



Marketing Science

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

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

Ryan Dew, Asim Ansari, Olivier Toubia

To cite this article:

Ryan Dew, Asim Ansari, Olivier Toubia (2022) Letting Logos Speak: Leveraging Multiview Representation Learning for Data-Driven Branding and Logo Design. Marketing Science 41(2):401-425. <https://doi.org/10.1287/mksc.2021.1326>

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

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

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

Copyright © 2021, INFORMS

Please scroll down for article—it is on subsequent pages



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

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

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

Ryan Dew,^a Asim Ansari,^b Olivier Toubia^b

^aThe Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104; ^bColumbia Business School, Columbia University, New York, New York 10027

Contact: ryandew@wharton.upenn.edu,  <https://orcid.org/0000-0001-7652-0369> (RD); maa48@gsb.columbia.edu,  <https://orcid.org/0000-0001-6964-6297> (AA); ot2107@gsb.columbia.edu,  <https://orcid.org/0000-0001-7493-9641> (OT)

Received: November 25, 2019

Revised: March 21, 2021

Accepted: July 14, 2021

Published Online in Articles in Advance:
December 28, 2021

<https://doi.org/10.1287/mksc.2021.1326>

Copyright: © 2021 INFORMS

Abstract. Logos serve a fundamental role as the visual figureheads of brands. Yet, because of the difficulty of using unstructured image data, prior research on logo design has largely been limited to nonquantitative 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, consumer ratings of brand personality, and other high-level tags describing firms. We apply this framework to a unique data set of brands to understand which brands use which logo features and how consumers evaluate these brands' personalities. Moreover, we show that manipulating the model's learned representations through what we term "brand arithmetic" yields new brand identities and can help with ideation. Finally, through an application to fast-food branding, we show how our model can be used as a decision support tool for suggesting typical logo features for a brand and for predicting consumers' reactions to new brands or rebranding efforts.

History: K. Sudhir served as the senior editor for this article.

Funding: The authors thank the Wharton Behavioral Lab for funding. This work was supported by a generous grant from Analytics at Wharton. The first author is grateful to have received financial support from the INFORMS Society for Marketing Science Doctoral Dissertation Proposal Competition and the American Statistical Association's Section on Statistics in Marketing's Doctoral Dissertation Award.

Supplemental Material: The online appendix and data are available at <https://doi.org/10.1287/mksc.2021.1326>.

Keywords: logos • branding • machine learning • multiview learning • representation learning • image processing • Bayesian estimation

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 redesign their logos to convey new ideas or communicate new positionings. Virtually every company has a logo, and logo redesigns are commonplace, often following company mergers, acquisitions, and divestitures, but also occurring periodically for many brands in an effort to maintain a modern look (Henderson and Cote 1998). Yet, despite the clear significance of logos, the substantial costs of logo redesigns, and the need for tools for managers to navigate this complex process, marketing scholars have paid relatively little attention to logo design. In part, this lack of attention may stem from a perception that the design process is more art than science. Modern image processing and machine learning

methods, however, allow us to work with complex, unstructured data, such as images, thereby enabling us to study creative processes from a data-driven perspective.

In this paper, we leverage these new technologies to build a decision support system for the logo design process. Specifically, 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 (BP). 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, in which 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 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 brand's personality is? This is the perspective that branding researchers focus on in much of the prior literature, which examines how particular logo features impact consumer perceptions about a firm. Distinct from the manager's perspective, consumers may only have a vague sense of what the firm does when making these evaluations.

Underpinning the answers to each of these questions are the interrelated processes by which consumers 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 (Kardes et al. 2008, Loken 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 data set of hundreds of brands, containing logo data, textual descriptions, tags describing basic firm features, 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 consumer'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, to our knowledge, it is the first paper to study real logos from a holistic and quantitative perspective. This is important because it adds a level of objectivity to the design process: although our model cannot replace the creative touch of designers, it does offer both designers and firms decision support tools that can guide the crafting of 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. 2020), 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 makes 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 other observed firm traits. Distinct from supervised deep-learning models that have been successfully employed in a number of recent marketing studies (e.g., Liu et al. 2019, Liu et al. 2020), our MVAE is a semisupervised 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. 2018) allows us to address design from each of the distinct perspectives outlined 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 and tags describing a brand 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 describe our conceptual framework and review the literature on logo design and aesthetics in marketing. In Section 3, we describe the unique data set we compiled to calibrate our model. In Section 4, we briefly describe our logo feature-extraction algorithm, leaving a more detailed description to our online 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, providing both predictive and consumer-based validation studies. Finally, in Sections 8 and 9, we show how the framework can be used in practice, 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 process information related to brands. Much of this literature describes how specific logo features generate different consumer reactions. Other papers discuss how these reactions vary across cultures or study the mechanisms that govern consumers' reactions to various visual stimuli. In this section, we first theoretically motivate our model using the literature on consumer information processing. Then, we briefly review the findings about how consumers perceive logos that inform our logo feature-extraction algorithm, described in Section 4.

2.1. Conceptual Framework

Our application of multiview representation learning to modeling the design process is rooted in the information-processing literature. Categorization theory suggests that, when confronted with an unfamiliar logo, consumers make inferences about this new brand based on the degree to which it activates existing mental categories (Loken et al. 2008). Studies of category-based inference suggest that consumers compare the features of a target stimulus (a new brand or a brand extension) with features of a category (a parent brand) 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.

In designing logos, designers often implicitly rely on categorization theory for ideation. Specifically, designers draw on the idea that brands and logos exist in a landscape that forms the basis of consumer mental categories, understanding that consumers evaluate new designs by the concepts that those designs activate in their minds. Designers then position new logos within this landscape. For instance, designers routinely use mood or image boarding, whereby visual elements, such as the logos, fonts, and colors of existing brands in the focal category are composed on a board to stimulate thinking about the visual associations that a logo can activate (McDonagh and Storer 2004, Stigliani and Ravasi 2012, Endrissat et al. 2016, Miller 2016). These mood boards then serve as the basis for ideation for a final logo design.

Our model-based approach to design seeks to mathematically recreate this landscape through representation learning by embedding brands in a learned latent vector space, in which a firm's position in that space jointly 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 consciousness, 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 studies logos. In seminal work, Henderson and Cote (1998) propose a framework for thinking about why logos matter and how to design them well. Specifically, they investigate how logo characteristics impact recognition and affective reactions of consumers, deriving from the NHE dimensions of design (*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) test these NHE dimensions across cultures and find them to be universally good descriptors of design.

Other behavioral researchers use experimental manipulation of fictional logos to study consumer reactions to design and the psychological mechanisms that underlie such reactions. Studying shape, for instance, Klink (2003) relates the color, size, and shape of logos to brand names; Walsh et al. (2010) studies the impact of moving from an angular logo to a round one; and Jiang et al. (2015) show that the circularity or angularity of the logo affects customer perceptions of hardness or softness, which, in turn, influence attribute judgments about products. Others study the orientation of logo elements, including Cian et al. (2014), who find 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) find that upward diagonals convey greater activity than downward diagonals, leading to more positive reactions. Researchers also analyze 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, including work by Doyle and Bottomley (2006) and Hagtvædt (2011). In the latter, they show that incomplete typeface can lead to perceptions of untrustworthiness and increased innovativeness.

2.3. Aesthetics

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

In terms of colors, Deng et al. (2010) study consumers' preferences for color combinations in product design. Kareklas et al. (2014) show that people exhibit an automatic preference for white over black in product choice and advertising, similar to the implicit bias observed in other studies in psychology. Relatedly, Semin and Palma (2014) find that white is perceived as more feminine, whereas black is perceived as more masculine. Psychological work looks at the effect of color on emotional response. For example, Valdez and Mehrabian (1994) find that, of the three key color dimensions, saturation and lightness drive emotional responses along the pleasure, arousal, and dominance dimensions.

Beyond colors, other work examines fonts, including Childers and Jass (2002), who explore how semantic connotations of typeface influence consumers' ratings of products, and Henderson et al. (2004), who analyze many fonts to uncover factors that describe typeface design and link them to consumer impressions. Still other work looks at more abstract and higher level features of design. For example, Navon (1977) finds that global features are processed more readily and fully than local ones. Pieters et al. (2010) use eye tracking to study two distinct aspects of visual complexity of advertisements. Relevant to how brand constructs relate to visual elements, Orth and Malkewitz (2008) decompose package design into five distinct types and prescriptively related these to brand personalities. Still more research shows that the orientation of stimuli can influence peoples' perception of products. Meyers-Levy and Peracchio (1992) show that the camera angle of an ad featuring a product can influence judgments of the product. Chae and Hoegg (2013) find that, in cultures in which reading is done from left to right, products are viewed more favorably when positioned congruently with this spatial orientation.

Collectively, these studies on logos and aesthetics imply that NHE dimensions and objective measures, such as the color, angularity, orientation, font, and typeface of the logo, are important for a quantitative modeling approach to support logo design. 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 practice. To that end, our work also differs from the aforementioned 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 data set consisting of four components: *logos*, *textual* descriptions of firms from their websites, *brand personality* ratings from consumers reacting to both the logo and textual description, and finally a set of basic descriptors or *tags* capturing high-level differences between firms in terms of their functions and markets.

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. A detailed description of our full data can be found in Online Appendix D, Tables 9 and 10.

3.1. Logos

Firms employ a variety of logos for different purposes. Broadly speaking, a logo may comprise three key features: marks, logotype, and subtext. Marks are the nontextual 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 although some logos have both, and some include subtext. Some firms use variants of their logo for different purposes. 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 was in use.

3.2. 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 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.

3.3. Brand Personality

To understand consumer perceptions of brands, we collected brand-personality ratings from consumers, following the framework of Aaker (1997). Specifically, we used Amazon Mechanical Turk to elicit brand personality perceptions from U.S.-based consumers by showing participants both the logo and the text describing the firm. We then asked them to rate the extent to which they thought each of a set of traits describes the focal firm based on the logo and text provided. We used the original set of 42 personality traits from Aaker (1997) as well as three reverse-coded attention check traits.¹ We gathered 20 responses per brand and use the average response on each of the 42 traits as our data. In some of the subsequent visualizations, we also group the brand-personality traits according to the factor structure outlined in Aaker (1997) by taking the average of all traits assigned to a given factor.

3.4. Basic Descriptors (Tags)

Finally, we also collected a number of basic descriptors of firms to characterize high-level patterns of heterogeneity between them. Specifically, to capture high-level differences in target markets, we had a research assistant label each firm as business to business (B2B) and/or business to consumer (B2C). Then, as a simple 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, in addition to the B2B/B2C labels. Because all of these variables are binary descriptors, we refer to them simply as tags.

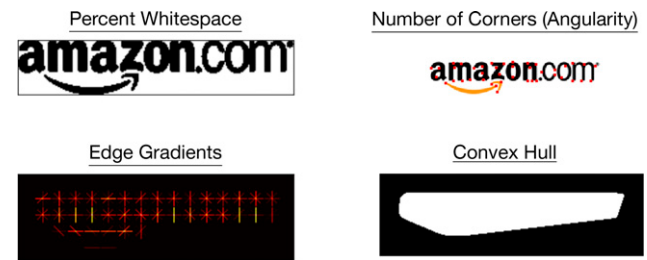
4. Logo Feature Extraction

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 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.

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 by either color or 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

Figure 1. (Color online) Examples of Global Features, Using Amazon's Logo as an Example



Notes. Percentage of 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 nonbackground pixels.

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 online 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 online appendix. Here, we briefly describe the logo features, grouping them into color, format, shape, font, and other features for expositional convenience.

4.2.1. 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

Figure 2. (Color online) Examples of the Segmentation Process, Using Amazon's Logo as an Example



Notes. The original logo is at top. Beneath that is the segmented logo, in which 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).

Figure 3. (Color online) Color Dictionary

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

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

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 nonbackground pixels) of all logo pixels. We also compute other summary statistics about color in the hue–saturation–value color space, including the mean and standard deviation of the saturation and lightness channels.

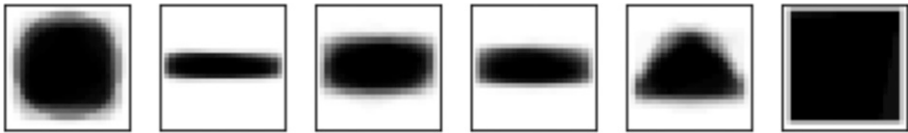
4.2.2. Format and Shape. These include features that capture the presence of a mark in the logo, the number of marks, and the aspect ratio of the logo. We also cluster the convex hulls of our logos to form a dictionary of logo shapes, shown in Figure 4. Similarly, we standardize the shape of each mark, convert it to grayscale, and then cluster all marks into 14 representative mark types. We give examples of these classes in Figure 5.

4.2.3. Font. Font is a crucial feature of logos. We, therefore, develop 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 curate 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.

4.2.4. 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 grayscale entropy (the average variance of pixel intensities across sliding windows). We also include measures of both horizontal and vertical symmetry, computed by looking at the correlation between halves of the image. For orientation, we compute both measures of position of the mark relative to the text and also edge-based metrics. Several of these features are illustrated in Figure 1, and more details are provided in the online appendix.

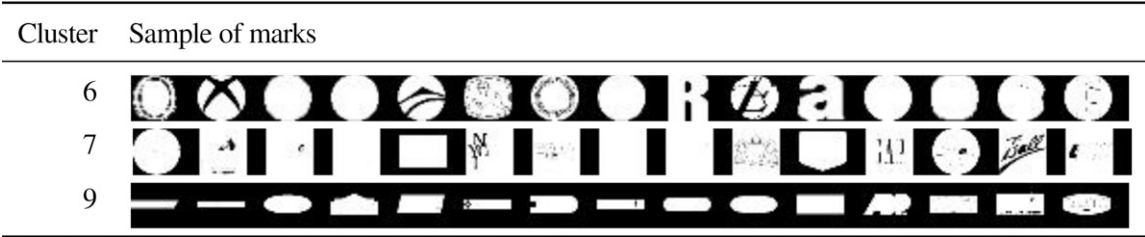
4.2.5. Discretizing Variables. Some of our logo features are real or integer-valued. We discretize each of these features into two binary variables, corresponding to whether the logo is in the bottom or top quartile of the data for that feature. This measures whether the logo is particularly low or high on a feature. For example, discretizing the *number of corners* variable gives us two binary variables: *low number of corners* and *high number of corners*. 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, two, three, or more than three colors. We

Figure 4. Hull Classes: The Six Typical Shapes of Logos as Characterized by Their Convex Hulls



Note. Each logo in our data set is assigned to one of these classes.

Figure 5. Mark Classes: Three Examples of Our Mark Classes with 10 Randomly Sampled Examples of Each



Note. Each mark is assigned to a single class.

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” or a “few” corners.

5. Exploring the Data

In this section, we present some model-free evidence to illustrate the interplay among logo features, firm function as captured by the tags, and brand-personality perceptions. These analyses motivate the full model by illustrating the complex relationship between logo design and firm identity. We use forest plots to visualize the linkages among variables in an intuitive and interactive fashion. These plots show how one focal outcome variable varies as a function of another (binary or binarized) explanatory variable. In the remainder of this section, we highlight a few of these plots. However, we also provide a web app that allows the reader to explore the full set of possible forest plots, and it can be accessed at https://dr19.shinyapps.io/explore_logo_data/.

In Figure 7, we present two forest plots that illustrate how the color of a logo relates to other features of the brand. 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 such as 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.

Apart from conveying brand image, firms may rely on logos to signal the kind of product or service that customers will receive. The second part of Figure 7 visualizes the variation in the dominant color of the logo across 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, although 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, such as manufacturing, are using black logos.

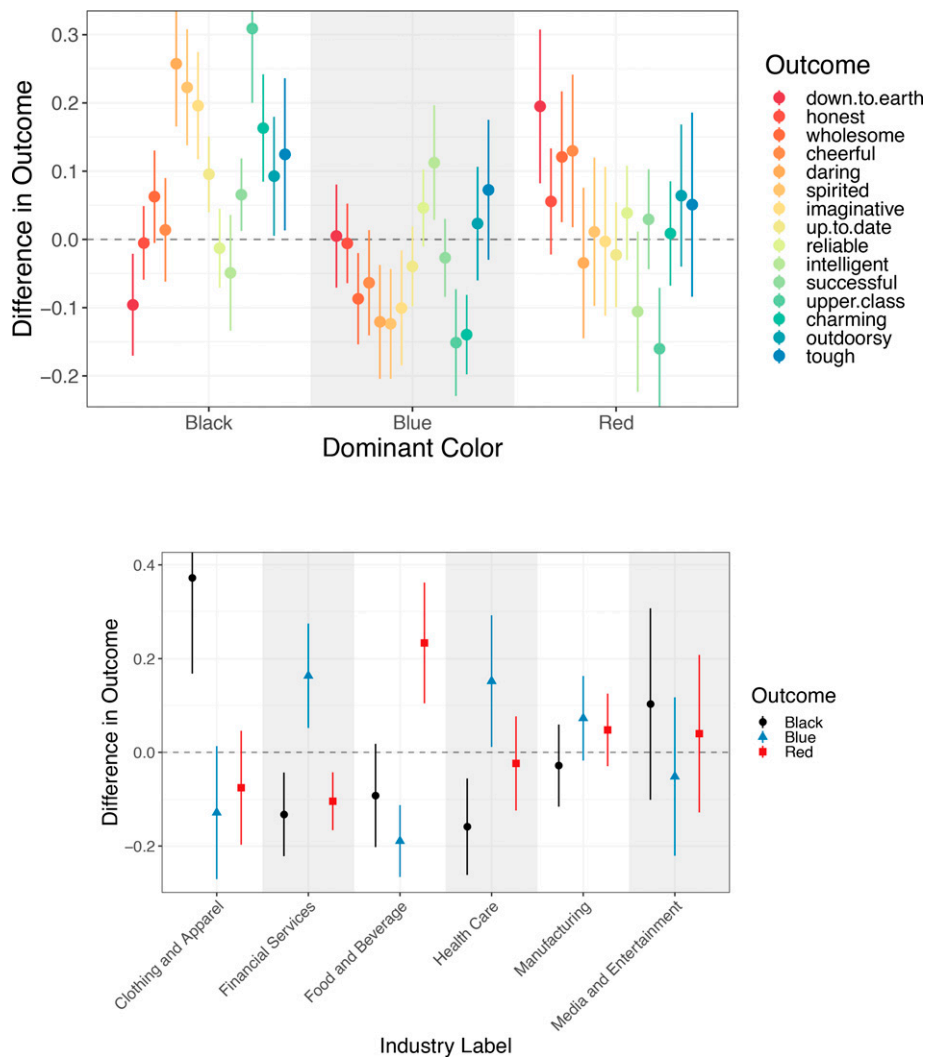
Beyond simple features such as color, our feature-extraction algorithm also isolates more complex features of logos inspired by past research. Although most prior work has not used brand personality as the dependent measure, we can nonetheless find support for many findings from the literature by examining the links between logo features and personality perceptions, including:

Figure 6. Font Classification System Employed by the Algorithm

Serif font classes: Clarendon (Clarendon) Didone (Bodoni) Oldstyle (Bembo) Slab (Rockwell) Transitional (Times)	Font weight: Original Light Bold
Sans-serif font classes: Geometric (Futura) Square (Eurostile) Grotesque (Helvetica) Humanist (Gill Sans)	Font style: Upright Italics
Calligraphic font classes: Casual (Nadianne) GLYPHIC (COPPERPLATE)	Font width: Normal Condensed Wide

Note. Fonts were matched to a font class, weight, style, and width.

Figure 7. (Color online) Forest Plots Describing Relationships Between Brand and Logo Traits



Notes. At top: Each color in the plot represents a different brand-personality factor denoted in the legend. On the x-axis are different dominant logo colors. On the y-axis is the difference in brand personality perception for firms that have that color versus firms that do not have that color. At bottom: A similar plot but showing instead how a logo is more or less likely to have a certain dominant color based on its brand's industry tag. In both plots, error bars around the points represent two standard errors.

- Horizontally symmetric logos tend to be perceived higher along almost all brand-personality dimensions except intelligent, perhaps reflecting the role of harmony in positive affect discussed in Henderson and Cote (1998).

- 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 most brand-personality traits.

- 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 find that upward diagonals convey activity.

- Placing the mark toward the right is associated with lower perceptions of down-to-earthness, honesty, and

wholesomeness but marginally higher intelligence. Although 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 (2009).

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

- A circular hull is positively associated with cheerful, daring, and 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 show that our features capture many of the aspects discussed in the

literature.³ Finally, although we only highlighted a few relationships here, we also provide a web app that allows the reader to explore the full set of possible relationships in our data using forest plots, and it can be accessed at https://dr19.shinyapps.io/explore_logo_data/.

These visual analyses are interesting but limited: they examine relationships between features in isolation but cannot be used to reason about the *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? To answer questions regarding combinations of features and to facilitate the use of unstructured, textual data that may more accurately reflect the nuances of a company, we need a model that leverages these types of data to simultaneously capture all aspects of brand identity.

6. Modeling Framework

We now describe our model for logo design. We draw on recent advances in deep generative modeling (Kingma and Welling 2013, Kingma et al. 2014, Ranganath et al. 2014, Rezende et al. 2014) and multiview learning (Suzuki et al. 2016, Li et al. 2018, Wu and Goodman 2018) to learn multimodal representations of brands in a joint latent space that is shared across our different data modalities.⁴ Specifically, we flexibly capture the linkages among the textual website descriptions, logo features, tags capturing heterogeneity between firms, and brand-personality metrics in a semisupervised 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., designer, brand manager, and consumer) without the need to specify one domain as the dependent variable and the others as independent variables.

6.1. Variational Autoencoders

We begin by briefly describing a simple 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.

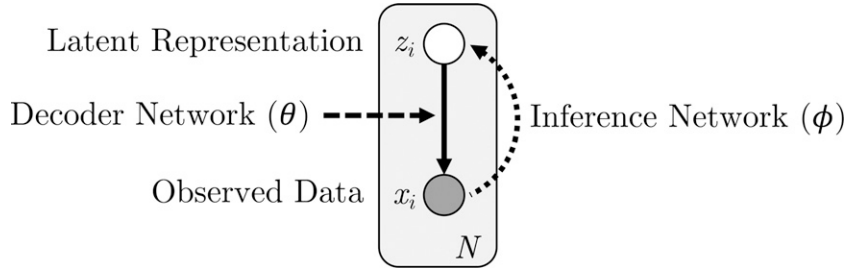
The generative model represents the probability distribution of the observed data, x_i for observation i , in terms of a multidimensional latent variable z_i . The mapping between z_i and the parameters of the probability distribution is specified using a multilayered neural network, called the *decoder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

To approximate the posterior of the latent variables, $p_\theta(z_i | x_i)$, VAEs rely on an amortized variational distribution $q_\phi(z_i | x_i)$, which 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, ϕ , are amortized (i.e., shared) across all observations, allowing for scalable variational 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 8.

6.2. Multimodal VAE

As we have data from multiple domains, we use a multimodal VAE, or MVAE, to learn a latent representation that is shared across domains (Suzuki et al. 2016, Wu and Goodman 2018). We have data on $i = 1, \dots, N$, brands across the four domains, indexed by $d \in \{\text{Text}, \text{Logo}, \text{Tags}, \text{BP}\}$. The observed data for brand i in domain d is written as x_i^d , and the complete observation is given by $x_i = \{x_i^{\text{Text}}, x_i^{\text{Logo}}, x_i^{\text{Tags}}, x_i^{\text{BP}}\}$. 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 $j = 1, \dots, V_d$ such that $x_i^d = \{x_{i1}^d, \dots, x_{iV_d}^d\}$. The generative model specifies the probability distribution of the observed data in each domain in terms of a shared latent variable vector 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 9 visually illustrates the modeling and inferential framework. Although we observe data for all domains for each brand in our data, the framework allows for missing domains. We

Figure 8. Graphical Model for a Standard VAE



Notes. Given x_i , the inference network with parameters ϕ specifies the approximate posterior for z_i . The decoder network with parameters θ transforms the latent representation z_i into the parameters of the likelihood for x_i .

now focus on the generative model for the domains before turning our attention to inference.

6.2.1. 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 $p(z_i) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. As in the standard VAE, the joint distribution of the data and the latent variables is given as $p_\theta(x_i, z_i) = p_\theta(x_i | z_i)p(z_i)$. However, the probability models for the different domains are independent, conditional on z_i , that is, $p_\theta(x_i | z_i) = \prod_d p_{\theta_d}(x_i^d | z_i)$. In turn, the probability model for each domain is specified using independent feature-level probability distributions such that $p_{\theta_d}(x_i^d | z_i) = \prod_j p_{\theta_d}^j(x_{ij}^d | z_i)$. Let μ_i^d contain the parameters for the different feature-level distributions associated with observation i within domain d . A domain-specific decoder network,

$\text{DNet}_d(z_i; \theta_d)$, captures the nonlinear relationship between μ_i^d and z_i such that $\mu_i^d = \text{DNet}_d(z_i; \theta_d)$. We first describe the different feature-level probability distributions and follow with a description of the domain-specific decoder networks.

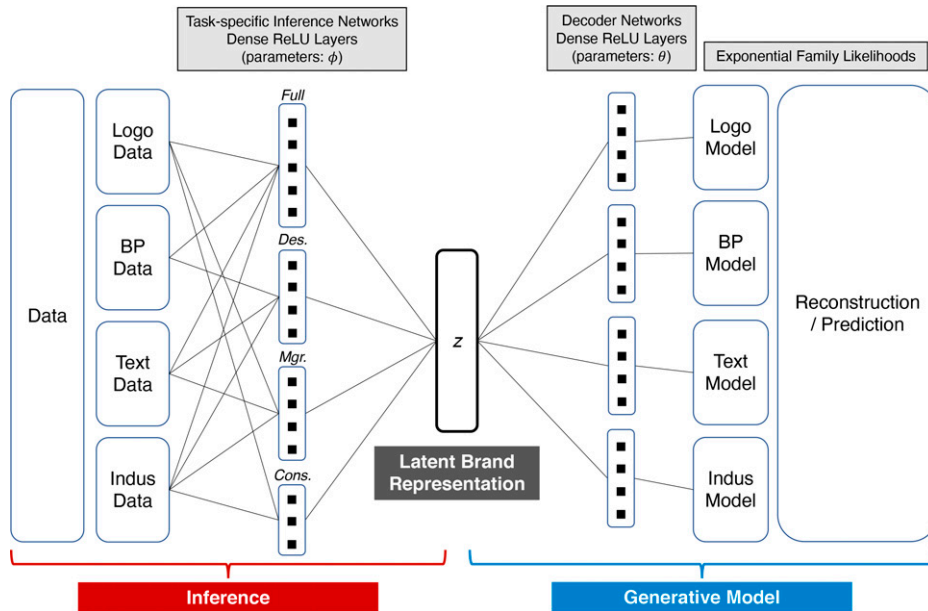
6.2.2. Feature-Level Distributions. Conditional on the joint representation z_i , each brand's features are modeled using independent domain- and feature-specific exponential-family distributions as follows:

- Text. A Bernoulli distribution captures whether a given word j is present in brand i 's textual description with probability⁵

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

- Logo features. The logo features are either binary or categorical. For binary features, such as whether the

Figure 9. (Color online) An Illustration of Our MVAE Framework



logo has a mark, we use a Bernoulli distribution. For categorical features consisting of $m = 1, \dots, M_j$ possible options, such as the dominant color, we use a categorical distribution, such that

$$x_{ij}^{\text{Logo}} \sim \text{Categorical}(\text{softmax}(\mu_{ij}^{\text{Logo}})), \quad (2)$$

$$\mu_{ij}^{\text{Logo}} = (\mu_{ij1}^{\text{Logo}}, \dots, \mu_{ijM_j}^{\text{Logo}}). \quad (3)$$

The probability vector of the categorical distribution is given by

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

- Brand personality. Brand personality is real-valued as it is the average of all respondent ratings, measured between 0 and 4. We, therefore, model it using a normal distribution, such that⁶

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

- Tags. The tags, including B2B/B2C and industry labels, are binary and follow a Bernoulli distribution.

In these feature-level distributions, the observation-specific distributional parameters (e.g., the mean μ_{ij1}^{BP} and the variance σ_{ij}^{BP} of the normal in Equation (4)) are specified nonlinearly in terms of the latent variable z_i for that observation using modality-specific decoder networks.

6.2.3. Decoder Network. We use a domain-specific feed-forward network $\mu_i^d = \text{DNet}_d(z_i; \theta_d)$ for the decoder. The network has L_d hidden layers composed of rectified linear activation units (ReLU) that apply the transformation $\text{ReLU}(x) = \max(0, x)$ to their input. In addition, the network allows for skip connections (Dieng et al. 2019) that connect the latent vector z_i directly to each layer. The skip connections help avoid latent variable collapse because of which models such as ours get stuck in uninformative local optima. Each layer ℓ computes a transformed representation of the brand through a set of $H_{\ell d}$ hidden units, whose activations are contained in the size $(H_{\ell d}, 1)$ vector $\mathbf{h}_{i\ell}^{\text{Dec},d}$. 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 hidden layers are connected sequentially to each other, resulting in the following sequence of computations:

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

and the \mathbf{a}_{ℓ}^d vectors contain the biases (intercepts) for

the hidden units in layer ℓ . The last layer, also known as the output layer, computes the parameters μ_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 θ_d to refer to all of the decoder network parameters within domain d across all the features j . Although the exact nature of the decoder network differs across domains, the preceding conveys the general structure. We describe the specifics of each domain's network architecture in a later section.

6.2.4. 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:

$$p_{\phi}(z_i | x_i) \approx q_{\phi}(z_i | \xi_i) = \mathcal{N}(\xi_i^m, \text{diag}(\xi_i^v)), \quad (6)$$

where, as in the standard VAE, an inference network computes the mean and variance terms of this normal distribution, $\xi_i = \{\xi_i^m, \xi_i^v\}$, from data x_i . The inference network $\xi_i = \text{INet}(x_i; \phi)$ is again a feed-forward neural network composed of L hidden layers, each composed of H_{ℓ} ReLU hidden units. Skip connections are not needed in this network. The neural networks associated with the decoder and inference networks are independent and do not share any parameters. The inference procedure consists of optimizing the decoder and inference network parameters θ and ϕ 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 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 in which we observe everything except the logo; (3) the manager's inference network, corresponding to the case in which we observe everything

except consumer's perceptions of brand personality; and (4) the consumer's inference network, corresponding to the case in which we observe the logo and the tags. That is, we learn four distinct inference networks, which we index by $t \in \{\text{Full}, \text{Des}, \text{Mgr}, \text{Con}\}$, where t stands for task, corresponding to four separate functions,

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

where \tilde{x}_i^t is shorthand for the data available for inference task t (for example, for $t = \text{Con}$, $\tilde{x}_i = \{x_i^{\text{Logo}}, x_i^{\text{Tags}}\}$). Intuitively, this function corresponds to the model's "best guess" at the posterior given data from the available domains for the particular task. We summarize these tasks in Table 1. Regardless of which inference network is used, the decoder network and probability models are shared across tasks. Hence, each inference network is forced to learn a coherent, unified representation, irrespective of the missing modalities. Finally, although we have assumed a set of tasks corresponding to our data setting, this structure can be easily adapted to include other tasks of interest.

6.2.5. Inference. Inference with this multimodal setup involves variational expectation maximization, adapted to allow for our multiple decoder and inference networks. This involves optimizing the parameters θ and ϕ of the decoder and inference networks, such that first encoding and then decoding data x_i leads to a prediction that is as close as possible to the original data.

The classic VAE, with one decoder and one inference network, minimizes the loss function:

$$\ell(\theta, \phi) = \sum_{i=1}^N -E_{z \sim q_\phi(z_i | \xi_i = \text{INet}(x_i; \phi))} [\log p_\theta(x_i | z_i)] + \text{KL}(q_\phi(z_i | x_i) \parallel p(z_i)), \quad (7)$$

where $\text{KL}(\cdot \parallel \cdot)$ is the Kulback–Leibler divergence between distributions. This loss is the negative of the standard evidence lower bound for doing variational inference on the latent parameters, z_i , but in which 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, and the second term regularizes the z_i estimates toward the prior.

In our multiview inference framework, $p_\theta(x_i | z_i)$ from Equation (7) decomposes into a product of the domain-specific decoder networks and feature-specific probability distributions. To this, we add 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 (7), this means that, in our optimization, at each iteration, the $q_\phi(z_i | \xi_i = \text{INet}(x_i; \phi))$ used for observation i depends on the bin to which brand i is assigned in that iteration. Let m index the iteration of the optimization, and $\delta_{itm} = 1$ if brand i is assigned to bin t on iteration m and zero otherwise. The resulting *per iteration* loss function is given by

$$\ell_m(\theta, \{\phi_t\}) = \sum_{i=1}^N \sum_{t=1}^4 \delta_{itm} \left\{ -E_{z_i \sim q_{\phi_t}} [\log p_\theta(x_i | z_i)] + \text{KL}[q_{\phi_t}(z_i | \tilde{x}_i^t) \parallel p(z_i)] \right\}. \quad (8)$$

where $q_{\phi_t} = q_{\phi_t}(z_i | \xi_{i,t} = \text{INet}_t(\tilde{x}_i^t; \phi_t))$.

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).

6.3. Implementation

We implemented our model using PyTorch and the Pyro probabilistic programming language (Bingham et al. 2019). We optimized the loss in Equation (8) using stochastic gradients and the Adam algorithm (Kingma and Ba 2014). To prevent overfitting, we utilized dropout for regularization (Goodfellow et al. 2016). We used cross-validation to determine all model hyperparameters, including the number of latent dimensions (K), number of hidden layers for each network (L), and number of hidden units per layer per network (H). This procedure suggested 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 most likely because our inputs are already high-level features, which, therefore, limits the usefulness of the increasing levels of abstraction enabled by adding more layers. We also found that it is important to mirror the complexity of the inference network to the complexity of the task (i.e., the number of inputs to that task): our final model architectures use a single hidden layer consisting of 400 hidden units for the full inference network, 200 hidden units for each of the manager's and designer's inference networks, and

Table 1. A Summary of the Tasks, Including What Each Uses as Inputs (i.e., What Data Are Provided) and What Each Is Intended to Predict (i.e., What Data Are Missing)

Task	Inputs	Predictions
Full	Text, logo, BP, tags	None
Designer	Text, BP, tags	Logo
Manager	Text, logo, tags	BP
Consumer	Logo, tags	Text, BP

Note. Note that for all of the tasks, we are also able to *reconstruct* the given inputs.

50 hidden units for the consumer's inference network. We use 400 hidden units in all of the decoders and find the model relatively insensitive to this choice. We include more details on implementation, including pseudocode, in the online appendix.

6.4. Relation to Prior Literature

Our framework is related to three other frameworks that have been proposed to extend VAEs to multimodal settings. First, our use of task-specific inference networks is a generalization of the per-domain inference network idea introduced by Suzuki et al. (2016). Although they focus only on two domains, ours covers the many modalities case, in which the modalities can be grouped into relevant tasks. More recently, Wu and Goodman (2018) introduce a product-of-experts formulation to handle more than two modalities. They also use a subsampled training procedure that is 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, Nazabal et al. (2020) develop 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

In this section, we present our results. We begin with model comparisons. We then describe the learned latent space and test how the learned representations correspond with consumer perceptions using two online studies. We then discuss implications for ideation and decision support in Sections 8 and 9.

7.1. Fit and Benchmarks

To assess model fit, we ran fourfold cross-validation. We summarize the out-of-sample fit of the model, averaged across folds and broken down by domain, in Tables 2 and 3. 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 brands, and (2) *predictive fit*, which shows the model's ability to predict missing domains for new brands, using the task-specific inference networks.⁷ Table 2 gives fit statistics for reconstruction, and Table 3 gives fit statistics for prediction. In computing both types of fit, the decoder networks remain the same, but the data given to the model and, thus, the inference network used for inferring z , change. Although both are out-of-sample

Table 2. Average Reconstruction Cross-Validation Error Using the Full Inference Network

Domain	Metric	TSI	POE	PPCA	NIR
BP	MSE	0.320	0.447	0.340	1.008
Logo: Binary	F1	0.272	0.218	0.172	0.051
Logo: Dominant color	F1	0.216	0.176	0.163	0.034
Logo: Hull shape	F1	0.219	0.182	0.170	0.102
Logo: Mark shape	F1	0.123	0.095	0.091	0.015
Logo: Font serifs	F1	0.405	0.320	0.308	0.297
Logo: Number of colors	F1	0.492	0.444	0.440	0.124
Tags	F1	0.282	0.329	0.116	0.031
Text	F1	0.114	0.065	0.051	0.005

Notes. Note that MSE is the mean squared error, for which higher numbers indicate *worse* fit, and F1 is the harmonic mean of recall and precision for which higher numbers indicate *better* fit. Each column is a different model and each row is a domain with a corresponding set of features being reconstructed.

statistics, 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. In contrast, 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 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 values indicate better fit. For the binary and categorical features, we use the F1 score, which is the harmonic mean of two measures of the success of a classifier, precision, and recall. 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. Intuitively, the F1 score is high for a model that is correctly able to distinguish positive cases from negative cases. We use these metrics as opposed to naive measures such as accuracy because of the highly imbalanced nature of many of our features. In Tables 2 and 3, we report the average of these statistics across features. We report the precision and recall statistics as well as holdout likelihood in Online Appendix F.

We compare our model, denoted as *task-specific inference* (TSI), to several benchmarks:

- **Product of experts (POE).** This uses the *product-of-experts* framework developed by Wu and Goodman (2018). Instead of our task-specific inference networks; here, each domain has its own latent representation z_d , and these representations are combined using a product-of-normals rule.

Table 3. Average Prediction Cross-Validation Error

Task	Domain	Metric	TSI	POE	PPCA	Designer	NIR
Designer	Logo: Binary	F1	0.132	0.106	0.089	0.131	0.051
	Logo: Dominant color	F1	0.096	0.096	0.095	0.086	0.034
	Logo: Hull shape	F1	0.160	0.149	0.146	0.154	0.102
	Logo: Mark shape	F1	0.064	0.064	0.065	0.059	0.015
	Logo: Font serifs	F1	0.319	0.311	0.297	0.306	0.297
	Logo: Number of colors	F1	0.265	0.244	0.245	0.258	0.124
Manager	BP	MSE	0.794	0.774	0.811		1.008
Consumer	BP	MSE	0.834	0.828	0.847		1.008
	Text	F1	0.014	0.017	0.011		0.005

Notes. Note that MSE is the mean squared error, for which higher numbers indicate *worse* fit, and F1 is the harmonic mean of recall and precision for which higher numbers indicate *better* fit. Each column is a different model and each row is a domain in a task.

• Probabilistic principal component analysis (PPCA). This is a version of our model with no nonlinearities in the generative model, which is equivalent to doing PPCA with task-specific (amortized) inference.

• Designer. This is 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.

• No information rate (NIR). This is the naive model in which each feature is predicted to have its mean value across all of the brands in the training set.

Each of these benchmarks (except NIR) is estimated analogously to the focal model with amortized inference and the same structure for the decoder networks.

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 basic 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), outperforming 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 with the Designer benchmark, in which 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 predictive validity of the model, we now turn to understanding the learned brand representations. In general, it is difficult to interpret the specific dimensions of our learned latent space as all of the features of the data are compressed to a 20-dimensional vector. Hence, each dimension of z simultaneously encodes different aspects of the data, and likewise, specific features tend to be encoded in a distributed way across the dimensions of z . Even though the z -space cannot be directly interpreted, distances within it are meaningful: if two brands are close 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.

Table 4 shows the two nearest neighbor brands in z -space for a set of four focal firms along with the distance of each neighbor to 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: they both are innovative social network platforms and both have simple, blue, bulky logos. Similarly, Gucci’s nearest neighbors are Dior and Cartier. Both operate in the luxury retail space, have similar sophisticated brand personalities to Gucci’s and black-and-white, sleek, high-whitespace logos.

It is not always possible to find a neighboring brand that matches a focal brand on all four domains. Consider Lowe’s, shown in the second row of the table.

Table 4. (Color online) Nearest Neighbors in z-Space

Focal Brand	Neighbors in z-space		Focal Brand	Neighbors in z-space	
					
Facebook	Twitter (3.00)	Uber (3.60)	Gucci	Dior (3.48)	Cartier (3.71)
					
Lowe's	TravelCenters (3.93)	Union Pacific (4.05)	McDonald's	Heinz (4.84)	Wells Fargo (4.87)

Note. 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.

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 on consumer perceptions and logo aesthetics. All three firms have logos with the same distinct medium blue color, a lack of white space, a bulky design, and even a similar sans-serif font. Because the learned representations capture all domains simultaneously, Lowe's is placed closer to these two brands as opposed to other aesthetically distinct competitors. We can tell a similar story for McDonald's. Based just on what McDonald's does, one might expect to find brands such as Burger King or Wendy's as its nearest neighbors. Instead, we find Heinz and Wells Fargo. The reason for the discrepancy becomes clear as we look across all aspects of the brands. McDonald's, Heinz, and Wells Fargo are all classic American brands. Heinz, like McDonald's, operates in food service. Moreover, all three brands have correlated brand-personality ratings, scoring relatively high on traits such as family-oriented, western, and small town. Most obviously, all three have very similar logo designs. Finally, we see the two examples in the first row have much lower distances to their neighbors than those in the second row, which emphasizes that firms are close together when they match on *all* of the dimensions.

7.3. Domain Importance Through Scaling

Given that our learned representations are derived from four distinct domains—text, logo, BP, and tags—a natural question is, to what degree do each of these domains contribute to z ? To our knowledge, there is

no easily derived decomposition of z into the variance explained by the four domains. Rather, the task-specific inference networks combine information from all of the modalities in a nonlinear fashion to produce the final z , making it very difficult to backout the contribution of each domain. Thus, in order to understand the contribution of each domain to the final representations, we develop a procedure that we call domain *scaling*. The intuition is simple: to understand how important domain d is to the final representation z , we reestimate the model but multiplicatively scale all of the likelihood terms for domain d by a small number ϵ (e.g., $\epsilon = 1e-8$). By doing so, we effectively remove the contribution of domain d from the model.⁸

Unfortunately, the z -vectors learned across the different scalings are impossible to compare. The dimensions of z are not uniquely identified, and hence, each run of the model may return different z 's. Therefore, rather than using z directly to compare scaling, we instead use the brand distances in the learned z -space, which do not depend on comparing the dimensions of z and, thus, are consistent and interpretable across different runs. We use these distances to compare the different scalings through two metrics:

1. Rank correlation. For each scaling and for each brand, we compute the distance between that brand and all the other brands in z -space. Then, we compute the Spearman rank correlation of these distance vectors across different scalings. In particular, we focus on the rank correlation between the distances learned in the scaled version and the distances learned in the full, unscaled model. If the correlation is high, it indicates that the scaled version is learning similar relationships between brands as the full model, which suggests that the scaled domain does not contribute much to the relationships learned by the full model.

Table 5. Results of the Scaling Analysis

Metric	Scaled text	Scaled BP	Scaled logo	Scaled tags
Rank correlation	0.69	0.72	0.85	0.91
Shared top 10 neighbors	2.72	3.69	4.67	5.85

Note. For both metrics, lower scores indicate that domain is contributing more to the learned representation in the full model.

2. Top 10 neighbors. For each scaling for each brand, we compute the 10 closest brands to that brand. We then compare, for each brand, how many of its top 10 neighbors are shared across different scalings. Again, our primary focus is on comparing each scaled version with the full version. This forms a metric with a maximum of 10 with higher values indicating more agreement between the scaled version and the full version. Similar to the rank correlation, higher values suggest that the scaled domain is not contributing.

The average of these two measures across brands for each of the different scalings are shown in Table 5. From the table, we see that textual data appears to contribute the most: when its contribution is scaled down, the representations change most dramatically as indicated by the relatively low rank correlation score (0.69) and the low number of average shared neighbors (2.72). Brand personality contributes the second most to the representation and logos third. Removing the descriptive tags does not seem to alter the learned representations as significantly. This is likely because the textual data are already a very rich source, characterizing observed heterogeneity between firms. That textual data contributes the most may also explain the relatively poor performance of the consumer's task in Section 7.1 as it is the only task that does not leverage this detailed information.

7.4. Validation Studies

We now describe two studies we ran to test whether our model's outputs are consistent with consumer perceptions.⁹

7.4.1. Study 1: Intrusion Test. One way of validating whether the model has learned meaningful representations is by generating random brands from the model and assessing their coherence. Our MVAE is, at its heart, a generative model, and new brands can be randomly generated simply by drawing a new \mathbf{z}_i vector from the prior, $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and propagating that vector down the decoder networks. If the model has learned a meaningful latent space, then brand identities generated in this fashion should be coherent. We give examples of randomly generated brands as well as more details of this process in Online Appendix G.

To test whether consumers find brands generated in this way coherent, we designed an experiment in which we first generated 24 random brands from the

model. For each of the brands, we designed a logo based on the model's suggested logo template by incorporating as many high-probability features as possible and with no knowledge of the other domains. We also generated summaries of the other three domains: a word cloud, corresponding to the most likely words from the text part of the model; the highest probability industry tag, excluding very common tags such as B2B and B2C; and the three highest and lowest brand-personality traits.

To assess whether consumers view these brands as coherent, we showed random subsets of 12 of these randomly generated brand profiles to a set of 226 participants from an online, university laboratory panel. To test coherence, for each brand profile shown, we randomly held out one of the domains and used an *intrusion task*. As a test, we gave participants two options to "fill in" the missing domain: either the truth or the same domain from one of the different randomly generated brands (the *intruder*). We selected the intruder randomly from the three random brands that were farthest away from the focal brand out of all the random brands not included in that participant's study.¹⁰ The results support our claim that the model generates coherent brands: out of the 12 given profiles, the average proportion of "correct" choices (i.e., where the intruder was *not* chosen) per respondent was 0.65 with a 95% credible interval of (0.63, 0.67) (obtained from a simple binomial model). Moreover, the distribution of correct choices was left-skewed with many respondents selecting correctly nearly 100% of the time (see Online Appendix A, Figure 14, for the distribution). Taken together, these results provide strong evidence that consumers are able to identify the matching domain at a rate significantly better than chance, indicating that the model is indeed generating coherent brands.

7.4.2. Study 2: z-Space Distances. A second aspect of model validity is whether the model has learned meaningful similarities and differences between brands. Under the MVAE, similarities between brands can be measured by distance in z-space. Thus, the goal of this study was to establish that distances in z-space are, in fact, meaningful for consumers. To establish that, we first randomly selected 50 of the real brands in our data. For each of those brands, we then randomly selected one brand from among the lowest quartile of distances from that brand in z-space (i.e., close brands) and another

brand from among the highest quartile of distances (i.e., far-away brands). For each brand, we compiled a profile, using that brand's real data, consisting again of the firm's logo, a word cloud of the text from the firm's website, all of the firm's industry tags (excluding B2B/B2C), and finally the top and bottom three brand-personality traits. We then showed a sample of 221 participants from a university-affiliated online panel a randomly selected set of 12 of these 50 focal brands together with the near and far brands for each and asked them to select which brand was closer to the focal brand.

The results again confirm that the model results are meaningful: the average proportion of correct responses per respondent was 0.77 with a 95% credible interval of (0.75, 0.78) (again coming from a simple binomial model). Moreover, the results are again left-skewed, indicating most people agreed with the model for the vast majority of the cases (again, see Online Appendix A, Figure 14, for the distribution). This provides strong evidence that respondents agree with our model's judgments of brand similarity.

8. 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 that are in the focal industry or that have similar identities to the new brand.¹¹ Elements of these brands' logos may then be mixed with visual features unique to the new brand. For instance, a designer for a new medical device company may start by looking at what logo design patterns are popular in healthcare 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.

8.1. Example: Medical Devices

We first consider the task of designing for a medical device company. As described, medical devices can be considered a fusion of technology and healthcare. In our data, we have an industry tag corresponding to healthcare 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 healthcare and

technology, we first need to define two average vectors

$$z_S = \frac{1}{N_S} \sum_{i \in S} z_i, \quad \bar{z}_S = \frac{L}{\|z_S\|} z_S, \quad (9)$$

where S refers to the set of brands belonging to some prespecified group of interest, N_S is the number of brands in that set, $\|\cdot\|$ denotes the Euclidean norm, and L is the average Euclidean norm of all of the learned vectors z_i . Intuitively, this average is just the average of all 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 canceled out with the relatively smaller components of others. Hence, to ensure comparability across all vectors in the space, we employ this *norm-preserving average*.

Returning to our example, then, we define two norm-preserving averages, \bar{z}_{Health} , which is the average of all brands tagged as healthcare companies, and \bar{z}_{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:

$$z_{\text{MedDevice}} = 0.5\bar{z}_{\text{Health}} + 0.5\bar{z}_{\text{Tech}}. \quad (10)$$

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 10 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 and 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 healthcare, biotechnology, software, information technology, and hardware. The model also makes a strong prediction that the company is B2B. When $z_{\text{MedDevice}}$ is propagated through the text decoder, the highest probability words include technology, patients, solutions, and innovation (a word cloud showing the most relevant terms is shown in Online Appendix A, Figure 15). For brand personality, the highest relative traits are technical, intelligent, and contemporary, and the lowest are outdoorsy, rugged, and masculine. Finally, we summarize the logo features we expect for this company in Online Appendix A (Table 7) and provide a simple rendering of a logo created by the authors using the suggested logo features in Figure 10.

Figure 10. (Color online) Rendering of a Logo Based on $z_{\text{MedDevice}}$



Note. A simple, nonprofessional rendering of a logo based on the visual profile in Online Appendix A, Table 7.

8.2. Making Fast Food More Daring

Brand arithmetic can also be used with personality traits. Consider the task of designing a daring fast-food (DFF) company. In general, fast-food brands are not perceived as particularly daring: in our data, on a scale from 0 to 4, the average consumer rating of McDonald's for "daring" was 1.0 and for Burger King 1.05, and 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 that scored in the top decile of daring. For fast food, we create $\bar{z}_{\text{FastFood}}$ by averaging together the z_i vectors of McDonald's, Burger King, and KFC. To create a new brand identity, DFF, we can simply interpolate between \bar{z}_{Daring} and $\bar{z}_{\text{FastFood}}$: $z_{\text{DFF}} = 0.5\bar{z}_{\text{Daring}} + 0.5\bar{z}_{\text{FastFood}}$.

Unlike the medical device case, in which 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 natural "daring fast food" brand to correspond to either of these new profiles. When we compute the nearest neighbors to z_{DFF} , they are Fanta, Dominos, and Yum Brands. 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 $\bar{z}_{\text{FastFood}}$. However, Shake Shack is not in our original data.

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, the highest three traits are cool, trendy, and spirited. Though not daring, these traits are correlated with daringness but may be more likely to occur in a fast-food context, and daring appears in the top 10.¹² More importantly, we can also use the model to predict what a daring fast-food firm would look like. Although red is still the highest probability dominant color (probability = 0.365), black is much more likely than for a normal fast-food firm with an increase in probability of 0.129, the highest of any color. Circular and square logos become much less likely with narrow and wide rectangular logos

gaining probability. Having just a single color becomes much more likely, and having more than three colors becomes much less likely. This list reflects just a few of the predicted changes but gives a sense of the utility of brand arithmetic: by interpolating in this way, we can begin to understand what changes can be made to a fast-food logo to make it look more daring.

8.3. Creating 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: $z_{\text{Mid}} = 0.5z_A + 0.5z_B$. We then consider which of our existing brands are closest to this midpoint. In many cases, the closest brands to z_{Mid} are simply the original two brands or their closest neighbors. However, by looking at which brands are close to z_{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 closest midpoint brands several very interesting case studies. The nearest midpoint brand is Zara, a European fast fashion brand, which is more upmarket than Old Navy but not quite as luxurious as the third and fourth closest brands, Ralph Lauren and Coach. We also find some interesting car companies close to the midpoint: Audi, likely because of its proximity to Mercedes-Benz, and also Kia, a more downmarket manufacturer.

- Louis Vuitton and Nike. When interpolating between luxury fashion brand Louis Vuitton and sporting apparel and footwear company Nike, we find the closest midpoint brand is Adidas, essentially similar to Nike. However, among the other brands close to the midpoint are Calvin Klein, a relatively upmarket fashion brand with a sporty look and with a logo that fuses elements of both Louis Vuitton and Nike. Other nearby brands include Under Armour, another sporty, upmarket retailer, and BMW, an innovative, sporty, luxury car manufacturer.

- Google and McKinsey. Finally, we interpolate between the tech company and search engine Google and the management consultancy McKinsey. The closest brands to the midpoint between these firms are SAP, Microsoft, and IBM. SAP and Microsoft are both business-oriented technology providers. Both are dominant B2B providers of software. Besides being a technology company, IBM also provides extensive IT consulting services.

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.

9. Decision Support

In all of the previous analyses, we use the full inference network and manipulate the learned z_i representations to aid in the brand ideation process. Yet perhaps the most important contribution of our framework is providing a decision support tool for both designers in the early stages of designing logos and managers who want to understand the impact of logo design on brand perception. Supporting data-driven decisions across these perspectives requires our task-specific inference networks, which allow us to make predictions about missing modalities. In this section, we illustrate the decision support provided by these networks through two applications: in application 1, we consider using the manager's inference network to guide an actual rebranding that took place during the course of our research. Then, in application 2, we consider using the designer's and manager's inference networks together to provide decision support for the creation of a new brand identity.

9.1. Application 1: Rebranding McDonald's

Rebranding is common for firms looking to update their image and keep pace with changing markets (Henderson and Cote 1998). 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 changed their logo at least once.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 11. 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 11. 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 12. These results suggest that consumers 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 such as 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 considerable whitespace. 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.

9.2. Application 2: Shake Shack and In-N-Out

In our second application, we illustrate a full-use case of the model's decision support capabilities in the typical setting of a firm designing a new logo. Typically, firms hire designers to produce potential options for their new (or redesigned) logo. A manager then chooses among these options or suggests possible changes to the proposal (Henderson and Cote 1998). We illustrate how our model can help a designer designing candidate logos and a manager evaluating potential logo options or changes to a proposed logo.

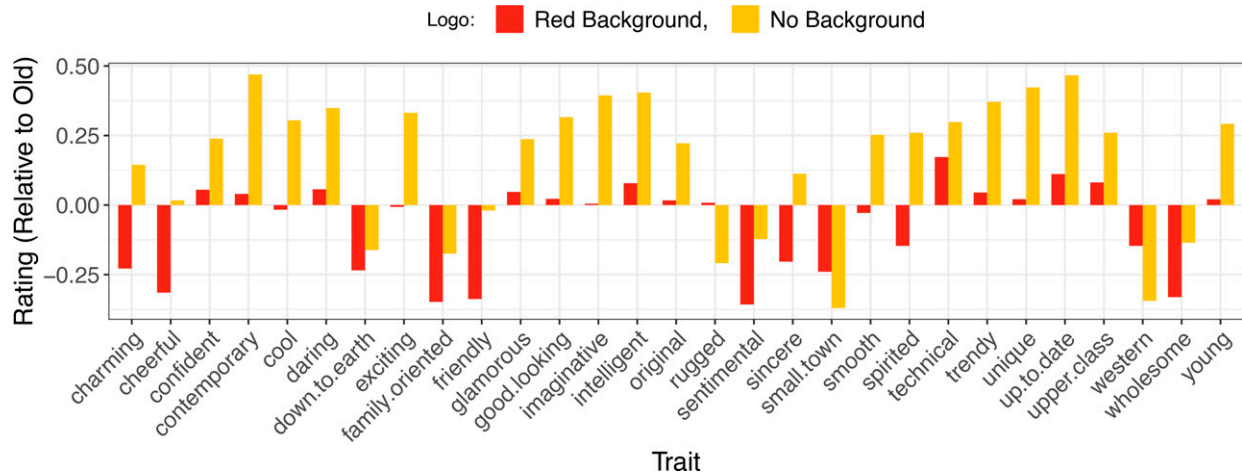
Building on our previous analysis, we consider designing a logo for a new fast-food firm. Specifically, we use as inspiration two fast-food firms that were not included in our data—Shake Shack and In-N-Out—and imagine the task of designing their logos from scratch. Structuring our analysis around existing brands that were not included in our training data allows us to ensure our model inputs are realistic and provides another benchmark to assess the validity of our results: although we are treating these brands as if

Figure 11. (Color online) Three Logos of McDonald's



Notes. (1) An older logo, popular from the 1990s (and on McDonald's roadside signs). (2) A newer version, introduced in the 2000s. (3) An even newer version with a red background.

Figure 12. (Color online) Comparison of the Predicted BP Evaluations Across the Two Candidate Logos Relative to the Old Logo Shown in Figure 11



Note. We omit traits for which both logos were expected to have approximately the same evaluation as the old logo.

they are new to the world, they do in fact exist, and we can, therefore, compare our model recommendations to their actual designs.

The two brands offer a compelling case study because their brand identities are quite distinct. Shake Shack is a relatively recent entrant to the fast-food space. Its origin in New York City and its focus on up-scale, urban markets is a fundamentally different positioning than competing fast-food chains. Its true logo is also quite different from the typical fast-food restaurant. 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. As the basis for a hypothetical new firm, we can use Shake Shack to test whether our model is able to suggest meaningful features even for a target that deviates from the norm. In-N-Out, on the other hand, draws on many of the classic fast food themes: it emphasizes its history, the classic roadside iconography of fast food firms, the typical red and yellow patterns in its true logo, and again, it serves an affordable menu similar to many mainstream fast food restaurants. Hence, in asking our model to design a logo for a hypothetical new firm based on In-N-Out, we can validate that our model is indeed able to guide the creation of more typical brand identities.








For the designer's task, decision support starts with the nonlogo inputs to the model: text describing the firm, a target brand personality profile, and the relevant set of tags describing the firm's high-level characteristics. To create our In-N-Out and Shake Shack lookalikes, we first gathered the same data for these two brands as for the brands in our calibration sample: we extracted

the words from their website and identified relevant tags. For brand personality, we used a *target* brand personality. For validation purposes, we also collected each brand's most typical logo. We process these data in an identical fashion as our training data, creating a new set of features that can be used by our model and, in particular, our task-specific inference networks.¹⁵ These features, minus the true logos, serve as the basis for our two lookalike brands.

9.2.1. Designing New Logos. To start, we consider decision support for the designer, tasked with developing the logo for our Shake Shack and In-N-Out lookalikes. Under our framework, this task is equivalent to using the website text, target brand personality, and tags as inputs to the designer's inference network from which we infer an approximate posterior for z_i . We then sample from that posterior to produce a distribution over the new brand's predicted logo features.¹⁶ For the designer, this process produces what is essentially a logo template: a set of likely and unlikely features that would be typical of a firm with that description. We summarize the model's outputs for the Shake Shack lookalike in Table 6.

Comparing these predictions to the actual logos (Online Figure 17), we find they are fairly accurate. The black colors, 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, 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 differences between the true logo and the

Table 6. 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

Categorical features			Binary features		
Feature	Likely values	Probability	Feature class	Likely values	Probability
Dominant color:	Black 	0.519	Accent color:	Light gray 	0.633
	Dark gray 	0.164		Dark gray 	0.303
	Light gray 	0.045	Contains color:	Light gray 	0.643
Hull shape:	Triangle	0.431	Font:	Black 	0.676
	Medium oval	0.259		Width: Original	0.904
	Thin oval	0.203		Style: No italics	0.946
Sans/serif font:	Sans	0.888		Weight: Bold	0.636
Mark class:	Detailed circular design	0.226	Other:	Class: Geometric	0.489
	Thin	0.219		Has a mark	0.967
	Narrow/vertical	0.142		High percentage whitespace	0.645
Number of colors:	Three colors	0.449		High perimetric complexity	0.521
	Two colors	0.308		Many downward diagonal edges	0.477

Note. For each of the binary features shown, we only report the highest values (e.g., the highest probability font class was geometric with probability 0.491).

prediction have to do with the font weight (light, not bold) and the accent colors (light green, not gray).¹⁷

In some sense, these fairly accurate predictions are not surprising: we have already demonstrated the predictive validity of the framework. The purpose of this example, however, is to illustrate how the framework can be used by a designer. The template shown in Figure 6 illustrates what the typical logo of a firm with the supplied features might look like. Similar to a designer's mood-boarding process, it is based on synthesizing existing logos of firms with similar features. Rather than manually synthesizing existing logos, the MVAE's designer inference network computationally synthesizes them through the latent z -space and produces the template in Figure 6. These features reflect the basic features a designer could use as a starting point to design a typical young, trendy, and glamorous food brand that describes itself using the same words Shake Shack uses on its website. Of course, the template is meant as a starting point, to aid the designer in the brainstorming process, not as a replacement for the designer. In Shake Shack's case, the fundamental role of the designer in improving on this template is evident in where the true logo departs from the template: although green was not a suggested color from our model, the neon green, thin burger in Shake Shack's logo is reminiscent of the signage at a typical 1950s "burger joint" with the burger explicitly indicating the industry. Similarly, the thin font is reminiscent of such signage.¹⁸

Finally, note that 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 (probability = 0.644). Although it also predicts two or three colors,

red and yellow occur with a much higher probability for In-N-Out with red being the most likely color in general and yellow being the most likely accent color. Although bold fonts are still predicted, sans-serif fonts are somewhat less likely with serif font being predicted with probability 0.234. Other relatively more likely visual features include high saturation in colors, high brightness, low perimetric complexity, and high vertical symmetry, most 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 emphases in the target brand personality as well as the different words emphasized on the two firms' websites (as shown in Online Appendix A, Figures 17 and 18). Taken together, the contrasting model recommendations show that the model is able to make meaningful distinctions between ostensibly similar firms in a way that could guide designers to crafting effective brand imagery.

9.2.2. Assessing Visual Changes. Finally, we again consider the task of assessing changes to a brand's logo from the perspective of a manager, 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. In this way, our model provides decision support for managers as a sandbox for experimenting with potential redesigns or for comparing several potential designs.

To illustrate this, suppose Shake Shack was considering changing its font from its current light font

weight to a bold font weight (as suggested by the preceding model). How might consumer perceptions about Shake Shack change? Such a prediction can be tested directly by adding the “bold font” feature to the binary logo features and removing “light font” and then using the MVAE’s manager’s inference network. When we do so, we find that personality traits such as western, small town, down to earth, and family-oriented likely go up, and traits such as up to date, intelligent, contemporary, and daring likely go down.¹⁹ Although the changes are relatively slight, in keeping with the relatively small proposed modification, we do notice a pattern: Shake Shack’s modern image may be negatively affected although its perception along classic fast-food dimensions such as family-oriented may be bolstered.

10. Conclusion

In this paper, we explore logo design and brand identity from a data-driven perspective. Leveraging a relatively large data set of prominent brands, a novel logo feature extraction algorithm, and both model-free and model-based analyses, we show 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 show 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 develop an automatic approach for extracting meaningful and manipulable features from logos. Second, we develop a multiview learning framework based on multimodal variational autoencoders with a novel approach to inference. Our inference procedure combines task-specific inference networks with stochastic data binning and is especially suitable for the simultaneous estimation of multiple inference networks that are geared toward 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.

More generally, we see much promise for multimodal learning methods, such as our MVAE, in other marketing contexts. Many sources of data in marketing are multimodal, including information about

customers in omni-channel settings, and user-generated data on online platforms, which often contain text, images, and numeric outcome metrics. Much of the content generated by firms is similarly multimodal: e-commerce platforms typically contain photos, videos, text, and numeric information (e.g., prices). Our MVAE framework specifically and multimodal representation learning more generally are ideal for tackling these modern data challenges to uncover deeper insights about customers, brands, and markets.

Finally, we note several areas for future research. 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. Although exploring optimality of designs may pose an interesting future research area, the task of moving from a typical 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).

Acknowledgments

The authors thank Sanjana Rosario and Nikhil Kona for excellent research assistance and the Wharton Behavioral Lab for support. This paper also benefited tremendously from feedback from numerous seminar and conference participants, including seminars at Stanford University, Temple University, the University of Maryland, Drexel University, Reykjavik University, the Cheung Kong Graduate School of Business Research Camp, and the annual Four Schools Conference.

Endnotes

¹ The reverse-coded traits were honest/dishonest, exciting/boring, and good looking/ugly. Any participant who answered that both traits are descriptive of the firm was automatically removed.

² See https://en.wikipedia.org/wiki/Vox-ATypL_classification.

³ We include a forest plot showing these relationships in Online Appendix A, Figure 13.

⁴ We use the terms modality, data source, and domain interchangeably.

⁵ This simple coding reflects whether 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.

⁶ The $\log(e^{\theta} + 1)$ structure in Equation (4) enforces positivity and is more numerically stable compared with simple exponentiation.

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 on which it was trained. Our model does exceptionally well on this in-sample measure, but we do not report it here, favoring the harder out-of-sample metrics.

⁸ There are two benefits of this likelihood scaling procedure over just simply removing the domains from the model: from a practical perspective, this procedure is significantly easier to implement, requiring adding only a simple predefined scalar to the likelihood. Second, although we do not explore this here, this scaling factor can be tuned, or continuously increased, to assess the contribution of the domain across a broad spectrum of weights.

⁹ Examples of all study stimuli are in Online Appendix H.

¹⁰ This distance restriction is necessary to ensure the choice is meaningful: without this restriction, a participant may be forced to choose between two very similar options (e.g., brand-personality descriptions that differ by only one word or word clouds that contain many of the same terms).

¹¹ See, for example, <https://99designs.com/blog/tips/logo-design-process-how-professionals-do-it/>.

¹² We show the traits that increased and decreased the most relative to the original $\bar{z}_{\text{FastFood}}$ in Online Appendix A, Figure 16.

¹³ In fact, since compiling our initial data set in 2016, at least three brands have changed their logos.

¹⁴ For an informal overview of the history of the McDonald's logo, see <https://www.digitaldoughnut.com/articles/2019/september/mcdonalds-history-and-evolution-of-a-famous-logo>.

¹⁵ The features of Shake Shack are summarized in Online Appendix A, Figure 17, and those for In-N-Out in Online Appendix A, Figure 18.

¹⁶ It is important to note that this operation is *entirely out-of-sample*: neither Shake Shack nor In-N-Out's logos were used in learning the parameters of any of the functions in our model, nor were they used in this case to compute the approximate posterior.

¹⁷ The light gray may be an artifact of the feature extraction process: when thin, black features are imposed on a white background, the color quantization procedure described in the online appendix nearly always erroneously detects a light gray color in addition to the black. This also accounts for the prediction of three colors.

¹⁸ See <https://www.fastcompany.com/3041777/the-untold-story-of-shake-shacks-16-billion-branding>.

¹⁹ The model's predictions of the 10 BP traits most positively and negatively affected are shown in Online Appendix A, Figure 19.

References

- Aaker JL (1997) Dimensions of brand personality. *J. Marketing Res.* 34(3):347–356.
- Bingham E, Chen JP, Jankowiak M, Obermeyer F, Pradhan N, Karaletsos T, Singh R, Szerlip P, Horsfall P, Goodman ND (2019) Pyro: Deep universal probabilistic programming. *J. Machine Learn. Res.* 20(1):973–978.
- Blei DM, Kucukelbir A, McAuliffe JD (2017) Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* 112(518):859–877.
- Chae BG, Hoegg J (2013) The future looks “right”: Effects of the horizontal location of advertising images on product attitude. *J. Consumer Res.* 40(2):223–238.
- Childers TL, Jass J (2002) All dressed up with something to say: Effects of typeface semantic associations on brand perceptions and consumer memory. *J. Consumer Psych.* 12(2):93–106.
- Cian L, Krishna A, Elder RS (2014) This logo moves me: Dynamic imagery from static images. *J. Marketing Res.* 51(2):184–197.
- Deng X, Kahn BE (2009) Is your product on the right side? The “location effect” on perceived product heaviness and package evaluation. *J. Marketing Res.* 46(6):725–738.
- Deng X, Hui SK, Hutchinson JW (2010) Consumer preferences for color combinations: An empirical analysis of similarity-based color relationships. *J. Consumer Psych.* 20(4):476–484.
- Dieng AB, Kim Y, Rush AM, Blei DM (2019) Avoiding latent variable collapse with generative skip models. *Proc. 22nd Internat. Conf. Artificial Intelligence Statist. Proc. Machine Learn. Res.*, vol. 89, 2397–2405.
- Doyle JR, Bottomley PA (2006) Dressed for the occasion: Font-product congruity in the perception of logotype. *J. Consumer Psych.* 16(2):112–123.
- Endrissat N, Islam G, Noppeney C (2016) Visual organizing: Balancing coordination and creative freedom via mood boards. *J. Bus. Res.* 69(7):2353–2362.
- Goodfellow I, Bengio Y, Courville A (2016) *Deep Learning* (MIT Press, Cambridge, MA). <https://www.deeplearningbook.org/>.
- Hagtvedt H (2011) The impact of incomplete typeface logos on perceptions of the firm. *J. Marketing* 75(4):86–93.
- Henderson PW, Cote JA (1998) Guidelines for selecting or modifying logos. *J. Marketing* 62(2):14–30.
- Henderson PW, Giese JL, Cote JA (2004) Impression management using typeface design. *J. Marketing* 68(4):60–72.
- Henderson PW, Cote JA, Leong SM, Schmitt B (2003) Building strong brands in Asia: Selecting the visual components of image to maximize brand strength. *Internat. J. Res. Marketing* 20(4):297–313.
- Jiang Y, Gorn GJ, Galli M, Chattopadhyay A (2015) Does your company have the right logo? How and why circular and angular logo shapes influence brand attribute judgments. *J. Consumer Res.* 42(5):706–726.
- Kardes FR, Posavac SS, Cronley ML, Herr PM (2008) Consumer inference. Hagtvedt CP, Herr PM, Kardes FP, eds. *Handbook of Consumer Psychology* (Taylor & Francis Group, New York), 165–192.
- Kareklas I, Brunel FF, Coulter RA (2014) Judgment is not color blind: The impact of automatic color preference on product and advertising preferences. *J. Consumer Psych.* 24(1):87–95.
- Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. Preprint, submitted December 22, <https://arxiv.org/abs/1412.6980>.
- Kingma DP, Welling M (2013) Auto-encoding variational Bayes. Preprint, submitted December 20, <https://arxiv.org/abs/1312.6114>.
- Kingma DP, Mohamed S, Rezende DJ, Welling M (2014) Semisupervised learning with deep generative models. Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger KQ, eds. *Advances in Neural Information Processing Systems, Montreal*, 3581–3589.
- Klink RR (2003) Creating meaningful brands: The relationship between brand name and brand mark. *Marketing Lett.* 14(3):143–157.
- Li Y, Yang M, Zhang Z (2018) A survey of multi-view representation learning. *IEEE Trans. Knowledge Data Engrg.* 31(10):1863–1883.
- Liu X, Lee D, Srinivasan K (2019) Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *J. Marketing Res.* 56(6):918–943.
- Liu L, Dzyabura D, Mizik N (2020) Visual listening in: Extracting brand image portrayed on social media. *Marketing Sci.* 39(4):669–686.
- Loken B, Barsalou LW, Joiner C (2008) Categorization theory and research in consumer psychology: Category representation and category-based inference. Hagtvedt CP, Herr PM, Kardes FP, eds. *Handbook of Consumer Psychology* (Taylor & Francis Group), 133–164.
- McDonagh D, Storer I (2004) Mood boards as a design catalyst and resource: Researching an under-researched area. *Design J.* 7(3):16–31.
- Meyers-Levy J, Peracchio LA (1992) Getting an angle in advertising: The effect of camera angle. *J. Marketing Res.* 29(4):454–461.
- Miller L (2016) Mood boarding: What it is and how it helps build design concepts. *Design Sensory*. <https://designsensory.com/>

- p>thinking/2016/12/06/mood-boarding-what-it-is-and-how-it-helps-build-design-concepts/.
- Navon D (1977) Forest before trees: The precedence of global features in visual perception. *Cognitive Psych.* 9(3):353–383.
- Nazabal A, Olmos PM, Ghahramani Z, Valera I (2020) Handling incomplete heterogeneous data using VAEs. *Pattern Recognition* 107:107501.
- Orth UR, Malkewitz K (2008) Holistic package design and consumer brand impressions. *J. Marketing* 72(3):64–81.
- Pieters R, Wedel M, Batra R (2010) The stopping power of advertising: Measures and effects of visual complexity. *J. Marketing* 74(5):48–60.
- Ranganath R, Tang L, Charlin L, Blei DM (2014) Deep exponential families. Preprint, submitted November 10, <https://arxiv.org/abs/1411.2581v1>.
- Rezende DJ, Mohamed S, Wierstra D (2014) Stochastic backpropagation and approximate inference in deep generative models. Xing EP, Jebara T, eds. *Proc. 31st Internat. Conf. Machine Learn., Beijing China*, II-1278–II-1286.
- Schlosser AE, Rikhi RR, Dagogo-Jack SW (2016) The ups and downs of visual orientation: The effects of diagonal orientation on product judgment. *J. Consumer Psych.* 26(4):496–509.
- Semin GR, Palma TA (2014) Why the bride wears white: Grounding gender with brightness. *J. Consumer Psych.* 24(2):217–225.
- Stigliani I, Ravasi D (2012) Organizing thoughts and connecting brains: Material practices and the transition from individual to group-level prospective sensemaking. *Acad. Management J.* 55(5):1232–1259.
- Suzuki M, Nakayama K, Matsuo Y (2016) Joint multimodal learning with deep generative models. Preprint, submitted November 7, <https://arxiv.org/abs/1611.01891>.
- Valdez P, Mehrabian A (1994) Effects of color on emotions. *J. Experiment. Psych. General* 123(4):394–409.
- van der Lans R, Cote JA, Cole CA, Leong SM, Smidts A, Henderson PW, Bluemelhuber C, et al. (2009) Cross-national logo evaluation analysis: An individual-level approach. *Marketing Sci.* 28(5):968–985.
- Walsh MF, Winterich KP, Mittal V (2010) Do logo redesigns help or hurt your brand? The role of brand commitment. *J. Product Brand Management* 19(2):76–84.
- Wu M, Goodman N (2018) Multimodal generative models for scalable weakly-supervised learning. Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, eds. *Advances in Neural Information Processing Systems, Montreal*, vol. 31, 5575–5585.
- Wu M, Goodman N (2019) Multimodal generative models for compositional representation learning. Preprint, submitted December 11, <https://arxiv.org/abs/1912.05075>.