Letting Logos Speak: Leveraging Multiview Representation Learning for Data-Driven Branding and Logo Design

Ryan Dew*1, Asim Ansari2, and Olivier Toubia2

1 The Wharton School, University of Pennsylvania
2 Columbia Business School, Columbia University

January 8, 2021

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 the interplay between logo design and brand identity creation 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 successful brands, to understand which brands select 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 paper, we build a multiview representation learning framework that captures the linkages between a brand’s function, its logo features, and consumer perceptions of its brand personality. This framework allows us to mathematically embed a diverse set of brands in a latent space, which in turn provides a mechanism for exploring many interrelated questions about design and branding from a data-driven, holistic perspective.

In particular, there are three intertwined perspectives, with their own unique questions, that we address through our representation learning framework:

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.

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 consumer’s perspective.** What associations exist among logo features, brand function, and brand perception? That is, given a logo, and some knowledge of what a firm does (i.e., its industry), what inferences will a consumer make about what the brand with that logo stands for, or what the personality of that brand is? This is the perspective that branding researchers have focused on in much of the prior literature, which has examined how particular logo features impact consumer perceptions about a firm.

Underpinning the answers to each of these questions are the interrelated processes by which con-
sumers perceive logos, and designers design them. When consumers encounter a new logo, the vast literature on consumer information processing suggests that they evaluate this new logo on the basis of logos they have encountered before (Loken et al., 2008; Kardes et al., 2008). Likewise, when designers design logos, they do so with this process implicitly in mind. For example, in mood boarding, one of the most common brainstorming techniques in practice (e.g. Endrissat et al., 2016; Miller, 2016), designers take concepts from a company-supplied brief, and generate visual elements that link to those concepts. Part of that process often involves thinking of existing brands that already draw on those concepts, or on common design elements that have been used historically to evoke those concepts.

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 employ a variant of the 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 develop 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 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. Background and Conceptual Framework

There is a sizable literature, especially in consumer behavior, on how consumers react to logos and marketing aesthetics, and how consumers process information related to brands. 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 first motivate our model theoretically by tying it to the literature on consumer information processing. Then, we briefly review the findings about how consumers perceive logos, with a specific focus of informing our logo feature extraction algorithm, described in Section 4.

2.1. Conceptual Framework

Theoretically, our application of multiview representation learning to modeling the branding process is rooted in the information processing literature. When confronted with an unfamiliar logo, categorization theory suggests that consumers make inferences about this new brand based on the degree to which it activates existing mental categories (Loken et al., 2008). Building on these ideas, studies of category-based inference suggest that consumers compare the features of a target stim-
ulus with features of a category, to determine if the stimulus is a member of that category (Kardes et al., 2008). If sufficient overlap of features exists, consumers imbue the stimulus with the typical associations of the category. Such effects have been frequently studied in branding contexts, often in the context of brand extensions or new product introductions, where it has been shown that the more overlap exists between the brand extension or product and the parent brand, the more associations will be transferred to the new stimulus.

In designing logos, designers often implicitly rely on categorization theory as the basis of ideation. A common technique employed by designers is mood boarding or image boarding whereby visual elements corresponding to a specific concept or theme are composed on a board, to stimulate thinking about the visual associations of a category, concept, or theme (McDonagh and Storer, 2004; Stigliani and Ravasi, 2012; Endrissat et al., 2016). In relying on these methods, designers implicitly draw on the idea that consumers will evaluate a new design by the concepts that the design activates in the consumer’s mind. In this context of logos, mood boards often include logos, or logo features like fonts and colors, drawn from existing brands in the focal categories, that designers believe will activate desired associations (e.g. Miller, 2016). These mood boards then serve as the basis for ideation for a final logo design.

The processes by which consumers evaluate logos and designers design them both draw on the idea that there exists a landscape of brands and logos, which forms the basis of consumers mental categories, and in which designers must position new logos. Our model-based approach to design seeks to mathematically recreate this landscape through representation learning, by embedding brands in a learned latent vector space, where a firm’s position in that space predicts what that firm does, how that firm describes its brand through text, the visual features of that firm’s logos, and how consumers perceive that firm’s brand personality. In this way, the learned space captures a semantic map of brands and their logos in the present-day consumer conscience, which serves as the basis for our data-driven approach to design ideation and decision support.

2.2. 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.

2.3. 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. 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.

---

1The 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.

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 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.
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).

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 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 expository 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 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 grayscale, and then cluster all marks into 14 representative mark types. We give examples of these classes in Figure 5.

**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

<table>
<thead>
<tr>
<th>Name</th>
<th>R</th>
<th>G</th>
<th>B</th>
<th>Color</th>
<th>Name</th>
<th>R</th>
<th>G</th>
<th>B</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>253</td>
<td>253</td>
<td>253</td>
<td>White</td>
<td>Dark Blue</td>
<td>30</td>
<td>42</td>
<td>124</td>
<td>Dark Blue</td>
</tr>
<tr>
<td>Black</td>
<td>20</td>
<td>18</td>
<td>10</td>
<td>Black</td>
<td>Light Gray</td>
<td>165</td>
<td>164</td>
<td>167</td>
<td>Light Gray</td>
</tr>
<tr>
<td>Red</td>
<td>226</td>
<td>33</td>
<td>41</td>
<td>Red</td>
<td>Light Blue</td>
<td>54</td>
<td>153</td>
<td>204</td>
<td>Light Blue</td>
</tr>
<tr>
<td>Blue</td>
<td>25</td>
<td>89</td>
<td>152</td>
<td>Blue</td>
<td>Light Green</td>
<td>99</td>
<td>178</td>
<td>67</td>
<td>Light Green</td>
</tr>
<tr>
<td>Dark Green</td>
<td>34</td>
<td>120</td>
<td>77</td>
<td>Dark Green</td>
<td>Yellow</td>
<td>245</td>
<td>202</td>
<td>36</td>
<td>Yellow</td>
</tr>
<tr>
<td>Orange</td>
<td>239</td>
<td>131</td>
<td>40</td>
<td>Orange</td>
<td>Tan</td>
<td>186</td>
<td>164</td>
<td>103</td>
<td>Tan</td>
</tr>
<tr>
<td>Dark Gray</td>
<td>116</td>
<td>111</td>
<td>111</td>
<td>Dark Gray</td>
<td>Dark Red</td>
<td>174</td>
<td>39</td>
<td>63</td>
<td>Dark Red</td>
</tr>
</tbody>
</table>

**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.

![Color dictionary](image)

**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.

![Hull classes](image)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Sample of marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td><img src="image" alt="Sample of marks" /></td>
</tr>
<tr>
<td>7</td>
<td><img src="image" alt="Sample of marks" /></td>
</tr>
<tr>
<td>9</td>
<td><img src="image" alt="Sample of marks" /></td>
</tr>
</tbody>
</table>

**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.

![Mark classes](image)

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

![Font classification system](image)
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 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/](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 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-
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.
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 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.
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.

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. 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.

3We use the terms modality, data source, and domain interchangeably.
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, $x_i$, for each observation $i$, in terms of a multidimensional latent variable $z_i$. The mapping between the latent variable $z_i$ 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 $\theta$. The joint distribution of the data and the latent variables is given as $p(\theta(x_i, z_i) = p(x_i|z_i)p(z_i)$, where the prior for $z_i$ is assumed to be isotropic Gaussian, $p(z_i) = N(0, I)$.

To approximate the posterior of the latent variables, $p(\theta(z_i|x_i)$, VAEs rely on amortized variational inference, where the approximating variational distribution $q(\phi(z_i|x_i)$ is specified using another neural network, called the encoder or inference network. Note that the inference network uses the available data $x_i$ as its input to specify the variational distribution for the observation-specific $z_i$. The weights and biases of this network, $\phi$, 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 $d \in \{\text{Text, Logo, Ind, BP}\}$, 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 \( x^d_i \) and the complete observation is given by \( x_i = \{ x_{i}^{\text{Text}}, x_{i}^{\text{Logo}}, x_{i}^{\text{Ind}}, x_{i}^{\text{BP}} \} \). 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 \( x^d_i = \{ x^d_{i1}, \ldots, x^d_{iV_d} \} \). The generative model specifies the probability distribution of the observed data in each domain in terms of a shared latent variable vector \( z_i \). 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 data \( x_i \) to infer the common latent variable \( z_i \). 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 \( x_i \) in terms of a shared multidimensional latent variable \( z_i \), which has an isotropic Gaussian prior \( p(z_i) = N(0, I) \). As in the standard VAE, the joint distribution of the data and the latent variables is given as \( p_\theta(x_i, z_i) = p_\theta(x_i | z_i)p(z_i) \). However, the probability models for the different domains are independent, conditional on \( z_i \) i.e., \( p_\theta(x_i | z_i) = \prod_d p_{\theta_d}(x^d_i | z_i) \).

In turn, the probability model for each domain is specified using independent feature-level probability distributions such that \( p_{\theta_d}(x^d_i | z_i) = \prod_j p^d_{\theta_{d_j}}(x^d_{ij} | z_i) \). Let \( \mu^d_i \) contain the parameters for the different feature-level distributions associated with observation \( i \) within domain \( d \). A domain-
specific decoder network, which we denote $DNet_d(z; \theta_d)$, captures the non-linear relationship between $\mu_d^j$ and $z_i$, such that $\mu_d^j = DNet_d(z; \theta_d)$. 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 $z_i$, 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:

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

  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:

\[
\begin{align*}
\mathbf{x}_{ij}^{\text{Logo}} & \sim \text{Categorical}(\text{softmax}(\mathbf{\mu}_{ij}^{\text{Logo}})), \quad (2) \\
\mathbf{\mu}_{ij}^{\text{Logo}} & = (\mu_{ij1}^{\text{Logo}}, \ldots, \mu_{ijM_j}^{\text{Logo}}), \quad (3)
\end{align*}
\]

where,

\[
\text{softmax}(\mathbf{\mu}_{ij}^{\text{Logo}}) = \left( \frac{\exp(\mu_{ij1}^{\text{Logo}})}{\sum_{n=1}^{M_j} \exp(\mu_{ijn}^{\text{Logo}})}, \ldots, \frac{\exp(\mu_{ijM_j}^{\text{Logo}})}{\sum_{n=1}^{M_j} \exp(\mu_{ijn}^{\text{Logo}})} \right)
\]

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:

\[
\begin{align*}
\mathbf{x}_{ij}^{\text{BP}} & \sim \mathcal{N}(\mu_{ij1}^{\text{BP}}, \sigma_{ij}^{\text{BP}}), \quad \sigma_{ij}^{\text{BP}} = \log(e^{\mu_{ij2}^{\text{BP}}}/2 - 1)). \quad (4)
\end{align*}
\]

In the above feature-level distributions, the observation-specific distributional parameters (e.g., the mean \( \mu_{ij1}^{\text{BP}} \) and the variance \( \sigma_{ij}^{\text{BP}} \) of the normal in Equation 4) are specified non-linearly in terms of the latent variable \( z_i \) for that observation using modality-specific decoder networks.

**Decoder Network**  We use a domain-specific decoder network, \( \mathbf{\mu}_i^d = \text{DNet}_d(z_i; \theta_d) \), to model the potentially non-linear relationship between \( \mathbf{\mu}_i^d \) and \( z_i \). In our application, we use dense, feed-forward layers with rectified linear activation units (ReLU) and skip connections to specify \( \log(e^y - 1) \) 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.
DNet\(_d()\) for each domain. This is equivalent to the following sequence of computations:

\[
\begin{align*}
\mathbf{h}_{1}^{\text{Dec},d} &= \text{ReLU}(a_0^d + W_0^{d,z}z_i), \\
\mathbf{h}_{2}^{\text{Dec},d} &= \text{ReLU}(a_1^d + W_1^{d,h_{Dec},d} + W_1^{d,z}z_i), \\
\vdots \\
\mathbf{h}_{L_d}^{\text{Dec},d} &= \text{ReLU}(a_{L_d-1}^d + W_{L_d-1}^{d,h_{Dec},d} + W_{L_d-1}^{d,z}z_i), \\
\mu_i^d &= a_{L_d}^d + W_{L_d}^{d,h_{Dec},d} + W_{L_d}^{d,z}z_i,
\end{align*}
\]

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 \(\ell\) computes a transformed representation of the brand through the hidden units, whose activations are contained in the vector \(\mathbf{h}_{\ell}^{\text{Dec},d}\), 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 \(z_i\). The \(a_{\ell}^d\) vectors contain the biases (intercepts) associated with the hidden units in layer \(\ell\). Note that we combine the hidden unit activations with the original representation \(z_i\), 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 \(\mu_i^d\) 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.

We use \(\theta_d\) 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.

**Task-specific Inference Networks** The key task in using the MVAE framework is to learn the joint latent representations \(z_i\). In our work, we follow the standard practice of assuming a mean-
field variational approximation for the posterior of $z_i$. The approximate posterior is given by the normal distribution:

$$p_\theta(z_i|x_i) \approx q_\phi(z_i|\xi_i) = \mathcal{N}(\xi^m_i, \text{diag}(\xi^v_i)), \quad (6)$$

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

$$
\begin{align*}
    h_{i1}^{\text{Inf}} &= \text{ReLU}(c_0 + V_0x_i), \\
    h_{i2}^{\text{Inf}} &= \text{ReLU}(c_1 + V_1h_{i1}^{\text{Inf}}), \\
    &\vdots \\
    h_{iL}^{\text{Inf}} &= \text{ReLU}(c_{L-1} + V_{L-1}h_{i(L-1)}^{\text{Inf}}), \\
    \bar{\xi}_i &= c_L + V_Lh_{iL}^{\text{Inf}}.
\end{align*}
\quad (7)
$$

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

In our application, it is important to be able to infer $z_i$ given information on only a subset of the domains. This involves using brand-specific data on some subset of the domains to compute $z_i$, 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 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 consumers’s inference network, corresponding to the case where we observe

\(^6\)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.
the logo and industry tags. That is, we learn four distinct inference networks, which we index by $t \in \{\text{Full, Des, Mgr, Res}\}$, where $t$ stands for task, corresponding to four separate functions,

$$\tilde{x}_{i,t} = \text{INet}_t(\tilde{x}_{i,t}; \phi_t),$$

where $\tilde{x}_{i,t}$ is shorthand for the data available for inference task $t$ (for example, for $t = R$, $\tilde{x}_{i} = x_i^\text{Logo}$). 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 $\theta$ and $\phi$, of the decoder and inference networks, such that encoding and then decoding data $x_i$ 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:

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

where KL$(\cdot \mid \cdot)$ 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, $z_i$, 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 $p_\theta(x_i \mid z_i)$ from Equation 8 decomposes into a prod-
uct 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 \( q_f(z_i | \xi_i = \text{INet}(x_i; \phi)) \) 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:

\[
\ell_m(\theta, \{\phi_i\}) = \sum_{i=1}^{N} \sum_{t} \delta_{itm} \left\{ - E_{z_i \sim q_{\theta_t}(z_i | \xi_i = \text{INet}(x'_i; \phi_t))} \left[ \log p_\theta(x_i | z_i) \right] + \right.
\]

\[
\left. \quad \text{KL} \left[ q_{\phi_t}(z_i | \tilde{x}_t^i) \middle| \| p(z_i) \right] \right\} \quad (9)
\]

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

We implement our model using PyTorch and the Pyro probabilistic programming language (Bingham et al., 2018). We optimize the loss in Equation 9 using stochastic gradients and the Adam algorithm (Kingma and Ba, 2014). To prevent overfitting, complex models such as ours typically rely on regularization methods. We utilize dropout for this purpose (Goodfellow et al., 2016). We also employ KL annealing, a strategy that has been shown to improve the representations learned by VAE-type models, wherein the KL term in the loss function is given a weight that is slowly annealed from zero to one during the first part of model training (Bowman et al., 2015). 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 an array of values, assessing model performance using cross-validation. From this procedure, we determined an optimal dimensionality of the latent space of \( K = 20 \). We 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. We also found that it is very important to mirror the complexity of the inference network and the complexity of the task (i.e., the number of inputs to that task): our final model architecture consists of 400 hidden units for the full inference network, 200 hidden units for each of the manager’s and designer’s inference networks, and 50 hidden units for the consumer’s inference network. We use 400 hidden units in all of the decoders, and found the model relatively unsensitive to this choice. We include more details on implementation, including pseudocode, in the web appendix.

6.5. Relation to Prior Literature

As noted previously, our framework is a generalization of the standard VAE framework introduced by Kingma and Welling (2013) and Rezende et al. (2014) to multimodal learning. However, ours is not the only multimodal extension of VAEs in the computer science literature. In particular, our framework is closely related to three of these extensions: first, our task-specific inference network framework can be seen as a generalization of the per-domain inference network idea introduced by Suzuki et al. (2016). While their framework is only developed for two domains, ours covers the many modalities case, where the modalities can be grouped into relevant tasks. More recently, Wu and Goodman (2018) introduced an alternate way to handle more than two modalities through a product-of-experts formulation. They also introduce a subsampled training procedure, which as previously noted, is quite similar to ours. Their framework is more general than ours, in the sense that it does not require the specification of specific tasks of interest. However, this generality comes at the cost of predictive performance, as we show subsequently in our benchmarks. Finally, even more recently, Nazabal et al. (2020) developed a framework for handling heterogeneous inputs in a VAE framework that echoes our per-domain likelihood structure, but uses a different training procedure and representation structure.
7. Model Results

We present the model results in five parts: first, we show the performance of our MVAE framework, on both fit and predictive tasks, and compare it to a wide variety of benchmarks. Second, we describe the learned latent space, by illustrating the similarities that are learned for a set of representative brands, and by showcasing examples of randomly generated brands. Third, 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 provide decision support for designers and managers, through two applications to fast food branding: in the first application, we study the potential consumer-level implications of an actual recently implemented redesign of McDonald’s logo. In the second application, we examine how our model can be used for understanding branding for a new entrant to a space, by analyzing the branding decisions of Shake Shack.

7.1. Fit and Benchmarks

To assess model fit, we ran 4-fold cross validation. We summarize the out-of-sample fit of the model, averaged across folds, and broken down by domain, in Tables 1 and 2. In our MVAE framework, there are important distinctions between two types of fit measures: (1) reconstruction fit, which is computed using the full inference network on the held-out set of brands, and captures how well the model does at recreating the inputs it is given for new; and (2) predictive fit, which shows the model’s ability to predict missing domains for new brands.\(^7\) Table 1 gives fit statistics for reconstruction, while Table 2 gives fit statistics for prediction. While both statistics are out-of-sample, they have distinct interpretations. Good performance on reconstruction indicates that a generative model is able to learn meaningful representations for new brands, which in turn indicates that the learned latent space is truly capturing the statistical signal of the inputs. On the other hand, good performance on prediction indicates the model is able to perform the tasks we specified, in the traditional supervised sense, by being able to successfully predict missing

\(^7\)There is also in-sample reconstruction fit, which is how well the model is able to reconstruct its inputs for the same set of brands it was trained on. Our model does exceptionally well on this in-sample measure, but we do not report it here, favoring the harder out-of-sample metrics.
features from the given features.

We measure fit using two metrics: for the real-valued brand personality features, we compute the mean squared error (MSE), which we then average across all personality traits. For MSE, lower scores indicate better fit. For the binary and categorical features, we use the F1 score. The F1 score is a way of integrating two measures of the success of a classifier, precision and recall, where precision is the fraction of true positives identified by the model out of all positives identified, and recall is the fraction of true positives identified by the model out of all true positives. The F1 score is then defined as the harmonic mean of these two statistics:

\[
F1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{10}
\]

Intuitively, the F1 score will be high for a model that is correctly able to distinguish positive cases from negative cases. We use these metrics, as opposed to naive measures like accuracy, due to the highly imbalanced nature of many of our features. In Tables 1 and 2, we report the average of these statistics across features. We report the precision and recall statistics in Web Appendix D.

We compare the performance of our model, which we denote in the tables as TSI (task-specific inference), to several benchmarks:

- **POE**: the product-of-experts framework developed by Wu and Goodman (2018). Specifically, rather than using our task-specific inference networks, we follow that paper’s product-of-experts strategy, where each domain has its own latent representation \( z_d \), and these representations are combined using a product-of-normals rule.

- **PPCA**: a version of our model with no nonlinearities in the generative model, which is equivalent to doing probabilistic PCA (PPCA) with task-specific (amortized) inference.

- **Designer**: an adaptation of our framework with only the designer’s task. This model is, essentially, a supervised model for predicting logo features from other features, and thus there is no reconstruction cross-validation fit, only prediction.

- **NIR**: the no information rate, which is the naive model where each feature is predicted to have its mean value across all of the brands in the training set.
### Table 1: Average reconstruction cross validation error using the full inference network, where each column is a different model. Note that MSE is the mean squared error, where higher numbers indicate worse fit, while F1 is the harmonic mean of recall and precision, where higher numbers indicate better fit.

<table>
<thead>
<tr>
<th>Task</th>
<th>Domain</th>
<th>Metric</th>
<th>TSI</th>
<th>POE</th>
<th>PPCA</th>
<th>Designer</th>
<th>NIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designer</td>
<td>Logo: Binary</td>
<td>F1</td>
<td>0.132</td>
<td>0.106</td>
<td>0.089</td>
<td>0.131</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>Logo: Dom. Color</td>
<td>F1</td>
<td>0.096</td>
<td>0.096</td>
<td>0.095</td>
<td>0.086</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>Logo: Hull Shape</td>
<td>F1</td>
<td>0.160</td>
<td>0.149</td>
<td>0.146</td>
<td>0.154</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>Logo: Mark Shape</td>
<td>F1</td>
<td>0.064</td>
<td>0.064</td>
<td>0.065</td>
<td>0.059</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>Logo: Font Serifs</td>
<td>F1</td>
<td>0.319</td>
<td>0.311</td>
<td>0.297</td>
<td>0.306</td>
<td>0.297</td>
</tr>
<tr>
<td></td>
<td>Logo: Num. Colors</td>
<td>F1</td>
<td>0.265</td>
<td>0.244</td>
<td>0.245</td>
<td>0.258</td>
<td>0.124</td>
</tr>
<tr>
<td>Manager</td>
<td>BP</td>
<td>MSE</td>
<td>0.794</td>
<td>0.774</td>
<td>0.811</td>
<td>1.008</td>
<td></td>
</tr>
<tr>
<td>Consumer</td>
<td>BP</td>
<td>MSE</td>
<td>0.834</td>
<td>0.828</td>
<td>0.847</td>
<td>1.008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Text</td>
<td>F1</td>
<td>0.014</td>
<td>0.017</td>
<td>0.011</td>
<td>0.005</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Average prediction cross validation error using the full inference network, where each column is a different model. Note that MSE is the mean squared error, where higher numbers indicate worse fit, while F1 is the harmonic mean of recall and precision, where higher numbers indicate better fit.

From the fit statistics, we note several things: first, all models, and especially our TSI framework, do significantly better than random (NIR) at explaining and predicting the data. We see, however, that some domains are more difficult to predict than others, with the text being the most difficult to predict. This difficulty is not particularly surprising: the F1 score for this domain is averaged over all of our textual tokens, treated separately. We also notice that the consumer’s task is quite challenging: in general, error rates in this task are relatively high, suggesting it is difficult to make predictions from a logo and industry tags alone, though we still perform better than chance.

Turning our attention to the more sophisticated benchmarks, we see that our proposed framework is competitive with the state-of-the-art framework proposed in the literature (POE), outper-
forming it on most metrics in both fit and prediction. Comparing our model and PPCA, we see that the nonlinearities in the generative model are especially important for reconstruction tasks, as well as for predicting binary logo features and brand personality. Most interesting, however, is the comparison to the Designer benchmark, where we see our multimodal framework slightly outperform the simpler, unidirectional task. This finding adds to a growing literature on the benefits of multimodal learning, suggesting that jointly learned representations can be tremendously valuable, even in supervised prediction tasks (e.g. Wu and Goodman, 2019).

7.2. Understanding the Latent Space

Having established the validity of the latent space in a predictive sense, we now turn to understanding what it represents. In general, it is difficult to interpret specific dimensions of our learned latent space. Intuitively, the \( z \) representation acts as an information bottleneck, wherein all of the features of the data are compressed to a 20-dimensional vector. Hence, each dimension of \( z \) simultaneously encodes much information, and likewise, specific features tend to be encoded in a distributed way across the dimensions of \( z \). But, 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 at where brands lie in this space, we can better understand what the learned representations are capturing.

In Table 3, we show the two nearest neighbor brands in \( z \)-space for a set of 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. Focusing on the firms in the first row of the table, Facebook’s closest neighbor is Twitter: not only are they both innovative social network platforms, but they both have simple, blue, bulky logos. Similarly, Gucci’s nearest neighbors are Dior and Cartier, who both operate in the luxury retail space, both have similar sophisticated brand personalities to Gucci’s, and both have similar black and white, sleek, high whitespace logos.

For each brand, there is not always another brand that matches it in all four domains of in-
Table 3: The two closest brands to each focal brand in z-space, including their logo, name, and, in parentheses, the distance between the focal brand and the neighbor in z-space.

<table>
<thead>
<tr>
<th>Focal Brand</th>
<th>Neighbors in z-space</th>
<th>Focal Brand</th>
<th>Neighbors in z-space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>Twitter (3.00) Uber (3.60)</td>
<td>Gucci</td>
<td>Dior (3.48) Cartier (3.71)</td>
</tr>
<tr>
<td>Lowe’s</td>
<td>TravelCenters (3.93) Union Pacific (4.05)</td>
<td>McDonald’s</td>
<td>Heinz (4.84) Wells Fargo (4.87)</td>
</tr>
</tbody>
</table>

Consider, for instance, Lowe’s, shown in the second row of the table. Based purely on what a firm does, we might expect the nearest neighbor of Lowe’s to be Home Depot, or another home improvement store, yet the model identifies TravelCenters of America and Union Pacific as its nearest neighbors. These two firms operate in related but distinct industries from Lowe’s, yet share much in common, both in terms of consumer perceptions, and aesthetically. All three firms have logos with the same distinct medium blue color, a lack of whitespace, a bulky design, and even a similar sans-serif font. Since the learned representations capture all domains simultaneously, we see it place Lowe’s closer to these two brands, as opposed to other aesthetically distinct competitors. We can tell a similar story for McDonald’s: the model places Heinz and Wells Fargo closest to McDonald’s in z-space. There are again several reasons for this: all three brands are classic American brands; Heinz operates in food service; they have correlated brand personality ratings, with all three scoring relatively high on traits like family-oriented, western, and small-town; but most obviously, with all three brands sharing a very similar logo design. Finally, focusing on the distances from each neighbor to its focal brand, we see the two examples in the first row have much lower distances to their neighbors than those in the second row. This difference emphasizes the intuition behind z-space: firms are close together when they match on all of the dimensions.
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 \( z_i \) vector from the prior, \( z_i \sim \mathcal{N}(0, I) \), 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 illustrate this, we generated a set of random random \( z_i \) vectors, then fed them through the decoder networks to create a corresponding set of brand concepts. In the remainder of this section, we describe one of those brands, and we include summaries of four others in the web appendix.

Starting first with the industry tags, the most likely tags corresponding to our focal random \( z_i \) are professional services (with probability 0.31, subsequently denoted \( p=0.31 \)), internet services (\( p=0.19 \)), and administrative services (\( p=0.18 \)). In our data, the highest probability tag, professional services is used primarily to describe service providers to other businesses, including consulting services, internet and technology providers, insurance providers, and shipping and logistics companies. Accordingly, the model also predict the firm to operate as both a B2C (\( p=0.78 \)) and B2C (\( p=0.65 \)) company.

To more concretely understand this brand’s function and identity, we can use the decoder network to generate what words would likely appear on this firm’s website. Many brands use many of the same words on their website (e.g., company, customers, product), and hence merely looking at high probability words can be undiagnostic. To isolate the set of words most relevant to our focal brand, we use the relevance score from Sievert and Shirley (2014), defined in our context as:

\[
r(w, i|\lambda) = \lambda \log(\mu_{iw}) + (1 - \lambda) \log\left(\frac{\mu_{iw}}{p_{iw}}\right),
\]

where \( \mu_{iw} \) is the probability of word \( w \) for brand \( i \) under our model, \( p_{iw} \) is the overall probability of a word appearing on any brand’s website in the data, and \( \lambda \) is a hyperparameter, which we set to be 0.6 following the advice of Sievert and Shirley (2014). Intuitively, this metric is designed to isolate words that are particular to our focal brand, that are not common to other brands. We display a word cloud summarizing the high relevance terms for this brand in the left panel of Figure 11.
The words are coherent, and suggest a company that may operate in shipping or logistics, with words like ships, central, capabilities, and territories appearing prominently. Similar, the words institutions, platform, clients, extensive, experts, solutions, and global are all words that suggest a B2B-oriented, professional services company.

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: masculine, western, tough, rugged, and corporate. The lowest relative personality traits are: family-oriented, cheerful, feminine, sincere, and charming. These traits make intuitive sense for the brand described previously.

Finally, in Table 4, we show the visual features corresponding to this randomly generated $z_i$. 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 the right panel of Figure 11. The dark, bold design suggested here again makes intuitive sense: it is easy to imagine these traits used in the logo of a tough, rugged, masculine logistics corporation. The accent color yellow is often used in tough industries, and the suggested design as a whole is reminiscent of companies like HD Supply, Caterpillar, and J.B. Hunt.

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. Elements of these brands’ logos may then be mixed with visual features unique to the new brand. For instance, a
Figure 11: At left, the word cloud describing the most relevant terms for our focal random brand. At right, a simple, nonprofessional rendering of a logo with the features described in Table 4.

<table>
<thead>
<tr>
<th>Categorical Features</th>
<th>Likely Values</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant color:</td>
<td>Black ■</td>
<td>0.631</td>
</tr>
<tr>
<td></td>
<td>Yellow ■</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td>Dark Blue ■</td>
<td>0.089</td>
</tr>
<tr>
<td>Hull shape:</td>
<td>Narrow Rect./Oval</td>
<td>0.440</td>
</tr>
<tr>
<td></td>
<td>Medium Rect./Oval</td>
<td>0.287</td>
</tr>
<tr>
<td>Sans/Serif Font:</td>
<td>Sans</td>
<td>0.444</td>
</tr>
<tr>
<td>Mark Class:</td>
<td>Long, Horizontal</td>
<td>0.178</td>
</tr>
<tr>
<td></td>
<td>Detailed Horizontal</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td>No Mark</td>
<td>0.129</td>
</tr>
<tr>
<td>Number of Colors:</td>
<td>Two colors</td>
<td>0.555</td>
</tr>
<tr>
<td></td>
<td>Three colors</td>
<td>0.236</td>
</tr>
<tr>
<td></td>
<td>One color</td>
<td>0.144</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Binary Features</th>
<th>Likely Values</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature Class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accent colors:</td>
<td>Light Gray ■</td>
<td>0.327</td>
</tr>
<tr>
<td></td>
<td>Dark Gray ■</td>
<td>0.205</td>
</tr>
<tr>
<td></td>
<td>Orange ■</td>
<td>0.155</td>
</tr>
<tr>
<td></td>
<td>Yellow ■</td>
<td>0.122</td>
</tr>
<tr>
<td>Font:</td>
<td>Weight: Bold</td>
<td>0.563</td>
</tr>
<tr>
<td></td>
<td>Width: Ordinary</td>
<td>0.464</td>
</tr>
<tr>
<td></td>
<td>Class: Grotesque</td>
<td>0.827</td>
</tr>
<tr>
<td>Other:</td>
<td>Few downward diagonal edges</td>
<td>0.810</td>
</tr>
<tr>
<td></td>
<td>Has a mark</td>
<td>0.802</td>
</tr>
<tr>
<td></td>
<td>Low % Whitespace</td>
<td>0.800</td>
</tr>
<tr>
<td></td>
<td>High variation in lightness</td>
<td>0.721</td>
</tr>
<tr>
<td></td>
<td>Few upward diagonal edges</td>
<td>0.710</td>
</tr>
</tbody>
</table>

Table 4: Visual profile for the focal randomly generated brand, illustrating the likely values of the categorical features at left, and a selection of high probability binary features at right. Note: gray appears very often as an accent color when the dominant color is black. This is because at the border between black and white, the algorithm often detects a faint line of gray.
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 averaging (or adding) together $z_i$ 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 need to define an average vector,

$$z_{\text{Tag}} = \frac{L}{N_{\text{Tag}}} \sum_{i \in \text{Tag}} z_i$$  \hspace{1cm} (12)

where Tag refers to the set of brands with a given tag (or, more generally, belonging to some pre-specified group of interest), $N_{\text{Tag}}$ is the number of brands in that set, and $L$ is the average Euclidean norm of all of the learned vectors $z_i$. Intuitively, this average is just the average of all of the $z_i$ vectors for all firms in some group, rescaled by the average norm of all of the brand representations. As more vectors are averaged together, their norm tends to become smaller, as the large components of one are cancelled out with the relatively smaller components of others. Hence, to ensure comparability across all vectors in the space, we employ this the *norm-preserving average*.

Returning to our example, then, we define two norm-preserving averages, $z_{\text{Health}}$, which is the average of all brands tagged as Health Care companies, and $z_{\text{Tech}}$, which is the average of all brands tagged as either Hardware, Consumer Electronics, or Software companies. We can then interpolate between these two vectors, to create a new representation for a medical device.
company:

\[ z_{\text{MedDevice}} = 0.5z_{\text{Health}} + 0.5z_{\text{Tech}}. \]  

(13)

To validate that this procedure indeed produces a reasonable representation, we first check which firms are close to the interpolated \( z_{\text{MedDevice}} \): among the ten nearest neighbors to \( z_{\text{MedDevice}} \) are medical device manufacturers Baxter International, Becton-Dickinson, McKesson, and ThermoFisher Scientific, medical IT company Cerner Corporation, and pharmaceutical companies AbbVie, Celgene.

We can also see what predictions the model makes about such a firm. Comfortingly, when we predict the industry tags from \( z_{\text{MedDevice}} \), the top five tags are Health Care, Biotechnology, Software, Information Technology, and Hardware. The model also makes a strong prediction that the company will be B2B. Moreover, a word cloud showing the most relevant terms when \( z_{\text{MedDevice}} \) is propagated through the text decoder is shown in Figure 12. For brand personality, the 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 5, and provide a simple rendering of a logo that contains many of those features in Figure 13.
<table>
<thead>
<tr>
<th>Categorical Features</th>
<th>Likely Values</th>
<th>Prob</th>
<th>Binary Features</th>
<th>Likely Values</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant color:</td>
<td>Medium Blue</td>
<td>0.547</td>
<td>Accent color:</td>
<td>Medium Blue</td>
<td>0.707</td>
</tr>
<tr>
<td></td>
<td>Dark Blue</td>
<td>0.124</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Light Blue</td>
<td>0.116</td>
<td>Font:</td>
<td>Width: Original</td>
<td>0.999</td>
</tr>
<tr>
<td>Hull shape:</td>
<td>Thin Oval</td>
<td>0.837</td>
<td></td>
<td>Style: No Italics</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>Medium Oval</td>
<td>0.145</td>
<td></td>
<td>Weight: Bold</td>
<td>0.920</td>
</tr>
<tr>
<td>Sans/Serif Font:</td>
<td>Sans</td>
<td>0.900</td>
<td>Other:</td>
<td>Has a Mark</td>
<td>0.980</td>
</tr>
<tr>
<td>Mark Class:</td>
<td>Wispy horizontal</td>
<td>0.313</td>
<td></td>
<td>Low # Regions</td>
<td>0.607</td>
</tr>
<tr>
<td></td>
<td>Detailed design</td>
<td>0.120</td>
<td></td>
<td>Low Entropy</td>
<td>0.308</td>
</tr>
<tr>
<td></td>
<td>Thin</td>
<td>0.095</td>
<td></td>
<td>Mark Position: Left</td>
<td>0.289</td>
</tr>
<tr>
<td>Number of Colors:</td>
<td>One color</td>
<td>0.859</td>
<td></td>
<td>Low % Horizontal Edges</td>
<td>0.262</td>
</tr>
<tr>
<td></td>
<td>Two colors</td>
<td>0.133</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Three colors</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Visual profile corresponding to $z_{MedDevice}$ 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.

**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 $z$-vectors for each of these concepts: for daring, we create an average $\bar{z}_{Daring}$ by averaging the $z_i$ vectors for all brands who scored in the top decile of daring. For fast food, we create $\bar{z}_{FastFood}$ by averaging together the $z_i$ 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:

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

The higher $\alpha$, the more of the daring personality will be added. The higher $\beta$, the more the resulting brand will resemble the typical fast food firm.$^9$

---

$^9$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.
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.

To illustrate this, we consider two combinations: $z_{DFF}(0.5, 0.5)$, which has the same interpolation weights as in the medical devices example, and $z_{DFF}(0.5, 1.0)$, which increases the degree of daring being added. Unlike the medical device case, where we could verify that the arithmetic had produced a reasonable result by computing the new $z$’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 $z_{DFF}$, they are simply Burger King, KFC, and Pizza Hut. Nonetheless, we can still make predictions for this previously unobserved brand identity. In both cases, the two highest industry labels associated with $z_{DFF}$ 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 $z_{Daring}$ to $z_{FastFood}$, we have morphed the fast food representation to be a bit more daring. When we increase $\beta$ to 1, we see this effect even more dramatically. We illustrate this contrast between the two values of $\beta$ and the average fast food firm in Figure 14.

The value of $\beta$ also determines to what degree the predicted visual features differ from the

---

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

11 As we illustrate in the next section, more recent entrants to the market do reflect the predicted personality: when Shake Shack’s $z_i$ is estimated using the full inference network, it falls closer to $z_{DFF}$ than it does to $z_{FastFood}$. However, Shake Shack is not in our original data.
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 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:\textsuperscript{12}

$$z_{\text{Mid}} = 0.5z_A + 0.5z_B.$$  

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

\textsuperscript{12}Just as before, the weights here need not be 0.5 for each; a more general formulation of $z_{\text{Mid}} = \alpha z_A + \beta z_B$ can also be used to adjust the emphasis of each original brand in the hybrid brand.
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. Decision Support Application 1: Rebranding of McDonald’s

In all of the previous analyses, we have used the full inference network, and manipulated the learned $z_i$ representations to aid in the brand ideation process. Now, we consider the task of using our task-specific inference networks to understand design, provide design decision support, and, in the next section, to support the branding process for new firms.
Figure 15: Three logos of McDonald’s: (1) an older logo, popular from the 1990s (and on McDonald’s roadside signs); (2) a newer version, introduced in the 2000s; and (3) an even newer version with a red background.

Rebranding is incredibly common for firms looking to update their image and keep pace with changing markets. Often, this rebranding involves a change (small or large) to the firm logo. Case in point: of the 389 firms in our data for which we were able to find information about the history of their logo, 137 (35%) of them have since changed their logo.13 One such firm is McDonald’s, whose golden arches have experienced many evolutions over time. Recently, McDonald’s has relied on a relatively simple design featuring just the golden arches, quite distinct from the version that was most common in the 1990s, as shown in Figure 15. Even more recently, McDonald’s has reintroduced a red background to the arches.14

To illustrate the utility of our model for aiding in rebranding, we explore how consumers may perceive each of McDonald’s candidate logos, using the older logo as a baseline. Specifically, we construct three hypothetical profiles for McDonalds by fixing the text and industry tags, and varying the logo design across the three designs shown in Figure 15. We then use the manager’s inference network to infer a $z_i$ for each of these three profiles, and finally use the brand personality decoder to infer how consumers may perceive each profile. The predicted brand personality ratings, relative to the old logo, are shown in Figure 16. These results suggest that consumers will perceive the red background logo as more similar to the older, more complex designs. This, however, may not necessarily be beneficial: by and large, the simple, arches-only design is expected to be perceived higher along many dimensions, including things like contemporary, good-looking, and up-to-date, which are likely target traits in a rebranding. In fact, the arches-only logo is predicted to fall short on only two dimensions: small-town and western. Intuitively, this makes sense: many modern logos feature relatively simple, single color designs, with considerably whitespace.

---

13 In fact, in the course of writing this paper, at least three brands changed their logos.
14 For an informal overview of the history of McDonald’s logo, see https://www.digitaldoughnut.com/articles/2019/september/mcdonalds-history-and-evolution-of-a-famous-logo
The arches-only logo is squarely in this mold, and thus is perceived as up-to-date and contemporary, but perhaps without the small-town charm of older logos.

7.6. Decision Support Application 2: Introduction of Shake Shack

In our final application, 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 17, 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 18.

**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 $z_i$. 

**Figure 17:** (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.
Figure 18: (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.

We then sample from that posterior to produce a distribution over Shake Shack’s predicted logo features.\(^{15}\) We summarize the predicted visual features in Table 6.

Comparing these predictions to the actual logo shown in Figure 17, 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

\(^{15}\)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.\textsuperscript{16} 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 \textsuperscript{17} and \textsuperscript{18}.

\textsuperscript{16}https://www.fastcompany.com/3041777/the-untold-story-of-shake-shacks-16-billion-branding

\begin{table}[h]
\centering
\begin{tabular}{|l|l|l|}
\hline
\textbf{Categorical Features} & \textbf{Likely Values} & \textbf{Prob} \\
\hline
Dominant color: & Black & 0.445 \\
& Dark Blue & 0.444 \\
& Red & 0.018 \\
Hull shape: & Medium Oval & 0.383 \\
& Thin Oval & 0.370 \\
& Triangle & 0.246 \\
Sans/Serif Font: & Sans & 0.999 \\
Mark Class: & Thin & 0.538 \\
& Detailed circular design & 0.307 \\
& Hollow circle & 0.083 \\
Number of Colors: & Three colors & 0.731 \\
& Many colors & 0.268 \\
\hline
\textbf{Binary Features} & \textbf{Likely Values} & \textbf{Prob} \\
\hline
Accent color: & Light Gray & 0.999 \\
& Black & 0.572 \\
Contains color: & Light Gray & 0.999 \\
& Black & 0.998 \\
Font: & Width: Original & 0.999 \\
& Style: No Italics & 0.999 \\
& Weight: Light & 0.968 \\
& Class: Geometric & 0.551 \\
Other: & High % Whitespace & 0.999 \\
& Has a Mark & 0.993 \\
& High Perimetric Complexity & 0.929 \\
& High # Regions & 0.718 \\
\hline
\end{tabular}
\caption{Visual profile corresponding to \(z_{\text{ShakeShack}}\), 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.}
\end{table}
Figure 19: 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.

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 $z_i$, 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 19, 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 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 17 and 18 are close to 0.8.
Assessing Visual Changes  Finally, we again consider the task of assessing changes to a brand’s visual identity, similar to what we did previously with the McDonald’s case study. 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


56
Table 7: Logo features with descriptions and links to past literature.

<table>
<thead>
<tr>
<th>Category</th>
<th>Feature</th>
<th>Description</th>
<th>Original Type</th>
<th>Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>Color</td>
<td>Whether a given color is present</td>
<td>Binary</td>
<td>Valsaz and Mehloobian (1994); Klink (2003);</td>
</tr>
<tr>
<td></td>
<td>Dominant Color</td>
<td>The color with the highest number of pixels</td>
<td>Categorical</td>
<td>Deng et al. (2010);</td>
</tr>
<tr>
<td></td>
<td>Accent Color</td>
<td>All colors that are not the dominant color</td>
<td>Binary</td>
<td>Semen and Palma (2014); Kareklas et al. (2014)</td>
</tr>
<tr>
<td></td>
<td>% Whitespace</td>
<td>How much of the logo (mark)'s convex hull is background (white space)?</td>
<td>Real</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Saturation</td>
<td>The mean value of the saturation channel across pixels in HSV colorspace</td>
<td>Real</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD Saturation</td>
<td>The standard deviation of the saturation channel</td>
<td>Real</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Lightness</td>
<td>The mean value of the value channel in HSV colorspace</td>
<td>Real</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD Lightness</td>
<td>The standard deviation of the value channel</td>
<td>Real</td>
<td></td>
</tr>
<tr>
<td>Format and Shape</td>
<td>Has Mark</td>
<td>Is there a mark?</td>
<td>Binary</td>
<td>Navori (1977); Klink (2003);</td>
</tr>
<tr>
<td></td>
<td>Size</td>
<td>How many marks there are</td>
<td>Count</td>
<td>Orth and Malkevitch (2009); Walsh et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>Number of Marks</td>
<td>Convex hull</td>
<td>Real</td>
<td>Spener (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The smallest convex polygon that fully contains the logo, classified into types</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standardized shape</td>
<td>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</td>
<td>Categorical</td>
<td></td>
</tr>
<tr>
<td></td>
<td># Corners</td>
<td>The number of corners found by the Harris corner detector</td>
<td>Count</td>
<td></td>
</tr>
<tr>
<td>Font</td>
<td># Characters</td>
<td>Number of logo segments classified as characters</td>
<td>Count</td>
<td>Doyle and Bottomley (2004)</td>
</tr>
<tr>
<td></td>
<td>Serif</td>
<td>Classification of characters into serif, sans-serif, or calligraphic fonts</td>
<td>See footnote 17</td>
<td>Henderson et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>Italic</td>
<td>Upright versus italic characters</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weight</td>
<td>Original, bold, or light characters</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Width</td>
<td>Original, condensed, or wide characters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td># Colors</td>
<td>How many distinct colors are there</td>
<td>Count</td>
<td>Henderson and Cote (1998); Jano-Owczarek et al. (2003); van der Lans et al. (2009); Pieters et al. (2010)</td>
</tr>
<tr>
<td></td>
<td># Segments</td>
<td>How many distinct regions are there</td>
<td>Count</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Perimetric complexity</td>
<td>Gaps between characters, computed by summing edge pixels</td>
<td>Real</td>
<td>Henderson and Cote (1998); van der Lans et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>Greyscale entropy</td>
<td>The local average variance of greyscale pixel intensity</td>
<td>Real</td>
<td></td>
</tr>
<tr>
<td>Symmetry</td>
<td>Horizontal Symmetry</td>
<td>The correlation in pixel values when the image (mark or logo) is split in half horizontally (i.e., left and right halves)</td>
<td>Real</td>
<td>Henderson and Cote (1998); van der Lans et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>Vertical Symmetry</td>
<td>The correlation in pixel values when the image (mark or logo) is split in half vertically (i.e., top and bottom halves)</td>
<td>Real</td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td>Position</td>
<td>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. The percentage of non-zero edge gradients classified as horizontal, vertical, up-diagonal, and down-diagonal, computed by traversing the binarized logo in both left-right and top-down directions and computing numerical gradients.</td>
<td>Binary</td>
<td>Chee and Hoog (2013); Cian et al. (2014); Deng and Kahn (2016); Schlosser et al. (2016)</td>
</tr>
<tr>
<td></td>
<td>Edge Gradients</td>
<td></td>
<td>Real</td>
<td></td>
</tr>
</tbody>
</table>

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 20. 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 21. Given this format, de-
terminating 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 $25 \times 25$ 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. 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.

18The number of clusters both in this step and others was determined by the researcher, using scree plots.
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 $25 \times 25$ standardized greyscale representation, and then apply ordinary k-means clustering over the resultant length 625 vectors, determining the optimal number of clusters via scree plots. The only challenge is for the marks: the standardization procedure discards information about size. Yet, we also want to capture the different sizes of marks: a mark that forms the background of, and thus takes up 80% of a logo is different than one that takes up only 10%. To take this into account, we include an additional term in the clustering of marks, that adds weight to the fraction of the the logo’s area taken up by the mark.

C. Implementation Details

We implement our model using PyTorch and the Pyro probabilistic programming language (Bingham et al., 2018). Recall that our training procedure consists of optimizing a per iteration loss function:

$$\ell_m(\theta, \{\phi_t\}) = \sum_{i=1}^{N} \sum_{\forall t} \delta_{it} \left\{ -E_{z_i \sim q_{\phi_t}(z_i|\tilde{x}_t, \tilde{f}_t)} \left[ \log p_{\theta}(x_i | z_i) \right] + \right. $$

$$\left. \text{KL} \left[ q_{\phi_t}(z_i|\tilde{x}_t) \parallel p(z_i) \right] \right\}. \quad (14)$$

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 perform stochastic gradient descent on this objective function, we use the Adam optimizer (Kingma and Ba, 2014), with learning rate 0.00001, and where the gradient is evaluated at each step of the optimization using a single observation. We run our randomized optimization routine for 5,000 iterations. In each iteration, we perform 10 optimization steps. Monitoring convergence is difficult in this setting, as we have many objectives: for each inference network, there are both reconstruction and predictive losses. We determined the number of steps by monitoring all of these out-of-sample losses on the first fold of our K-fold cross validation, stopping when all losses seemed to have reached their minima. We show pseudocode for our general procedure in Algorithm 1.

To prevent overfitting, complex models such as ours typically rely on regularization methods. For regularization, we rely on 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). Dropout has become the standard regularization method in the deep learning literature, and was implemented using the built-in PyTorch functionalities.

To determine all model hyperparameters, including the number of latent dimensions ($K$), layer sizes, and number of hidden layers, we performed grid search over an array of values, assessing model performance using the fit statistics described in the main body of the paper, computed on the first fold of our 4-fold cross validation. Through grid search, we found the optimal dimension-
Algoithm 1 Inference pseudocode: stochastic binned optimization of the ELBO

Set Adam target learning rate $\beta^*$
Initialize $\phi, \theta$ and learning rate $\beta$
Set $m = 0$
while not converged do
  $m = m + 1$
  Draw minibatch: $B_m = \text{sample(1:N_{Train}, 600, replace = FALSE)}$
  Randomize $i \in B_m$ to tasks $t = 1, \ldots, 4$
  for $s = 1, \ldots, 100$ do
    Draw single observation from $B_m$ to estimate the (stochastic) gradient $\nabla \ell_m(\theta, \{\phi_t\})$
    Update $\phi, \theta$ via single Adam step
    Update $\beta$
  end for
  Evaluate $\ell_m(\theta, \{\phi_t\})$ and compare to $\ell_{m-1}$
end while

ality of the latent space was $K = 20$. We also found no benefit to increasing the number of hidden layers of any of the networks above one (i.e., we used 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. We also found that it is very important to mirror the complexity of the inference network and the complexity of the task (i.e., the number of inputs to that task): our final model architecture consists of 400 hidden units for the full inference network, 200 hidden units for each of the manager’s and designer’s inference networks, and 50 hidden units for the consumer’s inference network. When employing larger inference networks, we found the model has a tendency to overfit overly complex domains. We use 400 hidden units in all of the decoder, and found the model relatively unsensitive to this choice.
D. Precision and Recall

<table>
<thead>
<tr>
<th>Metric</th>
<th>TSI</th>
<th>POE</th>
<th>PPCA</th>
<th>NIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>MSE</td>
<td>0.320</td>
<td>0.447</td>
<td>0.340</td>
</tr>
<tr>
<td>Logo: Binary</td>
<td>Prec.</td>
<td>0.420</td>
<td>0.397</td>
<td>0.367</td>
</tr>
<tr>
<td>Logo: Dom. Color</td>
<td>Prec.</td>
<td>0.250</td>
<td>0.191</td>
<td>0.195</td>
</tr>
<tr>
<td>Logo: Hull Shape</td>
<td>Prec.</td>
<td>0.266</td>
<td>0.201</td>
<td>0.275</td>
</tr>
<tr>
<td>Logo: Mark Shape</td>
<td>Prec.</td>
<td>0.139</td>
<td>0.103</td>
<td>0.105</td>
</tr>
<tr>
<td>Logo: Font Serifs</td>
<td>Prec.</td>
<td>0.652</td>
<td>0.379</td>
<td>0.435</td>
</tr>
<tr>
<td>Logo: Num. Colors</td>
<td>Prec.</td>
<td>0.554</td>
<td>0.543</td>
<td>0.499</td>
</tr>
<tr>
<td>Industry Tags</td>
<td>Prec.</td>
<td>0.423</td>
<td>0.474</td>
<td>0.233</td>
</tr>
<tr>
<td>Text</td>
<td>Prec.</td>
<td>0.283</td>
<td>0.186</td>
<td>0.143</td>
</tr>
</tbody>
</table>

**Table 8**: Average reconstruction cross validation error using the full inference network, where each column is a different model. Note that MSE is the mean squared error, where higher numbers indicate worse fit, while precision (Prec.) is the fraction of features the model predicted to be present that were actually present, where higher numbers indicate better fit.

<table>
<thead>
<tr>
<th>Task</th>
<th>Domain</th>
<th>Metric</th>
<th>TSI</th>
<th>POE</th>
<th>PPCA</th>
<th>Designer</th>
<th>NIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designer</td>
<td>Logo: Binary</td>
<td>Prec.</td>
<td>0.237</td>
<td>0.230</td>
<td>0.208</td>
<td>0.215</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>Logo: Dom. Color</td>
<td>Prec.</td>
<td>0.096</td>
<td>0.100</td>
<td>0.087</td>
<td>0.090</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>Logo: Hull Shape</td>
<td>Prec.</td>
<td>0.236</td>
<td>0.144</td>
<td>0.189</td>
<td>0.161</td>
<td>0.073</td>
</tr>
<tr>
<td></td>
<td>Logo: Mark Shape</td>
<td>Prec.</td>
<td>0.065</td>
<td>0.080</td>
<td>0.073</td>
<td>0.068</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>Logo: Font Serifs</td>
<td>Prec.</td>
<td>0.402</td>
<td>0.366</td>
<td>0.268</td>
<td>0.310</td>
<td>0.268</td>
</tr>
<tr>
<td></td>
<td>Logo: Num. Colors</td>
<td>Prec.</td>
<td>0.304</td>
<td>0.312</td>
<td>0.307</td>
<td>0.336</td>
<td>0.083</td>
</tr>
<tr>
<td>Manager</td>
<td>BP</td>
<td>MSE</td>
<td>0.794</td>
<td>0.774</td>
<td>0.811</td>
<td>1.008</td>
<td></td>
</tr>
<tr>
<td>Consumer</td>
<td>BP</td>
<td>MSE</td>
<td>0.834</td>
<td>0.828</td>
<td>0.847</td>
<td>1.008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Text</td>
<td>Prec.</td>
<td>0.031</td>
<td>0.041</td>
<td>0.017</td>
<td>0.004</td>
<td></td>
</tr>
</tbody>
</table>

**Table 9**: Average prediction cross validation error using the full inference network, where each column is a different model. Note that MSE is the mean squared error, where higher numbers indicate worse fit, while F1 is the harmonic mean of recall and precision, where higher numbers indicate better fit.
<table>
<thead>
<tr>
<th>Metric</th>
<th>TSI</th>
<th>POE</th>
<th>PPCA</th>
<th>NIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>0.320</td>
<td>0.447</td>
<td>0.340</td>
<td>1.008</td>
</tr>
<tr>
<td>Logo: Binary</td>
<td>0.230</td>
<td>0.177</td>
<td>0.140</td>
<td>0.060</td>
</tr>
<tr>
<td>Logo: Dom. Color</td>
<td>0.226</td>
<td>0.195</td>
<td>0.183</td>
<td>0.080</td>
</tr>
<tr>
<td>Logo: Hull Shape</td>
<td>0.225</td>
<td>0.204</td>
<td>0.193</td>
<td>0.167</td>
</tr>
<tr>
<td>Logo: Mark Shape</td>
<td>0.136</td>
<td>0.111</td>
<td>0.111</td>
<td>0.062</td>
</tr>
<tr>
<td>Logo: Font Serifs</td>
<td>0.397</td>
<td>0.343</td>
<td>0.339</td>
<td>0.333</td>
</tr>
<tr>
<td>Logo: Num. Colors</td>
<td>0.495</td>
<td>0.453</td>
<td>0.455</td>
<td>0.250</td>
</tr>
<tr>
<td>Industry Tags</td>
<td>0.242</td>
<td>0.286</td>
<td>0.094</td>
<td>0.042</td>
</tr>
<tr>
<td>Text</td>
<td>0.083</td>
<td>0.048</td>
<td>0.038</td>
<td>0.007</td>
</tr>
</tbody>
</table>

**Table 10:** Average reconstruction cross validation error using the full inference network, where each column is a different model. Note that MSE is the mean squared error, where higher numbers indicate worse fit, while F1 is the harmonic mean of recall and precision, where higher numbers indicate better fit.

<table>
<thead>
<tr>
<th>Task</th>
<th>Domain</th>
<th>Metric</th>
<th>TSI</th>
<th>POE</th>
<th>PPCA</th>
<th>Designer</th>
<th>NIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designer</td>
<td>Logo: Binary</td>
<td>Recall</td>
<td>0.112</td>
<td>0.091</td>
<td>0.079</td>
<td>0.113</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>Logo: Dom. Color</td>
<td>Recall</td>
<td>0.109</td>
<td>0.110</td>
<td>0.114</td>
<td>0.097</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>Logo: Hull Shape</td>
<td>Recall</td>
<td>0.180</td>
<td>0.178</td>
<td>0.177</td>
<td>0.177</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td>Logo: Mark Shape</td>
<td>Recall</td>
<td>0.078</td>
<td>0.074</td>
<td>0.086</td>
<td>0.074</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>Logo: Font Serifs</td>
<td>Recall</td>
<td>0.342</td>
<td>0.338</td>
<td>0.333</td>
<td>0.337</td>
<td>0.333</td>
</tr>
<tr>
<td></td>
<td>Logo: Num. Colors</td>
<td>Recall</td>
<td>0.272</td>
<td>0.259</td>
<td>0.261</td>
<td>0.263</td>
<td>0.250</td>
</tr>
<tr>
<td>Manager</td>
<td>BP</td>
<td>MSE</td>
<td>0.794</td>
<td>0.774</td>
<td>0.811</td>
<td>1.008</td>
<td></td>
</tr>
<tr>
<td>Consumer</td>
<td>BP</td>
<td>MSE</td>
<td>0.834</td>
<td>0.828</td>
<td>0.847</td>
<td>1.008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Text</td>
<td>Recall</td>
<td>0.012</td>
<td>0.014</td>
<td>0.010</td>
<td>0.007</td>
<td></td>
</tr>
</tbody>
</table>

**Table 11:** Average prediction cross validation error using the full inference network, where each column is a different model. Note that MSE is the mean squared error, where higher numbers indicate worse fit, while F1 is the harmonic mean of recall and precision, where higher numbers indicate better fit.
E. Additional Randomly Generated Brands

Here we show several more examples of random brands, generated identically to the focal brand discussed in Section 7.3. For all brands, we include a rendering of their logo that was produced without knowledge of the other domains, based just on the model’s suggested logo features.

E.1. Additional Brand 1

- High BP traits: original, imaginative, cool
- Low BP traits: corporate, upper-class, small-town
- Most likely industry tag: Internet Services (including companies that sell things online, that provide services like search or streaming entertainment, social networks, or provide internet service)

![Cloud and logo](image)

**Figure 22:** At left, the word cloud describing the most relevant terms for a random brand. At right, a simple, nonprofessional rendering of its logo, based on the model’s suggested logo features.
E.2. Additional Brand 2

- High BP traits: corporate, leader, technical
- Low BP traits: trendy, young, cool
- Most likely industry tag: Manufacturing (including manufacturers from many sectors, including auto parts, technology, materials and chemicals, hardware and equipment, and food)

![Figure 23: At left, the word cloud describing the most relevant terms for a random brand. At right, a simple, nonprofessional rendering of its logo, based on the model’s suggested logo features.](image)
E.3. Additional Brand 3

- High BP traits: sentimental, original, unique
- Low BP traits: corporate, western, rugged
- Most likely industry tag: Biotechnology (including pharmaceutical companies, medical devices and technology, and medical research companies)

Figure 24: At left, the word cloud describing the most relevant terms for a random brand. At right, a simple, nonprofessional rendering of its logo, based on the model’s suggested logo features.
E.4. Additional Brand 4

- High BP traits: feminine, trendy, cool
- Low BP traits: hard-working, secure, reliable
- Most likely industry tag: Consumer Goods (including companies that make and sell products to consumers, such as home goods, toys, and cosmetics)

Figure 25: At left, the word cloud describing the most relevant terms for a random brand. At right, a simple, nonprofessional rendering of its logo, based on the model’s suggested logo features.
E.5. Additional Brand 5

- High BP traits: cheerful, family-oriented, friendly
- Low BP traits: tough, masculine, hard-working
- Most likely industry tag: Travel and Tourism (including hotels, travel booking companies, restaurants and food service providers, airlines and cruiselines)

**Figure 26:** At left, the word cloud describing the most relevant terms for a random brand. At right, a simple, nonprofessional rendering of its logo, based on the model’s suggested logo features.