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

November 25, 2019

Abstract

Logos serve a fundamental role as the visual figureheads of brands. Yet, due to the difficulty of using unstructured image data, prior research on logo design has largely been limited to non-quantitative studies. In this work, we explore logo design from a data-driven perspective. We develop both a novel logo feature extraction algorithm that uses modern image processing tools to decompose pixel-level image data into meaningful features, and a multiview representation learning framework that links these visual features to textual descriptions of firms, industry tags, and consumer ratings of brand personality. We apply this framework to a unique dataset of hundreds of brands. Our model is able to predict which brands use which logo features, and how consumers evaluate these brands' personalities. Moreover, we show that manipulating the model's learned representations through what we term "brand arithmetic" yields new brand identities, and can help with ideation. Finally, through an application to fast food branding, we show how our model can be used as a decision support tool for suggesting typical logo features for a brand, and for predicting consumers' reactions to new brands or rebranding efforts.

Keywords: logos, branding, machine learning, multiview learning, image processing, Bayesian estimation

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

1. Introduction

Logos, which adorn everything from product packaging to advertising, are the most distinct marks used by brands. Designers create logos to represent the essence of brands, and firms motivate logo redesigns to convey new ideas or communicate a new positioning. Yet, despite the clear 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 the science underlying logo design can be captured mathematically, through a multiview representation learning framework that models the linkages between a brand's function, its logo features, and consumer perceptions of its brand personality. We then use this framework to study real logos of firms in different industries, to derive a semantic understanding of logos, and to show how a model can be used for aiding firms in logo design.

Our data-driven, multiview learning treatment of logos allows us to quantify the branding and design process from three related perspectives, which vary in terms of which aspects of brand identity are considered given, and which are the focal outcomes or outputs:

1. **The designer's perspective.** Given a description of a brand and a desired consumer-level perception, which logo features are most commonly used to achieve that identity? This question mirrors the design process, where a designer uses a company-supplied brief to design a logo. From our model-based perspective, answering this question relies on being able to use textual data and a target brand perception as inputs to predict logo features.

2. **The brand manager's perspective.** Given a newly designed logo, how will consumers perceive it? Or, given a set of candidate logos, that may vary on key design elements, which logo best matches a company's targeted brand perception? Answering such questions is of relevance to brand managers, and requires being able to use a logo and a brand profile as inputs to predict consumers' evaluations of the brand.

3. **The researcher's perspective.** What associations exist among logo features, brand function, and brand perception? Branding researchers are interested in understanding such broad linkages across different facets of brand identity. In addition, researchers with a focus on visual identity and marketing aesthetics are even more specifically interested in understanding

how particular logo features impact consumer perceptions about a firm's function and personality. By considering logo features as inputs, and text and brand personality as outcomes, we can understand how different logo features contribute to perceptions of firm function and brand identity.

As the above perspectives are intertwined, addressing them requires a flexible modeling approach that jointly represents all facets of brand identity. In this paper, we develop such a framework, using a novel logo feature extraction algorithm based on image processing techniques, in tandem with a flexible, deep generative model that distills multimodal data into manipulable, numeric vector representations.

Our results from applying this framework to a unique dataset of hundreds of successful brands, containing logo data, textual descriptions, industry tags, and consumer brand personality perceptions, indicate that the logo design process practiced by the firms in our study is quite systematic: from the designer's perspective, we find that a model-based approach can predict many logo features from text, industry, and brand personality descriptors. Similarly, from the manager's perspective, we find that by knowing brand function and the brand logo, we can predict how consumers will evaluate the brand. From the researcher's perspective, we find support for many findings from the literature on how aesthetics influence consumer judgments. Moreover, we find that our learned representations can, indeed, capture many intricate aspects of visual branding, and can be used for ideation and decision support. However, we also find that it is generally difficult to predict how consumers will evaluate brands based solely on logos.

Beyond these specific findings, our work makes several contributions. Foremost, it is the first paper to study real 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 subjective domain. Moreover, the design recommendations from the model can be used even by budget-strapped firms to thoughtfully design their logos. Finally, by representing all facets of a brand identity using a multidimensional latent space,

our framework allows designers to interpolate between different brands to yield novel combinations of existing identities, thus facilitating the creative process.

From a methodological perspective, ours is among the first papers in marketing to directly use image data, without relying on human coders. Distinct from recent work in marketing that has used deep learning frameworks to extract brand-relevant attributes from natural images (Liu et al., 2018), our work presents a novel image processing approach to automatically extract features from pixel-level image data, uniquely tailored to studying logos. Our feature extraction algorithm decomposes logos into *meaningful* features, which are driven by prior theory about logo semantics. These features form a "visual dictionary" that describes logos in a way that is meaningful to designers, and amenable to probabilistic modeling. The automatic nature of our feature extraction methods make them widely applicable and scalable, without the need for human coders.

Our work is also among the first in marketing to synthesize both unstructured text and image data. To do so, we develop a multimodal variational autoencoder (MVAE), which is an extension of the popular variational autoencoder (VAE), a deep learning framework used for learning representations of data (Kingma and Welling, 2013; Rezende et al., 2014). Our framework learns joint multiview representations of the different facets of brand identity present in our data: text, logo, brand personality, and industry. Distinct from supervised deep learning models that have been successfully employed in a number of recent marketing studies (e.g., Liu et al., 2018; Liu et al., 2019), our MVAE is a semi-supervised generative model (Kingma et al., 2014) that learns a posterior distribution over latent parameters that capture the joint statistical properties of all of these data modalities. This multiview representation learning approach (Li et al., 2016) allows us address design from each of the distinct perspectives outlined above, rather than limiting us to making unidirectional predictions.

To infer the latent representations of brands, we introduce task-specific inference networks that approximate the posterior distribution of a brand's latent representation using only a subset of the available data modalities. In doing so, our inference procedure mirrors the decision support contexts in which our model can be used. For example, to mirror the designer's task of designing a logo given a textual brief and a target personality, we learn a task-specific designer inference network, that takes as inputs text describing a brand, industry tags, and a target brand personality profile, and outputs a posterior distribution over that brand's representation, which can then be used to

generate a set of suitable logo features. This novel approach to inference aids in the relevance of our work to design and branding practice, as it provides a natural set of decision support tools that can be used to guide each of these distinct tasks.

The rest of the paper is organized as follows: in Section 2, we review the literature on logo design and aesthetics in marketing. In Section 3, we describe the unique dataset we compiled to calibrate our model. In Section 4, we briefly describe our logo feature extraction algorithm, leaving a more detailed description to our web appendix. In Section 5, we present descriptive "model-free" evidence of the links between design, brand personality, and firm function. In Section 6, we develop a multiview 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 how the learned representations can be used for ideation, and how the task-specific inference networks can be used as decision support tools in a data-driven design process. Finally, we conclude with a summary and directions for further study.

2. Literature

There is a sizable literature, especially in consumer behavior, on how consumers react to logos and marketing aesthetics. 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 study the mechanisms governing consumers' reactions to various visual stimuli. In this section we briefly review these findings, with the specific focus of informing our logo feature extraction algorithm, described in Section 4.

2.1. Logos

A limited amount of research in marketing has studied logos, starting with Henderson and Cote (1998), who investigated how logo characteristics impact recognition and affective reactions of consumers. In particular they studied the NHE dimensions (*naturalness*, the extent to which it contains natural shapes; *harmony*, the extent of its symmetry and balance; *elaborateness*, i.e., complexity as measured by the number and heterogeneity of logo elements). Subsequently, Henderson et al. (2003), and van der Lans et al. (2009), tested these NHE dimensions across cultures and found

them to be universally good descriptors of design.

Other behavioral researchers have used experimental manipulation of fictional logos to study consumer reactions and the psychological mechanisms that underlie such reactions. Klink (2003) related the color, size and shape of logos to brand names, Walsh et al. (2010) studied the impact of moving from an angular logo to a round one, and Jiang et al. (2015) showed that the circularity or angularity of the logo affects customer perceptions of hardness or softness, which in turn influence attribute judgments about products. Other studies have looked at the orientation of logos. Cian et al. (2014) found that the horizontal orientation of a logo or the positioning of its elements can evoke the idea of movement to influence consumers' engagement and attitudes. More recently, Schlosser et al. (2016) found that upward diagonals convey greater activity than downward diagonals, leading to more positive reactions. Researchers have also analyzed the impact of the font and typeface used in logos on consumer likelihood to choose a product, and the appropriateness of these characteristics for particular industries (Doyle and Bottomley, 2006). Hagtvedt (2011) showed that incomplete typeface can lead to perceptions of untrustworthiness and increased innovativeness.

Together, these studies imply that NHE dimensions and other objective measures such as the color, angularity, orientation, font and typeface within a logo are important to consider in developing a quantitative modeling approach to support logo design.

2.2. Aesthetics

There is a large body of work on aesthetics and perceptions within marketing and psychology. Here we selectively review results that are relevant to our focus on identifying features for logo design. Research in this domain has also emphasized the roles of color, font, orientation, and other factors on how humans perceive and respond to visual stimuli.

Deng et al. (2010) studied consumers' preferences for color combinations in product design. Their study shows that of the three common dimensions of color—hue, saturation, and lightness—people tend to de-emphasize lightness, relative to the other two. In addition, people prefer a small number of generally similar colors, but with a single contrasting color that highlights a single distinctive element. Kareklas et al. (2014) showed 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) found that white is perceived as more

feminine, whereas black is perceived as more masculine. Psychological work has looked at the effect of color on emotional response. For example, Valdez and Mehrabian (1994) found that of the three key color dimensions, saturation and lightness drive emotional responses along the pleasure, arousal, and dominance dimensions. They also found that shades of blue, green, and purple are experienced as being most pleasant, and shades of yellow as least pleasant.

Font and typeface have also been explored in advertising and impression management contexts. Childers and Jass (2002) explored the influence of semantic connotations of typeface on consumers' ratings of products. Henderson et al. (2004) analyzed many extant fonts based on the typology literature and ratings of experts to uncover factors—pleasing, engaging, reassuring, and prominent—that describe typeface impressions, and six factors—elaborate, harmony, natural, flourish, weight, and compressed—that describe typeface design. They concluded that there may be universal design elements that can help managers in impression management.

Other research has shown that the orientation of stimuli can influence peoples' perception of products. For example, Meyers-Levy and Peracchio (1992) showed that the camera angle of an ad featuring a product can influence judgments of the product. Chae and Hoegg (2013) found that in cultures where reading is done from left to right, products are viewed more favorably when positioned congruently with this spatial orientation. Deng and Kahn (2016) found that a product image's location on its packaging influences the item's perceived weight.

Many other aesthetic aspects that may be relevant for logo design have also been studied. For example, Navon (1977) found that global features are processed more readily and fully than local ones. More recently, Pieters et al. (2010) used eye-tracking to study two distinct aspects of visual complexity of advertisements: feature complexity and design complexity. Feature complexity refers to variation in basic features like color and edges, and is measured by variance at the pixel level, while design complexity pertains to variation in the elaborateness of the design, and is measured by six general principles: the quantity, irregularity, dissimilarity, and detail of objects, and the asymmetry and irregularity of object arrangement.

Relevant to how brand constructs relate to visual elements, Orth and Malkewitz (2008) decomposed package design into five distinct types and prescriptively related these to brand personalities. Spence (2012) discussed cross-modal effects such as visual perceptions associated with tastes and textures (e.g., the angularity of carbonation or bitterness), which could be relevant determinants of

logo design. Spence argued 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 by Patrick and Hagtvedt (2011) that discussed consumers preferences for congruity in the consumption experience (e.g., a fancy logo matching a fancy experience).

In summary, the literature described above has primarily used experimental approaches to identify a number of visual features that influence consumer perceptions and reactions. We use these features to guide the design of our logo feature extraction algorithm, which we describe subsequently. Unlike many of these studies, our work does not study the effects of single logo features in isolation on consumer perceptions, but rather examines logos holistically, exploring how visual features combine to convey meaning in logos in practice. To that end, our work also differs from the above literature in our use of a large number of real logos to understand and model the multimodal associations between logos, firm descriptors, and brand personality measures.

3. Data

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

Our modeling approach focuses on learning the links between existing logos and these other components; hence, for our approach to be meaningful for good design practices, we must ensure that the firms for which we gather data 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, based on 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 firm Brand Finance, or listed in the Forbes 500 in 2016. There was a large degree of overlap between the lists, leaving us with a sample of 715 brands. In data processing, we further eliminated firms with little textual data, resulting in a final set of 706 brands.

Logos: Firms 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 use 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 used the version that appeared most commonly on the firm's online marketing materials. When multiple logo versions were prevalent, we selected the logo with a white background, and with both logotype and mark, if such a logo is in use.

Text: To understand the link between logo features and how the firms think about themselves, we collected textual descriptions consisting of both functional and brand-relevant text taken directly from firms' websites. We collected this data in two batches: First, 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.

After gathering all this textual data, we applied standard text processing algorithms, to create a dictionary of brand and firm descriptors. We first tokenized and stemmed the words, removing stop words. We then removed all words that appeared in fewer than 20 of our 715 original brands. This left a dictionary of 852 words. Finally, we removed brands that contained fewer than 20 of these 852 words, leaving us with our final sample of 706 brands.

Industry Labels: In addition to the textual descriptions, as a simpler measure for capturing what firms do, we also collected *industry labels* from Crunchbase, a database commonly used by investors. Crunchbase offers a set of standard tags describing what firms do. For example, Uber has the labels

Customer Service, Mobile Apps, Public Transportation, Ride Sharing, and Transportation. We have 615 labels across our companies. These are further organized into category groups reflecting similar activities. For example, Public Transportation, Ride Sharing, and Transportation are all categorized under the group Transportation. We use these groups as our industry labels. We retain labels that apply to at least 10 firms, leaving us 34 industry labels.

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. We then asked 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.[1] We gathered 20 responses per brand, and use the average response on each of the 42 traits as our data. In some of the subsequent visualizations, we also group the brand personality traits according to the factor structure outlined in Aaker (1997) by taking the average of all traits assigned to a given factor.

4. Logo Feature Extraction

Modeling visual objects such as logos is difficult because of the need to work with unstructured image data. The computer vision and machine learning literatures have developed two broad approaches for incorporating images in models. The first approach uses raw pixel-level data as the input to a model. This is common, for example, in models of image recognition or image captioning, which typically use a neural network for supervised prediction. The second approach begins by processing the image to yield a "dictionary" of representative image features that are then used as inputs to a model. We follow the second approach: we first use our novel logo feature extraction algorithm, which is based on modern image processing methods, to process the logo images into logo features, and then incorporate these features in a model of design. Our feature extraction

[1]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.

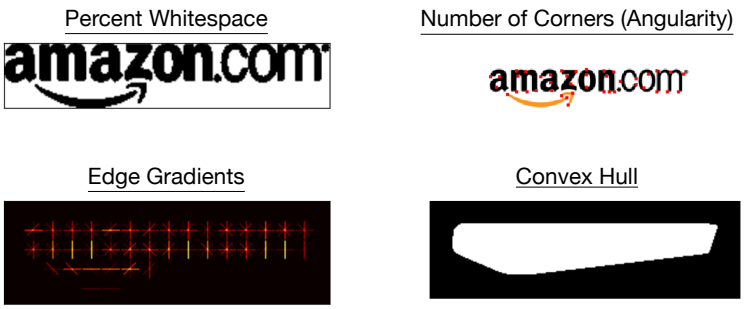


Figure 1: Examples of global features, using Amazon's logo as an example. Percent whitespace captures the percentage of pixels that are white (background), within the convex hull of the logo. The number of corners is a measure of angularity computed via the Harris corner detector. Edge gradients capture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

algorithm is rooted in the literature on logo design and consumers' responses to aesthetics, and distills logos into components that are meaningful for consumers and designers. When combined with the framework described in Section 6, this approach yields an interpretable machine learning framework, which is an important advantage over less structured approaches. Each of our logo features is human-interpretable, which is crucial for the model based on them to be useful in decision support.

4.1. Algorithm Overview

Our algorithm has four 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. One such computation involves density-based color quantization that gives the number of distinct colors in each logo. In the second stage of the algorithm, which we term *segmentation*, we assign each logo pixel to one of these colors and 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 then separate them into characters and marks. This third *character-identification* stage uses a template matching procedure to separate out characters from marks, and identify an approximate font used in the logo, if applicable. 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. A detailed description of these stages is available

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).

in the web appendix. We now describe the different logo features that we extracted.

4.2. Visual Features

A listing of all of our visual features, including their descriptions and connections to the previous literature, is available in the web appendix. Here, we briefly describe the logo features, grouping them into color, format, shape, font, and other features for expositional convenience.

Color: The full color dictionary, computed by clustering the colors across all our logos, is given in Figure 3. Apart from just computing which colors are present in a logo, our algorithm also identifies the dominant color (one per logo) and accent colors (all colors except the dominant color). It also computes the extent of white space within the convex hull (which is the smallest convex polygon that contains all of the non-background pixels) 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 include features that capture the presence of a mark in the logo, the number of marks, and the aspect ratio of the logo. We also cluster the the set of convex hulls across our logos to form a dictionary of logo shapes, shown in Figure 4. Similarly, we standardize the shape of each mark, convert it to greyscale, and then cluster all marks into 14 representative mark types. We give examples of these classes in Figure 5.

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: Color dictionary: 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.

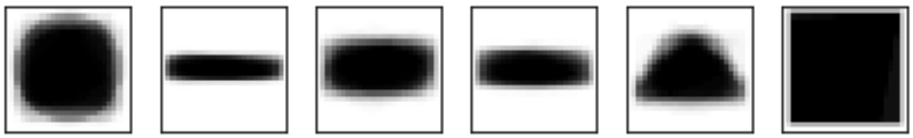


Figure 4: Hull classes: the six typical shapes of logos, as characterized by their convex hulls. Each logo in our dataset is assigned to one of these classes.

Cluster	Sample of marks
6	
7	
9	

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

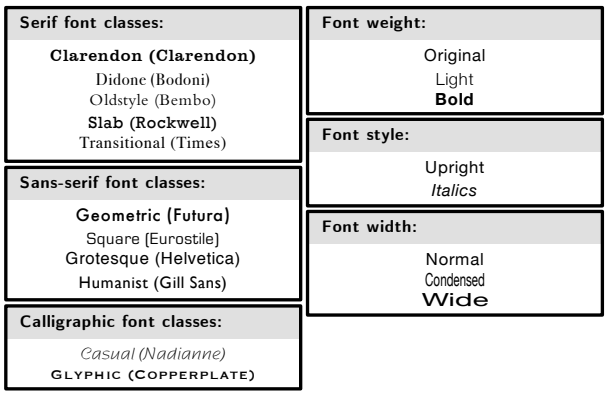


Figure 6: Font classification system employed by the algorithm: fonts were matched to a font class, weight, style, and width.

Font: Font is a crucial feature of logos. We therefore developed a procedure to identify and describe characters and their fonts. Specifically, we apply a template matching procedure to match each logo 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 styles, including fonts from all Vox-ATypI font classes, a standard font classification scheme used by font experts.[2] We illustrate our complete font typology in Figure 6.

Others: The literature review identified several other features that are important for logo design, such as complexity, symmetry, 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 use 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 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, and more details are provided in the web appendix.

Discretizing Variables: Some of our logo features are real or integer-valued. We discretize each of these features into two binary variables, corresponding to whether the logo is in the bottom or top quartile of the data for that feature. This measures whether the logo is particularly low or particularly high on a feature. For example, in discretizing the *number of corners* variable, we use two binary variables: *low number of corners*, which captures whether the logo is in the bottom quartile for number of corners, and *high number of corners*, which indicates whether the logo is in the top quartile. The only exception to this procedure is the number of colors in a logo: as the vast majority of logos have either one, two, or three colors, we convert this variable to a categorical variable with four levels: one color, two colors, three colors, or more than three colors. We have found that discretizing real and integer-valued variables improves the empirical performance of our model significantly, and also aids interpretability: it is difficult for a designer to attempt designing

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

a logo with 22 corners, but relatively easier to design one with "many" corners or "few" corners.

5. Exploring the Data

Before describing our modeling framework, we provide some model-free evidence to illustrate the interplay among logo features, firm function, as captured by the industry labels, and brand personality perceptions. This motivates the full model, by illustrating the complex relationship between logo design and firm identity. We use forest plots to visualize the linkages among these variables in an intuitive and interactive fashion. These plots show how one focal outcome variable varies as a function of another explanatory variable in binary form. 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://dr19.shinyapps.io/explore_logo_data/`.

In our data, brand personality (BP) provides an especially insightful portrait as to how consumers perceive the firm. In Figure 7, we present two forest plots that show how brand personality perceptions (the outcome variable) vary as a function of logo features. Both plots confirm with our intuition and relate to some of the findings from the literature on logos and aesthetics. The first plot compares BP perceptions (on the vertical axis) across three common dominant logo colors: black, blue, and red. The plot shows the difference in the outcome (e.g., perceived honesty of the brand) for firms that have a particular dominant color (e.g., blue) and firms that do not have that dominant color. 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 score high not only on upper class and charming, but also on outdoorsy and tough. This result, in isolation, seems surprising, as upper class and charming appear 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 simultaneously rely on many facets to build a personality-consistent logo.

The second plot of Figure 7 shows how some global features of the logo and its convex hull relate to brand personality. These features are less intuitive than color, but have been emphasized more in the literature. Moving from left to right in the plot, we find the following:

1. Horizontally symmetric logos tend to be perceived better along almost all dimensions, except

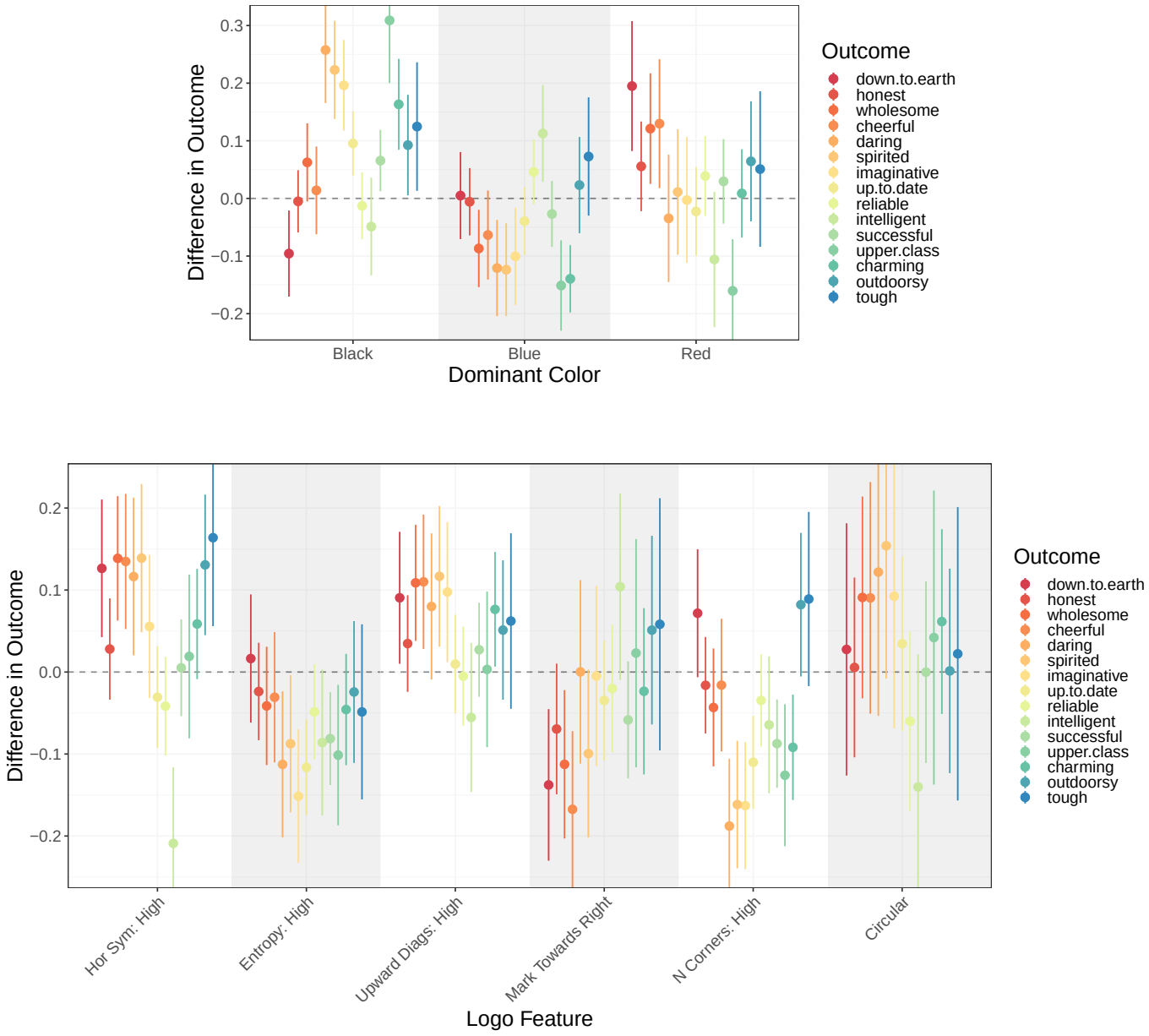


Figure 7: Each color in the plot represents a different brand personality factor, denoted in the legend. On the x-axis are features of the logo. On the y-axis is the difference in brand personality perception for firms that have a certain feature, versus firms that do not have that feature. Error bars around the points represent two standard errors.

intelligent, perhaps reflecting the role of harmony in positive affect discussed in Henderson and Cote (1998).

2. High entropy, a measure of complexity, that is similar to the concept of feature complexity in Pieters et al. (2010), is generally associated with low perceptions across the board.

3. A high proportion of upward diagonal edge gradients appears positively related with cheerful, spirited firms, which lends some support for the findings of Schlosser et al. (2016), who found that upward diagonals convey activity.

4. Placing the mark towards the right is associated with lower perceptions of down-to-earthness, honesty, and wholesomeness, but marginally higher intelligence. While not directly related to their findings, the idea that placement of the mark relative to the text matters for perceptions echoes the findings of Deng and Kahn (2016).

5. Angularity, as captured by the number of corners, is positively associated with down-to-earth and tough logos, and negatively related to the others. This appears consistent with Jiang et al. (2015), who found angularity to be associated with durability.

6. A circular hull is positively associated with cheerful, daring, spirited, but negatively associated with intelligence, supporting the findings of Jiang et al. (2015) that circularity is associated with comfortableness and customer sensitivity.

Taken together, these findings lend strong support to the idea that our features capture many of the aspects discussed in the literature.

Apart from conveying brand image, firms may rely on logos to signal the kind of product or service that customers will receive. As a simple measure of what a firm does, we use the industry labels from Crunchbase. Figure 8 shows another forest plot that visualizes the variation in the dominant color of the logo in terms of the industry labels. Again, we find that some of these relationships are quite strong and intuitive. For instance, blue is associated with financial services, but not with food and beverage, and the reverse is true for red. Black is associated with clothing and apparel, 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 again see

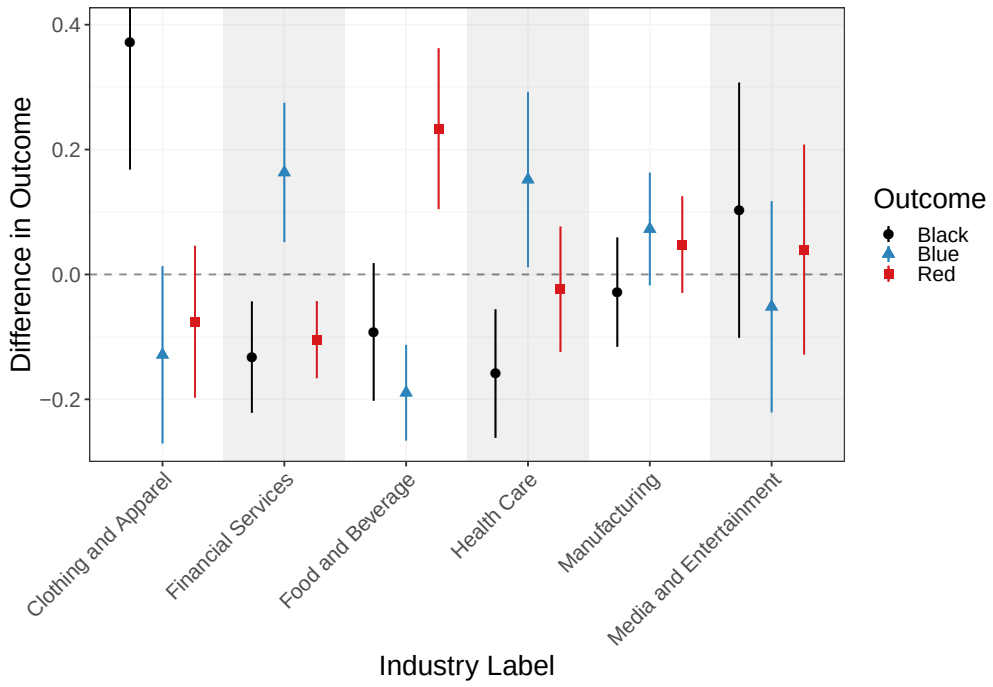


Figure 8: Forest plot for industry label and logo color: given a firm has a certain industry tag, the plots show whether that its logo is more or less likely to have one of three dominant colors: black, blue, and red. Error bars around the points represent two standard errors.

that the relationships are complex. 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 relationships in isolation: for example, how is industry related to color, or how is color related to brand personality? 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 simplified way of operationalizing what a firm does. To answer questions regarding combinations of features, and to facilitate the use of unstructured, textual data that may more accurately reflect nuances of a company, we need a model that leverages these type of data to simultaneously capture all aspects of brand identity.

6. Modeling Framework

We now describe our model for logo design. We draw on recent advances in deep generative modeling (Kingma and Welling, 2013; Ranganath et al., 2014; Rezende et al., 2014; Kingma et al., 2014) and multiview learning (Li et al., 2016; Suzuki et al., 2016; Wu and Goodman, 2018) to learn multimodal representations of brands in a joint latent space that is shared across our different data modalities.[3] Specifically, we flexibly capture the linkages among our four main data sources—the *textual* website descriptions, *logo* features, *industry labels*, and *brand personality* metrics—in a semi-supervised fashion, using a multimodal generalization of a variational autoencoder. Our representation learning approach enables us to answer questions from all three perspectives listed in the introduction (i.e., the designer's, brand manager's, and researcher's), without the need to specify one domain as the dependent variable and the others as independent variables.

6.1. Variational Autoencoders

We begin by briefly describing a simple variational autoencoder (VAE), before focusing on multimodal extensions that are relevant for our work. Variational autoencoders were proposed by Kingma and Welling (2013) and Rezende et al. (2014) as scalable mechanisms for estimating generative models of data. A variational autoencoder consists of two tightly integrated components: a *generative model* for the observed data that is specified in terms of latent variables, and an amortized *variational distribution* that approximates the posterior distribution of the observation-specific latent variables. The two components are jointly estimated from the data.

The generative model represents the probability distribution of the observed data, $\boldsymbol{x}_i$ for each observation i, in terms of a multidimensional latent variable $\boldsymbol{z}_i$. The mapping between the latent variable $\boldsymbol{z}_i$ and the parameters of the probability distribution is specified using a multilayered neural network, called the decoder network, whose parameters (weights and biases) are contained in the vector $\boldsymbol{\theta}$. The joint distribution of the data and the latent variables is given as $p_\theta(\boldsymbol{x}_i, \boldsymbol{z}_i) = p_\theta(\boldsymbol{x}_i|\boldsymbol{z}_i)p(\boldsymbol{z}_i)$, where the prior for $\boldsymbol{z}_i$ is assumed to be isotropic Gaussian, $p(\boldsymbol{z}_i) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$.

To approximate the posterior of the latent variables, $p_\theta(\boldsymbol{z}_i|\boldsymbol{x}_i)$, VAEs rely on amortized variational inference, where the approximating variational distribution $q_\phi(\boldsymbol{z}_i|\boldsymbol{x}_i)$ is specified using an-

[3]We use the terms modality, data source, and domain interchangeably.

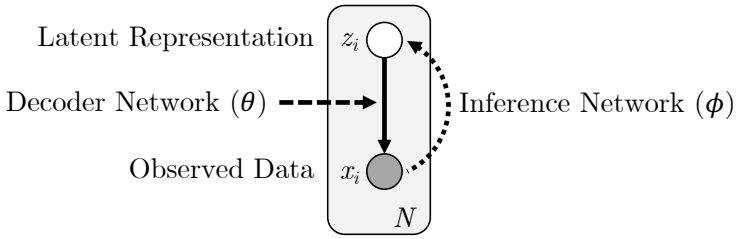


Figure 9: Graphical model for a standard VAE: the decoder network with parameters $\boldsymbol{\theta}$ transforms the latent representation $\boldsymbol{z}_i$ into the parameters of the likelihood for $\boldsymbol{x}_i$. Given $\boldsymbol{x}_i$, the inference network, with parameters $\boldsymbol{\phi}$, specifies the approximate posterior for $\boldsymbol{z}_i$.

other neural network, called the encoder or inference network. Note that the inference network uses the available data $\boldsymbol{x}_i$ as its input to specify the variational distribution for the observation-specific $\boldsymbol{z}_i$. The weights and biases of this network, $\boldsymbol{\phi}$, are amortized (i.e., shared) across all observations, allowing for scalable inference. Inference networks thus transform the inferential problem to that of learning a function, parameterized by a 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. The structure of such a standard VAE is illustrated in Figure 9.

6.2. Multimodal VAE

As we have data from multiple domains, we use a *multimodal* variational autoencoder (MVAE) to learn a latent representation that is shared across domains (Suzuki et al., 2016; Jaques et al., 2017; Wu and Goodman, 2018). We have data on $i = 1, \ldots, N$, brands across the four domains, indexed by $d \in \{\text{Text}, \text{Logo}, \text{Ind}, \text{BP}\}$, where Ind refers to the industry labels and BP indicates the brand personality. The observed data for brand i in domain d is written as $\boldsymbol{x}_i^d$ and the complete observation is given by $\boldsymbol{x}_i = \{\boldsymbol{x}_i^{\text{Text}}, \boldsymbol{x}_i^{\text{Logo}}, \boldsymbol{x}_i^{\text{Ind}}, \boldsymbol{x}_i^{\text{BP}}\}$. The domains differ in the number and type of features (e.g., words for text, logo features for logos, personality traits for brand personality). We index these features within domain d as $j = 1, \ldots, V_d$, such that $\boldsymbol{x}_i^d = \{x_{i1}^d, \ldots, x_{iV_d}^d\}$. The generative model specifies the probability distribution of the observed data in each domain in terms of a shared latent variable vector $\boldsymbol{z}_i$. Given our interest in analysis from multiple perspectives (e.g., the designer's perspective, which involves inferring the logo features from the other modalities, or the manager's perspective, which involves predicting consumer reactions from firm-generated content), we use multiple inference networks that condition on different subsets of the observed

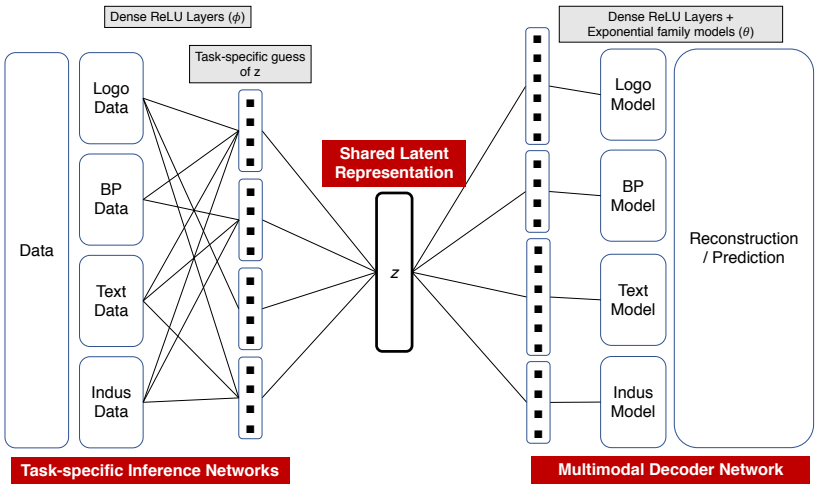


Figure 10: An illustration of the MVAE framework.

data $\boldsymbol{x}_i$ to infer the common latent variable $\boldsymbol{z}_i$. Figure 10 visually illustrates the modeling and inferential framework. While we observe data for all domains for each brand in our data, the framework allows for missing domains. We now focus on the generative model for the domains, before turning our attention to inference.

Multimodal Generative Model The generative model represents the probability distribution of the multimodal observed data $\boldsymbol{x}_i$ in terms of a shared multidimensional latent variable $\boldsymbol{z}_i$, which has an isotropic Gaussian prior $p(\boldsymbol{z}_i) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. As in the standard VAE, the joint distribution of the data and the latent variables is given as $p_\theta(\boldsymbol{x}_i, \boldsymbol{z}_i) = p_\theta(\boldsymbol{x}_i|\boldsymbol{z}_i)p(\boldsymbol{z}_i)$. However, the probability models for the different domains are independent, conditional on $\boldsymbol{z}_i$ i.e., $p_\theta(\boldsymbol{x}_i|\boldsymbol{z}_i) = \prod_d p_{\theta_d}(\boldsymbol{x}_i^d|\boldsymbol{z}_i)$. In turn, the probability model for each domain is specified using independent feature-level probability distributions such that $p_{\theta_d}(\boldsymbol{x}_i^d|\boldsymbol{z}_i) = \prod_j p_{\theta_d}^j(\boldsymbol{x}_{ij}^d|\boldsymbol{z}_i)$. Let $\boldsymbol{\mu}_i^d$ contain the parameters for the different feature-level distributions associated with observation i within domain d. A domain-specific decoder network, which we denote $\text{DNet}_d(\boldsymbol{z}_i; \boldsymbol{\theta}_d)$, captures the non-linear relationship between $\boldsymbol{\mu}_i^d$ and $\boldsymbol{z}_i$, such that $\boldsymbol{\mu}_i^d = \text{DNet}_d(\boldsymbol{z}_i; \boldsymbol{\theta}_d)$. We first describe the different feature-level probability distributions and follow with a description of the domain-specific decoder networks.

Feature-Level Distributions Conditional on the joint representation $\boldsymbol{z}_i$, each brand's features are modeled using independent domain- and feature-specific exponential-family distributions. The specific exponential-family distributions that we use for the different domain features are:

- *Text:* We use a Bernoulli distribution that captures whether or not a given word is present in a brand's textual description. That is, for each word j, we use the logistic-sigmoid transformation to model the probability that the word is present in brand i's description:

$$P(x_{ij}^{\text{Text}} = 1) = \frac{1}{1 + \exp(-\mu_{ij}^{\text{Text}})}. \tag{1}$$

This simple coding captures 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 be informative, it may also merely reflect the volume of text on the firm's website. Hence, we only model the presence or absence of a given word in the textual description.

- *Logo features:* Each of our logo features is either binary or categorical. For binary features, like whether the logo has a mark, we use a Bernoulli distribution. For categorical features consisting of $m = 1, \ldots, M_j$ possible options, like the dominant color, we use a categorical distribution, such that:

$$x_{ij}^{\text{Logo}} \sim \text{Categorical}(\text{softmax}(\boldsymbol{\mu}_{ij}^{\text{Logo}})), \tag{2}$$

$$\boldsymbol{\mu}_{ij}^{\text{Logo}} = (\mu_{ij1}^{\text{Logo}}, \ldots, \mu_{ijM_j}^{\text{Logo}}), \tag{3}$$

where,

$$\text{softmax}(\boldsymbol{\mu}_{ij}^{\text{Logo}}) = \left(\frac{\exp(\mu_{ij1}^{\text{Logo}})}{\sum_{n=1}^{M_j} \exp(\mu_{ijn}^{\text{Logo}})}, \cdots, \frac{\exp(\mu_{ijM_j}^{\text{Logo}})}{\sum_{n=1}^{M_j} \exp(\mu_{ijn}^{\text{Logo}})} \right)$$

gives the probability vector of the categorical distribution.

- *Industry labels:* Industry labels are binary variables and are modeled with a Bernoulli distribution.

- *Brand personality:* Brand personality is also real-valued, as it is the average of all respondents ratings, measured between 0-4. We therefore model it using a normal distribution, such that:[4]

$$x_{ij}^{\text{BP}} \sim \mathcal{N}(\mu_{ij1}^{\text{BP}}, \sigma_{ij}^{\text{BP}}),\ \sigma_{ij}^{\text{BP}} = \log(e^{\mu_{ij2}^{\text{BP}}} - 1)). \tag{4}$$

[4]The $\log(e^y - 1)$ structure in Equation 4 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 the above feature-level distributions, the observation-specific distributional parameters (e.g., the mean μ_{ij1}^{BP} and the variance σ_{ij}^{BP} of the normal in Equation 4) are specified non-linearly in terms of the latent variable $\boldsymbol{z}_i$ for that observation using modality-specific decoder networks.

Decoder Network We use a domain-specific decoder network, $\boldsymbol{\mu}_i^d = \text{DNet}_d(\boldsymbol{z}_i; \boldsymbol{\theta}_d)$, to model the potentially non-linear relationship between $\boldsymbol{\mu}_i^d$ and $\boldsymbol{z}_i$. In our application, we use dense, feed-forward layers with rectified linear activation units (ReLU) and skip connections to specify $\text{DNet}_d()$ for each domain. This is equivalent to the following sequence of computations:

$$
\begin{aligned}
\mathbf{h}_{i1}^{\text{Dec},d} &= \text{ReLU}(\mathbf{a}_0^d + W_0^{d,z}\boldsymbol{z}_i), \\
\mathbf{h}_{i2}^{\text{Dec},d} &= \text{ReLU}(\mathbf{a}_1^d + W_1^{d,h}\mathbf{h}_{i1}^{\text{Dec},d} + W_1^{d,z}\boldsymbol{z}_i), \\
&\vdots \\
\mathbf{h}_{iL_d}^{\text{Dec},d} &= \text{ReLU}(\mathbf{a}_{(L_d-1)}^d + W_{(L_d-1)}^{d,h}\mathbf{h}_{i(L_d-1)}^{\text{Dec},d} + W_{(L_d-1)}^{d,z}\boldsymbol{z}_i), \\
\boldsymbol{\mu}_i^d &= \mathbf{a}_{L_d}^d + W_{L_d}^{d,h}\mathbf{h}_{iL_d}^{\text{Dec},d} + W_{L_d}^{d,z}\boldsymbol{z}_i,
\end{aligned}
\tag{5}
$$

where $\text{ReLU}(x) = \max(0, x)$, applied componentwise. The above is equivalent to applying the ReLU operation sequentially, layer by layer, through the network. Each layer ℓ computes a transformed representation of the brand through the hidden units, whose activations are contained in the vector $\mathbf{h}_{i\ell}^{\text{Dec},d}$, of size equal to the number of hidden units in that layer. The weights associated with each layer are contained in the matrices, $W_\ell^{d,h}$ and $W_\ell^{d,z}$, where the latter is associated with the latent variables $\boldsymbol{z}_i$. The $\mathbf{a}_\ell^d$ vectors contain the biases (intercepts) associated with the hidden units in layer ℓ. Note that we combine the hidden unit activations with the original representation $\boldsymbol{z}_i$, in what is known as skip connections (Dieng et al., 2018), to inform the hidden units of the next layer.[5] This whole operation is repeated L_d times for the number of layers in the network for domain d. The output layer (i.e., the last layer) outputs the parameters $\boldsymbol{\mu}_i^d$ of the data likelihood. The use of multilayered feed-forward networks allows us to capture complex joint distributions involving the different domains, and the expressiveness of the model depends upon the number of hidden units and layers.

[5]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).

We use $\boldsymbol{\theta}_d$ to refer to all of the decoder network parameters within domain d across all the features j. While the exact nature of the decoder network differs across domains, the above conveys the general structure. We describe the specifics of each domain's network architecture in a later section.

Multiview Inference Networks The key task in using the MVAE framework is to learn the joint latent representations $\boldsymbol{z}_i$. In our work, we follow the standard practice of assuming a mean-field variational approximation for the posterior of $\boldsymbol{z}_i$. The approximate posterior is given by the normal distribution:

$$p_\theta(\boldsymbol{z}_i|\boldsymbol{x}_i) \approx q_\phi(\boldsymbol{z}_i|\boldsymbol{\xi}_i) = \mathcal{N}(\boldsymbol{\xi}_i^m, \text{diag}(\boldsymbol{\xi}_i^v)), \tag{6}$$

where, just as in the standard VAE, an inference network computes the mean and variance terms of this normal distribution, $\boldsymbol{\xi}_i = \{\boldsymbol{\xi}_i^m, \boldsymbol{\xi}_i^v\}$ from data $\boldsymbol{x}_i$. That is, $\boldsymbol{\xi}_i = \text{INet}(\boldsymbol{x}_i; \boldsymbol{\phi})$, which is a neural network given by:[6]

$$\begin{aligned}
\mathbf{h}_{i1}^{\text{Inf}} &= \text{ReLU}(\mathbf{c}_0 + V_0\boldsymbol{x}_i), \\
\mathbf{h}_{i2}^{\text{Inf}} &= \text{ReLU}(\mathbf{c}_1 + V_1\mathbf{h}_{i1}^{\text{Inf}}), \\
&\vdots \\
\mathbf{h}_{iL}^{\text{Inf}} &= \text{ReLU}(\mathbf{c}_{L-1} + V_{L-1}\mathbf{h}_{i(L-1)}^{\text{Inf}}), \\
\boldsymbol{\xi}_i &= \mathbf{c}_L + V_L\mathbf{h}_{iL}^{\text{Inf}}.
\end{aligned} \tag{7}$$

The inference procedure thus consists of optimizing the decoder and inference network parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ such that $q_\phi(\boldsymbol{z}_i|\boldsymbol{\xi}_i = \text{INet}(\boldsymbol{x}_i; \boldsymbol{\phi}))$ is as close to the true posterior $p_\theta(\boldsymbol{z}_i|\boldsymbol{x}_i)$ as possible.

In our application, it is important to be able to infer $\boldsymbol{z}_i$ given information on only a subset of the domains. This involves using brand-specific data on some subset of the domains to compute $\boldsymbol{z}_i$, which can then be used to make predictions on the missing domains. For example, when approaching the task of data-driven design (i.e., the designer's perspective), we have data on everything except the logo. Alternatively, a brand manager cares about how consumers will evaluate a brand

[6]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.

or brand-candidate, given a logo, text, and industry information. To tackle this challenge, we introduce the idea of task-specific inference networks: inference networks corresponding to different conditional posteriors, depending on the patterns of missingness that govern a particular context. Specifically, we implement four distinct inference networks: (1) the full data inference network, akin to that of the classical VAE; (2) the designer's inference network, corresponding to the case where we observe everything *except* the logo; (3) the manager's inference network, corresponding to the case where we observe everything *except* consumer's perceptions of brand personality; and (4) the researcher's inference network, corresponding to the case where we *just* observe the logo. That is, we learn four distinct inference networks, which we index by $t \in \{\text{Full}, \text{Des}, \text{Mgr}, \text{Res}\}$, where t stands for task, corresponding to four separate functions,

$$\boldsymbol{\xi}_{i,t} = \text{INet}_t(\tilde{\boldsymbol{x}}_i^t; \boldsymbol{\phi}_t),$$

where $\tilde{\boldsymbol{x}}_i^t$ is shorthand for the data available for inference task t (for example, for $t = R$, $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i^{\text{Logo}}$). Intuitively, this function corresponds to the model's "best guess" at the posterior distribution, given data from the available domains for the particular task. Note that, regardless of which inference network is used, the decoder network and probability models remain fixed. Hence, each inference network is forced to learn a coherent, unified representation, regardless of the missing modalities. Finally, we also note that, while we have assumed a set of tasks corresponding to our data setting, this structure can be easily adapted to include other tasks of interest.

6.3. Inference

Inference with this multimodal setup involves variational expectation maximization (variational EM), adapted to allow for our multiple decoder and inference networks. Intuitively, the goal of inference is to optimize the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, of the decoder and inference networks, such that encoding and then decoding data $\boldsymbol{x}_i$ leads to a prediction that is as close as possible to the original data.

In the classical VAE, with one decoder network and one inference network, the following loss

function is minimized:

$$\ell(\boldsymbol{\theta}, \boldsymbol{\phi}) = \sum_{i=1}^{N} -E_{z \sim q_\phi(\boldsymbol{z}_i | \boldsymbol{\xi}_i = \text{INet}(\boldsymbol{x}_i; \boldsymbol{\phi}))} [\log p_\theta(\boldsymbol{x_i} \,|\, \boldsymbol{z}_i)] + \text{KL}(q_\phi(\boldsymbol{z}_i \,|\, \boldsymbol{x}_i) \;||\; p(\boldsymbol{z}_i)), \tag{8}$$

where KL() is the Kulback-Leibler divergence between distributions. This loss is the negative of the standard evidence lower bound (ELBO) for doing variational inference on the latent parameters, $\boldsymbol{z}_i$, but where the parameters of the variational approximation are determined 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.

In our multiview inference framework, the $p_\theta(\boldsymbol{x}_i \,|\, \boldsymbol{z}_i)$ from Equation 8 decomposes into a product of the domain-specific decoder networks and feature-specific probability distributions. Moreover, we add to the above a stochastic binning procedure: for each iteration of our optimization, we split the data into four equally-sized bins, such that for each bin, we use a different one of our four inference networks, holding out the relevant data modalities. Returning to Equation 8, this means that, in our optimization, at each iteration, the $q_\phi(\boldsymbol{z}_i | \boldsymbol{\xi}_i = \text{INet}(\boldsymbol{x}_i; \boldsymbol{\phi}))$ used for observation i depends on the bin that brand i is assigned to in that iteration. Together, these two modifications imply the following *per iteration* loss function:

$$\ell_m(\boldsymbol{\theta}, \{\boldsymbol{\phi}_t\}) = \sum_{i=1}^{N} \sum_{\forall t} \delta_{itm} \Big\{ - E_{\boldsymbol{z}_i \sim q_{\phi_t}(\boldsymbol{z}_i \,|\, \boldsymbol{\xi}_{i,t} = \text{INet}_t(\tilde{\boldsymbol{x}}_i^t; \boldsymbol{\phi}_t))} [\log p_\theta(\boldsymbol{x}_i \,|\, \boldsymbol{z}_i)] + \text{KL}\left[q_{\phi_t}(\boldsymbol{z}_i \,|\, \tilde{\boldsymbol{x}}_i^t) \;||\; p(z_i)\right] \Big\} \tag{9}$$

where m indexes the iteration of the optimization, and $\delta_{itm} = 1$ if brand i is assigned to bind t on iteration m, and zero otherwise. Intuitively, this stochastic binning allows us to learn our task-specific inference networks simultaneously, by augmenting our complete data with incomplete instances of each of the original observations. Optimizing this loss is similar, but not exactly equivalent to the procedure suggested by Wu and Goodman (2018).

6.4. Implementation

Here, we briefly describe the specifics of how we implemented our model, including regularization, hyperparameter optimization, and model architecture details. We include more details on

implementation, including pseudocode, in the web appendix.

We implement our model using Tensorflow and the Edward probabilistic programming language (Tran et al., 2016). We optimize the loss in Equation 9 using stochastic gradient descent. To prevent overfitting, complex models such as ours typically rely on regularization methods. We utilize two regularization strategies: L2 regularization of the weights of the neural network, and dropout, both of which are standard approaches in the deep learning literature (Goodfellow et al., 2016). To determine all model hyperparameters, including the number of latent dimensions (K), number of hidden layers, layer sizes, and degree of regularization, we performed grid search over a wide array of values, assessing model performance using both cross-validation fit and posterior predictive checks. From this procedure, we determined an optimal dimensionality of the latent space of $K = 40$. We also found that using more than a single hidden layer in the neural networks did not improve model fit. This is likely because we are already working with highly processed inputs, thus limiting the usefulness of the increasing levels of abstraction enabled by adding more layers. Our final model architecture consists of 1024 hidden units in all of the inference networks, 1024 hidden units in the text decoder network, and 512 hidden units for all other decoders.

7. Model Results

We present the model results in five parts: first, we briefly describe the reconstruction error and predictive ability of our MVAE model. We then describe the learned latent space by showing how the model encodes features of brands (e.g., industry) and by illustrating the similarities that are learned for a set of representative brands. To further validate the learned generative model, we then showcase an example of a randomly generated brand. Next, we show how the learned representations can be used for ideation via a brand arithmetic approach in which a brand can be combined with another brand, or with specific features, to generate novel brand identities. Finally, we show how the different task-specific inference networks can be used to study new brands, and provide decision support for designers and managers.

	Full Data		Cross-Validation			
	NIR	Full	Full	Des	Mgr	Res
Brand Personality	0.406	0.011	0.268	0.279	**0.367**	**0.446**
Industry Labels	0.106	0.026	0.059	0.058	0.059	**0.081**
Logo: Binary	0.262	0.030	0.134	**0.229**	0.132	0.093
Logo: Dom. Color	0.128	0.011	0.062	**0.120**	0.061	0.038
Logo: Hull Shape	0.235	0.028	0.163	**0.224**	0.161	0.112
Logo: Mark Shape	0.115	0.020	0.096	**0.114**	0.094	0.070
Logo: Sans/Serif	0.215	0.020	0.117	**0.180**	0.111	0.067
Logo: # Colors	0.363	0.032	0.189	**0.347**	0.187	0.128
Text	0.158	0.016	0.111	0.109	0.111	**0.132**

Table 1: MAD statistics for the model using different inference networks, compared to the No Information Rate (NIR). Full stands for the full data inference network, Des for designer (i.e., given no logo data), Mgr for manager (i.e., given no brand personality data), and Res for researcher (i.e., given only logo data). The first two columns are using the full data; the remaining four columns are results from cross-validation, i.e., for firms not in the training data. The values in normal font weight are for *reconstruction* tasks, while the values in bold are for *predictive* tasks (i.e., predicting a heldout domain).

7.1. Fit

We summarize the fit of the model in Table 1 using Mean Absolute Deviation (MAD), computed separately for each of the domains. We define the MAD for domain d as:

$$\text{MAD}_d = \frac{1}{N}\frac{1}{J}\sum_{i=1}^{N}\sum_{j=1}^{J}|x_{ij}^d - E(x_{ij}^d)|, \tag{10}$$

where $E(x_{ij})$ is the expected value of x_{ij}^d under the model. We compare the predictions of our model to the no information rate (NIR), which is a natural simple benchmark, equivalent to using the empirical mean as the expected value in Equation 10, $E(x_{ij}^d) = \bar{x}_j^d$. Table 1 shows these statistics for both in-sample firms using the full inference network, and heldout firms via a 10-fold cross-validation involving all four inference networks.[7]

Table 1 showcases an important distinction between three types of fit measures: (1) *in-sample reconstruction error*, which is computed using the full inference network on the full data and captures how well the model does at recreating the inputs it is given during training; (2) *out-of-sample reconstruction error*, which represents how well the model is able to reproduce inputs it is given for *new* brands; and (3) *out-of-sample predictive error*, which shows the model's ability

[7]For visual features, we treat the full set of binary features as one group, and we compute separate MAD statistics for each of the categorical features.

to predict *missing* domains for *new* brands. Our model does exceptionally well on in-sample reconstruction error, shown in column two of the table for the full inference network, which is not surprising: with a high dimensional latent space, and expressive inference and decoder networks, the model is able to recreate the data it is trained on.[8] More importantly, the model is also able to learn meaningful representations for *new* brands, as shown by the out-of-sample reconstruction errors reported in column three for the full inference network, and by the out-of-sample reconstruction errors reported in non-bold font in the other columns for the task-specific inference networks. That is, given data for a brand that was not present during training, the inference network is able to output a latent representation for that brand, which can then be used by the decoder network to recreate the original data. That the model achieves relatively low error rates in this out-of-sample reconstruction task indicates that the learned latent space does, in fact, capture meaningful brand information.

The bold cells in the remaining three columns of Table 1 measure task-specific *predictive* accuracy. We can see that, relative to the reconstruction tasks, these predictive errors are much larger. Again, this is expected: the predictive task is harder than the reconstruction task, as it involves predicting held-out information for new brands. However, in nearly all cases, the model achieves better predictive error rates than the naive NIR benchmark. The only exception to this is the researcher network: given data on just the logo features, it is difficult for the model to predict the heldout domains.[9] Nonetheless, for the designer's task, the model is able to predict better than chance what features will be present in a firm's logo, given a text, brand personality, and industry tags, and for the manager's task, the model is able to predict how consumers will evaluate a brand's personality, given a brand's profile.

7.2. Understanding the Latent Space

Having established the validity of the latent space, we now turn to understanding what it represents. In general, it is difficult to interpret specific dimensions of our learned latent space. Consider, for

[8] We do not show the results for in-sample error for the other inference networks, as they are essentially identical to the full inference network: for firms in the training data, the model is able to learn highly correlated representations across all inference networks, leading to essentially equivalent predictions regardless of the inference network used.

[9] One explanation for the difficulty in predicting firm traits from logos alone is that the norms for logos vary by industry. Indeed, if the model is re-trained with the "researcher" network using both logo features and industry tags as inputs, predictive accuracy modestly improves.

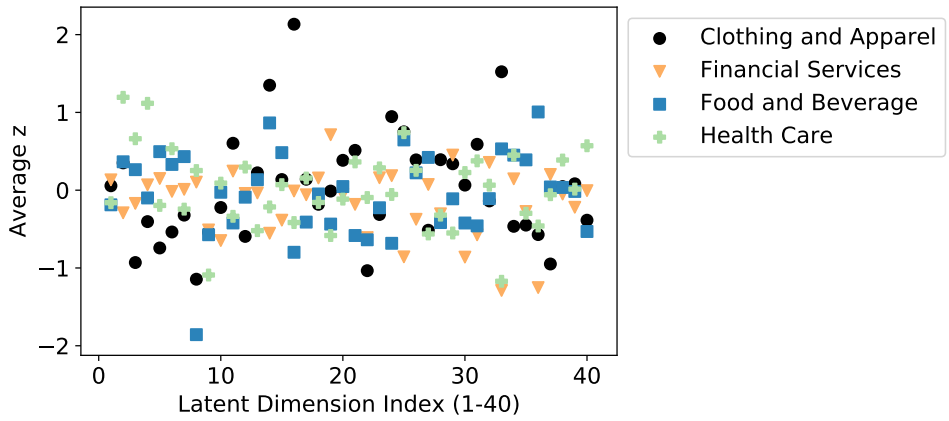


Figure 11: Average z_i values for brands with four example industry tags, illustrating how specific traits are encoded by several dimensions.

instance, how the model encodes industry tags. In Figure 11, we plot the average $\boldsymbol{z}_i$ values for brands that have one of four industry tags: Clothing and Apparel, Financial Services, Food and Beverage, and Health Care. We see that, in general, the average $\boldsymbol{z}_i$ value for a given tag is close to zero. For each tag, however, a few dimensions have extreme values, implying that industry is not encoded via a single dimension, but by a combination of latent dimensions.

Although the z-space cannot be directly interpreted, distances within it are meaningful: if two brands are closer together, they are predicted to share features. By looking where brands lie in this space, we can better understand what the learned representations are capturing. In Table 2, we show the three nearest neighbor brands in z-space for a set of representative focal firms, along with the distance each neighbor is from the focal firm. We see that, in general, a firm's neighbors are those brands that share many features: for example, they operate in a similar industry, have similar brand perceptions, and share similar logo features. Moreover, the more features two brands share, the closer they tend to be in terms of distance in z-space. For example, Facebook's closest neighbor is Twitter: not only are they both innovative social network platforms, but they both have simple, blue, bulky logos. Similarly, Old Navy's closest neighbor is Gap: both are owned by the same parent company, both sell clothing at affordable price points, and both have dark blue, again relatively simple logos. In other cases, the nearest neighbors are not such close matches: for example, in the case of 3M, we see that the closest neighbor in z-space, Becton Dickinson, is very similar in terms of firm function and perception, but not as similar aesthetically. The distance between firms reflects the degree to which they are similar in all dimensions, as can be seen by

Focal Brand	Neighbors in z-space		
Facebook Social network	**Twitter** Social network (8.36)	**Uber** Ride-sharing (8.98)	**Amgen** Biopharmaceuticals (9.92)
Old Navy Discount apparel	**Gap** Mid-market apparel (8.53)	**Ross Dress for Less** Discount apparel (9.45)	**VF Corporation** Apparel umbrella brand (9.5)
3M Materials and consumer goods	**Becton Dickinson** Medical technology (10.07)	**Illinois Tool Works** Components and equipment (10.4)	**Facebook** Social network (10.73)
Gucci Luxury goods	**MAC** Cosmetics (8.82)	**Dior** Luxury goods (8.96)	**Prada** Luxury goods (9.7)
KFC Fast food	**Pizza Hut** Fast food (9.38)	**Supervalu** Discount grocer (9.76)	**Burger King** Fast food (10.11)

Table 2: The 4 closest brands to each focal brand in z-space, including their logo, name, a brief description of what the firm does, and, in parentheses, the distance between the focal brand and the neighbor.

comparing the small distance between Facebook and Twitter (8.36) to the larger distance between 3M and Becton Dickinson (10.07). These comparisons also emphasize how distance in z-space is distinct from just simply clustering the logos themselves: $\boldsymbol{z}_i$ captures a holistic view of firm i, including aesthetics *and* other aspects of brand identity.

7.3. Generating Random Brands

As a final validation of the learned latent space and generative model, as well as to build familiarity with the outputs and predictions of the framework, we consider the task of generating random brands. Under the MVAE framework, this can be accomplished simply by drawing a new $\boldsymbol{z}_i$ vector

from the prior, $\boldsymbol{z}_i \sim \mathcal{N}(0, \boldsymbol{I})$, and propagating that vector down the decoder network. If the model has learned a meaningful latent space, then brand identities generated in this fashion should be coherent. To see this, we generate a single random $\boldsymbol{z}_i$ vector, then feed this vector through the decoder network to understand what features correspond to it.

Starting first with the industry tags, the most likely tags corresponding to this random $\boldsymbol{z}_i$ are: Software (probability 0.578), followed by Professional Services (0.220), Commerce and Shopping (0.176), and Consumer Goods (0.143). Hence, we infer that this is a software company, likely providing software to other companies, possibly in the retail or e-commerce space.

To understand the brand personality corresponding to this randomly-generated brand, we first note that each of the brand personality traits has a different overall mean in our data. For instance, the average score for "confident" across all brands is 2.56, while the average score for "feminine" is 0.827. Hence, a brand that scores 2.4 on confident but 1.6 on feminine is actually perceived as quite feminine, but slightly less confident, relative to the mean, despite its confident score being higher than its feminine score. For this reason, we consider personality scores *relative to the sample mean*. For our randomly generated brand, the highest relative personality traits are: young, trendy, and cool. The lowest relative personality traits are: masculine, tough, hard-working.

To more concretely understand this brand identity, we can use the decoder network to also generate what words would likely appear on this firm's website. Once again, we consider what words are more or less likely to appear *relative to the sample mean*, for the same reason as with brand personality: while many firms use words like "product" or "consumer," we want to find which words are most specific to the focal brand, and hence consider probabilities relative to mean rate of occurrence in our sample. For our randomly generated brand, the ten most likely words, relative to the mean, are: expert, keep, come, engineer, today, even, love, group, property, excite. These words are in keeping with the prior descriptions of the brand as a software company (expert, engineer), with a relatively young and hip brand identity (today, love, excite).

Finally, in Table 3, we show the visual features corresponding to this randomly generated $\boldsymbol{z}_i$. We break these visual features down into the five categorical variables, as well as a selection of the most likely binary variables, broken down into several categories. To better understand how these features could be translated into a logo, we also provide a simple, nonprofessional rendering of a logo based on this profile in Figure 12. Again, the visual design makes intuitive sense: the suggested

Categorical Features			Binary Features		
Feature	Likely Values	Prob	Feature Class	Likely Values	Prob
Dominant color:	Medium Blue	0.279	**Colors:**	Light Gray	0.800
	Black	0.239		Light Blue	0.582
	Tan	0.119	**Accent color:**	Light Green	0.707
Hull shape:	Medium Oval	0.976		Light Gray	0.708
	Thin Oval	0.023	**Font:**	Width: Original	0.999
Sans/Serif Font:	Sans	1.000		Style: No Italics	0.999
Mark Class:	Long, Horizontal	0.380		Class: Geometric	0.827
	Detailed Horizontal	0.204	**Other:**	Has a Mark	0.996
	Narrow, Vertical	0.087		Low # Corners	0.851
Number of Colors:	Many colors (> 3)	0.885		Low % Whitespace	0.835
	Three colors	0.112		High Vertical Symmetry	0.996
	Two colors	0.003		Low Perimetric Complexity	0.746

Table 3: Visual profile for a randomly generated brand, illustrating the likely values of the categorical features at left, and a selection of high probability binary features at right. We only report binary features that are relatively more likely than the population mean.

Figure 12: Simple, nonprofessional rendering of a logo with the features described in Table 3.

color scheme of blues, greens, and greys is coherent, in the sense that these three particular colors align with well known and commonly used "analogous" or "split" color schemes.[10] The lack of whitespace is consistent with many technology firms, including Samsung, Apple, and Twitter. Likewise, the geometric, sans-serif font is consistent with a modern, trendy image. In sum, the set of features corresponding to a randomly drawn $\boldsymbol{z}_i$ is coherent, lending additional support to the idea that our learned generative model has captured fundamental design principles.

7.4. Ideation through Brand Arithmetic

We now show how the learned representations can be leveraged for ideation purposes by brand managers or designers. The design process for new brands often begins by thinking of existing brands in the focal industry, or that have similar identities to the new brand.[11] Elements of these brands' logos may then be mixed with visual features unique to the new brand. For instance, a

[10] See, e.g., `https://www.tigercolor.com/color-lab/color-theory/color-harmonies.htm`
[11] See, e.g., `https://99designs.com/blog/tips/logo-design-process-how-professionals-do-it/`

designer for a new medical device company may start by looking at what logo design patterns are popular in health care, and in technology companies, and may then fuse these elements together to create a template for the new brand. Colloquially, it is also common to hear new brands, especially start-ups, described as the "X of Y" (e.g., the "Uber of grocery stores" for a grocery delivery service), or as a fusion of existing brands (e.g., a mix of Mercedes-Benz and Old Navy, for an accessible luxury car, or a mass market luxury fashion brand). In z-space, the idea of fusing brand traits or identities can be captured by adding together $\boldsymbol{z}_i$ vectors corresponding to specific traits or brands, an operation we refer to as brand arithmetic.

Medical Devices We first consider the task of designing for a medical device company. As described above, medical devices can be considered a fusion of technology and health care. In our data, we have an industry tag corresponding to Health Care, as well as the technology-related industry tags Hardware, Consumer Electronics, and Software. To understand what features we would expect in a brand that sits at the intersection of health care and technology, we first define two averages: $\bar{\boldsymbol{z}}_{\text{Health}}$, which is the average of all $\boldsymbol{z}_i$ vectors such that brand i was tagged as a Health Care company, and $\bar{\boldsymbol{z}}_{\text{Tech}}$, which is the average of all $\boldsymbol{z}_i$ vectors such that brand i was tagged as either a Hardware, Consumer Electronics, or Software company. We can then interpolate between these two vectors, to create a new representation for a medical device company:

$$\boldsymbol{z}_{\text{MedDevice}} = 0.5\bar{\boldsymbol{z}}_{\text{Health}} + 0.5\bar{\boldsymbol{z}}_{\text{Tech}}. \tag{11}$$

To validate that this procedure indeed produces a reasonable representation, we first check which firms are close to the interpolated $\boldsymbol{z}_{\text{MedDevice}}$: the five nearest neighbors include three companies that produce medical devices—Becton-Dickinson, Sony, and Stryker—as well as Raytheon, a defense technology company, and Allergan, a pharmaceutical company.

We can also see what predictions the model makes about such a firm. Comfortingly, when we predict the industry tags from $\boldsymbol{z}_{\text{MedDevice}}$, the top five tags are Health Care, Software, Information Technology, Manufacturing, and Hardware. Moreover, when $\boldsymbol{z}_{\text{MedDevice}}$ is propagated through the text decoder, the ten highest terms, relative to the population mean, are: patient, healthcare, technology, better, global, solution, health, outcome, innovate, and science. For brand personality, the

Categorical Features			Binary Features		
Feature	Likely Values	Prob	Feature Class	Likely Values	Prob
Dominant color:	Medium Blue ■	0.547	**Accent color:**	Medium Blue ■	0.707
	Dark Blue ■	0.124	**Font:**	Width: Original	0.999
	Light Blue ■	0.116		Style: No Italics	0.999
Hull shape:	Thin Oval	0.837		Weight: **Bold**	0.920
	Medium Oval	0.145		Width: Wide	0.267
Sans/Serif Font:	Sans	0.900	**Other:**	Has a Mark	0.980
Mark Class:	Wispy horizontal	0.313		Low # Regions	0.607
	Detailed design	0.120		Low Entropy	0.308
	Thin	0.095		Mark Position: Left	0.289
Number of Colors:	One color	0.859		Low % Horizontal Edges	0.262
	Two colors	0.133			
	Three colors	0.007			

Table 4: Visual profile corresponding to $\boldsymbol{z}_{\text{MedDevice}}$, illustrating the likely values of the categorical features at left, and a selection of high probability binary features at right. We only report binary features that are relatively more likely than the population mean.

Figure 13: A simple, nonprofessional rendering of a logo based on the visual profile in Figure 4.

highest relative traits are technical, intelligent, and contemporary, while the lowest are outdoorsy, rugged, and masculine. Finally, we summarize the logo features we expect for this company in Table 4, and provide a simple rendering of a logo that contains many of those features in Figure 13.

Daring Fast Food Brand arithmetic can also be used with personality traits. Consider the task of designing a daring fast food company. In general, fast food brands are not perceived as particularly daring: in our data, the average consumer rating of McDonald's for "daring" was 1.0, and for Burger King, 1.05, while the average "daring" rating across all firms is 1.6, with a max of 3.3. To mathematically represent combining "daring" and "fast food," we first create representative $\boldsymbol{z}$-vectors for each of these concepts: for daring, we create an average $\bar{\boldsymbol{z}}_{\text{Daring}}$ by averaging the $\boldsymbol{z}_i$ vectors for all brands who scored in the top decile of daring. For fast food, we create $\bar{\boldsymbol{z}}_{\text{FastFood}}$ by averaging together the $\boldsymbol{z}_i$ vectors of McDonald's, Burger King, and KFC. To create a new brand identity, daring fast food (DFF), we can then add the daring vector to the fast food vector. In this case, the intended outcome is to *add* an element of daring to the standard fast food firm, not

interpolate, and hence we consider a more general combination:

$$\boldsymbol{z}_{\text{DFF}}(\alpha, \beta) = \alpha \bar{\boldsymbol{z}}_{\text{FastFood}} + \beta \bar{\boldsymbol{z}}_{\text{Daring}}.$$

The higher α, the more of the daring personality will be added. The higher β, the more the resulting brand will resemble the typical fast food firm.[12]

To illustrate this, we consider two combinations: $z_{\text{DFF}}(0.5, 0.5)$, which has the same interpolation weights as in the medical devices example, and $z_{\text{DFF}}(0.5, 1.0)$, which increases the degree of daring being added.[13] Unlike the medical device case, where we could verify that the arithmetic had produced a reasonable result by computing the new $\boldsymbol{z}$'s nearest neighbors, in our data, there is no "daring fast food" brand to correspond to either of these new profiles. In both cases, when we compute the nearest neighbors to $\boldsymbol{z}_{\text{DFF}}$, they are simply Burger King, KFC, and Pizza Hut.[14] Nonetheless, we can still make predictions for this previously unobserved brand identity. In both cases, the two highest industry labels associated with $\boldsymbol{z}_{\text{DFF}}$ are Food and Beverage and Travel and Tourism, which are the two labels most often associated with fast food firms. For brand personality, when $\beta = 0.5$, the highest three traits are cheerful, family-oriented, and trendy, largely reflecting those traits that we expect in a fast food restaurant. However, when compared to the average fast food restaurant, the expected score for daring is 0.513 points higher. Related concepts, like exciting, glamorous, and contemporary, are also higher, illustrating the impact of the interpolation: by adding $\bar{\boldsymbol{z}}_{\text{Daring}}$ to $\bar{\boldsymbol{z}}_{\text{FastFood}}$, we have morphed the fast food representation to be a bit more daring. When we increase β to 1, we see this effect even more dramatically. We illustrate this contrast between the two values of β and the average fast food firm in Figure 14.

The value of β also determines to what degree the predicted visual features differ from the fast food norm. Consider, for instance, the predicted colors: for the average fast food firm, there are expected to be three colors (prob = 0.763), with dominant color red (prob = 0.99), and a yellow accent color (prob = 0.85). For the interpolation case, with $\beta = 0.5$, the probability of three colors

[12] The risk of relaxing the restriction that $\alpha + \beta = 1$ is that the resulting vectors may contain values significantly more extreme than would be implied by the $\mathcal{N}(0, 1)$ prior. We have found that such cases result in predictions that are more extreme in terms of probabilities or magnitudes assigned to features.

[13] Considering $\alpha = 1$ produces a brand that strongly resembles the typical fast food firm, even when $\beta = 1$; hence, we consider only $\alpha = 0.5$.

[14] As we illustrate in the next section, more recent entrants to the market do reflect the predicted personality: when Shake Shack's $\boldsymbol{z}_i$ is estimated using the full inference network, it falls closer to $\boldsymbol{z}_{\text{DFF}}$ than it does to $\bar{\boldsymbol{z}}_{\text{FastFood}}$. However Shake Shack is not in our original data.

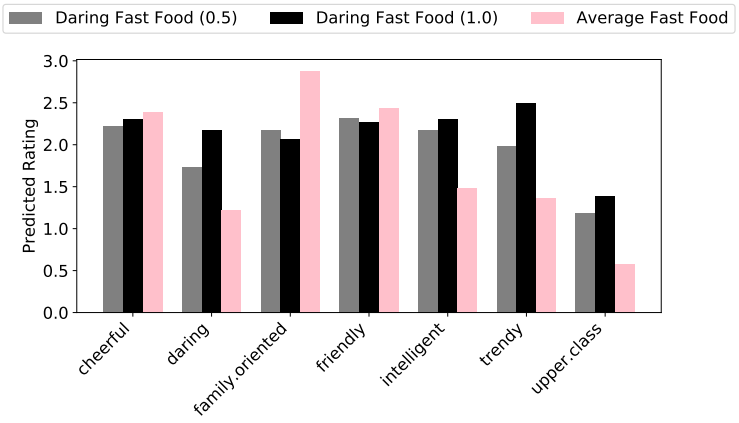


Figure 14: Contrasting brand personality predictions for a daring fast food restaurant, for $\beta = 0.5, 1.0$, and for the average fast food restaurant.

goes down to 0.425, with two and one color becoming much more likely (probs = 0.287 and 0.244 respectively). Red is still expected to be the dominant color (prob = 0.771), but dark blue and black are now possible (probs = 0.079 and 0.027 respectively). When we increase the degree of daring still further by setting $\beta = 1.0$, the probability of a black dominant color continues to rise (prob = 0.131). We see other features change as well: for example, the probability of seeing a bold font goes down, while the probability of having a low number of corners goes up. Together, these changes imply a set of candidate changes for developing a more daring visual identity for a fast food firm, illustrating how brand arithmetic allows for creative fusions of existing ideas.

Brand Hybrids As a final illustration of the brand arithmetic concept, we consider the idea of interpolating between specific brands. To interpolate between brands A and B, we find the midpoint between the two brands in z-space:[15]

$$\boldsymbol{z}_{\text{Mid}} = 0.5\boldsymbol{z}_A + 0.5\boldsymbol{z}_B.$$

We then consider which of our existing brands are closest to this midpoint. In many cases, the closest brands to $\boldsymbol{z}_{\text{Mid}}$ are simply the original two brands, or their closest neighbors. However, by looking at which brands are close to $\boldsymbol{z}_{\text{Mid}}$ but not close to either $\boldsymbol{z}_A$ or $\boldsymbol{z}_B$, we can understand better how the model interpolates between these two brands. We now describe three examples

[15] Just as before, the weights here need not be 0.5 for each; a more general formulation of $\boldsymbol{z}_{\text{Mid}} = \alpha\boldsymbol{z}_A + \beta\boldsymbol{z}_B$ can also be used to adjust the emphasis of each original brand in the hybrid brand.

interpolating between well-known brands:

- *Mercedes-Benz and Old Navy.* When interpolating between Mercedes-Benz, a luxury car brand, and Old Navy, an affordable apparel retailer, we find among the three closest midpoint brands two very interesting case studies: Landrover and Burberry. While Landrover is another luxury car brand, it notably is not one of the original ten closest neighbors in z-space of Mercedes. However, its logo shares many visual similarities to Old Navy, with both featuring dense, simple, oval-shaped designs. Hence, it is a natural fusion of Mercedes and Old Navy in terms of aesthetics and function. Burberry, on the other hand, represents a natural fusion of brand identity and firm function, taking the luxuriousness of Mercedes, and merging it with the apparel function of Old Navy.

- *Louis Vuitton and Nike.* When interpolating between luxury fashion brand Louis Vuitton, and sporting apparel and footwear company Nike, we again find interesting results. The closest midpoint brand is Calvin Klein, a relatively upmarket fashion brand with a sporty look, and with a logo that fuses elements of both Louis Vuitton and Nike. We also find the innovative and sporty luxury car company BMW falls close to the midpoint. While BMW is also a close neighbor to Louis Vuitton, other luxury brands like Dior fall much closer to Louis Vuitton's position in z-space. Yet, when Louis Vuitton is fused with Nike, this ordering reverses: BMW appears much closer to the midpoint, while brands like Dior fall away entirely.

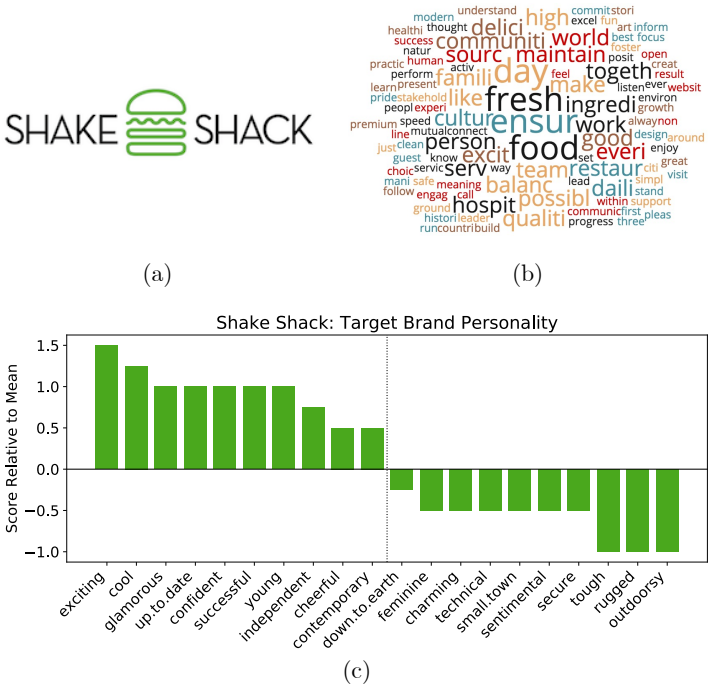
- *Google and McKinsey.* Finally, we interpolate between the tech company and search engine Google, and the management consultancy McKinsey. The two closest brands to the midpoint between these firms are IBM and Cognizant. Besides being a technology company, IBM also provides extensive IT consulting services. Likewise, Cognizant is a provider of IT services and consulting, an exact hybrid of the firm functions and brand identities of Google and McKinsey. Finally, further emphasizing the model's ability to pinpoint these brand fusions, the eighth closest brand to the midpoint is Tech Mahindra, another multinational IT consultancy, and a brand which is not even among the top 10 closest brands to either Google or McKinsey.

Taken together, these examples further emphasize the ability of brand arithmetic to meld together brand identities, and aid in the ideation process for new brands.

7.5. Task-specific Decision Support: The Case of Shake Shack

In all of the previous analyses, we have used the full inference network, and manipulated the learned $\boldsymbol{z}_i$ representations to aid in the brand ideation process. Now, we consider the task of using our task-specific inference networks to understand design and branding for new firms, and to provide design decision support. In particular, we focus on a case study of a relatively recent entrant to the fast food space, Shake Shack. Shake Shack makes a compelling case study for several reasons: first, its logo is quite different from the typical fast food restaurant. Second, its origin in New York City, and its focus on up-scale, urban markets is a fundamentally different positioning than competing fast food chains. Yet, despite these differences in aesthetics and brand, the functional aspect of the firm is essentially identical to other fast food restaurants: Shake Shack sells burgers, fries, and milkshakes, quickly, in a counter service format. Hence, Shake Shack is inherently drawing on existing branding concepts to create a new, hybrid brand.

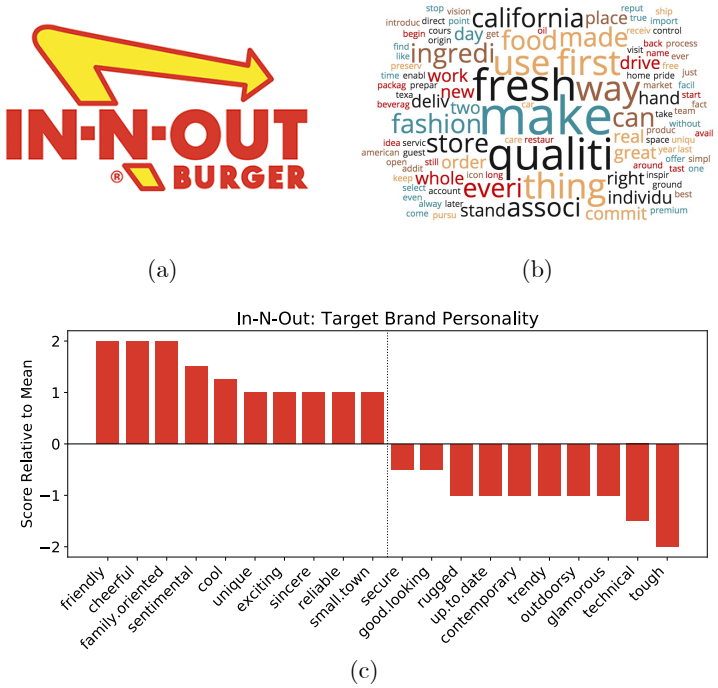
To establish in a data-driven fashion whether Shake Shack's identity is indeed typical of their desired market positioning, we first gather the same data for Shake Shack as we had for the brands in our calibration sample: we select Shake Shack's most typical logo, extract the words from their website, and identify relevant industry tags. For brand personality, rather than returning to MTurk to elicit personality perceptions, we instead approximate the personality that we think Shake Shack is trying to capture. This mirrors the design process, where personality would be something the brand is targeting, rather than something that is observed. We can then use this aspirational personality in suggesting logo features, and see if the actual Shake Shack logo achieves this perceptual goal. In Figure 15, we show Shake Shack's logo, the words from its website, represented as a word cloud, and our assumed target brand personality for Shake Shack (again, relative to the mean). We process these data in an identical fashion as our training data, creating a new set of features which can be used by our model, and in particular, our task-specific inference networks. We also gather and process data for another brand, In-N-Out, to provide a point of comparison in our analyses. In-N-Out also operates in the fast food space, but has a longer history than Shake Shack, and a more typical fast food brand identity. The features of In-N-Out are summarized in Figure 16.



(a) (b)

(c)

Figure 15: (a) Shake Shack's typical logo; (b) the processed words from Shake Shack's website, where the word size correlates with how often that word appeared; (c) a potential target brand personality for Shake Shack, showing the top 10 and bottom 10 personality traits.



(a) (b)

(c)

Figure 16: (a) In-N-Out's typical logo; (b) the processed words from In-N-Out's website, where the word size correlates with how often that word appeared; (c) a potential target brand personality for In-N-Out, showing the top 10 and bottom 10 personality traits.

Categorical Features			Binary Features		
Feature	Likely Values	Prob	Feature Class	Likely Values	Prob
Dominant color:	Black ■	0.445	**Accent color:**	Light Gray ■	0.999
	Dark Blue ■	0.444		Black ■	0.572
	Red ■	0.018	**Contains color:**	Light Gray ■	0.999
Hull shape:	Medium Oval	0.383		Black ■	0.998
	Thin Oval	0.370	**Font:**	Width: Original	0.999
	Triangle	0.246		Style: No Italics	0.999
Sans/Serif Font:	Sans	0.999		Weight: Light	0.968
Mark Class:	Thin	0.538		Class: **Geometric**	0.551
	Detailed circular design	0.307	**Other:**	High % Whitespace	0.999
	Hollow circle	0.083		Has a Mark	0.993
Number of Colors:	Three colors	0.731		High Perimetric Complexity	0.929
	Many colors	0.268		High # Regions	0.718

Table 5: Visual profile corresponding to $\boldsymbol{z}_{\text{ShakeShack}}$, as inferred from the designer's inference network, illustrating the likely values of the categorical features at left, and a selection of high probability binary features at right. We only report binary features that are relatively more likely than the population mean.

Designer's Task To start, we consider the task of designing Shake Shack's logo, based on their targeted brand personality, as well as a description of the brand. Under our framework, this task is equivalent to using Shake Shack's website text, industry tags, and target brand personality as inputs to the designer's inference network, from which we infer an approximate posterior for $\boldsymbol{z}_i$. We then sample from that posterior to produce a distribution over Shake Shack's predicted logo features.[16] We summarize the predicted visual features in Table 5.

Comparing these predictions to the actual logo shown in Figure 15, we see they are fairly accurate. The black colors, medium oval hull, sans-serif font, and detailed circular design of its mark are all spot on. Moreover, in terms of binary features, the true logo's font is indeed original width, no italics, light, and in the geometric font class. Especially relative to other fast food logos, there is a high amount of whitespace, it does have a mark, and the thin but complex features, particularly the mark, are of relatively high perimetric complexity. The only conspicuous difference between the true logo and the prediction have to do with the accent colors: the model predicts light gray with near certainty, while the true logo features neon green. The light gray is likely an artifact of the feature extraction process: when thin, black features are imposed on a white background, the color quantization procedure described in the web appendix nearly always erroneously detects a light gray color, in addition to the black. This also accounts for the prediction of three colors. Light

[16] It is important to note that this operation is *out-of-sample*: Shake Shack's logo is not used in learning the parameters of any of the functions in our model, nor is it used in this case to compute the approximate posterior.

green, on the other hand, is not predicted anywhere. The green burger icon emphasizes the crucial role of the designer, in going above and beyond the typical features suggested by the algorithm: the neon green, thin burger is reminiscent of the signage at a typical 1950's "burger joint," with the burger explicitly indicating the industry.[17] Taken together, these results imply that, while Shake Shack's visual identity is different from competitors in the fast food space, it is also, in some sense, typical: many of its visual features are predictable from its website text and a targeted young, trendy, and glamorous brand personality.

Shake Shack's predicted visual profile contrasts starkly with the model's predictions for In-N-Out: for In-N-Out, the model overwhelmingly predicts a red dominant color (prob = 0.917). Moreover, it predicts just one or two colors, with dark gray and yellow being predicted accent colors. Sans-serif fonts no longer completely dominate, with serif font being predicted with probability 0.347. Other visual features include high entropy, a low perimetric complexity, low percentage whitespace, and a low number of corners, all of which are accurate predictions, and reflect the fast food industry norms, rather than the edgier styling of Shake Shack. These differing predictions are driven by the differing emphasis in the target brand personality, as well as the different words emphasized on the two firms' websites, as captured in Figures 15 and 16.

Manager's Task Now, we consider the brand manager's problem: given the brand's logo, as well as website text and industry tags, how will consumers likely perceive that brand? Similar to the designer's task, to answer this question using our model framework, we use the manager's inference network to infer an approximate posterior distribution over the brand's latent $\boldsymbol{z}_i$, using the logo features, textual data, and industry tags. Then, we simulate a predictive distribution over brand personality perceptions, using this approximation.

For both Shake Shack and In-N-Out, the predicted perceptions are largely in line with our expectations. In Figure 17, we compare the two sets of predictions for the subset of brand personality traits that were predicted to be at least 0.5 points different from the overall trait mean, for at least one of the brands. We display the predictions relative to the population mean (e.g., both brands are predicted to be perceived as more cheerful than an average brand, but less technical). Notably, we see Shake Shack is predicted to excel on perceptions of cool, glamorous, good-looking, trendy, and

[17] https://www.fastcompany.com/3041777/the-untold-story-of-shake-shacks-16-billion-branding

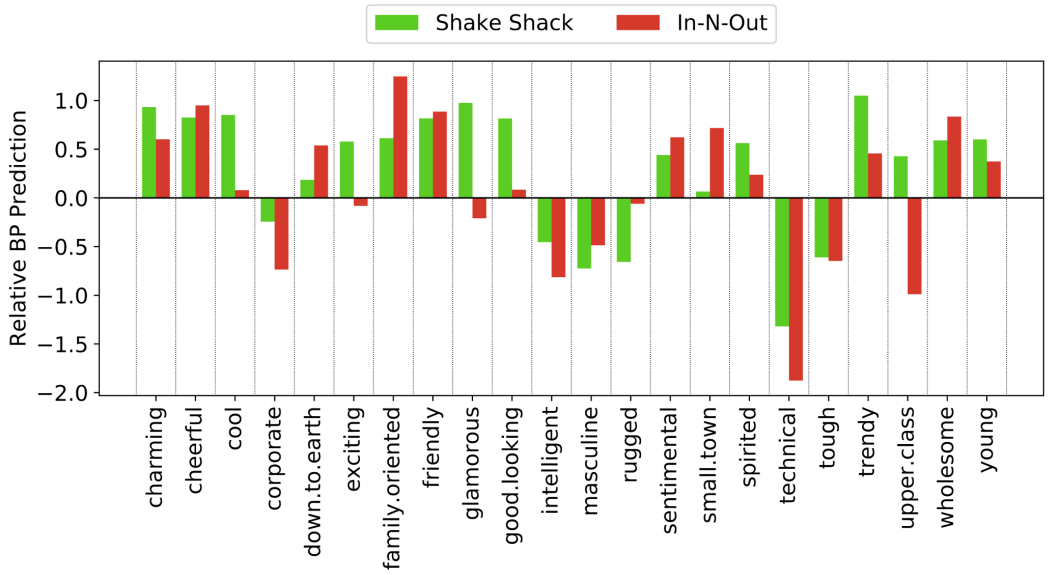


Figure 17: Predicted brand personality perceptions for both Shake Shack and In-N-Out, displayed as points different from the population mean (i.e., relative to the population mean). We show only traits that were predicted to be at least 0.5 points different from the overall trait mean for at least one of the brands.

upper class, while In-N-Out is predicted to be perceived as less corporate, more family-oriented, more small town, and substantially less upper class. These differences are very much in line with our expectations: in both cases, the correlations between the predicted BP profiles and the target BP profiles displayed in Figures 15 and 16 are close to 0.8.

Assessing Visual Changes Finally, we consider the task of assessing changes to a brand's visual identity, as are often considered when weighing competing designs for a new brand, or when an established brand is considering rebranding. The effect of proposed changes to a logo can again be assessed directly in our model framework, by using the manager's inference network to see how the model's predictions about consumer perceptions change with different logo feature inputs, conditional on the brand's textual description and industry tags.

To illustrate this, we consider a simple example: how would consumer perceptions about Shake Shack change if the firm had used a bold font weight, rather than a light font weight? Our model predicts that such a logo change would increase Shake Shack's perceptions along dimensions including family-oriented, technical, sincere, outdoorsy, down-to-earth, and wholesome, while decreasing perceptions along the glamorous, good-looking, daring, young, and smooth dimensions. In some cases, the effects are quite substantial in magnitude: for instance, the predicted positive change

in family-oriented is 0.37, compared to a standard deviation in family-oriented across brands of 0.72. Of a similar magnitude, the expected negative change in glamorous is -0.24, compared to a standard deviation in glamorous across brands of 0.66.

Notably, the model can also make predictions for more complicated changes in aesthetics and firm function. Consider, for instance, a proposed entry of Shake Shack into the consumer goods space, paired with a change in its logo featuring a new, bold font, and a new circular design. Similar to before, we can update Shake Shack's industry tags to include "Consumer Goods," we can change its font to bold, and its logo hull to circular, and then use the manager's inference network to understand how brand perceptions would change. In this case, the model predicts that the same dimensions of family-oriented, sincere, and technical would again rise, although family-oriented would rise by a much larger magnitude (0.63). On the other hand, the dimensions that would suffer now include independent, leader, and successful, each of which would be expected to fall substantially, by approximately one standard deviation. Together, these two examples illustrate the ability of our model to aid brand managers in assessing the potential impact of changes in aesthetics and brand positioning on consumers' perceptions of the brand.

8. Conclusion

In this paper, we explored logo design and brand identity from a data-driven perspective. Leveraging a relatively large dataset of prominent brands, a novel logo feature extraction algorithm, and both model-free and model-based analyses, we showed that many aspects of the design and branding processes can be predicted from data, including which features brands use in their logos, and how consumers perceive these brands' personalities. Moreover, we showed how our multiview representation learning approach yields both a mathematical framework for ideation through brand arithmetic, and a set of decision support tools that can be used to systematically approach the design process.

From a methodological perspective, our contributions are twofold: first, we developed an automatic approach for extracting meaningful and manipulable features from logos. Second, we developed a multiview learning framework based on multimodal variational autoencoders, with a novel approach to inference. Our inference procedure combines task-based inference networks with

stochastic data binning, and is especially suitable for the simultaneous estimation of multiple inference networks that are geared towards providing decision support tools for managers as well as designers. By combining these two methodological advances, we contribute to a nascent literature on interpretable machine learning: our feature extraction algorithm produces interpretable features, which, when combined with our complex, nonlinear generative model, produce interpretable recommendations and insights. Moreover, our model-free and model-based analyses facilitate a scalable understanding of how logo design patterns vary across different industries and brand personalities.

Finally, there are several important limitations of this study. Foremost, ours is a model of logo typicality, not optimality. We are able to capture what a typical firm does, not what is the best logo for a firm, given objectives other than typicality. While exploring optimality of designs may pose an interesting future research area, the task of moving from a typical logo to an optimal logo may also be better suited to a human designer, who can add the creative flair that characterizes the most successful logos (e.g., the FedEx arrow, the Amazon "a to z"), beyond what our model-based approach can suggest. Additionally, our model does not make strong claims about the causality of design: that is, it does not answer why existing logos are designed the way they are, but rather conditions on the existing design landscape. 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 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.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

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.

Jaques, N., Taylor, S., Sano, A., and Picard, R. (2017). Multimodal autoencoder: A deep learning approach to filling in missing sensor data and enabling better mood prediction. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 202–208. IEEE.

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 Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kingma, D. P., Mohamed, S., Rezende, D. J., and Welling, M. (2014). Semi-supervised learning with deep generative models. In *Advances in neural information processing systems*, pages 3581–3589.

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.

Liu, L., Dzyabura, D., and Mizik, N. (2018). Visual listening in: Extracting brand image portrayed on social media. In *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*.

Liu, X., Lee, D., and Srinivasan, K. (2019). Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *Journal of Marketing Research*, 56(6):918–943.

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.

Suzuki, M., Nakayama, K., and Matsuo, Y. (2016). Joint multimodal learning with deep generative models. *arXiv preprint arXiv:1611.01891*.

Tran, D., Kucukelbir, A., Dieng, A. B., Rudolph, M., Liang, D., and Blei, D. M. (2016). Edward: A library for probabilistic modeling, inference, and criticism. *arXiv preprint arXiv:1610.09787*.

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.

Web Appendix

A. Logo Feature Details

Table 6: Logo features with descriptions and links to past literature.

Category	Feature	Description	Original Type	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	H ow 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	
	# 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[18]	
	Class	Vox-ATypI font class		
	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	
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	
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); Deng and Kahn (2016); Schlosser et al. (2016)
	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	

Note that the features are grouped according to their theoretical basis in the literature. In the model, each feature is treated independently.

* The font variables are originally counts: we match every identified character 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). As this matching is just a noisy approximation of the font features, we further process them to form features: for example, we model sans versus serif as a categorical variable, where the outcome is the type with the highest count, and we model "Weight: Bold" as a binary variable, which equals one if at least 25% of the identified letters have the bold descriptor.

B. Technical Details on the Logo Feature Extraction Algorithm

We now give more of the technical details of our image processing algorithm. For specific features, see Web 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 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 use this and most data analysis software imports images in this form. An alternative representation, that we use 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 18. 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 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.

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 19. Given this format, determining

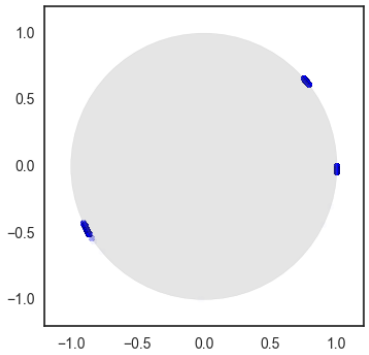


Figure 18: The three colors from Burger King's logo (blue, red, and yellow), plotted as the Hue value from HSV in polar coordinates. Here, red is the cluster of points at right, yellow is the cluster in the top-right, and blue is the cluster in the bottom-left. This is the space on which the DBSCAN clustering operates.

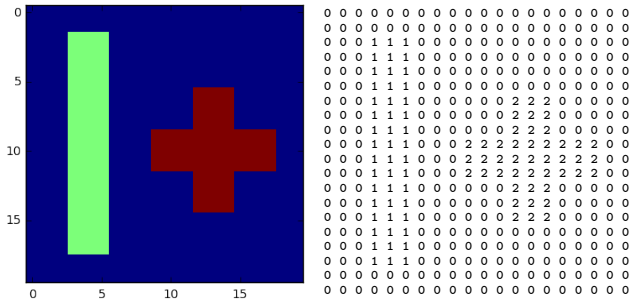


Figure 19: An example of color quantization: the image at left is quantized, yielding the matrix representation at right, where 0 corresponds to blue, 1 to red, and 2 to green.

distinct regions of the logo is as simple as identifying connected regions of this matrix. 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 close-by 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 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.

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 run another clustering algorithm on these triples across logos.[19] 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 and colors: we convert each hull and each mark to a 25×25 standardized greyscale representation,

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

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. Implementation Details

We implement this framework using Tensorflow and the Edward probabilistic programming language (Tran et al., 2016). Recall that our training procedure consists of optimizing a *per iteration* loss function:

$$\ell_m(\boldsymbol{\theta}, \{\boldsymbol{\phi}_t\}) = \sum_{i=1}^{N} \sum_{\forall t} \delta_{itm} \Big\{ - E_{\boldsymbol{z}_i \sim q_{\phi_t}(\boldsymbol{z}_i \,|\, \boldsymbol{\xi}_{i,t} = \text{INet}_t(\tilde{\boldsymbol{x}}_i^t; \boldsymbol{\phi}_t))} \left[\log p_\theta(\boldsymbol{x}_i \,|\, \boldsymbol{z}_i) \right] + \\ \text{KL} \left[q_{\phi_t}(\boldsymbol{z}_i | \tilde{\boldsymbol{x}}_i^t) \;||\; p(z_i) \right] \Big\}. \quad (12)$$

At each iteration, brands are randomly split across our four inference networks, which allows us to simultaneously learn the parameters of each of our task-specific inference networks. To simplify this batching procedure, for each iteration, we use "mini-batches" of size 600, which are evenly divided across the four inference networks, resulting in 150 observations per network.[20] To perform stochastic gradient descent on this objective function, we use the Adam optimizer (Kingma and Ba, 2014), with learning rate 0.0001, and where the gradient is evaluated at each step of the optimization using a single observation. We run our randomized optimization routine for 1000 iterations. In each iteration, we perform 100 optimization steps. We monitor log loss after each iteration to assess convergence, stopping when the log loss no longer substantially decreases. We show pseudocode for this procedure in Algorithm 1.

To prevent overfitting, complex models such as ours typically rely on regularization methods. We rely on two regularization strategies: first, we implement L2 regularization of the weights of the neural network, which is equivalent to adding a squared penalty function to the loss for the weight parameters (i.e, the loss becomes ELBO $+ \lambda \sum_n w_n^2$, where w_n is a single weight parameter). Second, we employ dropout, which randomly severs the connections between nodes of the neural network during training at a pre-specified rate, r, usually taken to be $r = 0.5$ (Goodfellow et al., 2016). Both of these are standard in the deep learning literature, and were implemented using the built-in Tensorflow functionalities.

To determine all model hyperparameters, including the number of latent dimensions (K), layer sizes, number of hidden layers, and degree of regularization, we performed grid search over a wide array of values, assessing model performance using three key metrics, two of which mirror the fit

[20] While these are technically "mini-batches," there are only 706 firms in the full data, which means that 600 observations is nearly the full data.

Algorithm 1 Inference pseudocode: stochastic binned optimization of the ELBO

```
Set Adam target learning rate β*
Initialize φ, θ and learning rate β
Set m = 0
while not converged do
    m = m + 1
    Draw minibatch: B_m = sample(1:N_Train, 600, replace = FALSE)
    Randomize i ∈ B_m to tasks t = 1, ..., 4
    for s = 1, ..., 100 do
        Draw single observation from B_m to estimate the (stochastic) gradient ∇ℓ_m(θ, {φ_t})
        Update φ, θ via single Adam step
        Update β (Edward automatically decays the learning rate toward the target)
    end for
    Evaluate ℓ_m(θ, {φ_t}) and compare to ℓ_{m−1}
end while
```

statistics described in the paper: (1) cross-validation *reconstruction error* using the full inference network, which is equivalent to model fit for new brands when using the full inference network; (2) cross-validation *predictive performance*, which is equivalent to how well the model predicts held-out domains for new brands using the task-specific inference networks; and (3) bias in the generative model. To assess (1) and (2), we compute the mean absolute deviation (MAD) for each domain, which, as described in the body of the paper, is given by:

$$\text{MAD}_d = \frac{1}{N}\frac{1}{J}\sum_{i=1}^{N}\sum_{j=1}^{J}|x_{ij}^d - E(x_{ij}^d)|, \tag{13}$$

where $E(x_{ij})$ is the expected value of x_{ij} under the model. We compare the predictions of our model to the no information rate (NIR), which is equivalent to using the empirical mean as the predicted value, $E(x_{ij}) = \bar{x}_{ij}$. To evaluate (3), bias in the generative model, we randomly simulate 1000 brands by drawing z_i randomly from the $\mathcal{N}(0, 1)$ prior, then see how the means and standard deviations of the features of the simulated brands compare to the full data empirical means and standard deviations observed in the data. This third procedure is effectively a posterior predictive check, which assesses whether the learned generative model in fact generates data that looks similar to the real data, up to the first and second moments. There is often a trade-off between these three metrics: better reconstruction error (1) may lead to worse predictions (2), or a biased generative model (3). Similarly, better predictions may lead to a worse generative model. Hence, we look for a specification that does well on all three metrics; the ultimate specification relies on researcher judgment, weighing all these outcomes.

Through grid search, we found the optimal dimensionality of the latent space was $K = 40$. We also found no benefit to increasing the number of hidden layers of any of the networks above one (i.e., single-layered, feed-forward networks). This is likely because we are already working with highly processed inputs, thus limiting the usefulness of the increasing levels of abstraction

enabled by adding more layers. The optimal model structure used 1024 hidden units in all of the inference networks, 1024 hidden units in the text decoder network, and 512 hidden units for all other decoders. Finally, for regularization, we found a mild L2 penalty for the decoder networks (penalty coefficient $\lambda = 10$), and a dropout rate of $r = 0.5$ applied to both the inference and decoder network performed best.