thin Rect./Oval (0.333)	Triangle (0.001)
Mark Class	Vertical Narrow (0.359)	Square (0.330)	Bulky Hollow Geom. (0.306)
Font Serifs	Sans-Serif (0.970)	Serif (0.030)	No Characters (0.000)

Table 11: Predicted categorical logo features for Shake Shack. For each feature, we list the top three most likely outcomes under the model, together with their probabilities. (Throughout, the abbreviation “Med.” stands for “Medium.”

features are interpretable.

Our multiview learning model provides a way of moving across three focal modalities—textual descriptions of brands, their logos, and consumers’ perceptions brand personality. In turn, this allows us to understand what features of logos convey which aspects of brand meaning, to aid in the design process, and to help managers understand the implications of various design patterns. Our inference procedure mimics the way we envision the model being used by both designers and managers: the encoder networks allow any feature to be used as an input, from which the model can predict all left out features. In applying the model to our data, we learned a latent space that is meaningful, and in which vector operations, like the interpolation between two brands, yields interesting insights to brand identity. This interpolation procedure illustrates another key benefit of the representation learning paradigm, insofar as it can assist in the creative process.

Finally, there are several important limitations of this study. Foremost, our model is a model of typicality, not optimality, as alluded to particularly with the example of Shake Shack. We are able to capture what a typical firm does, not what is the best logo for a firm to do, given certain objectives other than typicality. Additionally, our model does not make strong claims about the causality of design: that is, why are existing logos designed the way they are? Answering this question is difficult, and likely involves both temporal factors (e.g. mimicry of a successful brand) and functional factors (e.g. red is easy to see on a sign from far away, or red stimulates the appetite). We leave both of these issues as topics for future study.

References

Aaker, J. L. (1997). Dimensions of Brand Personality. *Journal of Marketing Research*, 34(3):347.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877.

Chae, B. G. and Hoegg, J. (2013). The Future Looks Right: Effects of the Horizontal Location of Advertising Images on Product Attitude. *Journal of Consumer Research*, 40(August):223–238.

Childers, T. L. and Jass, J. (2002). All Dressed Up With Something to Say: Effects of Typeface Semantic Associations on Brand Perceptions and Consumer Memory. *Journal of Consumer Psychology*, 12(2):93–106.

Cian, L., Krishna, A., and Elder, R. S. (2014). This logo moves me: Dynamic imagery from static images. *Journal of Marketing Research*, 51(2):84–197.

Deng, X., Hui, S. K., and Hutchinson, J. W. (2010). Consumer preferences for color combinations: An empirical analysis of similarity-based color relationships. *Journal of Consumer Psychology*, 20(4):476–484.

Deng, X. and Kahn, B. E. (2016). Is Your Product on the Right Side? The "Location Effect" on Perceived Product Heaviness and Package Evaluation. *Journal of Marketing Research*, (Forthcoming).

Dieng, A. B., Kim, Y., Rush, A. M., and Blei, D. M. (2018). Avoiding Latent Variable Collapse With Generative Skip Models.

Doyle, J. R. and Bottomley, P. A. (2004). Font appropriateness and brand choice. *Journal of Business Research*, 57(8):873–880.

Doyle, J. R. and Bottomley, P. A. (2006). Dressed for the Occasion: Font-Product Congruity in the Perception of Logotype. *Journal of Consumer Psychology*, 16(2):112–123.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 226–231.

Hagtvedt, H. (2011). The Impact of Incomplete Typeface Logos on Perceptions of the Firm. *Journal of Marketing*, 75(4):86–93.

Henderson, P. W. and Cote, J. A. (1998). Guidelines for Selecting or Modifying Logos. *Journal of Marketing*, 62:14–30.

Henderson, P. W., Cote, J. A., Leong, S. M., and Schmitt, B. (2003). Building strong brands in Asia: Selecting the visual components of image to maximize brand strength. *International Journal of Research in Marketing*, 20:297–313.

Henderson, P. W., Giese, J. L., and Cote, J. A. (2004). Impression Management Using Typeface Design. *Journal of Marketing*, 68(4):60–72.

Janiszewski, C. and Meyvis, T. (2001). Effects of Brand Logo Complexity, Repetition, and Spacing on Processing Fluency and Judgment. *Journal of Consumer Research*, 28(1):18–32.

Jiang, Y., Gorn, G. J., Galli, M., and Chattopadhyay, A. (2015). Does Your Company Have The Right Logo? How and Why Circular and Angular Logo Shapes Influence Brand Attribute Judgments. *Journal of Consumer Research*, 42:ucv049.

Kareklas, I., Brunel, F. F., and Coulter, R. A. (2014). Judgment is not color blind: The impact of automatic color preference on product and advertising preferences. *Journal of Consumer Psychology*, 24(1):87–95.

Kingma, D. P. and Welling, M. (2013). Auto-Encoding Variational Bayes. (Ml):1–14.

Klink, R. R. (2003). Creating Meaningful Brands: The Relationship Between Brand Name and Brand Mark. *Marketing Letters*, 14(3):143–157.

Li, Y., Yang, M., and Zhang, Z. (2016). Multi-View Representation Learning: A Survey from Shallow Methods to Deep Methods. 14(8):1–20.

McLaren, K. (1976). The Development of the CIE 1976 (L* a* b*) Uniform Colour Space and Colourdifference Formula. *Journal of the Society of Dyers and Colourists*, 92(9):338–341.

Meyers-Levy, J. and Peracchio, L. A. (1992). Getting an angle in advertising: The effect of camera angle. *Journal of Marketing Research*, 29(4):454–461.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9(3):353–383.

Orth, U. R. and Malkewitz, K. (2008). Holistic Package Design and Consumer Brand Impressions. *Journal of Marketing*, 72(3):64–81.

Patrick, V. M. and Hagtvedt, H. (2011). Aesthetic Incongruity Resolution. *Journal of Marketing Research (JMR)*, 48(2):393–402.

Pieters, R., Wedel, M., and Batra, R. (2010). The Stopping Power of Advertising: Measures and Effects of Visual Complexity. *Journal of Marketing*, 74(5):48–60.

Ranganath, R., Tang, L., Charlin, L., and Blei, D. M. (2014). Deep Exponential Families. *arXiv:1411.2581v1*, pages 2–4.

Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). Stochastic Backpropagation and Approximate Inference in Deep Generative Models. *Proceedings of the 31st International Conference on Machine Learning*.

Schlosser, A. E., Rikhi, R. R., and Dagogo-Jack, S. W. (2016). The Ups and downs of visual orientation: The effects of diagonal orientation on product judgment. *Journal of Consumer Psychology*.

Semin, G. R. and Palma, T. A. (2014). Why the bride wears white: Grounding gender with brightness. *Journal of Consumer Psychology*, 24(2):217–225.

Spence, C. (2012). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology*, 22(1):37–54.

Valdez, P. and Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General*, 123(4):394–409.

van der Lans, R., Cote, J. a., Cole, C. a., Leong, S. M., Smidts, A., Henderson, P. W., Bluemelhuber, C., Bottomley, P. a., Doyle, J. R., Fedorikhin, A., Moorthy, J., Ramaseshan, B., and Schmitt, B. H. (2009). Cross-National Logo Evaluation Analysis: An Individual-Level Approach. *Marketing Science*, 28(5):968–985.

Walsh, M. F., Winterich, K. P., and Mittal, V. (2010). Do logo redesigns help or hurt your brand? The role of brand commitment. *Journal of Product & Brand Management*, 19(2):76–84.

Wu, M. and Goodman, N. (2018). Multimodal generative models for scalable weakly-supervised learning. In *Advances in Neural Information Processing Systems*, pages 5575–5585.

Appendices

A. Logo Feature Details

In the following table, we explain all of our logo features, and their bases in the literature. Note that the features are grouped, according to their theoretical basis in the literature. In the model, each feature is treated independently.

Category	Feature	Description	Value	Literature
Color	Color	Whether a given color is present	Binary	Valdez and Mehrabian (1994); Klink (2003); Deng et al. (2010); Semin and Palma (2014); Kareklas et al. (2014)
	Dominant Color	The color with the highest number of pixels	Categorical	
	Accent Color	All colors that are not the dominant color	Binary	
	% Whitespace	How much of the logo (mark)'s convex hull is background (whitespace)?	Real	
	Mean Saturation	The mean value of the saturation channel across pixels in HSV colorspace	Real	
	SD Saturation	The standard deviation of the saturation channel	Real	
	Mean Lightness	The mean value of the value channel in HSV colorspace	Real	
	SD Lightness	The standard deviation of the value channel	Real	
Format and Shape	Has Mark	Is there a mark?	Binary	Navon (1977); Klink (2003); Orth and Malkewitz (2008); Walsh et al. (2010) Spence (2012)
	Size	How much of the logo does the mark take up	Real	
	Number of Marks	How many marks there are	Count	
	Convex hull	The smallest convex polygon that fully contains the logo, classified into types	Categorical	

	Standardized shape	The mark is standardized into a 25×25 pixel shape, then clustered pixelwise, weighted by size, which captures similarity in both shape and size of the mark	Categorical	
	Aspect Ratio	The ratio of the height and width	Real	
	# Corners	The number of corners found by the Harris corner detector	Count	
Font	# Characters	Number of logo segments classified as characters	Count	Doyle and Bottomley (2004); Henderson et al. (2004)
	Serif	Classification of characters into serif, sans-serif, or calligraphic fonts	See footnote[11]	
	Family	Vox-ATypI font families		
	Italics	Upright versus italic characters		
	Weight	Original, bold, or light characters		
	Width	Original, condensed, or wide characters		
Complexity	# Colors	How many distinct colors are there?	Count	Henderson and Cote (1998); Janiszewski et al. (2001); van der Lans et al. (2009); Pieters et al. (2010)
	# Segments	How many distinct regions are there?	Count	
	Perimetric complexity	A measure of shape complexity, given by the ratio of the number of edge pixels to interior pixels, where the edge pixels are computed via canny edge detection	Real	
	Greyscale entropy	The local average variance of greyscale pixel intensity	Real	

[11] The basic type of all font variables is a count: for every identified character, we match it to one element in our font dictionary, which then determines all of the font properties. Thus, the basic feature is a count of how many times each font feature appears (e.g. 5 bold letters, 4 geometric fonts). However, noting that this matching is just a noisy approximation of the font features, we also form features from these counts. For example, we may model the dominant font family, or sans versus serif, as a categorical variable, where the outcome is the family or type with the highest count.

Symmetry	Horizontal Symmetry	The correlation in pixel values when the image (mark or logo) is split in half horizontally (i.e. left and right halves)	Real	Henderson and Cote (1998); van der Lans et al. (2009)
	Vertical Symmetry	The correlation in pixel values when the image (mark or logo) is split in half vertically (i.e. top and bottom halves)	Real	
Repetition	Size Repitition	The standard deviation of the sizes of the subcomponents of the image (mark or logo)	Real	Henderson and Cote (1998); van der Lans et al. (2009)
	Complexity Repitition	The standard deviation of the perimetric complexity of the subcomponents of the image (mark or logo)	Real	
Orientation	Position	The position of the mark relative to the text. We compute both hard and soft versions of this metric: for example, hard left means the mark is entirely to the left of the text, whereas soft left means that the center of the mark is to the left of the center of the text.	Binary	Chae and Hoegg (2013); Cian et al. (2014);
	Edge Gradients	The percentage of non-zero edge gradients classified as horizontal, vertical, up-diagonal, or down-diagonal, computed by traversing the binarized logo in both left-right and top-down directions and computing numerical gradients.	Real	Deng and Kahn (2016); Schlosser et al. (2016)

B. Technical Details on the Logo Feature Extraction Algorithm

In this section, we give more of the technical details of our image processing algorithm. For specific features, see Appendix A. The basic data representation of images is the raster array, which defines an image by an $h \times w$ grid of color values. The grid cells are called pixels, and the colors are typically broken down according to an underlying color model. The most common color model is the red-green-blue (RGB) system, which defines the full spectrum of colors by intensities on red, green, and blue color channels. Most image analysis algorithms are based on this representation of an image, and most data analysis software imports images in this form. An alternative representation, which we make use of in our own image processing algorithms, is the hue-saturation-value (HSV) color model, which is a cylindrical coordinates transformation of the RGB color space. It defines colors in terms of their hue, meaning the basic color itself, saturation, meaning how "intense" the color is, and value, which refers to how bright the color is. Finally, greyscale images can be also represented through raster arrays as a single decimal value at each pixel, representing the intensity of light at that pixel.

B.1. Color Quantization through Density-based Clustering

The algorithm begins by learning how many distinct colors are in a given logo through a density-based clustering algorithm. Specifically, we employ the DBSCAN algorithm, which is a popular clustering algorithm which does not rely on a pre-specified number of clusters or distributional assumptions (Ester et al., 1996). Rather, it uses a density criterion to automatically determine both the number of clusters and cluster membership. DBSCAN is ideal for this application, as we know exactly the nature of the colorspace on which we are clustering, allowing us to specify a sensible density cutoff. Moreover, it is robust to noise.

We perform DBSCAN clustering on the HSV colorspace, which is a cylindrical coordinate transformation of the RGB colorspace that separates out the actual color value (hue) from other aspects of the color (saturation and lightness, also called value). Because of the cylindrical nature of the colorspace, hue (i.e. color) is represented along a circle, and hence the clustering must also operate over a circle, as shown in Figure 21. This is another benefit of DBSCAN: it does not rely on any assumptions about the distributions of the points or the geometry of the space, besides for being able to specify a suitable density metric.

A downside of DBSCAN is that it can be computationally inefficient, and the logos in our dataset can be quite large. Thus, we typically do DBSCAN on a random selection of pixels. Once we have identified the number of clusters through that, we use those same cluster centers in the standard k-means algorithm. The end result of the clustering is an assignment of each pixel in the original logo to a color cluster, or to the background. This is referred to as color quantization.

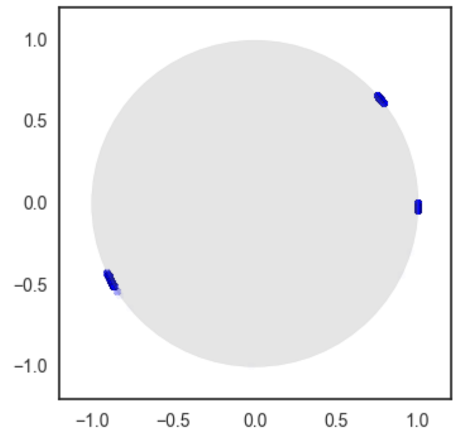
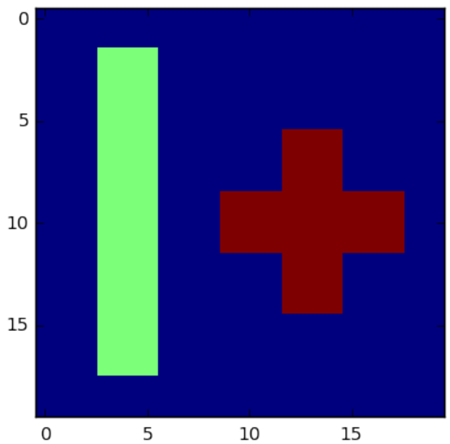


Figure 21: The three colors from Burger King's logo (blue, red, and yellow), plotted as the Hue value from HSV in polar coordinates. Here, red is the cluster of points at right, yellow is the cluster in the top-right, and blue is the cluster in the bottom-left. This is the space on which the DBSCAN clustering operates.

B.2. Region-based Segmentation

Computationally, quantizing the logo reduces the three dimensional raster array into a two dimensional matrix of cluster assignments. This is illustrated in Figure 22. Given this format, determining distinct regions of the logo is often as simple as identifying connected regions of this matrix, and this, plus some steps to filter out noise and very small image segments, is how our algorithm proceeds. However, there are two complications. The first relates to text: in practice, some fonts are condensed to the point that two letters are slightly joined, leading the algorithm to think there is only one connected region, when there are in fact two distinct letters. The second complication relates to the mark, and is in some sense the inverse of the first: sometimes, a single mark may consist of several very closeby regions.

To address the first concern, we employ mathematical morphology, specifically the erosion and dilation operations. Erosion is a standard image processing technique that works on binarized images (background = 0, foreground = 1), transforming that image by assigning each pixel in the transformed image the minimum value within a pre-defined neighborhood of that pixel in the original binary image. Dilation is similar, but employing the maximum. In practice, what this means is that in erosion, connected regions are typically shrunk, whereas in dilation, they are expanded. To use these operations to help separate barely connected letters, we employ the following three steps: first, for every region isolated in the basic segmentation, we apply erosion, and identify any subregions generated by that erosion. Second, we separate those subregions, and then dilate them to approximately their original form. Finally, we run each of these new features through the font identification system defined in the next section. If any of them is identified as a font, the old region is discarded in favor of the subregions.



$$
\begin{matrix}
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&2&2&2&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&2&2&2&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&2&2&2&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&2&2&2&2&2&2&2&2&2&0&0\\
0&0&0&1&1&1&0&0&0&2&2&2&2&2&2&2&2&2&0&0\\
0&0&0&1&1&1&0&0&0&2&2&2&2&2&2&2&2&2&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&2&2&2&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&2&2&2&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&2&2&2&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0
\end{matrix}
$$

Figure 22: An example of color quantization: the image at left is quantized, yielding the matrix representation at right, where 0 corresponds to blue, 1 to red, and 2 to green.

To address the second concern, we again apply DBSCAN clustering, this time using position on the logo as the quantity of interest. We set the density in the DBSCAN algorithm according to the size of the logo. This then finds mark pixels that are close together, regardless of whether or not they are actually connected.

B.3. Font Identification

For each of the segments identified through the above procedure, we first try to match them to a font. To do that, we standardize each segment to a grayscale 25×25 pixel representation, then apply template matching against our extensive collection of fonts, which have also been converted to the same representation. This representation is equivalent to representing each segment, and each font instance, as a length 625 vector, with values between 0 (black) and 1 (white). By template matching, we mean a simple distance calculation between the segment of interest, and each member of our font dictionary. In practice, this takes the form of a correlation between the entries in the segment vector and the entries in each font instance vector. We use a fairly simple heuristic to identify whether a segment represents a character: if the correlation between the segment and any font instance is greater than a certain cutoff, we say it is a match, and say that the segment matches the font with the highest correlation. We use different cutoffs, depending on the complexity of the segment, where complexity is measured by the perimetric complexity (the ratio of edge pixels to interior pixels). This is important because some letters, like i (which is represented without the dot), l, and o are very similar to commonly occurring mark features.

B.4. LAB Color Clustering

The colors within a given logo are represented in the continuous RGB color space. To convert these color triples to meaningful dictionary items, we then run another clustering algorithm on these triples across logos.[12] However, in order to cluster the colors, we need a sensible distance metric in this space. While RGB colors are the standard for computer representation, it is well established that distances in RGB color space do not correspond well to distances in human perceived distance. To rectify that, we employ another colorspace transformation, from RGB to the CIE-LAB (also just called LAB) colorspace, which is designed such that distances in colorspace correspond to differences in human perception of color (McLaren, 1976). Then we perform standard K-means clustering, resulting in the color dictionary shown in Figure 3.

B.5. Hull and Mark Clustering

To cluster both the hulls and the marks, we apply a similar procedure described above for fonts: we convert each hull and each mark to a 25×25 standardized greyscale representation, and then apply ordinary k-means clustering over the resultant length 625 vectors, determining the optimal number of clusters via scree plots. The only challenge is for the marks: the standardization procedure discards information about size. Yet, we also want to capture the different sizes of marks: a mark that forms the background of, and thus takes up 80% of a logo is different than one that takes up only 10%. To take this into account, we include an additional term in the clustering of marks, that adds weight to the fraction of the the logo's area taken up by the mark.

C. Additional Model-free Analyses

The goal of these analyses is to see whether or not the brand personality and the industry category of the brand explain anything about a firm's logo. To do that, we considered all logo features as real-valued outcomes, and ran naive OLS regressions, saving the adjusted R-squared value from each.[13] We did this analysis in three separate batches: (1) predicting logos from industry, (2) predicting logos from brand personality, and (3) predicting logos from both together.

In Tables 13 and 14, we present the results for the most and least explained variance features, from regressions 1 and 2. In general, we find that brand personality scores capture much more variance than the industry codes, though this may also be attributed to the greater variance in the continuous brand personality scores, versus the binary industry labels. We find that features pertaining to the color palette tend to be the easiest to

[12] The number of clusters both in this step and others was determined by the researcher, using scree plots.

[13] In many cases, the true variable is not real valued (see Appendix A), but rather binary, and thus this approach will sometimes be underpowered.

Most			Least		
Feature	R^2	Adjusted	Feature	R^2	Adjusted
SD: sat	0.249	0.201	Dom. color: grey dark	0.037	-0.025
Mean: sat	0.187	0.135	Width: mixed	0.042	-0.02
Perc. white	0.164	0.111	Mark class: thin vertical rectangle	0.044	-0.017
GPC	0.137	0.081	Mark pos: absright	0.049	-0.012
Hor. symmetry	0.132	0.076	Mark class: wispy horizontal lines	0.049	-0.012
Color: yellow	0.13	0.074	Mark pos: top	0.053	-0.008
Font weight: bold	0.121	0.065	style mixed	0.054	-0.007
Hull type: rectangle-oval thin	0.12	0.063	Mark class: simple shapes	0.054	-0.006
Color: black	0.119	0.062	Mark class: long horizonal	0.054	-0.006
Down diagonals	0.118	0.062	Mark class: bad letters	0.054	-0.006

Table 13: The ten logo features with the most and least variance explained by brand personality, as captured by simple OLS.

Most			Least		
Feature	R^2	Adjusted	Feature	R^2	Adjusted
Hor. symmetry	0.147	0.084	Mark class: bad letters	0.025	-0.046
SD: sat	0.141	0.078	Dom. color: brown	0.038	-0.033
Mean: light	0.14	0.078	Dom. color: red dark	0.044	-0.026
Horizontal edges	0.135	0.072	Mark pos: bottom	0.045	-0.025
Perc. white	0.117	0.052	Mark pos: bot	0.047	-0.023
Entropy	0.114	0.049	Color: red dark	0.047	-0.022
Dom. color: blue medium	0.113	0.048	Mark class: bulky hollow geometric	0.048	-0.022
SD: light	0.111	0.046	Mark class: hollow circles	0.048	-0.022
Color: blue medium	0.111	0.046	Dom. color: blue dark	0.049	-0.021
Hull type: rectangle-oval thin	0.105	0.039	Mark class: long horizonal	0.052	-0.018

Table 14: The ten logo features with the most and least variance explained by industry codes, as captured by simple OLS.

Feature	Industry	BP	Both	Feature	Industry	BP	Both
SD: sat	0.078	0.206	0.222	down diag	0.022	0.069	0.081
Mean: sat	0.03	0.145	0.159	SD: light	0.046	0.052	0.077
Perc. white	0.052	0.123	0.157	Color: grey dark	0.008	0.051	0.073
Hor. symmetry	0.084	0.074	0.128	Color: black	0.024	0.051	0.065
Mean: light	0.078	0.055	0.104	Entropy	0.049	0.026	0.06
Horizontal edges	0.072	0.055	0.103	# Chars	0.011	0.052	0.057
Color: yellow	0.005	0.08	0.093	Color: red	0.038	0.045	0.056
GPC	0.019	0.081	0.09	# Colors	0.028	0.027	0.049
Hull type: rectangle-oval thin	0.039	0.062	0.083	ar	0.033	0.032	0.046
Font weight: bold	0.01	0.07	0.083	# Regions	0.016	0.052	0.046

Table 15: The 20 highest *adjusted* R^2 values from predicting logo features with both brand personality and industry codes, compared to the same adjusted R^2 from just the industry code model, and just the BP model. We see in almost all cases, a modest increase in adjusted R^2 from considering both sets of predictors jointly. Note that the number in the BP column may be slightly different than in Table 13, as several firms were missing industry codes, and had to be excluded.

explain in both cases, including the mean and variance of the HSV colorspace's saturation and lightness (value) channels, the percentage whitespace, and a few of the color variables. Interestingly, in both cases, the degree of horizontal symmetry is also well explained, as do various aspects of complexity, including perimetric complexity and entropy. The variables that are least explained by BP and industry tend to be those that either relate to the mark class, or those that tend to have very few observations associated with them, like logos with mixed font styles, or logos with the mark at the bottom.

In Table 15, we show what happens to the adjusted R-squared in regression 3, when we include both brand personality and industry codes in simple OLS to predict logo features. This illustrates the importance of jointly considering both what the firm *does*, as well as the firm's *brand identity*: in almost all cases, we find that the adjusted R-squared of including both sets of predictors is higher than either of the models in isolation. As this is adjusted for the number of predictors, this indicates that there is explanatory power by considering both sets of variables jointly.

D. Simulating More Brand Identities

Generating identities from the model is straightforward. In this section, we present several additional simulations, albeit in less detail that the cold, modern corporation above. Each of these was generated simply by evaluating each of the decoder network at a vector of 10 standard normal draws.

Sophisticated Media The following corresponds to a brand identity with

$$z = (-1.60, 0.45, -0.71, -1.35, -1.29, 1.50, -1.36, 0.01, 1.23, -1.33):$$

- Relatively likely words: 'book', 'physic', 'televis', 'word', 'step', 'decemb', 'sophist', 'someth', 'pleas', 'readi'
- Relatively unlikely words: 'communiti', 'can', 'custom', 'compani', 'global', 'solut', 'servic', 'innov', 'work', 'provid'
- Top three relative brand personality traits: glamorous, trendy, exciting
- Bottom three relative brand personality traits: masculine, hard working, wholesome
- Some likely visual features: black dominant color, yellow and light green accent colors, light font, no italics, geometric font class, has a mark

From these traits, we label this a sophisticated media firm.

Family Friendly Food The following corresponds to a brand identity with

$$z = (-1.12, 0.22, 0.04, -1.22, 1.17, 0.56, 0.28, 0.91, 1.11, 0.83):$$

- Relatively likely words: 'www', 'televis', 'central', 'happen', 'mutual', 'dollar', 'ingredi', 'ultim', 'hand', 'kind'
- Relatively unlikely words: 'employe', 'technolog', 'solut', 'global', 'new', 'custom', 'work', 'innov', 'servic', 'provid'
- Top three relative brand personality traits: cheerful, friendly, family-oriented
- Bottom three relative brand personality traits: rugged, tough, masculine
- Some likely visual features: brown dominant color, red and yellow accent colors, bold font, geometric font class, has a mark

From these traits, we label this a family-friendly food firm.