INFORMATION DESIGN WITH BIG DATA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Shengli Hu

August 2019

© 2019 Shengli Hu

INFORMATION DESIGN WITH BIG DATA

Shengli Hu, Ph. D.

Cornell University 2019

This dissertation consists of three essays that investigate the effects of information design from different aspects and in different business contexts.

The first essay --- information design in storytelling --- studies the impact of information structure in the form of induced suspense and surprise on audience experience in the context of movie viewing, and identifies significant impacts of surprise on audience experience and evaluation. By formulating the information design problem as a constraint optimization problem as in Ely et al. (2015), we propose and validate textual measures of suspense and surprise based on a merged dataset of movie scripts, movie features, and market outcome statistics.

The second essay --- cognitive categorization, memorability, and likability --- explores what makes a visual design memorable and likable by proposing and providing scalable methods to descriptively quantify and evaluate two cognitive processes for logos: (1) perceptual categorization; and (2) functional categorization. With a dataset consisting of 125,270 logo designs from the U.S. market, spanning 39 industry categories, and annotated scores of memorability and likability, we evaluate both the absolute and relative impacts of two forms of cognitive categorization and logo features on design memorability and design likability. In addition, we explore multiple methods for logo clustering and analyze drivers of memorability and likability. To validate the cognitive interpretation of our proposed measures, we further gather and incorporate human perceptual templates into the algorithm. We discuss managerial

insights for logo design.

Third essay reviews the existing research on the intersection with a focus on visual image data and proposes a classification scheme for the diverse range of computer vision methods and constructs for visual marketing that has been developed in the literature. The classification criteria include the nature of research questions, application contexts, computational methods, and forms of big data. Additionally, the paper provides comparative evaluations of each criterion both horizontally and vertically, a normative guide on the use of these systems and results under different situations, and an agenda for future research.

BIOGRAPHICAL SKETCH

Shengli was lucky to be have been admitted and transferred to Cornell University, from University of Southern California, where she studied Economics. Prior to USC, she obtained a Bachelor degree of Management Science from Fudan University in Shanghai, China. She was born in a coastal city in Jiangsu province in east China.

To my mother Jing Zhang, and my ragamuffin, Dottie Hu.

ACKNOWLEDGMENTS

It has been a life-changing and eye-opening experience since the summer of 2014, when I moved across the country to the gorgeous Ithaca. I would not have been able to push through without the understanding, selfless support and guidance of many incredible people, to whom I am deeply indebted to.

First and foremost, my greatest thanks go to my advisor Vrinda Kadiyali, to whom I look up greatly. I have been extremely lucky to have her as my advisor. Her vision, integrity, broad research interests, knowledge and appreciation for various research areas, leadership and interpersonal skills, humor, kindness, and personal charisma, among others, made my PhD journey much more rewarding and fruitful than expected. I am deeply and forever grateful to her, and will strive to live up to the high standard she set as a role model.

I would like to thank Shawn Mankad and Bharath Hariharan for being on my dissertation committee, and for their guidance and help throughout my PhD studies. Shawn is an extremely kind, hard-working, and intelligent person, to whom I am greatly indebted to. It was Shawn who has guided me through the early years of graduate study when I was most disoriented. Bharath is an extremely fun, easy-going, and prolific scholar, from whom I have been fortunate to learn from. I am grateful to him for including me, the outsider, in the computer vision and graphics reading group and has always been open to discussions, which broadened my knowledge in computer vision and computer science in general.

I am also extremely indebted to the faculty groups of Operations Technology Information Management (OTIM), and Marketing, for their understanding and support throughout my graduate study, wandering and straddling in the blurry grounds in between. I thank Andrew Davis for his guidance, patience, effort, and understanding during my first year; Vithala Rao, for being extremely supportive and understanding,

and for endless discussions and inspirations; Nagesh Gavirneni for his advice and help over time; Yao Cui for his guidance, patience, and understanding during my third year; Li Chen and Clarence Lee, for discussions and encouragement over time.
Outside Cornell, I am also grateful to Sampath Rajagopalan, Ramandeep Randhawa, Sheldon M Ross, Isabelle Brocas, Shinichi Sakata(, and Dinesh Puranam) at USC, Weixin Shang at Lingnan University, Jeff Liu at CUHK, without the guidance and help of whom I would not have been admitted to PhD programs in the first place.
I am also grateful to many Johnson graduates for their endless support, time, and discussions, especially Chuchu Liang, Jialie Chen, Alan Kwan, Sungjin Kim, Sarah Lim, Chao Kang, Piyush Anand, Kate Volkova, Dayoung Kim, Huisi Li, Gaurav Kankanhalli, Zhen Lian, among many others. Additional thanks to Xanda Schofield and Tianze Shi across the street for helping with ACL presentations when my visa was denied. I will always miss them fondly.
I thank my parents: Jing Zhang and Zhen Hu, who have supported my decisions to pursue a PhD in Management, and my decision to pursue research positions outside academia through and through, financially, emotionally, and logistically. I would never have made it to where I am today without their love, understanding, and support.
Special thanks to my boyfriend Dylan Shinzaki, who, not only wined and dined the impoverished me, but also, and more importantly, has been extremely understanding and supportive through my most stressful periods during PhD and on the job market, being the most agreeable, level-headed, intelligent, and cerebral person I have met.
The same goes to my ragamuffin, Dottie Hu, who has been my rock since day one.

TABLE OF CONTENT

Information Design and Audience Experience

Shengli Hu

Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853

sh2264@cornell.edu

---

1. Introduction

In this paper, we examine how information sequencing in a story can influence the appeal of a story. We study this question in the context of information goods, defined as goods that can be digitized (Varian 2000) (e.g. books, movies, phone conversations etc.) Information goods now comprise the largest sector of the U.S. economy (*U.S. Bureau of Economic Analysis, U.S. Department of Commerce* 2017); this adds relevance to our research question.

Storytelling is a central component to many types of information goods and has been proven a powerful tool for audience engagement. A well-constructed story can be effective in communication (Heath and Heath 2007), and persuasion (Phillips and McQuarrie 2010, Bilandzic and Busselle 2013). Marketers have harnessed the power of storytelling to strengthen brand recognition and identification (Herskovitz and Crystal 2010), and foster stronger consumer connections to the brand (Escalas 2004, Papadatos 2006).

In this paper, we explore how information sequencing affects plot uncertainty, and whether plot uncertainty is successful in audience engagement. We follow Ely, Frankel and Kamenica (2015) in focusing on two fundamental information-sequencing concepts — suspense and surprise. Suspense is defined as the build-up tension of uncertainty. It is greater if there is greater uncertainty of next periods belief. Surprise is defined as the drastic shift of beliefs. Surprise is greater if the current belief realization is further apart from that of the last period. The audiences utility from the viewing experience is a concave function of the aggregate suspense and/or surprise, depending on individual preferences over the stochastic belief path. We build and expand on these theoretical ideas in Ely et al. (2015) by building a model of a producer of information goods building storytelling with appropriate suspense

and surprise, when faced with consumers with preferences over these two information sequencing concepts. We then test the empirical predictions of the model as follows. We analyze 1088 U.S. movie scripts released in 1928-2015. We are able to demonstrate how to measure information sequencing in story telling (suspense and surprise) in movie scripts and their impact on movie performance.

To preview results, we find that our surprise-related information measures exhibit significant positive effects on the aggregate audience rating whereas suspense-related information measures are not significant with coefficients literally equal to zero. Our results have implications for managers and consumers: when designing information goods given the objective of greater positive consumer response, it could be more efficient and effective for managers to focus surprising contents as opposed suspenseful contents.

We now turn to the following sections — Section 2 outlines where our study stands among multiple streams of literature in economics, information science, and marketing. Section **??** details the theoretical backbone we build our empirical measures upon. Section 5 describes our setting and the datasets in question. Section 6 lays out how we operationalize our empirical measures of information content based on Section **??**. Section 7 provides our model specification and results, followed by robustness tests in Section **??**, and conclusions in Section 9.

2. Literature Review

We discuss below several streams of literature related to our work, and highlight how we build on and extend current research.

As mentioned in the introduction, we study information good. Several economists and information science researchers have examined various aspects of these goods since the emergence of the Internet at the end of 90s. Research questions include pricing strategy (Varian 1997), versioning (Varian 1997), network effects (Parker and Alstyne 2005), bundling (Bakos and Brynjolfsson 1999, Bakos and Brynjolfsson 2000, Geng and Stinchcombe 2005, Wu, Hitt and Chen 2008), etc. We examine consumer preferences for time-varying information content, or in other words, design of an information good. This aspect has not been examined in these literatures before. However, there is an emerging literature in marketing on product design of information products. For instance, Halbheer, Stahl and Koenigsberg (2014)

design an optimal strategy for market sampling of online information goods, Netzer and Toubia (2014) propose the optimal design features for creative content. Such papers take a machine-learning optimization view of product design. In contrast, our work takes an economic theory-based approach to product design, as discussed in Section **??** below.

Our estimation of time-varying information content (surprise and suspense, as mentioned in the Introduction (Section 1)) is related to the following second stream of literature. Marketing researchers have used continuous consumer feedback ("moment-to-moment" data) to track and measure consumer experiences in a variety of settings such as advertising (Baumgartner, Sujan and Padgett 1997, Elpers, Wedel and Pieters 2003, Elpers and Mukherjee 2004) and TV show pilot testing (Hui, Meyvis and Assael 2014). The study closest to ours is (Teixeira, Wedel and Pieters 2012). They measure the moment-to-moment intensity of joy and surprise expressed by participants who were watching Internet video advertisements. The authors relate the moment-to-moment emotional intensity to attention and concentration as well as to viewing behavior and derive the optimal emotion trajectories to aid effective TV advertisement designs using a dynamic frailty model. Our work differs from Teixeira et al. (2012) in that we instead focus on the information content of the product itself (rather than the consumer reaction to the product). This way, we are able to provide the link between design features and consumer evaluation. We show how to measure the evoked emotions from product designs alone without eliciting real-time consumer responses.

The third stream of literature related to work is the marketing literature on motion pictures, especially that using textual data. Eliashberg, Hui and Zhang (2007) and Eliashberg, Hui and Zhang (2014) are among the few that apply natural language processing (hereafter, NLP) to the movie industry. They show that including textual information in movie synopsis leads to better predictions of box office outcomes. Toubia, Iyengar, Bunnell and Lemaire (2015) adopts a seeded LDA approach to understand and predict individual movie consumption patterns. While Toubia et al. (2015) are informed by theories grounded in positive and media psychology, we are mostly informed by theories in microeconomics on microfoundations of preferences under the mathematical framework of Bayesian updating. This affords us theoretical foundations that connects information structure and audience experience.

In a more general sense, our study is connected to Information Management (IM hereafter) literature. Specifically, IM studies that (1) design data mining methods of the content of online information goods, for instance, Adamopoulos, Ghose and Todri (2017), Ghose and Han (2011), Archak, Ghose and Ipeirotis (2011), Ghose and Ipeirotis (2011); (2) examine

online information consumption, for instance, Calin, Dellarocas, Palme and Sutanto (2013), Chiou and Tucker (2013), Calzada and Gil (2017), Athey, Mobius and Pál (2017), Sismeiro and Mahmood (2018); (3) prescribe optimal strategies for online content management, for instance, Caro and Martínez-de Albéniz (2018).

3. Theory and Hypotheses

We first detail the setup reproducing and combining mathematical formulations in Ely et al. (2015), which is a dynamic generalization of Bayesian updating models as in Kamenica and Gentzkow (2011).

An audience forms a series of beliefs about the state of the world $\omega \in \Omega$, based on the information revealed by the producer (principal) over time. A belief about the state ω is denoted $\mu^{\omega} \in \triangle(\Omega)$. Let $t \in \{0, 1, \ldots, T\}$ denote the period during which information that advances the storyline gets revealed by the producer.

We refer to such information as signals π's sent by the producer, consisting of a finite realization space S and a mapping from state space Ω to probability distributions over S: $\Delta(S)$. Given a signal, each realization s leads to a posterior belief $\mu_s \in \Delta(\Omega)$. Each signal leads to a distribution over posterior beliefs denoted $\tau \in \triangle(\triangle(\Omega))$. Formally, a signal π induces a posterior distribution τ and vice versa if

$$\mu_s(\omega|s) = \frac{\pi(s|\omega)\mu_0(\omega)}{\sum_{\omega' \in \Omega} \pi(s|\omega')\mu_0(\omega')}, \forall s, \forall \omega \tag{1}$$

$$\tau(\mu) = \sum_{s:\mu_s=\mu} \sum_{\omega' \in \Omega} \pi(s|\omega')\mu_0(\omega'), \forall \mu \tag{2}$$

where $\mu_0(\omega) \in \text{int}(\triangle(\omega))$ is the prior shared by both the producer and the audience. The producer first chooses an information revelation policy $\tilde{\pi} \in \tilde{\Pi}$ over the entire span that maps the current period and the current belief of the audience to a signal. All information revelation policies generate a stochastic path of beliefs about the state of the world, which is modeled as a belief martingale $\tilde{\mu}$.

Formally, a belief martingale $\tilde{\mu}$ is a sequence of induced beliefs $(\tilde{\mu}_t)_{t=0}^{T}$ such that (1) $\tilde{\mu}_t \in \Delta(\Delta(\Omega)), \forall t$; (2) $E[\tilde{\mu}_t|\mu_0, \ldots, \mu_{t-1}] = \mu_{t-1}, \forall t \in \{0, \ldots, T\}$. Also let $\eta = (\mu_t)_{t=0}^{T}$ denote

a belief path — the realization of a belief martingale. We also assume belief martingales are Markov, which means $P(\tilde{\mu}_{t+1}|(\mu_t)_{t=0}^{t}) = P(\tilde{\mu}_{t+1}|\mu_t)$. By law of iterated expectation, a signal induces a distribution of posteriors $\tilde{\mu}_{t+1}$ such that $E[\tilde{\mu}_{t+1}] = \mu_t$.

The audience's experienced utility from viewing the information good is constructed such that, if he has a preference for suspense, his utility function is

$$U_{susp}(\eta, \tilde{\mu}, T) = \sum_{t=0}^{T-1} u\left(E_t \sum_{\omega} (\tilde{\mu}_{t+1}^{\omega} - \mu_t^{\omega})^2\right) \tag{3}$$

and if he has a preference for surprise, his utility function is

$$U_{surp}(\eta, T) = \sum_{t=0}^{T} u\left(\sum_{\omega} (\mu_t^{\omega} - \mu_{t-1}^{\omega})^2\right) \tag{4}$$

for some increasing and concave utility function $u(\cdot)$ with $u(0) = 0$. When $u(x) = \sqrt{x}$, U_{susp} is the standard deviation of posterior beliefs over all states and U_{surp} is the Euclidean distance between μ_t and μ_{t-1}.

Intuitively, a period generates more suspense if, given the current belief of the state of the world and a history of all available information, there is even greater uncertainty of beliefs in the next period — greater variance of next period's beliefs; a period generates more surprise if there is a greater shift of the current belief from last periods belief — greater Euclidean distances between current and last periods beliefs. Figure 1 illustrates temporal relations of such intuitions on a time line from 0 to T.

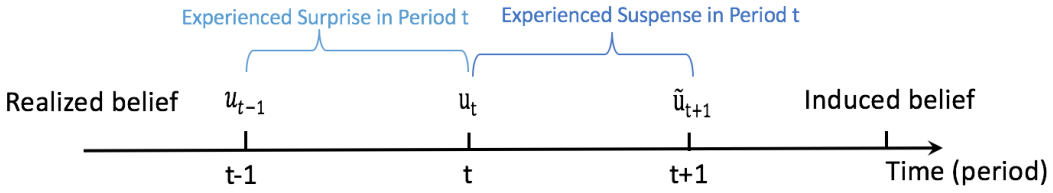


Figure 1: An illustration of Suspense and Surprise concepts on a time line

Now let the belief martingale induced by information policy $\tilde{\pi}$ given the prior belief μ_0 be $\langle\tilde{\pi}|\mu_0\rangle$. The producer chooses the information policy to maximize the expected experienced utility of consumers over the entire period, solving the optimization problem.

4. Optimal Information Policy

We first assume the number of periods and the common prior beliefs are exogenously given. In our context of movies, this is assuming that the length of a movie is relatively fixed within a range and audience come to the theatre with generic expectations based on prior exposures to the genre, the actors, the director(s), etc. We discuss how and when to allocate information over time so that cumulative suspense is maximized for the audience.

4.0.1 Suspense-optimal Information Policies

As is shown in Ely et al. (2015), when the number of periods and priors are fixed, the decisions of the producer boil down to a constrained optimization problem where the producer is faced with a budget of variances, or private information2. Since the audience utility function is concave, it is ideal to dole out private information evenly over time for maximal suspense. We formally state this intuition about how to allocation nformation as Hypothesis 1a.

Proposition 1 in Kamenica and Gentzkow (2011) (lemma 1 in Ely et al. (2015)) establishes the equivalence between the audiences induced belief martingale and the information policy chosen by the producer. We state this result in the following Assumption 1. Proposition 1 in Kamenica and Gentzkow (2011) (lemma 1 in Ely et al. (2015)) establishes the equivalence between the audience's induced belief martingale and the information policy chosen by the principal.[1] We state this result in the following assumption 1 that justifies our empirical measures in section 6:

Assumption 1 (Equivalence) *Given any Markov belief martingale $\tilde{\mu}$ and prior μ_0, there exists an information policy $\tilde{\pi}$ that induces $\tilde{\mu}$, denoted as $\tilde{\mu} = \langle \tilde{\pi} | \mu_0 \rangle$.*

Given assumption 1, the insights below follow immediately from Proposition 1 and Proposition 3 in Ely et al. (2015):

If the audience has a preference for suspense, the producer who maximizes the audience's experienced utility — expected suspense, adheres to the information policy that induces a belief martingale $\mu_t \in M_t, \forall t$, where $M_t = \{\mu | \Psi(\mu) = \frac{T-t}{T}\Psi(\mu_0)\}$. $\Psi(\mu)$ is the residual variance from full information revelation given current belief μ, which captures how much of

[1] Kamenica and Gentzkow (2011) show in a static model that when any posterior distribution can be induced by some signal given current belief. Ely et al. (2015) apply this result to a corresponding dynamic model and show its dynamic counterpart.

the producer's private information has not yet been revealed to the audience, analogous to its role in insider trading models (Ostrovsky 2012).

However, given the discrete nature of both time and private information in many settings, the optimal solution described in Hypothesis 1a is often not achievable. The assumption that the length of the good is exogenously fixed is not ideal, either, since sometimes there is indeed some uncertainty about the length of the experience to the audience. Under such circumstances, the suspense-optimal belief (martingale) is no longer unique as in Hypothesis 1a, therefore we introduce Hypothesis 1b that describes one refinement to rank among suspense-optimal beliefs.

Such a refinement mechanism operates in the following way. First, all the information bursts throughout the experience are identified; second, the length of the intervals between information bursts are calculated; third, the empirical expectation of intervals is summarized; finally, the greater the expected interval in between information bursts, the more suspenseful the experience feels like to the audience.

Hypothesis 1a (Suspense-optimal Strategy by Ely et al. (2015)) *The more evenly private information is revealed over time, the more suspenseful the viewing experience is to the audience.*

Hypothesis 1b (Suspense-optimal Strategy) *The greater the expectation of intervals between information revelation peaks, the more suspenseful the viewing experience is to the audience.*

4.1 Surprise-optimal Information Policies

Deriving the solution for optimal surprise requires further analysis — there is the commitment problem facing a surprise-seeking producer: the surprise-optimal martingale has paths that generate little surprise. In order to implement the optimal policy, the producer needs commitment power to follow through. Otherwise the audience would expect potential deviations and the chosen path would no longer be surprising. Illustrative examples are given in Ely et al. (2015). To circumvent such concerns, we move away from Assumption 1 and introduce another equivalence result based on the concept of Doob martingale, independent of previous discussions of suspense-optimal information policies.

Assumption 2 (Doob/Levy) *Given any Markov belief martingale $\tilde{\mu}$ and prior μ_0, there exists a unique Markov emotion martingale $\tilde{\epsilon}$ that is induced by $\tilde{\mu}$, denoted as $\tilde{\epsilon} = \langle \tilde{\mu} | \mu_0 \rangle$.*

Given Assumption 2, we measure multi-dimensional emotion martingales and identify surprise-optimal paths by simulation based on the definition of surprise in Equation 4. We detail such an intuition in Hypothesis 2a.

Hypothesis 2a (Surprise-optimal Strategy) *The greater the cumulative Euclidean distances between emotion paths over time, the more surprising the experience is to the audience.*

On the other hand, when commitment requirement is satisfied, the producer trying to optimize surprise adheres to the information policy that induces a belief martingale μ_t such that for all $\epsilon > 0$, if there are enough remaining periods, then $|\mu_{t+1} - \mu_t| < \epsilon$. Further, building on Mertens and Zamir (1977), $\forall \mu, \lim_{T \to \infty} \frac{W_T(\mu)}{\sqrt{(T)}} = \phi(\mu)$, where $W_T(\mu)$ is the value function and $\phi(\mu)$ is the pdf of standard normal at μth quantile. We detail the intuition in Hypothesis 2b.

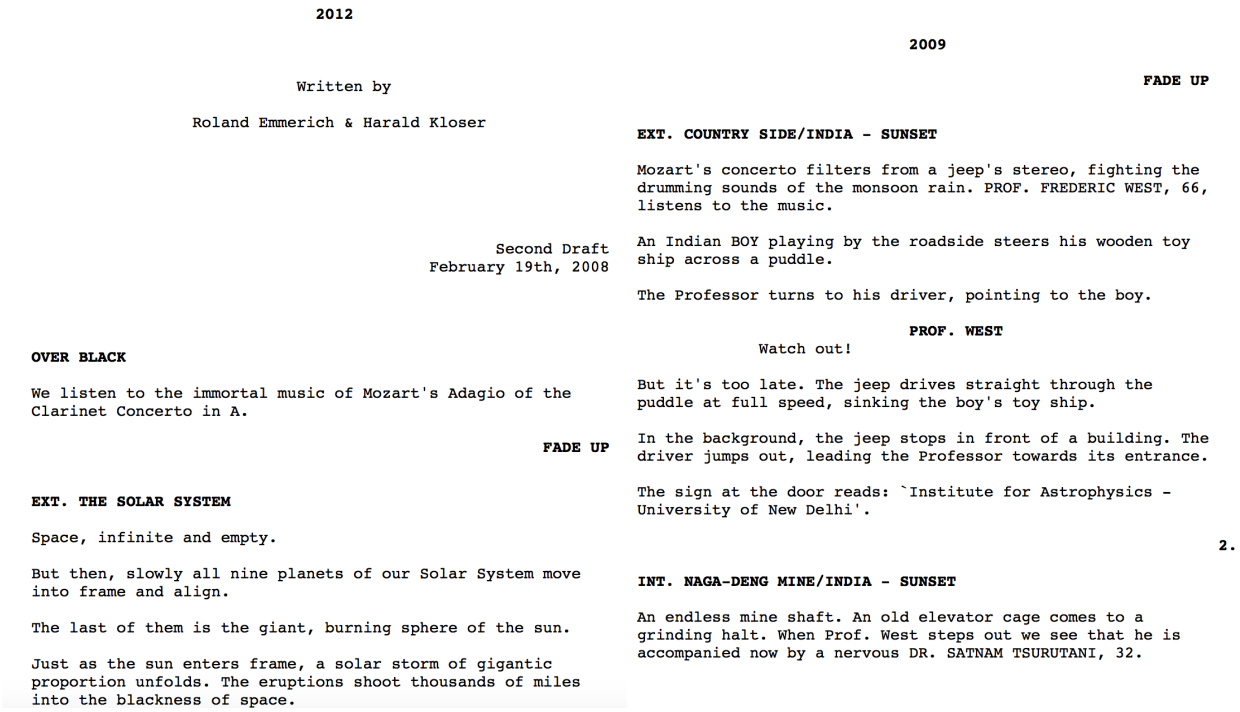
Hypothesis 2b (Surprise-optimal Strategy by Ely et al. (2015)) *The spikier the belief path is as it proceeds, the more surprising the viewing experience can be to the audience.*

5. Problem Domains and Datasets

5.1 Movie Scripts

We obtained the movie scripts from the Internet Movie Script Database (imsdb.com) using text mining packages in Python and R. We treat each movie script as a document. There are 18 genres of 1088 movie scripts in our dataset. The 18 genres are: action, adventure, animation, comedy, crime, drama, family, fantasy, film-noir, horror, musical, mystery, romance, sci-fi, short, thriller, war and western, representing a broad cross-sectional collections of film product categories. Figure 2 shows snapshots of a sample script (Movie *2012*) in our dataset. After excluding scripts that are shorter than 500 words or are not in easily accessible format, we are left with 1,108 distinct movie scripts of 16 genres (short, film-noir excluded). We also obtained movie metadata that includes release date, script date, screenwriter, genre, producer, cast, budget, open-week box office, total revenue, consumer average rating, etc., from the Internet Movie Database (imdb.com/interfaces), Rotten Tomatoes (rottentomatoes.com)

and The Numbers website (the-numbers.com), and matched these datasets with the script dataset.



2012

Written by

Roland Emmerich & Harald Kloser

Second Draft
February 19th, 2008

OVER BLACK

We listen to the immortal music of Mozart's Adagio of the Clarinet Concerto in A.

FADE UP

EXT. THE SOLAR SYSTEM

Space, infinite and empty.

But then, slowly all nine planets of our Solar System move into frame and align.

The last of them is the giant, burning sphere of the sun.

Just as the sun enters frame, a solar storm of gigantic proportion unfolds. The eruptions shoot thousands of miles into the blackness of space.

2009

FADE UP

EXT. COUNTRY SIDE/INDIA - SUNSET

Mozart's concerto filters from a jeep's stereo, fighting the drumming sounds of the monsoon rain. PROF. FREDERIC WEST, 66, listens to the music.

An Indian BOY playing by the roadside steers his wooden toy ship across a puddle.

The Professor turns to his driver, pointing to the boy.

PROF. WEST
Watch out!

But it's too late. The jeep drives straight through the puddle at full speed, sinking the boy's toy ship.

In the background, the jeep stops in front of a building. The driver jumps out, leading the Professor towards its entrance.

The sign at the door reads: `Institute for Astrophysics - University of New Delhi'.

2.

INT. NAGA-DENG MINE/INDIA - SUNSET

An endless mine shaft. An old elevator cage comes to a grinding halt. When Prof. West steps out we see that he is accompanied now by a nervous DR. SATNAM TSURUTANI, 32.

Figure 2: The beginning of a sample script

5.2 Summary Statistics

5.2.1 Summary Statistics of Movie Scripts and Movie Metadata

Release dates of the movie scripts in our sample range from April, 21, 1928 the earliest to November, 14, 2014 the latest. Figure 3 shows the distribution of release dates of our movie scripts. Across the 1088 scripts in our sample, multiple genres can be assigned to one movie (for instance, *Godfather* belongs to Crime and Drama; *Wall-E* belongs to Animation, Adventure, Comedy, Drama, Family, Romance and Sci-Fi.) and on average each movie is assigned 2.67 genres. The genre distribution is shown in Figure 4. Over 86% of the scripts in our sample was authored by one or two screenwriters. The maximum number of screenwriters associated with one script is 6 in our sample and the average is 1.66. The average movie script contains 23051.98 words after removing English stopwords (e.g., “the”,

"go", "we", etc.) and names[2]. The longest script of over 51,000 words in our sample is *JFK* released in 1991, written by Oliver Stone and Zachary Sklar (based on books by Jim Marrs and Jim Garrison), and the shortest of fewer than 2,000 words is *The Things My Father Never Taught Me* written by Burleigh Smith. We use aggregate movie ratings as a proxy for audience experienced utility levels, averaging IMDB ratings and "Tomatometer" from RottenTomatoes.com and rescaling to the range of 0 to 100, the distribution of which is shown in histogram 5. Descriptive statistics of selected variables are provided in table 3 in the specification section 6.3.

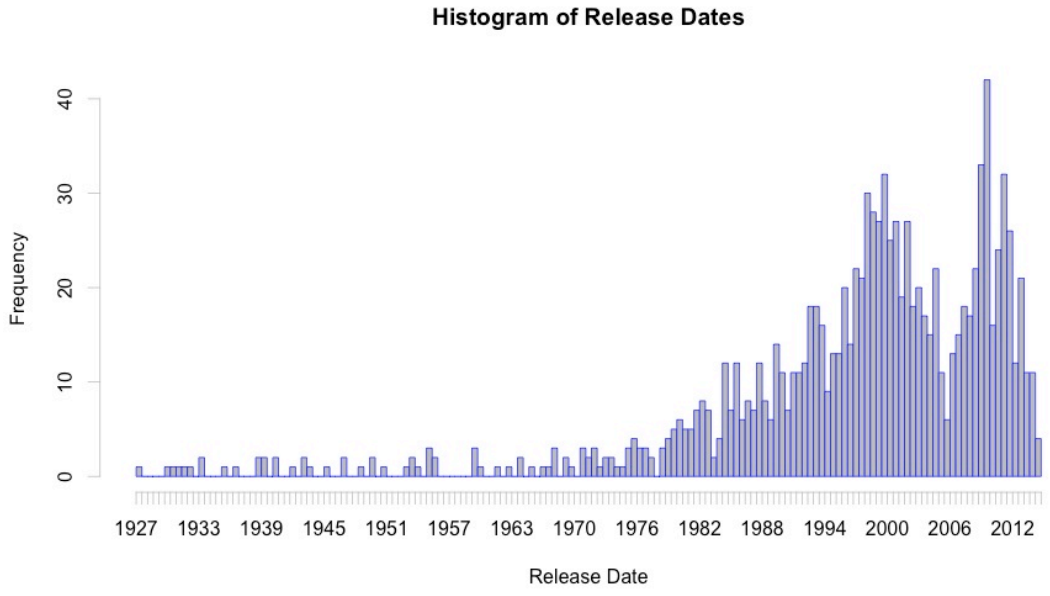


Figure 3: A histogram of script release dates

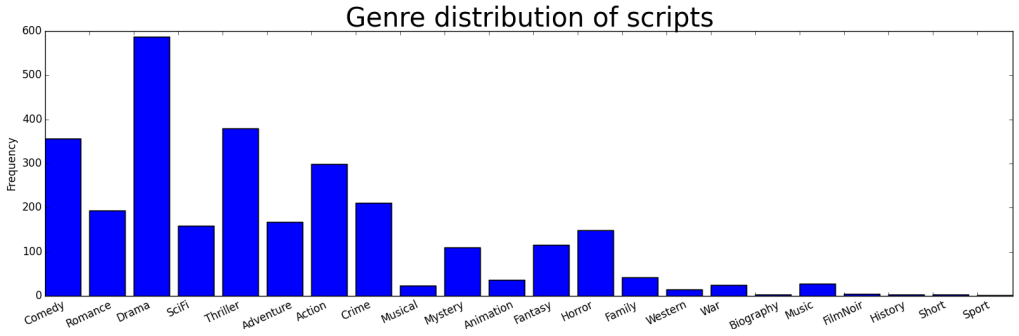


Figure 4: A histogram of script genres

[2]We included common names and main character names in the stopword list

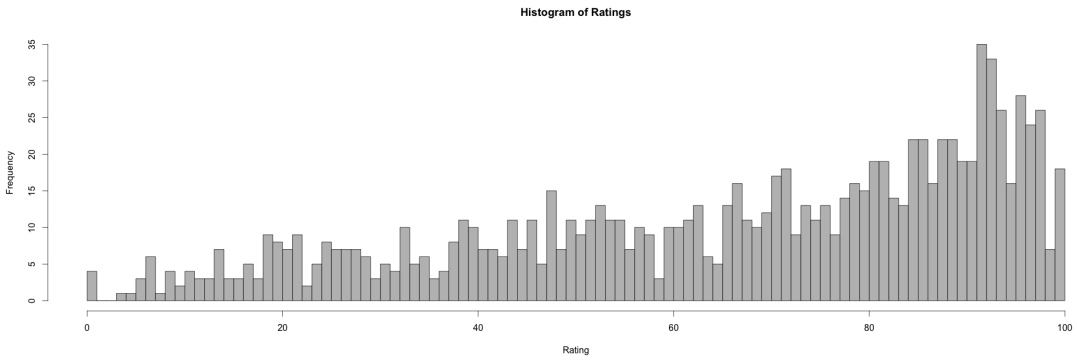


Figure 5: A histogram of movie ratings

6. Empirical Analysis

6.1 Information Content Measures

We measure the information content of a movie (an information good) over time in two aspects, through which the focal good affects the viewing experience of consumers: behavioral and cognitive. Behavioral information refers to the tangible actions of characters in the story that can be explicitly observed in movie scenes and reflected in scripts mostly through "linguistically Transitive" clauses (details in section 6.1.1). Emotional information refers to the intangible affects and feelings of characters in the story that are reflected in scripts mostly through monologues and dialogues.

The behavioral measure and emotional measure also test for different aspects of the theory, in the sense that the behavioral measure corresponds to the sequence of signals chosen by the principal whereas the evoked emotional measure corresponds to the audience's belief path (belief realization) after forming beliefs induced by signals sent by the principal.

We draw on concepts and techniques from computational linguistics and natural language processing to operationalize these two distinct measures, and analyze their main effects and potential interaction effects.

6.1.1 Behavioral Measure: Clause Transitivity

The problem of detecting information revelation from movie scripts is closely related to existing work in NLP and CL that has looked at automatically detecting spoilers in social media to prevent potential sabotage of the joy of an entertainment experience. These studies

have used computational methods including manual annotation and supervised classification (Guo and Ramakrishnan 2010), filtering heuristics (Golbeck 2012), supervised learning based on crowdsourced datasets (Boyd-Graber, Glasgow and Zajac 2013).

Our behavioral measure builds on Boyd-Graber et al. (2013)'s discussion of spoiler detection in the sense that, more broadly speaking, the problem of detecting information revelation from text is connected to core problems in linguistics and natural language processing: Transitivity. Distinctive from *grammatical transitivity* (whether a verb takes a direct object or not), *Transitivity* (with a capital letter) is a linguistic property of a clause that measures how impactful, deliberate and complete the action it describes is. Hopper and Thompson (1980) identify a number of components of Transitivity, only one of which is the presence of an object of the verb, that are concerned with the effectiveness with which an action takes place, e.g., the punctuality and telicity of the verb, the conscious activity of the agent, and the referentiality and degree of affectedness of the object. For instance, an action such as "kill" is much more Transitive than an action like "think", because "kill" is kinetic, completed, volitional, done with agency, actual, and greatly affects its object, whereas "think" is far less telic (without a clear start or end point), nor is it certain to affect its object or done with conscious volition. "Transitive actions advance the narrative, causally linking actors, actions and outcomes in recognizable schemata" (Schank and Abelson 1977). To our best knowledge, there is currently no accessible dictionary of linguistically Transitive features, which might be due to the ambiguous definition that makes annotation tasks difficult. Madnani, Boyd-Graber and Resnik (2010) documents the first attempt in the CL/NLP community that crowdsourced annotations of Transitivity for Wikipedia pages from Amazon Mechanical Turk (mTurk).

We obtained annotated datasets collected from online social media as in Madnani et al. (2010) and mTurk as in Boyd-Graber et al. (2013) — each observation is a movie plot description with a tag "spoiler" or "non-spoiler" — to train a linear-kernel support vector machine (SVM) (Cortes and Vapnik 1995, Joachims 1997, Joachims 1998, Joachims 2002). We represent sentences as points in a d-dimensional vector space where each dimension represents one feature (uni-gram, stems, bigrams, etc.), and learn a SVM predictor. This produces a function f that takes an arbitrary sentence and outputs whether it reveals information information (is a spoiler) or not. We tune the cost parameter and the number of active features for recursive feature selection using grid search. We multiply the vector of support vector (SV) coefficients by the SV-term matrix (intuitively, support vectors are "important"

data-points/scripts for prediction) to obtain the vector of term coefficients, which can be interpreted as the importance of each term for the prediction task due to kernel linearity. Features with the highest coefficients when training with stems of uni-grams are presented in Table 1. We construct a dictionary of linguistic Transitivity assigning corresponding term coefficients to words as their "linguistic Transitivity score". Features that are overly specific to movie or TV series such as episode, season, season finale, etc. are removed after manual inspection.

Rank	Feature	Rank	Feature
1	end	26	daughter
2	turn	27	brother
3	death	28	fall
4	kill	29	gun
5	dead	30	victim
6	father	31	final
7	reveal	32	averted
8	child	33	break
9	real	34	alive
10	shot	35	fight
11	subverted	36	dying
12	die	37	god
13	play	38	married
14	trope	39	escape
15	save	40	evil
16	family	41	doctor
17	finally	42	wanted
18	killer	43	due
19	wife	44	show
20	universe	45	suicide
21	shoot	46	entire
22	eventually	47	sex
23	girl	48	cut
24	murder	49	return
25	start	50	heart

(a) 50 Most Informative Features

Rank	Feature	Rank	Feature
1	abc	26	admire
2	abate	27	adrenaline
3	aboriginal	28	adulthood
4	absentia	29	advance
5	absinthe	30	adverted
6	absorb	31	advertise
7	abstemious	32	advise
8	absurdaly	33	affected
9	abundant	34	afterall
10	accelerator	35	afterschool
11	accented	36	aesthetic
12	accentuate	37	aimed
13	acceptable	38	aerial
14	accommodation	39	airline
15	accompaniment	40	airport
16	accompanying	41	aladdin
17	accounting	42	album
18	accuracy	43	alumnus
19	accosted	44	alright
20	acre	45	allergic
21	actually	46	alligator
22	acrimoniously	47	allied
23	adapt	48	allegedly
24	adept	49	allude
25	adjusted	50	almighty

(b) 50 Least Informative Features (Tied)

Table 1: Most and Least Informative Textual Features

After obtaining the dictionary of Transitive features, we divide each script into a temporal sequence of small chunks, each of which consists of 500[3] words. We tally the number of

[3]Multiple window sizes ranging from 200 words to 600 words have been used, all yielding rather similar results in section 6

Transitive features in all chunks and construct a panel dataset of Transitivity evolution for all the movie scripts. Figure 6 show such Transitivity patterns of three (randomly chosen) movie scripts in our sample.

Let $TRANS_t$ denote number of Transitivity features in period t, a proxy for the amount of information revelation in period t — the amount of private information of the principal released to the audience in period t. We calculate the empirical level and spikiness of revealed information over time by mean and standard deviation, denoted as $TRANS_{AVG}$ and $TRANS_{SD}$, respectively. Note that $TRANS_{SD}$ relates directly to the spikiness of the information revelation pattern manifest in section section 4.

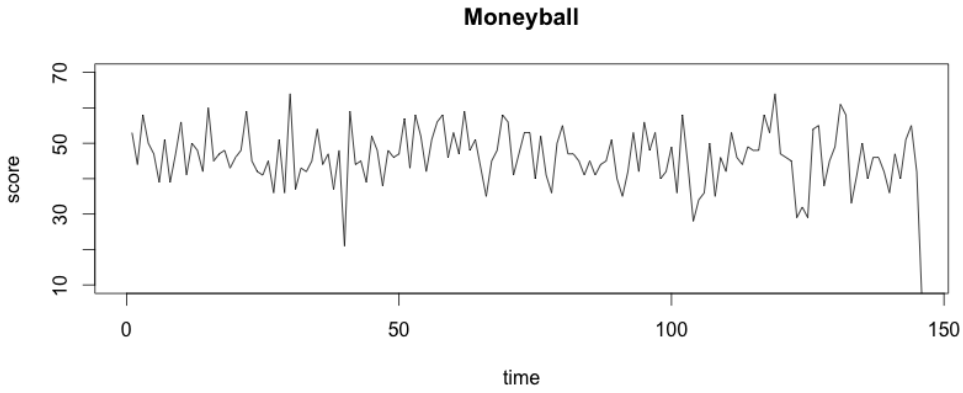
6.1.2 Cognitive Measure: Evoked Emotions

Following Proposition 1 in Kamenica and Gentzkow (2011) and lemma 1 in Ely et al. (2015), we assume further, that there exists a one-to-one mapping between audience's beliefs $\tilde{\mu}$ of the state of the world ω and his evoked emotions denoted as γ:

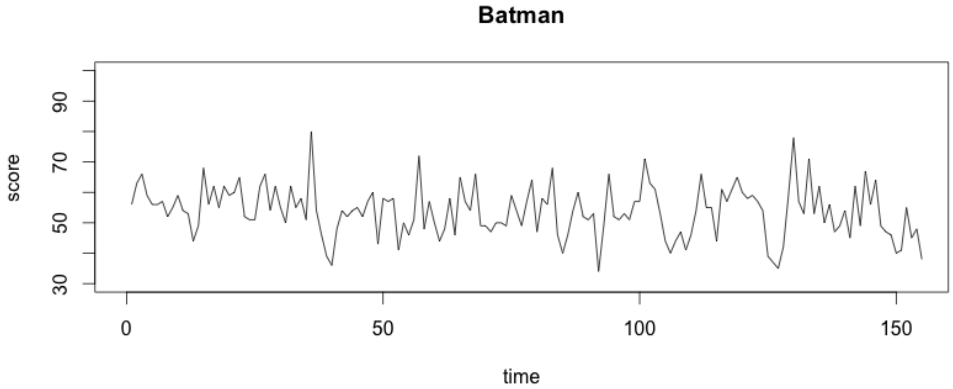
Assumption 3 *There exists a bijective mapping $\phi : \Delta(\Delta(\Omega)) \to \Gamma$ from beliefs of the state of the world $\tilde{\mu}$ to evoked emotions γ.*

By assumption 3, we measure the evoked emotion trajectories from movies as a proxies for belief paths of audiences. To measure evoked emotional contents as revealed information in scripts that affects consumer viewing experience, we use one of the largest evoked emotion lexicons created using Amazon Mechanical Turk (Mohammad and Turney 2010) by focusing on the eight emotions proposed by Plutchik (1980) — joy, sadness, anger, fear, trust, disgust, anticipation and surprise — which comprise the other commonly used framework of six basic emotions (joy, sadness, anger, fear, disgust and surprise) introduced by (Ekman 1992). Figure 7 display evoked emotions patterns of two (randomly chosen) movie in our sample. Different lines and colors represent different emotions and valences: anger, anticipation, disgust, fear, joy, trust, negative, positive and valence (positive − negative).

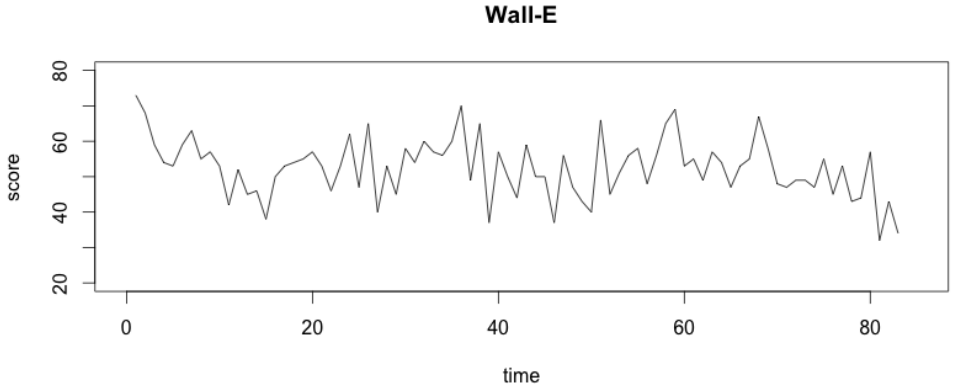
Let $BELIEF_{i,t}$ denote our measure of evoked emotion i in period t as detailed above, and $BELIEF_t$ denote the vector of realized beliefs of states in period t, which we use as a proxy for audience belief paths in period t. According to Equation 4, we take the Euclidean distance evoked emotion vectors $BELIEF_t$ between adjacent time points, written as $\triangle BELIEF_t$. We measure the level and spikiness of belief paths by mean and standard deviation, denoted



(a) Transitivity Trajectory of Movie Moneyball

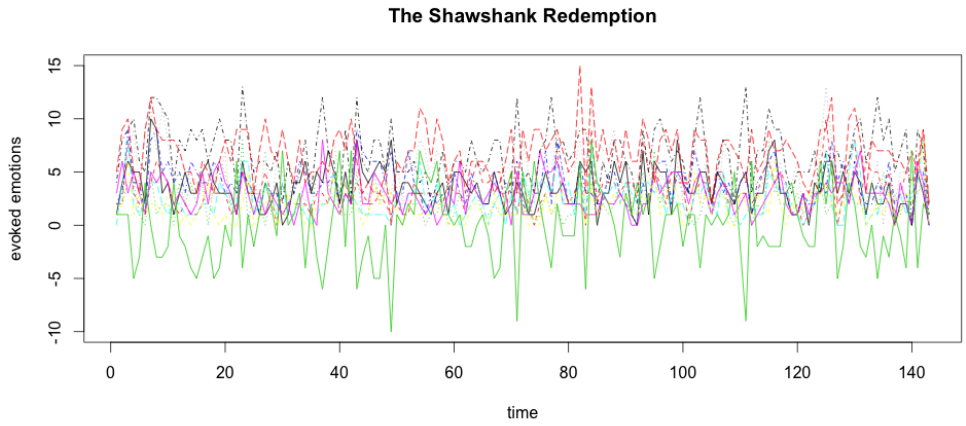


(b) Transitivity Trajectory of Movie Batman

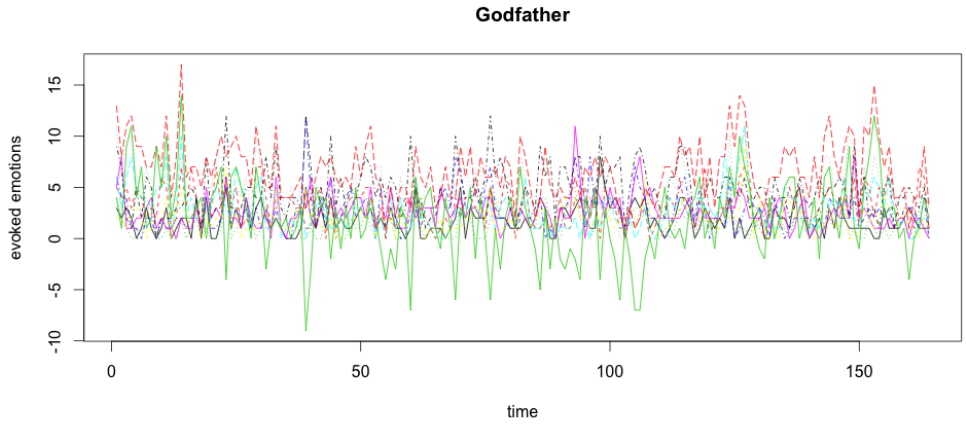


(c) Transitivity Trajectory of Movie Wall-E

Figure 6: Transitivity Time Series of Two Movies



(a) Evoked Emotion Trajectories of The Shawshank Redemption



(b) Evoked Emotion Trajectories of Godfather

Figure 7: Evoked Emotion Plots of Two Movies

as $\triangle BELIEF_{AVG}$ and $\triangle BELIEF_{SD}$, respectively. Note that $\triangle BELIEF_{AVG}$ directly corresponds to the audience experienced utility level associated with surprises according to Equation 4. We derive the same metrics by processing only the beginning of each script (the first 10% of each script) and denote them as $\triangle BELIEF0_{AVG}$ and $\triangle BELIEF0_{SD}$.

6.2 Validation

6.2.1 A Test of the Martingale Assumption

We test if our measured audience beliefs (evoked emotions) violate the martingale assumption using a statistical test introduced by Park and Whang (2005) (further modified by

Phillips and Jin (2014)), which is designed to test if a given time series is a martingale process against certain non-martingale alternatives. The class of alternative non-martingale processes against which the test has power is more general than that for some existing spectral-based tests (Durlauf 1991), and the test is not sensitive to any smoothing parameters as is in Hong (1999). Further, when it comes to martingale within the class of first-order Markovian processes, which applies to our setting, the proposed test is simple to implement.

The null hypothesis of interest is that a given time series y_t is a martingale process with respect to the filtration $\mathcal{F}_t$, which is taken to be the natural filtration of y_t — the sigma-field generated by all previous observations, i.e.,:

$$H_0 : \mathrm{P}(\mathrm{E}(y_t|\mathcal{F}_{t-1}) = y_{t-1}) = 1 \tag{5}$$

for each $t \geq 1$.

We implement the Cramer-von Mises type statistic given by

$$T_n = \int Q_n^2(x)\mu_n(dx) \tag{6}$$

where $Q_n(x)$, the basis of the test statistic is given by

$$Q_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \triangle y_t 1\{y_{t-1} \leq x\}. \tag{7}$$

We take y_t to be each of the time series of evoked emotions for all movies in our sample. For 94.13% of all trajectories (movie $\times$ evoked emotions), it fails to reject the null hypothesis of martingale assumption at a significance level of 5%.

6.2.2 Face Validity with Online Reviews

To establish face validity with human raters, we scraped all the top movie critics' reviews for each movie in our sample available at RottenTomatoes.com. We automatically identify reviews mentioning suspense concepts (e.g., "suspense", "suspenseful", "tension", "gripping", etc.) in reviews and count the number of such reviews for each movie. If there exist more than 2 such reviews (written by "top movie critics") for a movie, we assign the movie a suspense label.

For instance, the movie with the lowest $TRANS_{SD}$ score (lower $TRANS_{SD}$ relates to more even information revelation over time, thus more suspense according to the theory) in our sample turns out to be *Frozen River*, of which the top critics' reviews that we use to construct "human labels" are shown in Figure 8[4]. This movie was assigned a Suspense label because multiple top critics mentioned "suspense" in their reviews: "All in all, Frozen River is gripping stuff." — David Edelstein (New York Magazine/Vulture); "An impressive first feature by writer/director Courtney Hunt, Frozen River boasts considerable suspense-movie tension and a compelling emotional journey for its foreground characters."— Colin Covert (Minneapolis Star Tribune); etc.

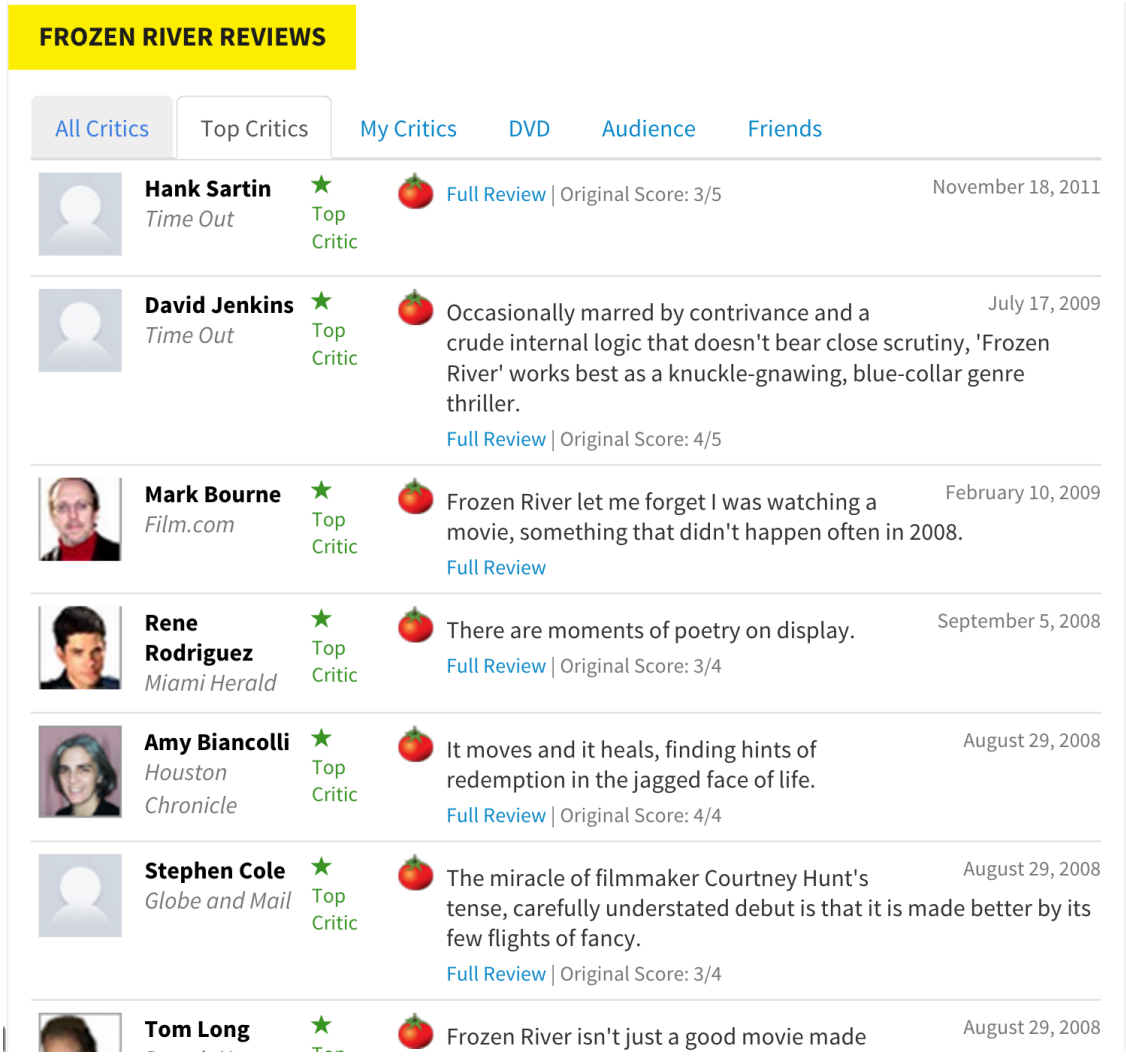


Figure 8: Top Critics' Reviews of Frozen River

[4]available at http://www.rottentomatoes.com/m/frozen_river/reviews/?type=top_critics, last accessed: Nov 11,2015

Meanwhile we construct another set of suspenseful movies by sorting them according to the value of $TRANS_{SD}$ and assign movies with a value lower than an αth percentile a Suspense label. The Cohen's Kappa Coefficient for labels derived from top critics' online reviews and the ones derived from our method reaches 53.4% when $\alpha = 20$. With regards to surprise, the same process applies with a threshold (cut from below as opposed to from above for suspense) of βth percentile based on the value of $\triangle BELIEF_{SD}$ to assign surprise labels based on our method. The corresponding Cohen's Kappa Coefficient reaches 51.6% when $\beta = 30$. Given the ambiguity and subjectivity of suspense and surprise evaluations and the uncertainty of being mentioned in critics' reviews, these figures indicate a moderate to substantial agreement between our method and "human raters" (annotations from top movie critics active in the RottenTomatoes online community).

6.3 Specification

To examine the relationship between information revelation patterns and audience experienced utility (proxied by aggregate audience ratings), we can specify the model as follows with a vector of control variables $CTRL_j$ for movie j,

$$\log(Y_j) = f(U_{susp}(TRANS_j), U_{surp}(\Delta BELIEF_j), CTRL_j) + \epsilon_j \tag{8}$$

$$Y_j = \{\text{Rating}_j, \text{Volume}_j\} \tag{9}$$

where $TRANS_j$ is a vector of information revelation measures including $TRANS_{AVG}$ and $TRANS_{SD}$ of movie j as detailed in section 6.1.1, and $\triangle BELIEF_j$ is a vector of our measures for audience belief paths including $\triangle BELIEF_{AVG}$, $\triangle BELIEF0_{AVG}$, $\triangle BELIEF0_{SD}$ and $\triangle BELIEF_{SD}$ of movie j as detailed in section 6.1.2. We take the dependent variable to be either the aggregate audience ratings Rating_j, or the total number of audience reviews Volume_j.

The control variables are as follows. On the product (movie/information good) level, we follow the literature (Narayan and Kadiyali 2015, Brown, Camerer and Lovallo 2013, Liu, Mazumdar and Li 2015, Luo 2014, etc.**?**) and control for production budget, running time, screenwriter, director, actor, genre, seasonality and era. With regards to screenwriters, directors and actors, we count both the total number involved in the production process

($WRICOUNT$, $DIRCOUNT$, $ACTCOUNT$) and the number of well-recognized individuals involved according to ranked lists of prominent screenwriters, directors and actors contributed by experienced users on IMDB.com[5] ($TOPWRI$, $TOPDIR$, $TOPACT$). We include dummy variables for genres ($COMEDY$, $DRAMA$, $ROMANCE$, $SCIFI$, $CRIME$, $THRILLER$, $ADVENT$, $MUSICAL$, $ANIMA$, $FANTASY$, $FAMILY$, $HORROR$, WAR, $WESTERN$, etc.), months and decades of the release date (JAN, FEB, MAR, ..., $70s$, $80s$, $90s$, ...). On the script level, we control for script length (total number of words: $LENGTH$ and total number of sentences: $SENTENCE$) and positive/negative sentiments (POS, NEG), following previous literature (Eliashberg et al. 2007, Eliashberg et al. 2014).

We start with functional forms specified by OLS, GLM (Poisson regression). We observe ratings are overly clustered at the higher end of the scale from 0 to 100 and the shape resembles a normal distribution censored from above — critics could only rate a movie at 100 even if they raved about it and would love to rate it well above the ceiling — therefore, we also fit a Tobit model to the data. All the details are provided in section 7.

Variables of interest (independent, dependent and control variables) are summarized in table 2. Descriptive statistics for selected variables are summarized in Table 3.

7. Results

We report preliminary results by specifying three functional forms of equation eq. (8): (1) OLS, with the dependent variable being consumer ratings; (2) GLM (Poisson), with the dependent variable being the volume of online review; (3) Tobit (censored from above), with the dependent variable being consumer ratings; shown in table 4 (created by *stargazer* (Hlavac 2015)).

We examine the relationship between the audiences' experienced utility (proxied by aggregate ratings) and information revelation strategies of the producers (manifest in information revelation and belief path trajectories) by estimating Equation 8. In Table 4 our measures for information revelation patterns ($TRANS_{AVG}$, $TRANS_{SD}$) and belief realizations (belief paths) ($\triangle BELIEF_{SD}$, $\triangle BELIEF0_{SD}$) have significant effects in all three models. Recall that $TRANS_{SD}$ relates to the spikiness of information revelation patterns

[5]Examples include http://www.imdb.com/list/ls059866155/, http://www.imdb.com/list/ls054809794/, etc.

Variable Name	Label	Source
$PROD$	production budget	The-Numbers.com
$RUNTIME$	running time	RottenTomatoes.com
$WRICOUNT$	screenwriter count	IMSDB
$TOPWRI$	top screenwriter count	IMDB
$DIRCOUNT$	director count	IMDB
$TOPDIR$	top director count	IMDB
$ACTCOUNT$	actor count	IMDB
$TOPACT$	top actor count	IMDB
$SEASON$	dummies for season based on release month	IMSDB
$DECADE$	dummies for decade based on release year	IMSDB
$COMEDY$	dummy variable for comedy	IMSDB
$ROMANCE$	dummy variable for romance	IMSDB
$DRAMA$	dummy variable for drama	IMSDB
$SCIFI$	dummy variable for sci-fi	IMSDB
$THRILLER$	dummy variable for thriller	IMSDB
$ADVENT$	dummy variable for adventure	IMSDB
$CRIME$	dummy variable for crime	IMSDB
$MUSICAL$	dummy variable for musical	IMSDB
$MYSTERY$	dummy variable for mystery	IMSDB
$ANIMA$	dummy variable for animation	IMSDB
$FANTASY$	dummy variable for fantasy	IMSDB
$FAMILY$	dummy variable for family	IMSDB
$LENGTH$	script length	IMSDB
$\triangle BELIEF_{AVG}$	mean of proxy for belief path	detailed in section 6.1.2
$\triangle BELIEF_{SD}$	std. dev. of proxy for belief path	detailed in section 6.1.2
$\triangle BELIEF0_{AVG}$	mean of proxy for starting belief path	detailed in section 6.1.2
$\triangle BELIEF0_{SD}$	std.dev. of proxy for starting belief path	detailed in section 6.1.2
$TRANS_{AVG}$	mean of proxy for information revelation	detailed in section 6.1.1
$TRANS_{SD}$	std. dev. of proxy for information revelation	detailed in section 6.1.1
$ANTICIPATE$	evoked anticipation	detailed in section 6.1.2
JOY	evoked joy	detailed in section 6.1.2
SAD	evoked sadness	detailed in section 6.1.2
$SURPRISE$	evoked surprise	detailed in section 6.1.2
$TRUST$	evoked trust	detailed in section 6.1.2
NEG	negative valence	detailed in section 6.1.2
POS	positive valence	detailed in section 6.1.2
$EMOTIONS$	sum of std. dev. of all the evoked emotions	detailed in section 6.1.2
RATING	proxy for audience utility	RottenTomatoes.com & IMDB

Table 2: Specification Table

measured by the number of "clause Transitive" features over time. Thus, the positive effect of $TRANS_{SD}$ ($p < 0.01$) suggests that the spikier the information revelation pattern is,

Statistic	N	Mean	St. Dev.	Min	Max
NEG	1,088	2.565	0.362	1.575	4.143
POS	1,088	2.708	0.334	1.609	4.880
$\triangle BELIEF_{SD}$	1,088	3.227	0.476	2.003	5.072
$\triangle BELIEF_{AVG}$	1,088	6,267.014	1,803.489	274	17,243
$LENGTH$	1,088	114.375	28.747	1	256
$IMDB$	1,088	6.977	1.002	2.300	9.300
$TOMATOMETER$	1,042	66.648	25.831	0	100
$WRICOUNT$	1,088	1.662	0.855	1	6
$TOPWRI$	1,088	0.132	0.394	0	2
$RUNTIME$	1,088	112.392	22.837	19	300
$TOPACT$	1,088	2.941	2.282	0	14
$TOPDIR$	1,088	0.292	0.481	0	3
$PROD$	1,088	38,020,970.000	43,593,501.000	7,000	425,000,000
$\triangle BELIEF0_{SD}$	1,088	2.226	0.382	1.233	3.912
$\triangle BELIEF0_{AVG}$	1,088	649.021	207.753	23	1,931
$TRANS_{AVG}$	1,088	27.290	4.915	8.0013	58.909
$TRANS_{SD}$	1,088	7.672	1.609	4.252	18.237
$COMEDY$	1,088	0.327	0.469	0	1
$ROMANCE$	1,088	0.177	0.382	0	1
$DRAMA$	1,088	0.540	0.499	0	1
$SCIFI$	1,088	0.145	0.352	0	1
$THRILLER$	1,088	0.348	0.477	0	1
$ADVENT$	1,088	0.275	0.447	0	1
$CRIME$	1,088	0.193	0.395	0	1
$MUSICAL$	1,088	0.021	0.144	0	1
$MYSTERY$	1,088	0.100	0.300	0	1
$ANIMATION$	1,088	0.033	0.179	0	1
$FANTASY$	1,088	0.107	0.309	0	1
$20s30s40s$	1,088	0.022	0.147	0	1
$50s60s$	1,088	0.027	0.161	0	1
$70s$	1,088	0.041	0.199	0	1
$80s$	1,088	0.127	0.333	0	1
$90s$	1,088	0.296	0.457	0	1
$00-04$	1,088	0.178	0.383	0	1
$05-10$	1,088	0.167	0.373	0	1
$10-14$	1,088	0.142	0.349	0	1

Table 3: Descriptive Statistics for Selected Variables

the higher the aggregate audience rating is, and the higher the proxied expected audience experienced utility gets, holding other variables fixed. $\triangle BELIEF_{SD}$ and $\triangle BELIEF0_{SD}$

are our spikiness measures for belief paths overall and at the beginning, respectively, both of which have significant effects under all three model specifications. The positive significant effect of $\triangle BELIEF0_{SD}$ indicates that the spikier the belief paths are, the higher the aggregate audience rating for the information good, and the greater the expected utility level an audience experiences.$\triangle BELIEF0_{SD}$ is negatively associated with the aggregate audience rating, which means that the less spiky the belief paths are at the beginning (when there are many periods remaining), the higher the aggregate audience rating is, and the higher the proxied expected utility level an audience experiences. In addition, we also included measures of total evoked suspense and total evoked anticipation based on feature counts according to NRC emotion lexicons (Mohammad and Turney 2010, Mohammad and Turney 2013) detailed in section 6.1.2 as $SURPRISE$ and $ANTICIPATION$. In all three models, $SURPRISE$ exhibits a significant positive effect on the aggregate audience rating whereas $ANTICIPATION$ is not significant with its coefficient literally equal to zero.

8. Additional Analysis for Robustness

8.1 Selection

8.1.1 Selection on Observables

Though our behavioral and cognitive measures for suspense and surprise are derived from recent advances microeconomic theories, it is still possible that observable differences between movies could drive differences in our measures. We therefore ran analyses to determine if other observables (genre, budget, screenwriter count, director count, top director count, actor count, top actor count, top screenwriter count, seasonality, year effect, runtime, script length, etc.) could predict our NLP measures. We found that these variables did not significantly predict our behavioral and cognitive measures for suspense and surprise in a logit model, as is shown in Table 6.

We also conducted a matching analysis using a "kernel matching estimator" (Heckman and Ichimura 1998) generalized to continuous treatment regimes (Imai and Dyk 2012, Fong and Imai 2014). We matched observations on other observable characteristics and computed the effect of our behavioral and cognitive measures for suspense and surprise on consumer evaluations. The results of the analysis is consistent with preliminary analysis in Table 4

and is shown in Table 6, showing that the suspense estimate is significant whereas the surprise estimate less so. We also replicated the result using a "nearest neighbor" matching estimator (Heckman, Ichimura, Smith and Todd 1998, Abadie and Imbens 2006) under generalized treatment regimes (Imai and Dyk 2012, Fong and Imai 2014), which produced a significant estimate for the suspense measure. The results were also insensitive to the choice of kernel. Taken together, these analyses suggest that observable differences do not predict our behavioral and cognitive measures for suspense and surprise derived from micro-economic theories and natural language processing techniques.

8.2 Endogeneity

Though our behavioral and cognitive measures are derived from economic and linguistic theories that most screenwriters are oblivious to, one could potentially argue that experienced screenwriters might anticipate consumers reactions to their particular choices of words and the distributions thereof within scripts and strategize in response as the stories unfold in their minds. This "rational expectation" idea of screenwriters as they compose justify concerns for endogeneity issues that could arise in this context. We present two strategies below to correct for potential biases introduced by such endogeneity. Though when we apply the Hausmen Test for endogeneity to the proposed suspense and surprise metrics, the H statistic is smaller than the corresponding critical value of chi-squared distribution of degree 2 ($3.785 < 5.991$), indicating that we fail to reject the null hypothesis that the regressor is exogenous.

8.2.1 Subsampling on screenwriters

We focus on a sub-sample of independent film scripts and screen-writers known for being untraditional and never outlining (for instance, the Coen brothers, Scott Neustadter, Michael H. Weber). Within this sub-sample the reverse link between storyline and consumer reaction through screenwriters' rational expectation that justifies the endogeneity issue becomes non-existent. We show in Table 7 that results from this subsample are consistent with results from the whole sample.

8.2.2 Instrument with Writers Guild Strikes

We introduce dummy variables for the period around which the Writers Guild of America (WGA) went on strike: 1960, 1981, 1985, 1988, and 2007-2008. If a movie script is dated within the periods 1959-1960, 1980-1981, 1984-1985, 1987-1988, or 2006-2008 and at least one of the screenwriters are members of the Writers Guild of America (WGA), the dummy variable $Strike_i$ for movie script i is set to 1, and 0 otherwise. It is orthogonal to consumer evaluations after the movie has been released and it is reasonably correlated with the writing processes of affected screenwriters. When regressed on the strike dummy variable, the F-statistics for the proposed suspense and surprise metrics are 11.91 and 8.14, respectively. By the rule of thumb of thresholding at the value of 10, the strike dummy variable is marginally weak for the surprise metric and yet less so for the suspense metric.

9. Conclusion

We empirically examine the relationship between consumer experienced utilities and information revelation patterns based on a consumer demand model for non-instrumental information goods proposed in Ely et al. (2015) that describes how suspense and surprise (and possibly other aspects of belief dynamics) shape the experienced utility function of consumers. Drawing on the linguistic concept of "clause Transitivity", the recent advancement in emotion analysis — NRC evoked emotion lexicon and machine learning techniques for classification tasks, we attempt to teach computers to understand syuzhet (storyline) with respect to suspense and surprise. We identify strong(er) empirical evidences in our sample of 1088 movie scripts and corresponding consumer data in support of consumer preferences for surprise (than for suspense). One potential interpretation might be grounded in the idea expressed in Neil Postman's *Amusing Ourselves to Death: Public Discourse in the Age of Show Business* in the sense that the faculties requisite for rational inquiry are weakened by televised viewing and consumers therefore respond more to surprise than suspense which requires greater cognitive effort on the part of consumers. This work represents (one of) the first step(s) in understanding the drivers of consumer demand for non-instrumental information goods empirically and its results and implications may well be applicable in settings other than movie industry, such as the design of pitch video clips for start-ups, how to optimally tell a story in a commercial to suspense-seeking or surprise-seeking consumers,

etc.

Possible extensions (or limitations) of the current study include deriving the optimal information revelation strategies when consumers have a preference for both suspense and surprise (specified by respective weights) and testing the corresponding analytical predictions. Other publicly available datasets such as public speech scripts and venture capital or crowd-funding pitching scripts along with respective outcome measures can be tested on to strengthen (the external validity of) or falsify the present consumer demand theory for information goods.

References

Abadie, A and GW Imbens, "Large sample properties of matching estimators for average treatment effects," *Econometrica*, 2006.

Adamopoulos, Panagiotis, Anindya Ghose, and Vilma Todri, "The Business Value of the Internet-of-Things: Evidence from an Online Retailer," 2017.

Archak, Nikolay, Anindya Ghose, and Panagiotis G Ipeirotis, "Deriving the pricing power of product features by mining consumer reviews," *Management Science*, 2011, *57* (8), 1485–1509.

Athey, Susan, Markus M Mobius, and Jenő Pál, "The impact of aggregators on internet news consumption," 2017.

Bakos, Y and E Brynjolfsson, "Bundling information goods: Pricing, profits, and efficiency," *Management Science*, 1999.

___ **and** ___ , "Bundling and Competition on the Internet," *Management Science*, 2000.

Baumgartner, H, M Sujan, and D Padgett, "Patterns of affective reactions to advertisements: The integration of moment-to-moment responses into overall judgments," *Journal of Marketing Research*, 1997.

Bilandzic, Helena and Rick Busselle, "Narrative persuasion," *The Sage handbook of persuasion: Developments in theory and practice*, 2013, pp. 200–219.

Boyd-Graber, Jordan, Kimberly Glasgow, and Jackie Sauter Zajac, "Spoiler Alert: Machine Learning Approaches to Detect Social Media Posts with Revelatory Information," in "ASIST 2013: The 76th Annual Meeting of the American Society for Information Science and Technology" 2013.

Brown, Alexander L, Colin F Camerer, and Dan Lovallo, "Estimating structural models of equilibrium and cognitive hierarchy thinking in the field: The case of withheld movie critic reviews," *Management Science*, 2013, *59* (3), 733–747.

Calin, Mihai, Chrysanthos Dellarocas, Elia Palme, and Juliana Sutanto, "Attention allocation in information-rich environments: The case of news aggregators," 2013.

Calzada, Joan and Ricard Gil, "What Do News Aggregators Do? Evidence from Google News in Spain and Germany," 2017.

Caro, Felipe and Victor Martínez de Albéniz, "Managing Online Content to Build a Follower Base," 2018.

Chiou, Lesley and Catherine Tucker, "Paywalls and the demand for news," *Information Economics and Policy*, 2013, *25* (2), 61–69.

Cortes, Corinna and Vladimir Vapnik, "Support-vector networks," *Machine learning*, 1995, *20* (3), 273–297.

Durlauf, SN, "Spectral based testing of the martingale hypothesis," *Journal of Econometrics*, 1991, (50).

Ekman, P, "An argument for basic emotions," 1992.

Eliashberg, Johoshua, Sam K Hui, and John Z. Zhang, "From Story Line to Box Office: A New Approach for Green-Lighting Movie Scripts," *Management Science*, 2007, *6* (53), 881–893.

____ , ____ , and ____ , "Assessing Box Office Performance Using Movie Scripts: A Kernel-Based Approach," *IEEE Transactions On Knowledge And Data Engineering*, 2014, *11* (26), 2639–2648.

Elpers, JLCMW and A Mukherjee, "Humor in television advertising: A moment-to-moment analysis," *Journal of Consumer Research*, 2004.

Elpers, Josephine LCM Woltman, Michel Wedel, and Rik GM Pieters, "Why Do Consumers Stop Viewing Television Commercials? Two Experiments on the Influence of Moment-to-Moment Entertainment and Information Value," *Journal of Marketing Research*, 2003, *40* (4), 437–453.

Ely, Jeffrey, Alexander Frankel, and Emir Kamenica, "Suspense and Surprise," *Journal of Political Economy*, 2015, *123* (1), 215–260.

Escalas, Jennifer Edson, "Narrative processing: Building consumer connections to brands," *Journal of consumer psychology*, 2004, *14* (1-2), 168–180.

Fong, C and K Imai, "Covariate balancing propensity score for general treatment regimes," *Princeton Manuscript*, 2014.

Geng, X and MB Stinchcombe, "Bundling information goods of decreasing value," *Management Science*, 2005.

Ghose, Anindya and Panagiotis G Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE Transactions on Knowledge and Data Engineering*, 2011, *23* (10), 1498–1512.

___ **and Sang Pil Han**, "An empirical analysis of user content generation and usage behavior on the mobile Internet," *Management Science*, 2011, *57* (9), 1671–1691.

Golbeck, J, "The twitter mute button: a web filtering challenge," 2012.

Guo, S and N Ramakrishnan, "Finding the storyteller: automatic spoiler tagging using linguistic cues," in "Proceedings of the 23rd International Conference on Computational Linguitics" 2010.

Halbheer, D, F Stahl, and O Koenigsberg, "Choosing a digital content strategy: How much should be free?," *International Journal of* , 2014.

Heath, Chip and Dan Heath, *Made to stick: Why some ideas survive and others die*, Random House, 2007.

Heckman, J, H Ichimura, J Smith, and P Todd, "Characterizing selection bias using experimental data," 1998.

Heckman, JJ and H Ichimura, "Matching as an econometric evaluation estimator," *The Review of* , 1998.

Herskovitz, Stephen and Malcolm Crystal, "The essential brand persona: storytelling and branding," *Journal of business strategy*, 2010, *31* (3), 21–28.

Hlavac, Marek, "stargazer: Well-Formatted Regression and Summary Statistics Tables.," 2015, (R package version 5.2.).

Hong, Y, "Hypothesis testing in time series via the empirical haracteristic function: A generalized spectral density approach," *Journal of the American Statistical Association*, 1999, (84).

Hopper, PJ and SA Thompson, "Transitivity in grammar and discourse," 1980.

Hui, Sam K, Tom Meyvis, and Henry Assael, "Analyzing Moment-to-Moment Data Using a Bayesian Functional Linear Model: Application to TV Show Pilot Testing," 2014, *33* (2), 222–240.

Imai, K and Van DA Dyk, "Causal inference with general treatment regimes," *Journal of the American Statistical* , 2012.

Joachims, T., "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," Technical Report 23, Universität Dortmund, LS VIII-Report 1997.

___ , "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," in "European Conference on Machine Learning (ECML)" Springer Berlin 1998, pp. 137–142.

___ , *Learning to Classify Text Using Support Vector Machines – Methods, Theory, and Algorithms*, Kluwer/Springer, 2002.

Kamenica, Emir and Matthew Gentzkow, "Bayesian Persuasion," *American Economic Review*, 2011, *October* (101), 2590–2615.

Liu, Xia, Tridib Mazumdar, and Bo Li, "Counterfactual Decomposition of Movie Star Effects with Star Selection," *Management Science*, 2015, *61* (7), 1704–1721.

Luo, H, "When to sell your idea: Theory and evidence from the movie industry," *Management Science*, 2014.

Madnani, Nitin, Jordan Boyd-Graber, and Philip Resnik, "Measuring Transitivity Using Untrained Annotators," in "Creating Speech and Language Data With Amazon's Mechanical Turk" 2010.

Mohammad, Saif M and Peter D Turney, "Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon," 2010, pp. 26–34.

Mohammad, Saif M. and Peter D. Turney, "Crowdsourcing a Word-Emotion Association Lexicon," 2013, *29* (3), 436–465.

Narayan, Vishal and Vrinda Kadiyali, "Repeated Interactions and Improved Outcomes: An Empirical Analysis of Movie Production in the United States," *Management Science*, 2015.

Netzer, Oded and Olivier Toubia, "Idea Generation, Creativity and Prototypicality," *Working Paper*, 2014.

Ostrovsky, M, "Information aggregation in dynamic markets with strategic traders," *Econometrica*, 2012.

Papadatos, Caroline, "The art of storytelling: how loyalty marketers can build emotional connections to their brands," *Journal of Consumer Marketing*, 2006, *23* (7), 382–384.

Park, Joon Y and Yoon-Jae Whang, "Testing for the Martingale Hypothesis," *Studies in Nonlinear Dynamics and Econometrics*, 2005, (9).

Parker, Geoffrey G. and Marshall W. Van Alstyne, "Two-Sided Network Effects: A Theory of Information Product Design," *Management Science*, 2005, *51* (10), 1494–1504.

Phillips, Barbara J and Edward F McQuarrie, "Narrative and persuasion in fashion advertising," *Journal of Consumer Research*, 2010, *37* (3), 368–392.

Phillips, Peter CB and Sainan Jin, "Testing the Martingale Hypothesis," *Journal of Business Economic Statistics*, 2014.

Plutchik, R, "A general psychoevolutionary theory of emotion," in "Emotion: Theory, research, and experience," Vol. 1, University of Texas Press, 1980.

Schank, RC and RP Abelson, "Scripts, Plans, Goals, and Understanding: An Inquiry Into Human Knowledge Structures.," 1977.

Sismeiro, Catarina and Ammara Mahmood, "Competitive vs. Complementary Effects in Online Social Networks and News Consumption: A Natural Experiment," *Management Science*, 2018.

Teixeira, Thales, Michel Wedel, and Rik Pieters, "Emotion-Induced Engagement in Internet Video Advertisements," *Journal of Marketing Research*, 2012, *49* (2), 144–159.

Toubia, Olivier, Garud Iyengar, Renee Bunnell, and Alain Lemaire, "Positive Psychology and the Consumption of Movies," *Working Paper*, 2015.

U.S. Bureau of Economic Analysis, U.S. Department of Commerce, 2017.

Varian, HR, "Versioning information goods," 1997.

Varian, HR, "Markets for information goods," in K Okina and T Inoue, eds., *Monetary Policy in a World of Knowledge-Based Growth, Quality Change, and Uncertainty Measurement*, Palgrave Macmillan, 2000.

Wu, S, LM Hitt, and P Chen, "Customized bundle pricing for information goods: A nonlinear mixed-integer programming approach," *Management Science*, 2008.

A. Supplemental Material

	Dependent variable:		
	log(Audience Rating)	*Volume*	*log*(Audience Rating)
	OLS	*GLM(Poisson)*	*Tobit*
	(1)	(2)	(3)
$PROD$	-0.00000^{**}	-0.000^{***}	0.001^{***}
	(0.00000)	(0.000)	(0.00000)
$RUNTIME$	0.239^{***}	0.004^{***}	0.000^{***}
	(0.044)	(0.0003)	(0.048)
$TRANS_{AVG}$	-0.745^{***}	-0.016^{***}	-1.064^{***}
	(0.231)	(0.002)	(0.271)
$TRANS_{SD}$	1.321^{**}	0.027^{***}	1.739^{***}
	(0.591)	(0.003)	(0.635)
$\triangle BELIEF_{SD}$	8.887^{**}	0.089^{***}	5.898^{**}
	(3.679)	(0.017)	(4.765)
$\triangle BELIEF_{AVG}$	-0.001	-0.00002^{*}	-0.000
	(0.002)	(0.00001)	(0.002)
$\triangle BELIEF0_{SD}$	-16.870^{***}	-0.160^{***}	-10.410^{**}
	(5.126)	(0.026)	(4.765)
$\triangle BELIEF0_{AVG}$	0.037^{**}	0.0003^{***}	0.017^{**}
	(0.015)	(0.00005)	(0.008)
$WRITER$	3.238^{*}	0.051^{***}	4.406^{**}
	(1.909)	(0.010)	(2.090)
$DIRECTOR$	11.480^{***}	0.205^{***}	14.780^{***}
	(1.659)	(0.009)	(1.693)
$ANTICIPATION$	-0.876	-0.0001	-0.004
	(0.022)	(0.0001)	(0.022)
$SURPRISE$	0.061^{**}	0.001^{***}	0.066^{***}
	(0.025)	(0.0001)	(0.025)
$DRAMA$	6.710^{***}	0.106^{***}	6.710^{***}
	(1.788)	(0.010)	(1.788)
$THRILLER$	-7.023^{***}	-0.107^{***}	-7.023^{***}
	(1.879)	(0.010)	(1.879)
$ADVENT$	-6.229^{***}	-0.101^{***}	-6.229^{***}
	(2.130)	(0.012)	(2.130)
$ANIMATION$	20.993^{***}	0.304^{***}	20.993^{***}
	(4.677)	(0.024)	(4.677)
Constant	416.600	6.803^{***}	56.670^{***}
	(453.600)	(2.462)	(18.030)
Observations	1,088	1,088	1,088
R^2	0.356		
Adjusted R^2	0.328		
Akaike Inf. Crit.		14,802.370	

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4: Regression Results

	Dependent variable:
	Suspense Measure
PROD	0.000
	(0.000)
RUNTIME	−0.001
	(0.003)
lenscript	−0.003
	(0.002)
WRITERCOUNT	−0.014
	(0.059)
TOPWRITER	−0.050
	(0.138)
TOPACTOR	−0.002
	(0.025)
DIRECTOR	0.180
	(0.120)
COMEDY	−0.040
	(0.132)
ROMANCE	−0.230*
	(0.137)
DRAMA	−0.316***
	(0.119)
SCI-FI	0.041
	(0.154)
THRILLER	0.231*
	(0.125)
ADVENT	0.423***
	(0.130)
CRIME	−0.109
	(0.135)
ANIMATION	0.586*
	(0.308)
Constant	8.497***
	(0.509)
Observations	998
R^2	0.094
Adjusted R^2	0.062
Residual Std. Error	1.543 (df = 963)
F Statistic	2.935*** (df = 34; 963)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Matching Results

Table 6: Propensity Score Estimates

	Dependent variable:	
	Suspense Measure	Surprise Measure
	(1)	(2)
PROD	0.000	−0.000
	(0.000)	(0.000)
RUNTIME	−0.001	0.001
	(0.003)	(0.003)
LENGTH	−0.003	−0.0002
	(0.002)	(0.0002)
WRITERCOUNT	−0.014	−0.004
	(0.059)	(0.006)
TOPWRITER	−0.050	0.021
	(0.138)	(0.013)
TOPACTOR	−0.002	−0.001
	(0.025)	(0.002)
DIRECTOR	0.180	0.003
	(0.120)	(0.011)
COMEDY	−0.040	−0.023*
	(0.132)	(0.012)
ROMANCE	−0.230*	−0.011
	(0.137)	(0.013)
DRAMA	−0.316**	−0.007
	(0.119)	(0.011)
SCI-FI	0.041	−0.001
	(0.154)	(0.015)
THRILLER	0.231*	−0.028**
	(0.125)	(0.011)
ADVENT	0.423**	0.009
	(0.130)	(0.012)
CRIME	−0.109	−0.011
	(0.135)	(0.013)
ANIMATION	0.586*	0.036
	(0.308)	(0.029)
Constant	8.497***	0.800***
	(0.509)	(0.049)
Observations	998	997
R^2	0.094	0.048
Adjusted R^2	0.062	0.015
Residual Std. Error (df = 963)	1.543	0.147
F Statistic	2.935*** (df = 34; 963)	1.458** (df = 33; 963)
Note:		*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:	
	log(Audience Rating)	Volume
	OLS	*GLM(Poisson)*
	(1)	(2)
PROD	−0.00000	−0.000**
	(0.00000)	(0.000)
RUNTIME	0.288***	0.003***
	(0.079)	(0.0004)
$TRANS_{AVG}$	−1.367***	−0.013***
	(0.423)	(0.002)
$TRANS_{SD}$	3.457***	0.020***
	(1.026)	(0.006)
$\triangle BELIEF_{SD}$	7.068	0.161***
	(6.353)	(0.039)
$\triangle BELIEF_{AVG}$	−0.003	−0.00004
	(0.004)	(0.00002)
$\triangle BELIEF0_{SD}$	−13.987	−0.468***
	(8.843)	(0.052)
$\triangle BELIEF0_{AVG}$	0.039	0.001***
	(0.025)	(0.0002)
TOPWRITER	3.925	0.036**
	(3.392)	(0.017)
DIRECTOR	13.672***	0.199***
	(3.103)	(0.016)
ANTICIPATION	−0.007	0.0001
	(0.034)	(0.0002)
SURPRISE	0.005	0.001***
	(0.040)	(0.0002)
DRAMA	6.151**	0.103***
	(3.058)	(0.018)
THRILLER	−3.759	−0.136***
	(3.291)	(0.019)
ADVENT	−11.488***	−0.027
	(3.204)	(0.020)
CRIME	0.084	0.045**
	(3.224)	(0.020)
ANIMATION	29.766***	0.177***
	(7.094)	(0.046)
Constant	71.772***	4.735***
	(25.776)	(0.158)
Observations	311	330
R^2	0.489	
Adjusted R^2	0.414	
Log Likelihood	35	-2,281.620
Akaike Inf. Crit.		4,645.241
Residual Std. Error	20.609 (df = 270)	
F Statistic	6.464*** (df = 40; 270)	

Cognitive Categorization, Memorability, and Likability: The Case of Logo Designs

Shengli Hu

Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853

sh2264@cornell.edu

We propose and provide scalable methods to descriptively quantify and evaluate two cognitive processes for logos: (1) perceptual categorization; and (2) functional categorization. We draw on research in perception, computer vision, and information theory. We measure via MTurk participants consumer memory and liking of logos. Our dataset consists of 125,270 logo designs from the U.S. market, spanning 39 industry categories.. This dataset enables us to evaluate both the absolute and relative impacts of two forms of cognitive categorization and logo features on design memorability and design likability. In addition, we explore multiple methods for logo clustering and analyze drivers of memorability and likability. To validate the cognitive interpretation of our proposed measures, we further gather and incorporate human perceptual templates into the algorithm. We discuss managerial insights for logo design.

Keywords: Logo Design, Memory, Liking, Cognitive Categorization, Convolutional Neural Networks, Deep Embedding Clustering, Visualization, Visual Templates.

1. Introduction

Logos are visual representations of organizations, companies and brands (Henderson and Cote 1998, MacInnis, Shapiro and Mani 1999, Swartz 1983). While logos should be aesthetically appealing to target consumers, companies do spend a great deal of time and money crafting logo designs that reveal central messages about the brand. These efforts are not unwarranted since studies show that consumers attribute their inferences from logo designs to the associated companies and brands (Jiang, Gorn, Galli and Chattopadhyay 2016), or even the broader environment (Hagtvedt 2011, Rahinel and Nelson 2016). As a result, logos could potentially influence the reputation of the focal companies and brands (Baker and Balmer 1997, Olins 1990, Van den Bosch, De Jong and Elving 2005), consumers' willingness to pay (Jun, Cho and Kwon 2008), and their brand loyalty (Müller, Kocher and Crettaz 2013). While the marketing studies above investigate various aspects of logos (Henderson and Cote 1998, Müller et al. 2013, Jiang, Gorn, Galli and Chattopadhyay 2015, Jiang et al. 2016, Dew, Ansari and Toubia 2018a), we study the following understudied element of logo design— what visual elements of logo designs affect logo memorability and liking? Are logos that are cognitively easier to categorize liked and remembered more? Do these design and categorization effects on memory and liking vary by type of logo?

We answer these questions by the following steps. First, we collect 125,270 U.S. logos from 39 industry groups, including both for-profit and non-profit industries. Then, we use Amazon MTurk to measure liking and memory scores for these logos. Next, we extract more than two dozen image features of logos using current methods in computer vision. These features include font, color, shape, aesthetics, among others. Then, we use measures in information theory to define two types of categorization (Rosch 1973, Mervis and Rosch 1981, Jones 2016). Research in cognitive psychology (Rosch and Lloyd 1978, Simon 1996) refers to cognitive categorization as one of the concepts for how people perceive and understand complex objects. We use two measures of categorization. The first measure — that we term functional categorization — measures how close a logo is to logos in its industry class. The second measure — that we term perceptual categorization— measures how close a logo is to objects that are encountered in life.

We then run a descriptive regression analysis to estimate the effects of several visual features and the two categorization variables. Further, we use several clustering algorithms to uncover stable groupings of logos and their visual and categorization characteristics. Primar-

ily, we are interested in uncovering whether there are differences in drivers of memory and liking across clusters. We discuss managerial implications of our results. the classification image procedures to incorporate perceptual templates from human inputs into our measures. We discuss the managerial implications of our results.

To preview the main results, we (Section 6) show that the effects of two categorization measures are nonlinear and divergent regarding both memorability and likability, so do multiple other lower-level image features, such as symmetry, the presence of texts, polygons, among others. Further results on image clusters are documented in Section 7, where we compare and contrast three classes of image clustering methods. We detail the procedure to incorporate human biases into algorithms and how it shifts the previous results and interpretations in Section 8.

Our paper's contributions are as follows. Substantively, it is the first study to evaluate both the absolute and relative impacts of two forms of cognitive categorization and logo features on design memorability and design likability. Methodologically, we explore multiple methods for logo clustering and analyze what distinguishes between different logo clusters, and how different resulting clusters relate to design memorability and likability. In addition, we incorporate human perceptual templates into our measures using the classification image procedure. Managerially, we provide managers with empirical evidence of necessary tradeoffs to be considered in logo design processes with respect to memory and liking.

The rest of the paper is organized as follows. In Section 2, we review multiple streams of literature in marketing, computer vision, and cognitive psychology, without which the current study would not come into being. In Section 3, we describe our data collection and annotation processes, summarize our dataset, detail how we extract image features using established computer vision and graphics methods, as well as both of cognitive categorization measures using DCNNs. In Section 4, we detail theoretical hypotheses based on cognitive categorization theories from previous literature in psychology and consumer behavior. Correspondingly, we present and interpret the results in Section 5. We detail three classes of image clustering methods and the resulting clusters in Section 7, followed by holistic visualization presentation and links in Section 7.6. We validate and modify the measures by deriving human perceptual and conceptual templates and incorporating them into categorization measures in Section 8. Lastly, we close by conclusions and future research ideas in Section 10.

2. Background and Related Literature

We review literature in three key areas — marketing, computer vision, and cognitive psychology. We also discuss how our work builds on and extends existing research.

2.1 Logos

There exist few relevant studies similar to the current research. More generally however, among early research on logos is dating back to Henderson and Cote (1998), who analyze a small set of logo features with factor analysis. They propose a set of visual constructs, among which natural, harmonious, and elaborate are shown predictive of outcome measures. Furthermore, they find the effects of robustly consistent cross-culturally. Van der Lans, Cote, Cole, Leong, Smidts, Henderson, Bluemelhuber, Bottomley, Doyle, Fedorikhin et al. (2009) extend Henderson and Cote (1998) by adding repetition, proportion, and parallelism into the mix and confirming the robustness of the cross-cultural results.

Other consumer studies have focused on various aspects of logo shapes, colors, and fonts. Klink (2003) explores the relevance of popular linguistic theory linking phonetic features to shapes in the context of brand names and logo features such as color and angularity. Walsh, Page Winterich and Mittal (2010) further show that consumers with different levels of loyalty to the focal brand respond differently when logo features evolve from angular to round. Jiang et al. (2015) investigate the same problem and demonstrate the underlying mechanism: consumers respond to the shape of company logos through perceived hardness versus softness, which in turn influences attribute inferences on the part of perceivers. There also exists studies on the perceived orientation of designs that in turn affects perceivers' engagement and attitudes towards the focal company and brand (Cian, Krishna and Elder 2014). Doyle and Bottomley (2006) survey typeface research on intended consumer product or brand. One major finding is that congruence in derived abstract connotations between the logo font and the consumer product is associated with the more indicated purchase of the product. Hagtvedt (2011) provide evidence that incomplete typeface could lead consumers to associate the focal brand with certain brand personalities such as untrustworthiness and innovativeness. Most recently, Dew, Ansari and Toubia (2018b) study the connection between brand personality traits and logo features using various image processing algorithms and propose a new generative model for logo generation given intended brand personality traits. While

Dew et al. (2018b) and the current study both algorithmically extract logo features, the research objectives, techniques, and outcome variables are indeed disparate — Dew et al. (2018b) focus on the link between logo features and brand personality traits, whereas the current study focuses on the link between cognitive categorization and memorability versus likability; Dew et al. (2018b) use traditional image processing and generative algorithms, whereas the current study uses deep neural networks for image classification and classification image techniques; Dew et al. (2018b) provide insights into what kinds of logo features induce perceived brand personality traits, whereas the present study what visual elements contribute to memorability versus likability.

The current study complements this body of work with a focus on design memorability and likability of logos at a large scale with the help of methods from computer vision and information theory research. We complement this body of work in that we focus on two cognitive categorization processes and their operationalization, which have not been studied before in the context of logos. Our results regarding other lower-level image features such as color and shape, echo and complement the results of previous work in this stream of research. Our main findings regarding the counteracting effects of visual elements on memorability versus likability in designs have not been indicated or documented in this literature, shedding new lights on nuanced factors to be taken into consideration when designing a business logo. In addition, the current study is, to our best knowledge, of the largest scale in this research stream, with much less of a concern on sample selection biases.

2.2 Perception, Aesthetics, and Vision Science

There exists a large body of research on perception and aesthetics in vision science which broadly encompass research efforts in multiple disciplines to answer fundamental questions that relate to how humans perceive and respond to colors and juxtapositions of colors under various circumstances.

Valdez and Mehrabian (1994) identify the links between colors and evoked emotions. They find direct human emotional responses along dimensions of pleasure, arousal, and dominance, per color dimensions such as saturation and lightness. Specifically, colors of blue, green, and purple appear to be most frequently associated with pleasant feelings, whereas yellow and orange the last few colors considered pleasant. Researchers have also identified the link between gender indication, and lightness — light colors such as white are more likely

associated with feminine features, whereas dark colors such as black are considered more masculine, regardless of cultural environments (Semin and Palma 2014). In the same vein, Kareklas, Brunel and Coulter (2014) apply the light-dark analysis to marketing contexts and find that consumers automatically prefer white over black when faced with product and advertising choices. Similarly, Deng, Hui and Hutchinson (2010) research the broader spectrum of color combinations and consumer preferences, and find that consumers are more sensitive to hue and saturation than lightness, prefer similar colors with a single contrasting color, and a small number of colors.

The current research complements this body of research by testing the effects of various image features such as different aspects of color, shape, text, and aesthetics on logo memorability, as well as logo likability. We find effects on memorability and likability that echo the results of separate studies in this stream literature. In addition, we evaluate both the absolute effects and relative effects of image features and cognitive categorization measures, while introducing the effects of cognitive categorization processes into the mix. While relying on cognitive theories in these realms to form hypotheses and validate conjectures, the focus of the current study is on teasing out effects of different modes of cognitive categorization on memorability and likability, alongside other low-level images features such as color and shape. Our estimation, in comparison, is also of a much greater scale with automatic algorithms that incorporate human perceptual templates. It falls within the scope of vision science, but completing it with immediate managerial implications of practical relevance in the commercial world.

2.3 Cognitive Categorization

Categorization has long been studied by cognitive psychologists and neuroscientists as one of the most fundamental cognitive processes human engage in whenever new perceptual stimuli are encountered (Rosch 1999, Simon 1996).

A large body of research has been devoted to (1) basic principles of cognitive categorization; (2) structures and definitions of categorization; (3) neural and psychophysical bases of cognitive categorization; (4) the debate between prototype categorization and logical categorization; (5) categorization under context; among many other important sub-topics. We clarify some basic terms following the literature in cognition before reviewing some of the fundamental theories and associated empirical evidence most relevant to the current study.

A category refers to a number of objects that are considered equivalent. Categories are generally designated by names such as cat, mammal, etc. *A taxonomy* is a system by which categories are related to one another through class inclusion. The higher the inclusiveness of a category within a taxonomy, the higher the level of abstraction. Thus the term level of abstraction within a taxonomy refers to a particular *level of inclusiveness*.

There are two basic principles proposed for the formation of categories (Rosch and Lloyd 1978): the cognitive economy and perceived world structure. Cognitive economy speaks of the function of category systems and asserts that the task of category systems is to provide *maximum information with the least cognitive effort*. Perceived world structure speaks of the structure of information provided and states that the perceived structure comes as *structured information* rather than arbitrary or unpredictable attributes. Such fundamental principles are consistent with the most basic constructs such as entropy and cross-entropy in information theory. One particular stream of research revolves around the discussion of natural categories.

Rosch (1973) asserts that the domains of color and form are structured into nonarbitrary, semantic categories which develop around perceptually salient "natural prototypes." Categories which reflected such an organization and categories which violated the organization were consistently easier to learn than the "distorted" categories. Even when not central, natural prototype stimuli tended to be more rapidly learned and more often chosen as the most typical example of the category than were other stimuli. In the same vein, categorizations which humans make of the concrete world are not arbitrary but highly determined. In taxonomies of concrete objects, there is one level of abstraction at which the most basic category cuts are made. Basic categories are those who carry the most information and are the most differentiated from one another. Rosch, Mervis, Gray, Johnson and Boyes-Braem (1976) establish four intuitive principles of basic objects which make up the most inclusive categories. The current study follows this school of thought by adopting natural categorization into basic categories as "perceptual categorization" in our context.

Mervis and Rosch (1981) review empirical findings that have established that categories are internally structured by gradients of representativeness, category boundaries are not necessarily definite, and there is a close relation between attribute clusters and the structure and formation of categories. Despite the implied debate about well-defined versus fuzzy boundaries of categories, there appears to be a consensus that categories tend to be viewed as being as separate from each other and as clear-cut as possible. One way to achieve

separateness and clarity of continuous categories is by conceiving of each category in terms of its clear cases rather than its boundaries —– prototype categorization. The current study conceptually follows the cognitive process established by this framework. Existing research shows that visual representations tend to be stored in memory as deviations from category prototypes (e.g., Estes (1986), Hintzman (1990)). As people are exposed through their lifetimes to more exemplars of a category, they develop a prototype in memory. When new exemplars are encountered, the visual memory system checks for and encodes differences from the prototype rather than remember every detail. For visual design, cognitive categorization could result from two sources — first, the perceptual categorization of the design and second, the functional categorization of the design.

Therefore the current study builds upon this body of research by adopting the prototype categorization framework, while complementing it by introducing another layer of categorization process on top of the broadly studied and debated framework. We present functional categorization based on the idea that logo designs are intended to signal functionality, and explore the link between different categorization processes and logo memorability or likability. Unlike most studies in this realm, our work features a large-scale dataset collected from public sources, and automatic visual categorization and clustering methods that incorporate categorical prototypes from human perception. In this sense, our work complements this body of work both methodologically and theoretically in that we identified tradeoffs of categorization mechanisms with respect to memory and liking in the process of logo design, that were unexplored before.

2.4 Image Memorability and Popularity in Computer Vision

Earlier research in computer vision used insights from cognitive science on face recognition and evaluation, although it is not until recently that larger-scale and more objective assessment has emerged. For instance, Bainbridge, Isola and Oliva (2013) collect face memorability scores for multiple face datasets and study what facial attributes contribute the most in predicting memorability using multilinear regression models. They find the most significant predictors include responsible, uncertain, kind, introverted, intelligent, atypical, trustworthy, attractive, familiar, unemotional, caring, boring, etc.

The stream most relevant to ours starts with Isola, Xiao, Torralba and Oliva (2011). The authors introduce an experimental procedure to measure human memory objectively,

and collect memorability annotations for the widely used SUN dataset (Xiao, Hays, Ehinger, Oliva and Torralba 2010). They find color features to be weakly correlated with memorability scores at best, which is consistent with findings of the current study. They perform image segmentation and object recognition on each segment, and thus semantic regressions analysis. They demonstrate that if a system knows which objects an image contains, it can predict memorability with a performance not too far from human consistency. Further, predict memorability using only global features algorithmically extracted from images — GIST (Oliva and Torralba 2001), SIFT (Lazebnik, Schmid and Ponce 2006), HOG2x2 (Dalal and Triggs 2005, Felzenszwalb, Girshick, McAllester and Ramanan 2010), and SSIM (Shechtman and Irani 2007) — and find the performance comparable to that with manual object labeling.

Building on Isola et al. (2011), Khosla, Raju, Torralba and Oliva (2015a) which directly inspired the current study, use the same experimental procedure to objectively measure human memory, and build a large annotated image memorability dataset of 60,000 images. They first establish that image memorability is an intrinsic and stable property of an image, and proceed to train DCNNs to predict memorability, the result of which correlates reasonably well with human validations. They find that image memorability is correlated with popularity (non-linearly), saliency, and evoked emotions such as disgust (positively), amusement (positively), awe (negatively), and contentment (negatively). Surprisingly, aesthetics exhibits no correlation with memorability.

Dubey, Peterson, Khosla, Yang and Ghanem (2015) extend Isola et al. (2011) and Khosla et al. (2015a) along a different dimension than the current study in that they investigate what makes an object inside an image memorable, whose results in part corroborate with ours, albeit under a different context. They uncover the effects of various factors including color saliency, color, object categories, as well as the relationship between object and image memorability. They find visual saliency and the number of objects to be positively correlated with memorability, whereas color features do not. Some object categories are found to be more memorable than others, such as animal, person, and vehicle categories as opposed to furniture, nature, building, and device categories which tend to have a vast majority of objects with very low memorability scores. The correlation between object memorability and image memorability also suggests that the most memorable object in an image plays a crucial role in determining the overall memorability of an image.

Besides memorability as the prediction objective, Khosla, Das Sarma and Hamid (2014) study what makes a photograph popular — image content or social cues — by gathering

a dataset of user-uploaded images from Flickr and keeping track of the total number of likes for each image. They extract various image features (50-class color descriptors from color histograms (Khan, Van de Weijer, Shahbaz Khan, Muselet, Ducottet and Barat 2013), GIST (Oliva and Torralba 2001), Local Binary Pattern for texture (Ojala, Pietikainen and Maenpaa 2002), HOG (Dalal and Triggs 2005), image embeddings from training DCNNs (Donahue, Jia, Vinyals, Hoffman, Zhang, Tzeng and Darrell 2014), detected objects) and social features (mean views, photo count by user, number of contacts of a user, number of groups the user belongs to, the average group size of groups the user belongs to, member duration, etc.). They find that image texture, gradient, and embeddings from deep learning to be reasonably predictive of popularity, as do several social features, whereas color features are not. In the same vein, Deza and Parikh (2015) study image virality by introducing three image datasets from Reddit and defining virality scores based on Reddit metadata. They train support vector machine classifiers using deep-learned image features to predict virality of individual images, relative virality in pairs of images, and the dominant topic of a viral picture. They find the following five critical visual attributes to be most predictive of internet virality: Animal, Synthetically Generated, (Not) Beautiful, Explicit and Sexual. The current study is closely related to the studies above in that we examine the likability of an image, which intuitively correlates with image popularity or virality, despite in different contexts.

The current study complements this stream of research in that (1) we investigate memorability and likability of logo (visual) designs, as opposed to existing results on memorability or popularity of natural scenes, images, photographs, faces, and visualizations; (2) we connect DCNN methods with categorization theories in cognitive psychology, introducing new interpretations of image classification problems and casting them in the context of consumer perception and thus business applications; (3) we bring together image memorability and likability that have been separately studied previously, and uncover interesting effects by juxtaposition; (4) the current study warrants greater external validity at a larger scale, providing specific evidences for organizational design choices and practical managerial insights. For instance, for product managers of marketing and design, our findings shed light on important trade-offs in the design process. Certain visual elements such as high complexity and color contrasts exhibit conflicting effects on memorability and likability, whereas some others prove effective independently. Lastly, unlike other studies in this body of research, our paper further incorporates perceptual templates from human inputs, bridging this literature with the emerging stream of human-in-the-loop machine learning (Cui, Zhou, Lin and

Belongie 2016).

2.5 Marketing Research Using Vision Methods

There exists few relevant studies in marketing. However there has been a growing body of marketing research that leverages the scalability and accuracy of computer vision methods.

This stream perhaps started with face evaluations by Xiao and Ding (2014), who study the effect of non-celebrity faces in print advertising with established face recognition methods from computer vision. Lu, Xiao and Ding (2016) create an automatic and scalable garment recommender system that identifies shoppers' preferences based on their reactions and uses that information to make meaningful, personalized recommendations. Todorov (2018) models social perception of faces using data-driven approaches whose objective is to identify quantitative relationships between high-dimensional variables (e.g., visual images) and behaviors (e.g., perceptual decisions) with as little bias as possible.

Product designs have been one of the main focuses. Dzyabura, Ibragimov and Kihal (2018) use machine learning models to predict demand in online and offline retail channels, as well as returns for new products. They demonstrate the superior prediction performance when product image features — color histograms, texture, learned representation by training AlexNet (Krizhevsky, Sutskever and Hinton 2012a) — are included. Zhang, Lee, Singh and Srinivasan (2018) estimate the economic impact of images and low-level image features on property demand in Airbnb. By classifying property photos with computer vision and deep learning models, they show that 48.9% of the effect of verified images boosting demand comes from the high image quality. They also identify 12 image attributes based on marketing and photography literature and demonstrate the direct impacts on demand after controlling for many observables, thus prescribing optimal product image strategies to increase demand for housing and lodging managers. Tkachenko, Ansari and Toubia (2018) apply deep learning techniques for computer-aided exploration of visual product designs. They use images and attributes of products sold on Amazon.com as well as human feedback from Amazon Turk workers as a basis for experiments. Their results imply that deep generative models offer a promising avenue for partial automation of the visual product design process.

User-generated visual content and brand images have been investigated as well. Liu and Mayzlin (2018) propose a "visual listening in" approach to measuring how brands are portrayed on Instagram by mining visual content posted by users. They use supervised

machine learning methods, traditional support vector machine classifiers, and deep convolutional neural networks, to measure brand attributes (glamorous, rugged, healthy, fun) from images. Then they apply the classifiers to brand-related photos posted on social media to gauge what consumers are visually communicating about brands. They find key differences between how consumers and firms portray the brands on visual social media, and how the average consumer perceives the brands. Papatla (2018) investigates whether the presence of faces in Visual UGC could be less detrimental if they are less prominent, based on findings that faces and bodies of humans in the visual field are processed holistically even if they are seen as distinct stimuli. They analyze consumer response to about 12,000 photos of 800 different products in six categories displayed by 35 online retailers. Shi, Lee, Singh and Srinivasan (2018) study the substitutability between the brand value and the style value in the fashion market. They quantify the style value by employing deep learning based computer vision techniques to create style features, including clothing style (e.g., compatibility between clothing items, creativity), model style (e.g., facial and body attractiveness), and photo style. These style features are incorporated in a dynamic structural model to estimate a dynamic structural model to analyze the content creation and consumption behavior of influencers in a fashion social network community. They find significant effects of brands and style features on the trendiness of a fashion look, as well as substitutability patterns between style features and brand levels.

Perhaps most relevant to ours is Dew et al. (2018a). They explore the visual elements in logos that express brand personality traits, based on which they introduce a logo tokenization algorithm that decomposes logos into theory-based and human-meaningful visual features. Applied to a small dataset of logos, matched with textual data from firms' websites, consumer evaluations of brands, third-party descriptions of the companies, they uncover links that exist between a brand's logo, description, and personality, and thereby facilitate a better understanding of the underpinnings of good design, and inform the design of new logos. While Dew et al. (2018b) and the current study both algorithmically extract logo features, the research objectives, techniques, and outcome variables are indeed disparate — Dew et al. (2018b) focus on the link between logo features and brand personality traits, whereas the current study focuses on the link between cognitive categorization and memorability versus likability; Dew et al. (2018b) use traditional image processing and generative algorithms, whereas the current study uses deep neural networks for image classification and classification image techniques; Dew et al. (2018b) provide insights into what kinds of logo features induce

perceived brand personality traits, whereas the present study what visual elements contribute to memorability versus likability.

We complement this body of research by focusing on the memorability and likability of logo designs, bridging literature in marketing, computer vision, and cognition. Our results facilitate a better understanding of the cognitive processes perceivers engage in, and how these processes, together with other design elements influence the design of logos for either memorability or likability. Our results uncover trade-offs due to divergent effects of different design elements on memory and liking that warrant consideration of the design of new logos. Unlike most studies in the body of literature, we proposed large-scale automatic methods to proxy for cognitive processes while incorporating perceptual templates elicited from human subjects using the image classification procedure (Eckstein and Ahumada 2002, Murray 2011) and the orientation constraint method (Vondrick, Pirsiavash, Oliva and Torralba 2015), enabling large-scale estimation without falling victim to small sample biases that have plagued many studies in this research stream.

3. Large-scale Memorability and Likability Logo Design Dataset

We gather a dataset of $125,270$ logo designs annotated with memorability and likability scores. It is a sub-sample of $543,758$ logo designs. We present in Section 5 a preliminary analysis of the data.

3.1 Business Logo Dataset

We collected vector logos from various sources on the Internet courtesy of Brands of the World, Logo Types, World Vector Logo, Vector Me. We merged vector logos using fuzzy matching based upon image similarity scores and available meta-data.

3.2 Low-level Image Features

We explore a variety of image features of our dataset, including file size, hue (mean and standard deviation), saturation (mean and standard deviation), value (i.e. lightness) (mean

and standard deviation), the number of edges, the number of straight lines, the number of corners, the number of circles, number of polygons, the presence of texts, the presence of marks, the average size of marks, the average width/height of detected marks, horizontal symmetry, vertical symmetry, among others. We tabulate extracted image features and corresponding methods in Table 1.

Category	Feature	Description
Color	Color	Whether a given color is present
	Dominant Color	The color with the highest number of pixels
	% Whitespace	Logo versus background ratio
	Mean Saturation	The mean value of the saturation channel across pixels in HSV
	SD Saturation	The standard deviation of the saturation channel
	Mean Lightness	The mean value of the value channel in HSV colorspace
	SD Lightness	The standard deviation of the value channel
Shape	Has Mark	Is there a mark?
	Mark Size	How much of the logo does the mark take up
	Edge Count	The number of edges detected with Canny edge detector
	Line Count	The number of lines detected with probabilistic Hough transform
	Corner Count	The number of corners detected with Harris corner detector
	Circle Count	The number of circles detected with Hough transform
	Polygon Count	The number of polygons detected with Hough transform
Complexity	# Colors	The number of distinct colors
	Entropy	The local average variance of greyscale pixel intensity
Symmetry	$\text{Count}_{\text{Horizontal}}$	The number of matched key points when split horizontally
	$\text{Distance}_{\text{Horizontal}}$	The correlation of horizontally matched key points in pixel values
	$\text{Count}_{\text{Vertical}}$	The number of matched key points when split vertically
	$\text{Distance}_{\text{Vertical}}$	The correlation of vertically matched key points in pixel values
Text	# Text	The number of detected texts

Table 1: Table of Low-level Features

We detail some low-level features in Table 1 which are less intuitive and require more involved algorithms:

- Mark detection: we pose it as an object detection task adapted to the artwork as

opposed to natural scenes common in computer vision. We define marks as either design patterns or recognizable fonts and typefaces. For the particular purpose of identifying marks in logos, we choose to use the YOLO algorithm (Redmon, Divvala, Girshick and Farhadi 2016) on pre-processed logo designs after applications of image erosion.

You Only Look Once (YOLO) algorithm is an object detection system that boasts real-time processing. For each input image, the YOLO algorithm predicts, for each grid cell prespecified for each image, a fixed number of bounding boxes, box confidence scores associated with each bounding boxes, and the conditional object class probabilities assigned to each identified bounding box. We use the default 7×7 grids, 2 bounding boxes, and 20 classes, following the original YOLO implementation. Therefore, the major concept of YOLO is to build a Convolutional Neural Network to predict a $(7, 7, 2 \times 5 + 20) = (7, 7, 30)$ tensor. We use statistics of the predicted bounding boxes for mark detection, with a confidence threshold of 0.25. The architecture of the YOLO algorithm consists of twenty-four convolutional layers followed by two fully connected layers (FC). The third, fourth, and the fifth layers use 1×1 convolution layer to reduce the number of channels. The mechanism and rationale as to why the 1×1 reduction layer serves the purpose are detailed in Lin, Chen and Yan (2013). The output from the last convolutional layer is flattened and passed through the two fully connected layers, which function in the form of linear regressions, to generate a vector of length $7 \times 7 \times 30$. We cast the vector to a $(7, 7, 30)$ tensor and parse out statistics of bounding boxes. For each grid cell, we select the bounding box with the highest Intersection Over Union (IoU) ratio. The training process of YOLO algorithm uses sum of squared errors of the predictions and the ground truths for calculating the loss function, which includes the object classification loss, the localization loss — the distance between predicted bounding boxes and the ground truths, and the confidence loss with respect to whether an object is indeed in the bounding box.

We choose to use YOLO because it is one of the fastest and simple, classic object detection algorithms that predict specifications of object bounding boxes. It provides benefits over more traditional methods of object detection such as sliding window and region proposal-based techniques, in that it:

- reasons globally about the image when identifying bounding boxes and predicting

object classes. It encodes contextual information about classes along with the appearance, so that unlike Fast R-CNN (Girshick 2015), a top detection method, it makes much fewer background errors comparatively. This is particularly important for our application as the most difficult part of identifying patterns in the logo designs are separating mark and background, which is more difficult than natural images;

- generalizes object representations, since when YOLO was trained on natural images and tested on artwork, it outperforms alternative top detection methods like DPM (Felzenszwalb et al. 2010) and R-CNN (Girshick 2015) by a large margin. The fact that it is highly generalizable and therefore much less likely to break down when applied to new domains or unexpected inputs makes it a good match for our application in the context of business logos;
- it is fast and simple, predicting multiple bounding boxes and classes with single convolutional network architecture.

Compared to some other state-of-the-art detection systems, YOLO makes more localization errors but is less likely to predict false positives on the background. In addition, YOLO learns very general representations of objects, which is particularly suitable for our context. It outperforms some other detection methods, including DPM (Felzenszwalb et al. 2010) and R-CNN (Girshick, Donahue, Darrell and Malik 2014), when generalizing from natural images to other domains like artwork.

- To expand on image erosion: erosion is a standard image processing method that works on binarized images (background = 0, foreground = 1), transforming that image by assigning each pixel in the transformed image the minimum value within a predefined neighborhood of that pixel in the original binary image. Intuitively, connected regions are typically shrunk after erosion, separating previously connected patterns that supposedly belong to distinct marks.

- Symmetry detection: we follow Loy and Eklundh (2006) to detect symmetry by grouping feature points based on their underlying symmetry. The symmetric pairs of features can be efficiently identified, and the symmetry bonding each pair is extracted and evaluated, which are grouped into symmetric constellations that specify the dominant symmetries present in the image. The method is well-suited for all orientations and

radii, and the method can detect local or global symmetries in complex backgrounds, bilateral or rotational symmetry, as well as multiple incidences of symmetry.

- Text detection: we implement the text detector pipeline proposed by Zhou, Yao, Wen, Wang, Zhou, He and Liang (2017), which directly predicts words or text lines of arbitrary orientations and quadrilateral shapes in full images, eliminating unnecessary intermediate steps (e.g., candidate aggregation and word partitioning), with a single neural network. It is relatively more straightforward and more efficient than some other state-of-the-art methods.

- Color clustering: for color quantization, we follow (Dew et al. 2018a) and implement Density-Based Clustering of Applications with Noise (DBSCAN) on pre-processed HSV colorspace of logo images. DBSCAN uses a density criterion to determine both the number of clusters and cluster membership automatically. Therefore, unlike many other clustering algorithms, it does not require a pre-specified number of clusters (such as K-means) or distributional assumptions (such as Gaussian mixtures). Since the HSV colorspace can be precisely described empirically, a density cutoff, which we pre-specify, is more sensible than the number of clusters. Our application also benefits from the robustness to background noise.

  To mitigate the computational inefficiency, we also estimate the number of clusters using a random subsample ($1,000$ out of $125,223$) using DBSCAN and proceed to use the cluster centroids in standard K-means clustering algorithms. We extract color features such as the number of distinct colors and dominant colors based on the color clusters resulting from implementing the DBSCAN algorithm followed by K-Means on the entire dataset in parallel.

- Canny edge detection: we implement the Canny edge detector (Canny 1986) to count the number of edges. We summarize the major steps as below:

  1. We first convert the images from RGB to grayscale;
  2. We remove background noise with a low-pass filter. Specifically, we apply a Gaussian blur on each image with a σ value of 0.5. It convolves the image with a Gaussian function, which reduces the image's high-frequency components and attenuates high-frequency signals. We tune the σ value with a small subsample of size 100;

3. We then determine the gradient magnitudes by applying a Sobel filter (Kanopoulos, Vasanthavada and Baker 1988), from which edges emerge in that when the color (in grayscale) of an image shifts, the pixel intensity shifts accordingly;
4. Lastly, we use non-maximum suppression to pick the best pixel for edges when there are multiple possibilities in a local neighborhood. It finds the pixel with the maximum value in an identified edge;
5. We post-process to remove additional noise from previous steps by filtering pixels through a pair of threshold ratios — a high threshold ratio, applied to the maximum pixel value of an image to establish a high threshold value, defines the lower bound of a strong edge, and a low threshold ratio, applied to the high threshold value, defines the upper bound of non-edges. We identify the edges falling between the high and low thresholds to be weak edges;
6. As the final post-processing step, we recursively identify the weak edges that are connected to strong edges and set them to strong edges, as well as removing the ones disconnected.

We extract detected edges as one of the image features because edge detection is one of the most fundamental operations in computer vision and the choice of Canny edge detector is because it is the most widely used edge detector and one of the most rigorously defined operator. The popularity of the Canny edge detector could be because of its good localization results, and easy operationalization.

- Harris corner detection: we implement the Harris corner detector (Harris and Stephens 1988) to count the number of corners in logo designs to complement Canny edge detectors which perform poorly around corners. It calculates the second-moment matrices once image filtering is done, and the magnitudes of gradients are obtained. It is based on the idea that image gradient is ill-defined around a corner, but has two or more different values. We detail the major steps of applying the Harris corner detection algorithm:

  1. We first compute horizontal and vertical derivatives of the image;
  2. For each pixel value, we compute second moments (products of derivatives) and the sums of second moments, which we use to define the second moment matrix for each pixel value;

3. Define for every pixel the second-moment matrix $H(x, y)$ as:

$$\begin{pmatrix} S_{x2} & S_{xy} \\ S_{xy} & S_{y2} \end{pmatrix} \tag{1}$$

where x, y represent the horizontal and vertical directions of derivatives respectively.

4. We compute the response of the Harris corner detector at each pixel as the disparity between the determinant of $H(x, y)$ and the squared value of the trace of $H(x, y)$ scaled by k, an empirically determined constant at 0.05;
5. We screen every pixel with a default threshold value on the resulting disparity;
6. Lastly, as before, we apply non-maximum suppresion to select the most prominent pixels as detected corners.

We choose to extract corners as it complements other low-level image features and is a commonly used image feature in pre-processing images for subsequent applications in computer vision. We use the Harris corner detector because of its popularity and abundant applications in traditional computer vision tasks such as motion detection, object recognition, image stitching, image retrieval and so forth.

- Probabilistic Hough transform for detecting lines, circles, polygons, etc.: building on results from Canny edge detection, we further extract the number of lines, circles, and polygons using Hough transform (Duda and Hart 1971, Kiryati, Eldar and Bruckstein 1991, Illingworth and Kittler 1988), which is robust under noise and partial occlusion compared to alternatives such as morphology and regression. It is based on the simple idea that points lying approximately on a line in the sample space will form a "cluster of crossings" in the parameter space. Therefore, once transformed to the parameter space, the detection problem becomes counting the peaks in the so-called "accumulator space" resulting from quantized parameter space.

We plot the original file size distribution of the annotated dataset by category. Figure 1 shows the relative distribution of the ten most concentrated categories. Interestingly, logos of the technology category appear to be much more concentrated on the lower end of the

size spectrum, whereas media logos and food & drinks logos show the greatest variations in terms of size.

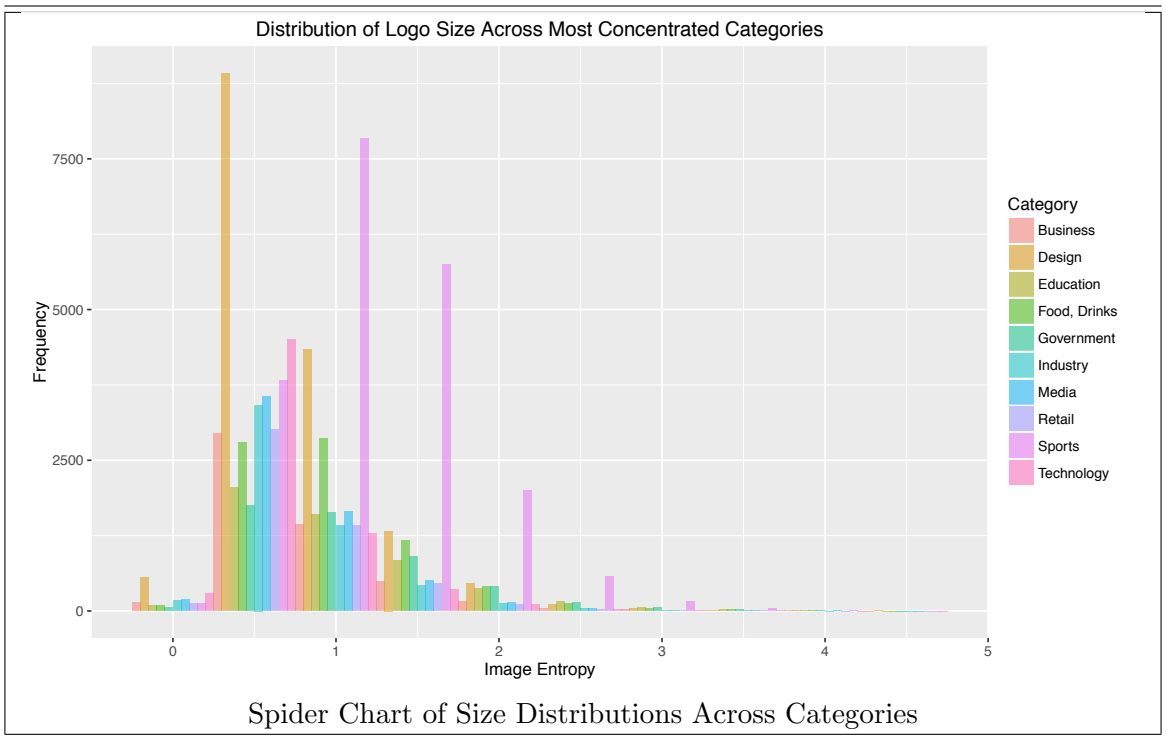


Figure 1: Histograms of Image Size Distributions Across Most Concentrated Countries and Categories

3D plots of representative categories are shown in Figure 2, where the hue, saturation, and value (lightness) values are the averages taken across all pixels of an image. Some descriptive patterns include: sports logos are distributionally more saturated than design logos.

We also measure the number of edges with the Canny detector, the number of corners with the Harris detector, the number of straight lines, circles, and polygons with probabilistic Hough line transformation. We plot a selection of empirical densities of the resulting number of edges, circles, straight lines and polygons in Figure 3 and Figure 4. Some interesting patterns emerged: (1) Logos of the Design industry appear to adopt less straight lines and polygons, whereas Sports logos appear to exhibit polygons much more frequently; (2) Sports, Media, and Food & Drinks Logos appear to showcase more circles whereas the opposite goes for Retail and Technology logos; among others.

3.3 Perceptual and Functional Categorization

Ample research in cognitive psychology refers to *cognitive categorization* as one of the cornerstones (Simon 1996) of research devoted to exploring how people perceive and understand complex objects. In essence, humans simplify understanding of complex objects by categorizing them into groups. This process yields category archetypes and thereby reduces cognitive load efficiently (Rosch and Lloyd 1978, Porac and Thomas 1994). As has been detailed in Section 2.3, existing research shows that visual representations tend to be stored in memory as deviations from category prototypes (e.g., Estes (1986), Hintzman (1990)). As people are exposed through their lifetimes to more exemplars of a category, they develop a prototype in memory. When new exemplars are encountered, the visual memory system checks for and encodes differences from the prototype rather than remember every detail. For visual design, cognitive categorization could result from two sources — first, the perceptual categorization of the design and second, the functional categorization of the design.

By perceptual categorization of a design, we refer to the purely visual elements of the design as well as the autonomous cognitive processes driven by the visual perception of similarities and differences (Holland, Holyoak, Nisbett and Thagard 1989, Rosch and Lloyd 1978). In the context of patterns of visual designs, such attributes are based both on individual features and their configuration (Stacey 2006)[1]; observers are intuitively able to make judgments of visual similarity on a holistic basis such that a single overall visual impression of similarity is reached (Goldstone 1994). Hence we propose a holistic scalable framework that simulates the process and measures that quantify the extent to which the resulting categorization might deviate from a visual prototype for the category, for instance in colors and their combinations (Palmer 1978). As people are exposed through their lifetimes to more exemplars of a category they develop a prototype in memory. When new exemplars are encountered, the visual memory system checks for and encodes differences from the prototype rather than remember all visual elements of the exemplar. This is the reasoning why, for instance, people of the same race are better at recognizing faces whereas across races faces may look more similar. We posit that a high level of visual category prototypicality is likely to attract attention and facilitate encoding of the stimulus compared to low prototypicality that could cause a visual overload and hamper meaningful visual encoding.

The second aspect of cognitive categorization that can affect memory for it lies in func-

[1]Patterns being analogous to styles in this context.

tionality. People remember information by associating its meaning to what they already know (Flavell (1979), Wang, Haertel and Walberg (1990), Veenman, Van Hout-Wolters and Afflerbach (2006)). When the meaning of an attitude object[2] is very clear, it is associated readily into the knowledge a consumer already has in the area. But when the attitude object draws little elaboration and its meaning is instantaneously incorporated into a gist of what the consumer already knows. The lack of elaboration might reduce the distinct memory for the object or the recognition probability.

How might these two distinct processes of cognitive categorization — perceptual and functional — impact liking for the logo? Research in cognitive psychology suggests that consumers like attitude objects that feel easy to process (Langlois and Roggman (1990), Bornstein and D'Agostino (1994), Halberstadt and Rhodes (2000), Lee and Labroo (2004), and Landwehr et al. (2011)). The idea is that information that feels easy to process because its information is perceptually clear and stands out or because its meaning is easy to elaborate on is liked more. These findings imply that logos that are visually prototypical are likely to pop out and be liked more, and if they encourage meaningful elaboration, they will be liked even more. But if the meaning is very ambiguous, they will be liked less.

To operationalize and quantify these two constructs, we distribute a half fraction of our bigger dataset including the annotated subset into five random train and test splits. We run a series of experiments predicting labeled categories based on raw pixels varying three aspects:

1. The network architectures:
   AlexNet (Krizhevsky, Sutskever and Hinton 2012b), VGG16 (Simonyan and Zisserman 2014), VGG19 (Simonyan and Zisserman 2014)), Inception-V3 (Szegedy, Liu, Jia, Sermanet, Reed, Anguelov, Erhan, Vanhoucke and Rabinovich 2015), and ResNet-50 (He, Zhang, Ren and Sun 2016);

2. Pre-training: ImageNet (Deng, Dong, Socher, Li, Li and Fei-Fei 2009) pre-trained model with transfer learning (only the last layer was re-trained), ImageNet pre-trained model followed by fine-tuning (the last three layers were re-trained for 15 and 20 epochs for memorability and likability prediction, respectively), training from scratch without pre-training; a Euclidean loss layer is used since memorability or likability is a single real-valued output;

[2]Attitude objects are what you make a judgment about or have a positive or negative feeling toward.

3. Weight initialization: random initialization drawn from Gaussian distributions (Krizhevsky, Sutskever and Hinton 2012c) or robust initialization proposed in (He, Zhang, Ren and Sun 2015);

3.4 Functional Categorization

We specify functional categorization measures to quantify if/how the underlying brand identity is perceived, for which we propose two information-theory based measures using DCNN prediction results: entropy and Kullback-Leibler divergence. Specifically, we use the readily available industrial category information as outcome labels for prediction varying other aspects detailed in Section 3.3. We pick the deep residual network (He et al. 2016), ImageNet pre-training with fine-tuning and random initialization based on comparative prediction results on pilot batches on a random subsample of size $10,000$. The output of the last fully-connected layer in fine-tuning is fed to a 39-way softmax layer at the end, representing the 39 different industrial identities. Given the output distribution of our fine-tuned ResNet for logo category classification, denoted as Q, and the true category label, denoted as P, we define our measure regarding functional categorization as the Shannon entropy of Q, $H(Q)$, the output distribution of the deep network:

$$H(Q) = -\sum_{c=1}^{C} Q_c \log(Q_c) \tag{2}$$

where Q represents the inferred category label distribution given by the DCNN. We plot the empirical distributions of proposed measures of functional categorization in Figure 5. Intuitively, entropy on functional categorization results reflects only the perceived distributional identity. Thus Entropy measure adheres to Law of Large Numbers, free of industrial constraints.

3.5 Perceptual Categorization

We measure perceptual categorization in the same way as functional categorization except that the entropy is calculated based on the predicted distribution across 1000 concrete objects for each design (denoted as Q, the same Shannon Entropy of Q is detailed in Equation 2, obtained from a fine-tuned ResNet-50 (He et al. 2016) based on ImageNet initializations.

We propose and measure these constructs based on previous results that draw on fundamental findings in cognitive sciences about the way the visual system operates — the idea that people's preference for any design depends on the extent to which its visual processing is surprisingly fluent. This processing arguably depends on two uncorrelated aspects of a visual design — processing expectation ex-ante and processing efficiency ex-post. When a design is more prototypical, fewer neural resources are recruited, and therefore the focal design is processed quickly. Such efficient processing results in an unconscious and automatic positive affective response (Winkielman and Cacioppo 2001). When people expect a certain level of processing difficulty, this fast and automatic affect increases perceived likability of design because processing expectation hinges on the visual complexity of the design. When a design is visually intricate, people are unable to attribute the positive affect evoked by efficient processing to specific design characteristics, and they therefore subconsciously infer that the positive reaction must imply that they like the design. On the other hand, when the processing expectation is low because a design is visually simple, people attribute the affective response arising from processing efficiency to design simplicity, and they correct for an increase of affect on their preference toward the design (Winkielman, Halberstadt, Fazendeiro and Catty 2006, Landwehr et al. 2011).

Intuitively, this metric implies the extent to which the logo visual elements overlap with recognizable object categories as prototypes, and therefore, we refer to it as a Perceptual Prototypicality measure, based on theories in perception research that we have detailed in Section 1. This metric might also appeal to the concreteness (abstractness) of the design pattern, which appears to be another open question yet to be resolved.

3.6 Memorability and Likability Scores

We obtained memorability scores for a subset of the entire dataset using an online experimental procedure that is adapted from the efficient visual memory game developed in Khosla, Xiao and Torralba (2012), Khosla, Bainbridge and Torralba (2013), Khosla et al. (2014), and Khosla et al. (2015b). We elicit liking scores by asking subjects to rate how much do they like the logos being shown on a scale from 0 to 7.

We recruited $38,542$ US-based subjects from Amazon Mechanical Turk and obtained 66 scores of both memorability and likability from each Turker, resulting in around 20 scores per logo. We calculated the estimated memorability and likability scores for each logo following

the optimization procedure adopted in Khosla et al. (2015b). In the game, Turkers view a sequence of logos, each of which is displayed for 1 second, with a 1.4 second gap in between logo representations. Their task is to click on a button whenever they see a repeat of a logo. Each task is designed to last about 4.5 minutes consisting of a total of 186 logos divided into 66 targets, 30 filters, and 12 vigilance repeats. Targets are repeated after at least 35 logos, and at most 150 logos. Vigilance repeats are shown within 7 logos from the first showing. The vigilance repeats ensure that Amazon Mechanical Turkers are indeed paying attention. Turkers failing more than 25% of the vigilance repeats are blocked, and all their answers are discarded. Figure 6 illustrates the experimental procedure in a flow-chart. We refer readers to Khosla et al. (2012), and Khosla et al. (2015b) for more procedural details.

Fig. 7 shows sample designs from our dataset arranged by annotated scores.

To guard against potential biases introduced by consumers' familiarity with the brand or the company associated with the logo, we first explicitly asked them not to base their decision on how much they like the brand for reasons other than aesthetics. Secondly, likability scores were elicited as part of a post-experiment questionnaire, that included a question about the subject's familiarity of the brands being shown. Surprisingly, the percentage of encounters of familiar logos or brands account for less than 3%, we set the recognition threshold to 40% to reduce noise) in our annotated set, indicating that such a bias is secondary in our setting. Besides, even if there were still some residual effects due to omitted variable bias (OVB, hereafter) from familiarity, we argue that, based on the statistical theory of (generalized) linear models, it would not alter our results on the effects of perceptual and functional categorization on memorability or likability, because (1) if we were to view it as an OVB, that familiarity is unlikely to correlate with fluency measures in our setting would make it trouble-free (even if it were correlated, marketing expenses of the brand would be a valid instrumental, among others, to mediate the bias, because it only influences memorability or likability through increased exposure — familiarity); (2) if we were to view it as a measurement error (exact value + noise) in the dependent variable of likability, it wouldn't bias the results either — theoretical derivations are in Chapter 4, or Page 63 of the 6th edition of Greene (2000).

3.7 Image Pre-processing

We briefly describe the pre-processing steps we take to transform logo design images (JPEG files) into numerical data (matrices of floating points).

1. We convert all images to RGB (if not already after reading images with Python packages OpenCV and skimage) for further processing at different stages;

2. We then resize images into the same height and width (sometimes including image registration);

3. Simple rescaling: we rescale pixel values in the range [0,255] to [0,1] by dividing the data by 255;

4. ZCA (Zero Components Analysis) Whitening: we remove second-order correlations by applying a whitening matrix to image data. The hypothesis is that this might make the model more likely to discover interesting regularities in the images rather than merely learn that nearby pixels are similar. Predictably, the transformation preserves edge information but sets to zero pixels in regions of relatively uniform color;

3.8 Summary Statistics

We summarize the industrial category distribution of our business logo dataset in Figure 8.

We detail the distribution of logos in our annotated dataset across the most categories in the spider charts of Figure 9. US logos are more evenly distributed across categories including food and drinks, technology, media, automobile, music and finance in our annotated sample.

4. Theory and Hypotheses

Continuing on discussions in Section 1 and Section 3.3, we take cues from perception research on how the visual system operates — the idea that peoples' preference for any design depends on the extent to which its visual processing is surprisingly fluent, which in turn, depends on two aspects of a visual design — processing expectation and processing efficiency. Processing efficiency results whenever a design is more prototypical, with fewer neural resources recruited, and is processed quickly. Such quick, efficient processing results in an

automatic subconscious positive affective response (Winkielman and Cacioppo 2001), which increases the perceived likability of a design. For logo design, processing efficiency could come from either perceptual categorization or functional categorization, both of which work in the same direction via similar mechanisms. In addition, when the logo contains texts, which helps to reveal the identity of the company or brand the focal logo represents, thus making functional categorization processes more efficient, it becomes more likable to perceivers via the same mechanism. Hence we hypothesize that information processing efficiency resulting from either categorization mechanism increases liking for the focal design, as are detailed in Hypothesis 1a and Hypothesis 2, where the effect of functional categorization is further expanded by Hypothesis 1b.

Hypothesis 1a (Functional Categorization) *Information processing difficulty caused by functional categorization processes reduces liking and memory for the logo.*

Hypothesis 1b (Presence of Texts) *The presence of texts increase liking for the logo, as opposed to non-text logos, but has no effects on memory for the logo.*

Hypothesis 2 (Perceptual Categorization) *Information processing difficulty caused by perceptual categorization processes reduces liking and memory for the logo.*

On the other hand, processing expectation hinges on the visual complexity of the design. When people expect difficulty in processing, they are unable to attribute any automatically occurring subconscious affect evoked by efficient processing to specific design characteristics, and they therefore subconsciously infer that the positive reaction must imply that they like the design.

Contrarily, when the processing expectation is low because a design is visually simple, people attribute the autonomous affective response arising from processing efficiency to design simplicity, and they correct for an increase of affect on their preference toward the design (Winkielman et al. 2006, Landwehr et al. 2011). The complexity of a visual design manifests in various ways such as color variation, shape characteristics, etc. The more complex a design is, the more likely it exhibits a more significant variation in color and shape. Therefore, we hypothesize that a more significant variation in colors increases liking, and as do a greater variation in shape — a higher degree of angularity, and a more significant number of corners. We detail them in Hypothesis 3a and Hypothesis 3b.

Hypothesis 3a (Color Variation) *Greater variation in colors increase liking but reduces memory for the logo.*

Hypothesis 3b (Angularity) *Greater angularity in shape (the number of corners) increase liking but reduces memory for the logo.*

Regarding particular design patterns such as symmetry and boundedness, we hypothesize that the Gestalt phenomena in perception and cognition (i.e., Zeigarnik effect) manifest in the context of logo design as greater memory and affinity towards designs that appear to be more symmetric and complete. We state these ideas more formally in Hypothesis 4a and Hypothesis 4b. Per Section 3.2, we assume that more perceptually bounded and complete designs are more likely to show more detected circles, polygons and possibly fewer detected lines.

Hypothesis 4a (Boundedness) *Complete and Bounded design patterns increase memory and liking for the logo as opposed to incomplete and unbounded design patterns.*

Hypothesis 4b (Symmetry) *More symmetric design patterns increase memory and liking for the logo as opposed to asymmetric design patterns.*

Given the fact that guidance from existing literature is limited, we acknowledge there is a gap between extracted low-level image features and relevant hypotheses detailed above. We proceed to test such hypotheses as a starting point to inspire more future work on this front.

5. Empirical Models

5.1 Variables for Empirical Analysis

- Dependent variables Memory_i and Liking_i: we use Memory)i to denote the memorability score of logo i collected from our online experiments and calculated based on the optimization algorithm following Isola, Xiao, Parikh, Torralba and Oliva (2014), while suppressing time subscripts for simplicity; we use Liking_i to denote the average likability score of logo i at the end of each online session;

- Category_{ij}: the dummy variable of industrial category j of logo i;

- Color$_{ij}$: the dummy variable of color cluster j of logo i, as detailed in Section 3.2;

- Perceptual_Categorization_i, Perceptual_Categorization_i (Sq): we use these to denote the calculated entropy values (and the squared values) of logo i based on perceptual classification results — the classification of 1000 daily object categories detailed in Section 3.5;

- Functional_Categorization_i, Functional_Categorization_i (Sq): we use these to denote the calcualted entropy values (and the squared values) of logo i based on functional classification results — the classification of 39 industry membership categories detailed in Section 3.4;

- Lines_i, Circles_i, Polygons_i, Texts_i, Corners_i: we denote these as the number of detected lines, circles, polygons, text boxes, and corners of logo i using corresponding vision algorithms which are detailed in Section 3.2;

- Distinct_Colors_i, Whitespace_i: referring to the number of resulting different color clusters and white space ratios of logo i, as detailed in Section 3.2;

- Symmetry_i, Symmetry_Distance_i: referring to the number of detected matching key points and average distances between matched key points of logo i, using the symmetry detection algorithm detailed in Section 3.2. The greater Symmetry_i is, and the smaller Symmetry_Distance_i is, the more symmetric the logo image is;

- Mark_Size_i, Mark_Widths_i: these refer the average size of detected bounding boxes of marks and the average widths thereof, of logo i, based on YOLO algorithm detailed in Section 3.2;

- Hue_i, Saturation_i, Lightness_i, Hue_Sd_i, Saturaion_Sd_i, Lightness_Sd_i: means and standard deviations of the corresponding dimension in HSV colorspace, of logo i.

5.2 Empirical Model

The full specification of our model is given by Equations 3.

$$
\begin{aligned}
\mathrm{Y}_i = \alpha_i + \textstyle\sum_{j=1}^{39} \beta_{1j}\mathrm{Category}_{ij} + \\
\beta_2\mathrm{Perceptual\_Categorization}_i + \beta_3\mathrm{Perceptual\_Categorization}_i^2 + \\
\beta_4\mathrm{Functional\_Categorization}_i + \beta_5\mathrm{Functional\_Categorization}_i^2 + \\
\beta_6\mathrm{Lines}_i + \beta_7\mathrm{Circles}_i + \beta_8\mathrm{Polygons}_i + \beta_9\mathrm{Texts}_i + \\
+\beta_{10}\mathrm{Corners\_}i + \beta_{11}\mathrm{Distinct\_Colors}_i + \beta_{12}\mathrm{Whitespace}_i + \\
\beta_{13}\mathrm{Symmetry}_i + \beta_{14}\mathrm{Symmetry\_Distance}_i + \beta_{15}\mathrm{Mark\_Size}_i + \beta_{16}\mathrm{Mark\_Widths}_i + \\
\beta_{17}\mathrm{Hue}_i + \beta_{18}\mathrm{Hue\_}Sd_i + \beta_{19}\mathrm{Saturation}_i + \beta_{20}\mathrm{Saturation\_}Sd_i + \beta_{21}\mathrm{Lightness}_i + \beta_{22}\mathrm{Lightness\_}Sd_i + \\
\sum_{j=1}^{225} \beta_{23j}\mathrm{Color}_{ij} + \\
\epsilon_i
\end{aligned}
\tag{3}
$$

where $Y_i \in \{\mathrm{Memory}_i, \mathrm{Liking}_i\}$ and ϵ_i is assumed to be Gaussian.

Our main results tabulated in Section 6 are based on the model that only includes variables from the first to the fourth line in Equation 3 (excluding or not excluding the quadratic terms), whereas full models are of the whole Equation 3 (excluding or not excluding the quadratic terms and the dummy variables for colors and categories).

6. Results

6.1 Regression Results

We summarize the regression results in Tables tables 6, 8 and 10, where Table 6 shows results from regression models in which only focal variables of cognitive categorization, basic low-level features, and the industrial category dummies, Table 8 shows results from models in which all non-correlated variables of low-level features included. Table 10 shows results

Table 2: Ordinary Least Squares Estimation Predicting Memorability and Likability

	Dependent variable:	
	Memory	Liking
Perceptual_Categorization	-0.005^{***}	0.038^{***}
	(0.002)	(0.010)
Perceptual_Categorization (Sq)	0.001^{***}	-0.008^{***}
	(0.0002)	(0.001)
Functional_Categorization	-0.027^{***}	-0.209^{***}
	(0.002)	(0.011)
Functional_Categorization (Sq)	-0.0002	0.053^{***}
	(0.001)	(0.004)
Lines	-0.003^{***}	-0.034^{***}
	(0.0001)	(0.001)
Circles	0.019^{***}	0.144^{***}
	(0.001)	(0.004)
Polygons	0.023^{***}	0.011^{***}
	(0.001)	(0.004)
Texts	0.00003	0.010^{***}
	(0.00002)	(0.0002)
Constant	0.768^{***}	2.893^{***}
	(0.006)	(0.041)
Observations	125,270	125,270
R^2	0.091	0.100
Adjusted R^2	0.091	0.099
Residual Std. Error (df = 125223)	0.151	0.973
F Statistic (df = 46; 125223)	272.155^{***}	301.013^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 3: Elasticities of Results in Table 2

	Memory	Liking
Perceptual Categorization	-0.019	0.149
Perceptual Categorization (Sq)	0.011	-0.135
Functional Categorization	-0.024	-0.189
Functional Categorization (Sq)	-0.0003	0.090
Lines	-0.005	-0.069
Circles	0.023	0.172
Polygons	0.068	0.031
Texts	0.001	0.212

Table 4: Ordinary Least Squares Estimation Predicting Memorability and Likability

	Dependent variable:	
	Memory	Liking
Perceptual_Categorization	−0.003	0.039
	(0.005)	(0.033)
Perceptual_Categorization (Sq)	−0.0001	−0.013**
	(0.001)	(0.005)
Functional_Categorization	−0.001	−0.329***
	(0.006)	(0.039)
Functional_Categorization (Sq)	−0.009***	0.102***
	(0.002)	(0.015)
Distinct_Colors	−0.00000	−0.00001
	(0.00000)	(0.00001)
Whitespace	0.00000	0.00000
	(0.00000)	(0.00000)
Mark_Size	−0.00000	0.00001
	(0.00000)	(0.00001)
Corners	−0.0001***	0.0001
	(0.00001)	(0.0001)
Symmetry	0.0002***	0.003***
	(0.00005)	(0.0003)
Symmetry_Distance	0.00004*	0.001***
	(0.00002)	(0.0001)
Mark_Widths	0.00005	−0.001**
	(0.0001)	(0.001)
Hue	0.056**	−0.209
	(0.025)	(0.169)
Hue_Sd	−0.057**	0.400**
	(0.028)	(0.188)
Saturation	−0.020	−0.081
	(0.028)	(0.188)
Saturation_Sd	0.018	−0.024
	(0.027)	(0.181)
Lightness	−0.012	0.099
	(0.029)	(0.194)
Lightness_Sd	0.016	0.016
	(0.024)	(0.159)
Lines	−0.003***	−0.058***
	(0.001)	(0.003)
Circles	0.040***	0.202***
	(0.002)	(0.016)
Polygons	0.019***	0.042***
	(0.002)	(0.013)
Texts	0.002	0.119***
	(0.006)	(0.042)
Constant	0.768***	1.863***
	(0.039)	(0.264)
Observations	125,270	125,270
R^2	0.129	0.146
Adjusted R^2	0.107	0.124
Residual Std. Error (df = 124998)	0.148	1.004
F Statistic (df = 271; 124998)	50.707***	66.594***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 5: Elasticities of Results in Table 4

	Memory	Liking
Perceptual_Categorization	-0.010	0.147
Perceptual_Categorization (Sq)	-0.001	-0.204
Functional_Categorization	-0.001	-0.263
Functional_Categorization (Sq)	-0.012	0.145
Distinct_Colors	-0.00002	-0.003
Whitespace	0.0004	0.012
Mark_Size	-0.001	0.005
Corners	-0.027	0.022
Symmetry	0.013	0.271
Symmetry_Distance	0.011	0.338
Mark_Widths	0.001	-0.027
Hue	0.019	-0.071
Hue_Sd	-0.009	0.066
Saturation	-0.003	-0.011
Saturation_Sd	0.004	-0.005
Lightness	-0.011	0.088
Lightness_Sd	0.003	0.003
Lines	-0.005	-0.107
Circles	0.041	0.208
Polygons	0.059	0.128
Texts	0.0004	0.029

from models exactly the same as Table 8 except dummy variables for colors.

We add quadratic terms of our focal cognitive categorization variables, as are shown in Table 2 in accordance with Table 6, Table 12 in accordance with Table 8, and Table 4 in accordance with Table 10. A heatmap of the correlation matrix encompassing most included variables is shown in Figure 10.

As are shown in our results, the effects of two cognitive categorization constructs are robust across various model specifications and are significantly negative, in the sense that the greater the entropy measures — the more ambiguous the design pattern appears — the less memorable and likable the the focal design looks to perceivers.

The effects of the number of detected lines, circles, and polygons are consistently significant — more lines correspond to greater memorability and likability and the reverse is true for the number of circles and polygons, in accordance with Hypothesis 4a.

The robust effect of symmetry detected in designs is also expected in that greater symmetry increases memory and liking.

Variables associated with perceived image complexity, such as variation in color and shape (the number of detected corners) appear less consistent in terms of their effects on memorability and liking. However, by removing collinear variables, their effects appear significant and consistent with Hypothesis 3a and Hypothesis 3b.

The effect of text presence proves robust and consistent across situations, adding validity to the argument grounded in functional categorization, as is detailed in Hypothesis 1b and Hypothesis 1a.

6.2 Elasticity Results

We tabulate resulting elasticities in accordance with regression tables in Table 7 (based on results in Table 6), Table 9 (based on results in Table 8), Table 11 (based on results in Table 10), Table 3 (based on results in Table 2), Table 13 (based on results in Table 12), and lastly, Table 5 (based on results in Table 4).

Elasticities of entropies based on functional categorization consistently outweigh those of perceptual categorization to memory by a large margin across all model specifications as are shown in all the elasticity tables. This implies that changes in functional categorization have a much more significant impact on design memorability than perceptual categorization. The only image features that consistently exhibit elasticities of comparable magnitude are

the number of detected corners, the number of detected circles, and the number of identified polygons.

On the other hand, functional categorization appears to have a greater impact than perceptual categorization on liking in most cases but not all. The image features that appear to have comparable effects on liking based on elasticities are symmetry measures, color features — means and standard deviations of hue, saturation, and lightness, the number of detected circles, the number of detected polygons, and the number of detected text boxes.

Notably, the elasticity of the number of circles could be as large as more than three times that of either categorization entropy metrics regarding liking, but about the same as that of functional categorization metric regarding memory. It is the same with variables associated with symmetry. In contrast, the elasticity of the number of polygons regarding memory is about greater than that regarding liking, both of which are comparable to the elasticities of functional categorization measure. Such observations could provide extra evidence for Hypothesis 4a and Hypothesis 4b.

We propose three hypotheses with supporting theories to help explain such elasticity results.

The first hypothesis follows our discussion in Section 4. In addition, Reber, Schwarz and Winkielman (2004) showed that the impact of fluency is moderated by expectations and attribution. On one hand, fluency has a particularly strong impact on affective experience if its source is unknown and fluent processing comes as a surprise. On the other hand, the fluency-based affective experience is discounted as a source of relevant information when the perceiver attributes the experience to an irrelevant source. We conjecture that functional categorization processes involve uncertainty regarding information sources, thus positively moderating the fluency effects on affection, making it more important compared to perceptual categorization, which does not involve information about sources.

The second hypothesis involves a dual-process theory — the elaboration likelihood theory (Petty and Cacioppo 1986, Petty, Cacioppo and Goldman 1981, Bhattacherjee and Sanford 2006). Elaboration Likelihood Theory posits that attitude change among individuals may be caused by two routes of influence, the central route and the peripheral route, which differ in the amount of thoughtful information processing or elaboration demanded of individual subjects (Petty et al. 1981, Petty and Cacioppo 1986). The central route requires a person to think critically about issue-related arguments in an informational message and scrutinize the relative merits and relevance of those arguments prior to forming an informed

judgment about the target behavior. The peripheral route involves less cognitive effort, where subjects rely on cues regarding the target behavior. As Petty, Wegener and Fabrigar (1997) put it, "The term 'elaboration' is used to suggest that people add something of their own to the specific information provided in the communication ... beyond mere verbatim encoding of the information provided. (p. 46)" We hypothesize that functional categorization involves greater elaboration and information processing than perceptual categorization, since it contains first identifying the physical patterns the same as perceptual categorization and making further inferences about the underlying meaning. Because the external expectation is held the same for both categorization mechanisms, the discrepancy that influences likability therefore is greater for functional categorization than perceptual categorization. However, when it comes to memory, the finer-granularity of perceptual categorization than functional categorization make it more costly as a mental process and therefore more prominent than functional categorization. This echoes the third hypothesis based on construal level theory below.

The third hypothesis is based on the Construal Level Theory (Liberman, Trope and Wakslak 2007, Liberman et al. 2007). Trope, Liberman and Wakslak (2007), from the perspective of consumer psychology, provide an excellent review of Construal level theory (CLT) — an account of how psychological distance influences individuals' thoughts and behavior. CLT assumes that people mentally construe objects that are psychologically near in terms of low-level, detailed, and contextualized features, whereas at a distance they construe the same objects or events in terms of high-level, abstract, and stable characteristics. Research has shown that different dimensions of psychological distance (time, space, social distance, and hypotheticality) affect mental construal and that these construals, in turn, guide prediction, evaluation, and behavior. We hypothesize that functional categorization is a high-level construal construct, general, abstract, average, and therefore more difficult to remember and more likable; whereas perceptual categorization a low-level construal construct, detailed, concrete, extreme, and therefore easier to remember, and less likable, all else being equal.

7. Image Clustering

In order to address the potential problem of heterogeneity inherent in analyses in Section 5, we explore multiple image clustering methods to identify distinct clusters in our samples and

Table 6: Ordinary Least Squares Estimation Predicting Memorability and Likability

	Dependent variable:	
	Memory	Liking
Perceptual_Categozation	-0.027^{***}	-0.074^{***}
	(0.001)	(0.003)
Functional_Categorization	-0.001^{*}	-0.016^{***}
	(0.0004)	(0.002)
Lines	-0.003^{***}	-0.034^{***}
	(0.0001)	(0.001)
Circles	0.019^{***}	0.146^{***}
	(0.001)	(0.004)
Polygons	0.023^{***}	0.013^{***}
	(0.001)	(0.004)
Texts	0.00003	0.010^{***}
	(0.00002)	(0.0002)
Constant	0.763^{***}	2.928^{***}
	(0.006)	(0.040)
Observations	125,270	125,270
R^2	0.091	0.098
Adjusted R^2	0.091	0.098
Residual Std. Error (df = 125225)	0.151	0.974
F Statistic (df = 44; 125225)	284.322^{***}	309.472^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 7: Elasticities of Results in Table 6

	Memory	Liking
Perceptual Categorization	-0.003	-0.063
Functional Categorization	-0.024	-0.066
Lines	-0.005	-0.068
Circles	0.023	0.174
Polygons	0.067	0.039
Texts	0.001	0.214

Table 8: Ordinary Least Squares Estimation Predicting Memorability and Likability

	Dependent variable:	
	Memory	Liking
Perceptual_Categorization	-0.003^{**}	-0.045^{***}
	(0.001)	(0.008)
Functional_Categorization	-0.021^{***}	-0.074^{***}
	(0.002)	(0.012)
Distinct_Colors	0.00000	-0.00001
	(0.00000)	(0.00001)
Whitespace	0.00000	0.00000
	(0.00000)	(0.00000)
Mark_Size	-0.00000	0.00001
	(0.00000)	(0.00001)
Corners	-0.0001^{***}	0.0001
	(0.00001)	(0.0001)
Symmetry	0.0002^{***}	0.003^{***}
	(0.00005)	(0.0003)
Symmmetry_Distance	0.00003^{*}	0.001^{***}
	(0.00002)	(0.0001)
Mark_Widths	0.00005	-0.001^{**}
	(0.0001)	(0.001)
Hue	0.051^{**}	-0.219
	(0.025)	(0.170)
Hue_Sd	-0.057^{**}	0.396^{**}
	(0.028)	(0.189)
Saturation	-0.019	-0.094
	(0.028)	(0.189)
Saturation_Sd	0.016	-0.002
	(0.027)	(0.182)
Lightness	-0.012	0.105
	(0.029)	(0.195)
Lightness_Sd	0.017	0.008
	(0.024)	(0.160)
Lines	-0.003^{***}	-0.058^{***}
	(0.001)	(0.003)
Circles	0.040^{***}	0.203^{***}
	(0.002)	(0.016)
Polygons	0.019^{***}	0.051^{***}
	(0.002)	(0.013)
Texts	0.002	0.125^{***}
	(0.006)	(0.042)
Constant	0.774^{***}	1.871^{***}
	(0.039)	(0.262)
Observations	125,270	125,270
R^2	0.128	0.142
Adjusted R^2	0.115	0.121
Residual Std. Error (df = 125000)	0.148	1.006
F Statistic (df = 269; 125000)	65.682^{***}	66.406^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 9: Elasticities of Results in Table 8

	Memory	Liking
Perceptual_Categorization	-0.011	-0.167
Functional_Categorization	-0.018	-0.061
Distinct_Colors	0.0002	-0.006
Whitespace	0.001	0.005
Mark_Size	-0.001	0.003
Corners	-0.027	0.030
Symmetry	0.012	0.271
Symmetry_Distance	0.009	0.351
Mark_Widths	0.001	-0.025
Hue	0.007	0.003
Hue_Sd	-0.006	0.045
Saturation	-0.001	-0.008
Saturation_Sd	0.002	-0.005
Lightness	0.006	0.029
Lightness_Sd	0.005	-0.006
Lines	-0.005	-0.106
Circles	0.042	0.212
Polygons	0.057	0.163
Texts	0.0005	0.028

Table 10: Ordinary Least Squares Estimation Predicting Memorability and Likability

	Dependent variable:	
	Memory	Liking
Perceptual_Categorization	-0.003^{**}	-0.044^{***}
	(0.001)	(0.008)
Functional_Categorization	-0.022^{***}	-0.076^{***}
	(0.002)	(0.012)
Distinct_Colors	0.00000	-0.00002
	(0.00000)	(0.00001)
Whitespace	0.00000	0.00000
	(0.00000)	(0.00000)
Mark_Size	-0.00000	0.00000
	(0.00000)	(0.00001)
Corners	-0.0001^{***}	0.0001^{*}
	(0.00001)	(0.0001)
Symmetry	0.0001^{***}	0.003^{***}
	(0.00005)	(0.0003)
Symmetry_Distance	0.00003	0.001^{***}
	(0.00002)	(0.0001)
Mark_Widths	0.00004	-0.001^{**}
	(0.0001)	(0.001)
Hue	0.021	0.009
	(0.019)	(0.130)
Hue_Sd	-0.034	0.273
	(0.025)	(0.169)
Saturation	-0.005	-0.060
	(0.019)	(0.129)
Saturation_Sd	0.009	-0.025
	(0.022)	(0.147)
Lightness	0.007	0.032
	(0.020)	(0.138)
Lightness_Sd	0.027	-0.034
	(0.021)	(0.140)
Lines	-0.003^{***}	-0.058^{***}
	(0.001)	(0.003)
Circles	0.040^{***}	0.205^{***}
	(0.002)	(0.016)
Polygons	0.019^{***}	0.054^{***}
	(0.002)	(0.013)
Texts	0.002	0.112^{***}
	(0.006)	(0.042)
Constant	0.756^{***}	1.970^{***}
	(0.034)	(0.232)
Observations	125,270	125,270
R^2	0.107	0.122
Adjusted R^2	0.102	0.118
Residual Std. Error (df = 125213)	0.149	1.007
F Statistic (df = 56; 125213)	222.775^{***}	226.489^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 11: Elasticities of Results in Table 10

	Memory	Liking
Perceptual_Categorization	-0.010	-0.169
Functional_Categorization	-0.018	-0.061
Distinct_Colors	0.00002	-0.003
Whitespace	0.001	0.010
Mark_Size	-0.001	0.005
Corners	-0.027	0.027
Symmetry	0.013	0.272
Symmetry_Distance	0.010	0.351
Mark_Widths	0.001	-0.030
Hue	0.019	-0.074
Hue_Sd	-0.009	0.065
Saturation	-0.002	-0.012
Saturation_Sd	0.003	-0.0004
Lightness	-0.011	0.094
Lightness_Sd	0.003	0.002
Lines	-0.005	-0.106
Circles	0.041	0.210
Polygons	0.058	0.156
Texts	0.0004	0.031

Table 12: Ordinary Least Squares Estimation Predicting Memorability and Likability

	Dependent variable:	
	Memory	Liking
Perceptual_Categorization	−0.003	0.045
	(0.005)	(0.033)
Perceptual_Categorization (Sq)	−0.0001	−0.014***
	(0.001)	(0.005)
Functional_Categorization	0.001	−0.321***
	(0.006)	(0.038)
Functional_Categorization (Sq)	−0.009***	0.099***
	(0.002)	(0.015)
Distinct_Colors	0.00000	−0.00002
	(0.00000)	(0.00001)
Whitespace	0.00000	0.00000
	(0.00000)	(0.00000)
Mark_Size	−0.00000	0.00000
	(0.00000)	(0.00001)
Corners	−0.0001***	0.0001
	(0.00001)	(0.0001)
Symmetry	0.0002***	0.003***
	(0.00005)	(0.0003)
Symmetry_Distance	0.00003	0.001***
	(0.00002)	(0.0001)
Mark_Widths	0.00004	−0.001*
	(0.0001)	(0.001)
Hue	0.019	0.019
	(0.019)	(0.130)
Hue_Sd	−0.033	0.272
	(0.025)	(0.168)
Saturation	−0.005	−0.055
	(0.019)	(0.129)
Saturation_Sd	0.010	−0.038
	(0.022)	(0.147)
Lightness	0.006	0.038
	(0.020)	(0.138)
Lightness_Sd	0.025	−0.018
	(0.021)	(0.140)
Lines	−0.003***	−0.058***
	(0.001)	(0.003)
Circles	0.040***	0.204***
	(0.002)	(0.016)
Polygons	0.019***	0.044***
	(0.002)	(0.013)
Texts	0.002	0.105**
	(0.006)	(0.042)
Constant	0.750***	1.949***
	(0.035)	(0.234)
Observations	125,270	125,270
R^2	0.109	0.127
Adjusted R^2	0.104	0.122
Residual Std. Error (df = 125211)	0.149	1.005
F Statistic (df = 58; 125211)	222.341***	226.635***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 13: Elasticities of Results in Table 12

	Memory	Liking
Perceptual_Categorization	-0.010	0.169
Perceptual_Categorization (Sq)	-0.002	-0.218
Functional_Categorization	0.0004	-0.257
Functional_Categorization (Sq)	-0.013	0.141
Distinct_Colors	0.0001	-0.006
Whitespace	0.001	0.005
Mark_Size	-0.001	0.003
Corners	-0.027	0.025
Symmetry	0.012	0.271
Symmetry_Distance	0.009	0.337
Mark_Widths	0.001	-0.023
Hue	0.006	0.006
Hue_Sd	-0.005	0.045
Saturation	-0.001	-0.007
Saturation_Sd	0.002	-0.008
Lightness	0.005	0.034
Lightness_Sd	0.005	-0.003
Lines	-0.005	-0.107
Cricles	0.042	0.211
Polygons	0.058	0.134
Texts	0.0005	0.026

test for robustness of our cognitive measures.

Image clustering has been extensively studied in the computer vision community. Based on the proliferation of image clustering literature during the past few decades, fueled by most recent advances in deep learning, we identify and experiment with three classes of image clustering methods:

1. Traditional clustering methods applied to low-level image features;
2. Unsupervised image feature learning followed by traditional clustering methods;
3. Joint unsupervised image representation learning and clustering methods,

which we detail in the following subsections.

7.1 Traditional Clustering Methods on Low-level Image Features

Traditionally, various clustering methods have been explored for images, including K-means (Wang, Wang, Song, Xu, Shen and Li 2015), agglomerative clustering (Gowda and Krishna 1978), and so forth. It is acknowledged that the fact that such traditional clustering methods depend on pre-specified distance metrics, which are difficult to identify for images, makes the effectiveness of traditional methods ambiguous when applied to image clustering problems. Therefore, for completeness, we apply the K-means clustering method to extracted low-level features (detailed in Section 3.2). All the low-level image features were properly pre-processed via one-hot encoding and scaling before clustering. We refer to this method as Method 1 in the following evaluation and comparison section.

7.2 Traditional Clustering Methods on Learned Image Representation

With the explosion of deep learning, various deep unsupervised feature learning methods have been proposed to learn image representations that are potentially informative and explainable. Methods such as autoencoder and its various variants (sparse autoencoder (Ng 2011), denoising autoencoder (Vincent, Larochelle, Lajoie, Bengio and Manzagol 2010)), were proposed to maximize the similarities between reconstructed and original images. In addition, deep generative models such as the autoencoding variational Bayes (Kingma and

Welling 2013) and the generative adversarial network (GAN) (Goodfellow et al. 2014a) were popular for image synthesis, encoding, and generation. As has been documented in Chang, Wang, Meng, Xiang and Pan (2017), the learned representations from generative adversarial models (GANs) outperform other unsupervised learning methods, such as various autoencoders and deconvolutional neural nets, by a large margin, when fed to downstream clustering algorithms. Therefore, we train a vanilla generative adversarial network (Goodfellow et al. 2014a) on the entire logo design dataset to obtain a distributed representation in the feature space. Following Chang et al. (2017), we apply K-means to cluster images as post-processing. We refer to this method as Method 2 in the following evaluation and comparison section.

Another widely used method to obtain image representations for clustering is to train supervised deep convolutional neural network (DCNN), as in Section 3.3, and retain weights from the (second) last fully-connected layer as inputs to clustering algorithms. We explore this method in Section **??** of image cluster visualization.

7.3 Joint Unsupervised Representation Learning and Clustering

Several methods have been proposed in the computer vision community to combine unsupervised feature learning with clustering. For instance, inspired by the parametric t-SNE (Maaten and Hinton 2008, Wattenberg, Viégas and Johnson 2016, Van Der Maaten 2014), Xie, Girshick and Farhadi (2016) introduced deep embedded clustering (DEC) for learning cluster centers, which left the question of effectively pre-training neural networks in the context of clustering unresolved. Yang, Parikh and Batra (2016) provides one of the first solutions by integrating a representation learning Convolutional Neural Network with an agglomerative clustering algorithm in a recurrent framework. They term it JULE — Joint Unsupervised Learning of deep representations and image clusters. By assimilating two processes into a single model with a unified weighted triplet loss, and optimizing end-to-end, the joint model improves on the performances of both the representation learning and the clustering tasks. In the same vein but vastly different implementation and design, Chang et al. (2017) introduced Deep Adaptive Clustering method that adaptively learns pairwise similarities between label features generated by a deep Convolutional Neural Network from paired images.

One potential drawback of Yang et al. (2016), as pointed out and resolved by Chang et al.

(2017), depends on its initialization — over-clustering initialized with K Nearest Neighbors. Since the distances between different images are difficult to define, beginning with the over-clustering may degrade its performance, especially when images are too complicated. We choose to implement Yang et al. (2016) for our application due to both the easily available source codes (unlike Chang et al. (2017), who have yet to make source codes available) and the fact that logo images are in-between MNIST (Deng 2012) and ImageNet (Deng et al. 2009) in terms of complexity and therefore the major pitfall of Yang et al. (2016) largely averted.

We implement DEC (Xie et al. 2016) and JULE (Yang et al. 2016) on our logo design dataset using the Keras and Torch codes released by authors on Github, respectively. We inherit all the default parameter values from Xie et al. (2016), and adopt most of the hyper-parameters used in Yang et al. (2016), except the unrolling rate for the recurrent process, which we used 0.9 rather than 0.2, as we wish to update less frequently than Yang et al. (2016) to compensate efficiency with some reasonable loss in performance. We refer to this method as Method 3 in the following evaluation and comparison section.

7.4 Comparing Clustering Algorithms and Results

Since we do not have any ground truths about the real clusters of logo images to objectively evaluate the resulting clusters from three classes of image clustering algorithms, we attempt to compare and contrast the relationships between different clustering results across a different number of clusters, using popular evaluation metrics of performance.

Following Chang et al. (2017) with metrics expansion and tailoring, we use four popular measures in the literature to evaluate the performance of clustering methods. They are Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), Normalized Information Distance (NID), and Normalized Variation Information (NVI). All four measures range in [0, 1], and higher scores indicate more similar clustering results. We use R package aricode for such efficient computations of clustering comparison methods.

In Table 14, we document all the popular evaluation metrics (the four major ones bolded) for clustering methods applied to results from Method 1 and Method 2, varying the number of clusters from 2 to 50. Other supporting evaluation metrics, which are not normalized or adjusted, include Rand Index (RI), Mutual Information (MI), Variation Information (VI), and Information Distance (ID).

	RI	**ARI**	MI	VI	**NVI**	ID	**NID**	**NMI**
2 Clusters	0.55	0.11	0.20	1.22	0.86	0.75	0.79	0.21
4 Clusters	0.62	0.11	0.23	2.21	0.91	1.16	0.84	0.16
5 Clusters	0.64	0.25	0.39	1.97	0.83	1.40	0.78	0.22
10 Clusters	0.67	0.17	0.25	2.64	0.91	1.54	0.86	0.14
15 Clusters	0.69	0.15	0.44	3.07	0.88	1.86	0.81	0.19
20 Clusters	0.76	0.15	0.39	3.64	0.90	1.83	0.82	0.18
50 Clusters	0.89	0.15	1.09	4.11	0.79	2.21	0.67	0.33

Table 14: Comparative Evaluation of Method 1 and Method 2

In Table 15, we document all the popular evaluation metrics (the four major ones bolded) for clustering methods applied to results from Method 2 and Method 3, varying the number of clusters from 2 to 50.

	RI	**ARI**	MI	VI	**NVI**	ID	**NID**	**NMI**
2 Clusters	0.51	0.01	0.10	0.96	0.99	0.72	0.90	0.02
4 Clusters	0.89	0.76	0.86	0.80	0.48	0.41	0.32	0.67
5 Clusters	0.89	0.75	0.94	0.93	0.49	0.53	0.36	0.64
10 Clusters	0.73	0.34	0.68	1.94	0.73	0.98	0.58	0.41
15 Clusters	0.84	0.44	0.98	2.37	0.70	1.21	0.55	0.44
20 Clusters	0.93	0.29	1.75	3.17	0.64	1.63	0.48	0.51
50 Clusters	0.95	0.46	2.05	2.59	0.55	1.35	0.39	0.60

Table 15: Comparative Evaluation of Method 2 and Method 3

It appears that clustering results from Method 2 are more similar to those from Method 3 than Method 1, especially when the number of clusters ranges between 4 and 10 or of large values. This seems intuitive in that both Method 2 and Method 3 are based on image embeddings whereas Method 1 does not.

7.5 Regressing Memorability and Likability on Cluster Identities

We replicate the analyses detailed in Section 5 with the inclusion of class memberships from different clustering methods as additional image features across a different number of clusters. Specifically, we regress logo memorability and likability scores on all the image features, categorization variables, and cluster membership assignments, for each clustering

method, and for each number of clusters ranging from 2 to $50 - 100$.

Table 16 shows the regression coefficients with respect to cluster memberships obtained from Method 1 (detailed in Section 7.1) when the number of clusters was fixed to be 15. Category variables were excluded due to additional correlations. Interestingly, we identify several clusters that independently and significantly influence memorability and likability. For instance, cluster 3, cluster 6, cluster 12, cluster 13, and cluster 14 have significant effects on memorability but not likability, whereas cluster 4, cluster 10, cluster 11, and cluster 15 exhibit significant effects on likability but not memorability. On the other hand, cluster 2, cluster 5, cluster 7, cluster 8, and cluster 9 show significant effects on both memorability and likability. We visualize the corresponding clusters in Section 7.6.

We tabulate the corresponding summary statistics of clusters reported in Table 16 in Table 17 and Table 18.

Table 19 shows the regression coefficients with respect to cluster memberships obtained from Method 2 (detailed in Section 7.1) when the number of clusters was fixed to be 5. Interestingly, we identify several clusters that significantly influence both memorability and likability, but in opposite directions. For instance, cluster 2 has a significant negative effect on memorability but positive effect on likability, whereas cluster 3 and cluster 5 both show significant positive effects on memorability and negative effects on likability. We collage a random subsample of logo images in cluster 2, cluster 3, and cluster 5 in Figure 11, corresponding to summary statistics provided in Table 20.

Table **??** shows the regression coefficients with respect to cluster memberships obtained from Method 2 fixing the number of clusters at 10. Here we identify clusters that significantly influence both memorability and likability, but in opposite directions, clusters that independently and significantly influence memorability and likability, as well as clusters that significantly affect both memorability and likability in the same direction. Likewise, We collage a random subsample of logo images in cluster 2, cluster 4, cluster 5, cluster 7, and cluster 9 in Figure 12, corresponding to the summary statistics in Table 21.

7.6 Visualization

We use Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction and visualization, in favor of t-SNE (Maaten and Hinton 2008) and its recent variants (e.g. Barnes-Hut t-SNE Van Der Maaten (2014)). t-Distributed Stochastic Neighbor Embed-

	Dependent variable:	
	Memory	Liking
	(1)	(2)
...	...	...
Cluster2	0.014**	0.133***
	(0.006)	(0.038)
Cluster3	0.152***	−0.185
	(0.043)	(0.286)
Cluster4	−0.0002	0.170***
	(0.006)	(0.039)
Cluster5	0.019***	0.121***
	(0.006)	(0.037)
Cluster6	0.035***	0.036
	(0.005)	(0.034)
Cluster7	0.010*	0.301***
	(0.005)	(0.034)
Cluster8	0.051***	−0.070*
	(0.006)	(0.040)
Cluster9	0.019***	0.175***
	(0.006)	(0.039)
Cluster10	0.008	−0.061*
	(0.006)	(0.036)
Cluster11	0.0001	0.163***
	(0.007)	(0.045)
Cluster12	0.024***	0.068
	(0.009)	(0.056)
Cluster13	0.036***	0.004
	(0.007)	(0.045)
Cluster14	0.035***	0.022
	(0.005)	(0.033)
Cluster15	0.002	0.240***
	(0.006)	(0.038)
Constant	0.907***	2.082***
	(0.008)	(0.053)
Observations	123,924	123,924
R^2	0.115	0.079
Adjusted R^2	0.115	0.078
Residual Std. Error (df = 123897)	0.149	0.984
F Statistic (df = 26; 123897)	619.256***	406.422***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 16: Regression Coefficients of Cluster Memberships from Method 1, 15 Clusters

Clusters	1	2	3	4	5	6	7
Hue (Mean)	61.800	60.635	46.558	60.609	60.354	60.903	60.764
Hue (Std)	29.010	29.934	32.351	29.831	29.429	29.596	29.488
Saturation (Mean)	34.308	34.806	40.849	34.027	33.197	33.795	33.928
Saturation (Std)	55.006	54.504	59.156	54.690	54.301	54.138	54.373
Value (Mean)	227.306	227.708	236.736	227.211	228.357	227.863	227.832
228.055	228.267	227.859	228.587	225.890	225.213	227.919	227.869
Value (Std)	47.333	46.096	40.879	46.695	45.569	46.015	46.170
Complexity	0.810	0.714	0.778	0.894	0.700	0.652	1.419
Greyscale (Std)	1.190	1.140	1.139	1.235	1.102	1.079	1.458
Edges	2,299.261	2,047.224	2,216.667	2,498.754	2,032.441	1,847.286	3,870.028
Lines	2.094	1.650	1.833	1.847	1.921	1.665	2.955
Circless	1.150	1.048	1.417	1.226	1.063	1.073	1.566
Polygons	2.827	2.708	3.167	2.918	2.849	2.720	3.475
Perceptual Categorization	4.056	3.909	4.167	3.832	4.081	4.195	3.332
Functional Categorization	1.048	1.194	2.424	0.874	1.140	1.117	0.319

Table 17: Summary Statistics of Clusters from Method 1, 15 Clusters

Clusters	8	9	10	11	12	13	14	15
Hue (Mean)	61.960	60.665	59.936	60.071	60.918	61.234	60.585	60.466
Hue (Std)	29.827	29.734	29.399	28.904	30.810	29.554	29.547	29.899
Saturation (Mean)	34.941	33.328	33.685	33.092	31.002	34.885	33.892	33.550
Saturation (Std)	55.877	53.760	54.702	54.382	51.988	55.011	54.437	53.944
Value (Mean)	228.055	228.267	227.859	228.587	225.890	225.213	227.919	227.869
Value (Std)	45.986	46.435	46.099	44.931	48.174	48.053	46.268	46.238
Complexity	0.695	0.842	0.778	0.885	0.903	0.739	0.628	0.809
Greyscale (Std)	1.129	1.188	1.188	1.221	1.205	1.139	1.094	1.167
Edges	1,943.172	2,522.549	2,133.115	2,427.860	2,494.489	1,977.355	1,823.622	2,433.025
Lines	2.092	1.898	2.456	2.089	2.027	1.316	1.531	1.969
Circles	0.969	1.209	0.990	1.189	1.190	1.213	1.032	1.142
Polygons	2.679	2.984	2.733	2.986	3.045	2.707	2.673	3.013
Perceptual Categorization	4.066	3.652	4.115	3.879	3.872	4.231	4.190	3.993
Functional Categorization	1.394	0.816	1.071	0.904	0.801	0.957	1.155	0.943

Table 18: Summary Statistics of Clusters from Method 1, 15 Clusters (Cont'd)

	Dependent variable:	
	Memorability	Likability
	(1)	(2)
...	...	...
Cluster2	-0.012^{***}	0.107^{***}
	(0.002)	(0.010)
Cluster3	0.011^{***}	-0.088^{***}
	(0.002)	(0.012)
Cluster4	0.003	-0.013
	(0.002)	(0.012)
Cluster5	0.007^{***}	-0.029^{***}
	(0.002)	(0.011)
Constant	0.930^{***}	2.132^{***}
	(0.006)	(0.041)
Observations	123,924	123,924
R^2	0.112	0.074
Adjusted R^2	0.112	0.073
Residual Std. Error (df = 123907)	0.150	0.987
F Statistic (df = 16; 123907)	976.281^{***}	614.468^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 19: Regression Coefficients of Cluster Memberships from Method 2, 5 Clusters

Clusters	1	2	3	4	5
Hue (Mean)	59.280	58.006	76.129	59.376	55.058
Hue (Std)	29.476	27.988	41.871	28.932	24.587
Saturation (Mean)	32.594	27.152	75.892	31.832	17.978
Saturation (Std)	54.573	51.259	81.722	52.966	43.148
Value (Mean)	229.867	233.909	195.264	229.275	239.533
Value (Std)	45.764	43.086	66.698	44.326	38.357
Complexity	0.728	1.452	0.732	0.882	0.608
Greyscale (Std)	1.115	1.474	1.152	1.206	1.081
Edges	2,097.272	3,951.730	2,085.417	2,571.577	1,770.203
Lines	2.034	2.992	1.792	2.056	1.520
Circles	1.091	1.589	1.125	1.140	0.998
Polygons	2.918	3.486	2.788	3.055	2.647
Functional Categorization	1.011	0.316	0.996	0.977	1.203
Perceptual Categorization	4.057	3.298	4.080	3.867	4.201

Table 20: Summary Statistics of Clusters from Method 2, 5 Clusters

ding (t-SNE) and its scalable variants have been a class of widely-used and state-of-the-art technique for dimensionality reduction particularly well-suited for the visualization of high-dimensional datasets. Despite the success of t-SNE, its shortcomings are well-documented: t-SNE struggles to preserve the global structure of the dataset (Wattenberg et al. 2016), unless an extremely large perplexity value is chosen, in which case, t-SNE approximates MDS (Kruskal 1964); t-SNE does not scale well especially when the size exceeds 100,000, even in parallel (Ulyanov 2016) or with approximation such as the Barnes-Hut algorithms (Van Der Maaten 2014). On the other hand, UMAP is a manifold learning technique competitive with t-SNEs for visualization quality, and arguably preserves more of the global structure with superior run time performance (McInnes, Healy and Melville 2018) and greater generality. We modified codes hosted in the repository of the Yale Digital Humanities Lab to render visualizations of the entire logo dataset in a two-dimensional projection within which similar images are clustered together, where the visualization layer uses a custom WebGL viewer.

We visualize the logo images on `https://meredithhu.github.io/research/logo_visualization.html`. They were first reduced by UMAP and then clustered by K-means. Figure 13, Figure 14, and Figure 15 show visualizations of 5, 10, and 20 clusters, respectively. The screenshots are sub-optimal in providing intuitions about how clusters differ from one another, relative to interactive web browser sessions. It can be easily seen that different clusters exhibit distinct color, shape, contrast, and ratio features.

However, it should be noted that, given the fact that neither t-SNE nor UMAP preserves distances or density of data points in high dimension, and only to some extent preserves nearest-neighbors, there exists mixed evidence as to whether clustering based on t-SNE or UMAP results is warranted.

Clusters	1	2	3	4	5	6	7	8	9	10
Hue (Mean)	75.782	59.825	60.566	60.130	55.078	59.965	59.517	60.005	60.356	58.026
Hue (Std)	41.799	29.455	29.374	29.565	24.577	28.916	28.818	29.277	29.506	28.012
Saturation (Mean)	75.538	34.085	34.025	32.566	17.887	33.140	30.776	33.448	34.882	27.286
Saturation (Std)	81.712	54.366	54.510	54.164	43.029	54.412	51.735	54.141	55.872	51.365
Value (Mean)	195.535	229.320	227.596	230.074	239.548	228.591	228.599	227.916	227.529	233.786
Value (Std)	66.643	45.002	45.845	45.390	38.380	44.915	46.911	46.550	45.523	43.165
Complexity	0.726	0.883	0.931	0.806	0.605	0.884	0.768	0.850	1.017	1.447
Greyscale (Std)	1.147	1.195	1.264	1.173	1.078	1.219	1.174	1.189	1.271	1.469
Edges	2,067.240	2,420.802	2,609.047	2,315.057	1,762.030	2,424.552	2,346.962	2,513.289	2,849.499	3,939.792
Lines	1.800	2.335	2.120	1.950	1.526	2.082	1.528	2.344	2.452	2.968
Circles	1.117	1.179	1.277	1.172	0.995	1.186	1.195	1.203	1.208	1.581
Npolygon	2.791	2.990	2.857	2.871	2.653	2.987	2.788	2.979	3.077	3.495
Functional Categorization	1.007	0.894	0.854	0.941	1.207	0.904	1.028	0.882	0.792	0.315
Perceptual Categorization	4.086	3.800	3.900	3.939	4.209	3.880	3.839	3.830	3.700	3.314

Table 21: Summary Statistics of Clusters from Method 2, 10 Clusters

8. Incorporating Human Visual Biases

One of major concerns with our approach and measures is that it might not even capture human perception. Therefore it is imperative to validate our proposed measures for cognitive categorization. One way to validate our proposed measures for cognitive categorization is to incorporate human visual biases into the process explicitly. Computer vision scientists often dedicate to removing dataset biases from models (Torralba and Efros 2011, Kulis, Saenko and Darrell 2011). Nevertheless, sometimes human biases could be beneficial to recognition systems, as they make the recognition process more efficient with less available resources. For instance, the canonical perspective bias makes the recognition task easier as it restricts objects to certain perspectives, and the Gestalt laws of grouping make the recognition less error-prone, as a collective perspective is imposed on the objects.

In the context of our application, incorporating human biases could be particularly beneficial in that we are using the algorithm to provide for proxies of human cognitive categorization processes. Therefore, inspired by the popular classification images procedure in human psychophysics that attempts to estimate the internal template that the human visual system might use for recognition of a category (Eckstein and Ahumada 2002, Murray 2011), which was further introduced into computer vision literature by Vondrick et al. (2015), we use a novel method to learn biases from the human visual system and incorporate them into computer vision systems for classification. We first provide a quick review of the classification images procedure that is essential to incorporating human cognitive biases in Section 8.1.

8.1 A Review of Classification Images Procedure

The Classification Images procedure (Eckstein and Ahumada 2002, Murray 2011) has been widely used in human psychophysics to estimate cognitive templates that human cognitive systems generate for categorization processes. Vondrick et al. (2015) was one of the first to introduce it into computer vision literature.

To approximate the template t that a human viewer uses to distinguish category A from category B, such as cat vs. dog, we sample white noise $\epsilon \sim N(0^d, I_d)$ and ask the viewer to indicate the category label for $a + \epsilon$, where a is an example of category A, $a \in A$. Most of the time the viewer will answer with the correct category, but sometimes ϵ might cast an effect large enough to cause the viewer to mistake $a + \epsilon$ for an example of category B, b.

The rationale of classification images is that, if we execute a large number of such trials, we will be able to estimate a categorization function $f(\cdot)$ that distinguishes between category A and category B, making the same mistakes as the viewer. From the errors and correct guesses of this approximating function $f(\cdot)$, we can obtain an estimate of the cognitive template of the viewer for distinguishing different categories. We can gain additional insights into how human cognitive systems work by simulating, analyzing, and generalizing the approximate functions derived this way.

Linear approximation functions are one of the most widely used in the literature due to its interpretability and simplicity. One of the most common way to estimate a cognitive template is the sum of individual images:

$$t = (\mu_{AA} + \mu_{BA}) - (\mu_{AB} + \mu_{BB}) \tag{4}$$

where the first underscript represents the true category and the second underscript the predicted category by the viewer. μ_{AB} represents the average image (generated in the same as an average face in psychology literature) of images that the viewer identifies to be of category B but are in fact of category A. The linear approximation is intuitive: it will aggregate and accentuate the parts of the images identified to be of category A, and mask the parts of the images identified to be of category B. Murray (2011) provides a comprehensive review of the procedure and its applications.

8.2 Estimating Human Cognitive Biases

Following the simplified procedure introduced in Vondrick et al. (2015), we estimate the cognitive templates by sampling white noise in feature spaces, such as Histogram of Gradients (HOG) (Dalal and Triggs 2005), and image embeddings from deep convolutional neural networks (Krizhevsky et al. 2012a). More specifically, we first sample white noise from a multi-variate Gaussian distribution with zero mean and unit variance as instances from the feature space. We then invert the noise feature back to pixels using image invertion tools introduced in Vondrick, Khosla, Malisiewicz and Torralba (2013). These artificial images are then shown to humans to indicate whether they see a particular category in our context. In the form of an online game that tests the participants't sixth senses, we ask viewers to "hallucinate" each one of the 39 business industrial categories that the "intentionally"

blurred image is representing, from we derive the functional categorization templates; for the perceptual categorization templates, we ask viewers to "hallucinate" 1000 objects categories the picture contains. Having obtained these annotations, we build a linear template t that approximates humans' cognitive templates by calculating:

$$t = \mu_A - \mu_{nA} \tag{5}$$

where μ_A is the average white noise in feature space that viewers hallucinated to be of category A, and μ_{nA} is the average of white noise in feature space that viewers incorrectly believed to be not of category A, but something else.

We randomly sampled $31,941$ points from a multivariate normal distribution with zero mean and unit variance and inverted each sample with HOGgles (Vondrick et al. 2013). We asked the same number of workers from Amazon Mechanical Turk to indicate whether they see the industrial category the image represents or not. In each task, we included 100 real image labeling instances and 10 easy image recognition tasks, with which we screen to ensure quality.

We visualize some of the cognitive templates from the classification images procedure and the associated estimation method. Unlike object templates, albeit blurred, which show significant object details and emerging sensible patterns, estimated templates of industrial categories for cognitive categorization processes do not exhibit explicitly visible patterns or details. Estimated templates of four categories ranging from food and drinks, education, technology, to dating are shown in Figure 16. The lack of detail could be because of the abstract and aggregate nature of the particular task, even though some subtle patterns do seem existent. For instance, the template of Dating appears to be brighter, the template of Food and Drinks busier, and the template of Technology darker than other categories on average. Figure 17 shows the screen shot of one such task distributed online via Amazon Mechanical Turk, restricting workers to U.S. locations.

8.3 Learning with Cognitive Biases from Templates

We adopt the orientation constraint method introduced by Vondrick et al. (2015) to incorporate cognitive templates into our deep Convolutional Neural Network classification systems.

The idea behind the orientation constraint method applied to support vector machines

(Vondrick et al. 2015) is simple: a standard SVM finds a separating hyperplane ω that maximizes the margin between positive and negative examples. To incorporate human cognitive biases from the estimated templates, we insert the additional constraint that the SVM hyperplane ω must be within a certain distance from the bias template t. Vondrick et al. (2015) specify this distance constraint by imposing a maximal amount of angel θ away from the template such that

$$\theta \leq \frac{\omega_T t}{\omega^T \omega} \tag{6}$$

in addition to the standard SVM objective and constraint.

To intergrate it into our DCNN classification system, we start with the last convolutional layer from Residual Net as detailed in Section 3.3 as input features into a simple linear support vector machine, and adopt the orientation constraint implementation provided by Vondrick et al. (2015), using the estimated category templates for each business industrial category and object category. The resulting predicted distributions were used to calculate proxies for functional and perceptual categorization in the same way as detailed in Section 3.4 and Section 3.5.

8.4 Regression Results from Classification Systems that Incoporate Human Biases

We replicate the same analyses as in Section 5 and tabulate the results in Table 22.

Most results appear qualitatively the same except for perceptual categorization coefficients, which become even less significant and of smaller magnitude, which appear to corroborate the main findings in Section 5, that functional categorization is more critical to viewer liking (and memory).

The linear model that underpins the most straightforward interpretation of classification images has turned out not to put a strong limit on the usefulness of these methods, as the linearity assumption can be tested (Murray 2011). And it often turns out to be valid, at least over the very small range of stimuli used in the typical classification image experiment (Abbey and Eckstein 2002, Murray 2002). Furthermore, departures from linearity are sometimes unimportant, as when we draw conclusions simply from the fact that a classification image shows strong correlations between a stimulus region of interest and the viewer's responses. Additionally, there are many ways of modifying the method to incorporate nonlinearities

	Dependent variable:	
	Memory	Liking
Functional Categorization	-0.026^{***}	-0.015^{***}
	(0.001)	(0.003)
Perceptual Categorization	-0.0004	-0.001
	(0.0004)	(0.002)
Texts	-0.0001^{***}	0.003^{***}
	(0.00003)	(0.0002)
Edges	0.00000^{***}	0.0002^{***}
	(0.00000)	(0.00000)
Lines	-0.003^{***}	-0.045^{***}
	(0.0001)	(0.001)
Circles	0.018^{***}	0.036^{***}
	(0.001)	(0.005)
Polygons	0.022^{***}	-0.058^{***}
	(0.001)	(0.004)
Constant	0.761^{***}	2.772^{***}
	(0.006)	(0.039)
Observations	125,270	125,270
R^2	0.091	0.128
Adjusted R^2	0.091	0.128
Residual Std. Error (df = 125224)	0.151	0.958
F Statistic (df = 45; 125224)	279.096^{***}	408.339^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 22: Ordinary Least Squares Estimation Predicting Memorability and Likability where Cognitive Variables Incorporated Human Visual Templates

in visual processing, including the general-purpose Volterra (Volterra 2005, Schetzen 1980) and Wiener kernel frameworks (Wiener 1966, Schetzen 1980) and more specific modifications based on models of nonlinearities in visual processing.

9. Managerial Relevance

The present study suggests necessary yet overlooked trade-offs in logo design processes.

For a product manager of a design team, design objectives will need to be specified — optimizing for memorability, likability, or both at the same time. Some visual features such as pattern complexity, use of distinct, saturated, and bright colors have positive effects on likability but could prove detrimental on memorability. Moderately complex or even low complex patterns, uniform colors are sometimes associated with greater memorability but have no impact on likability. Therefore, it is the manager's responsibility to scrutinize and decide on visual features aligned with established objectives. For instance, if the objective is to increase consumer awareness, sometimes uniform and straightforward patterns work better, whereas if the objective is to increase consumer affinity, perhaps complex and saturated patterns work better. Without loss of generality, memorability might be prioritized when consumer awareness, brand recognition (Lee 2002), new customer acquisition, inclusion into 'consumers' consideration set (Nedungadi 1990), among other factors, are of greater concern; whereas likability might be prioritized when customer engagement (Lee, Hosanagar and Nair 2018), customer retention, willingness to pay, and reputation are of greater concern. When memorability is prioritized, functional categorization matters more than perceptual categorization, and vice versa.

For a graphic designer, both the perceptual and conceptual meanings of design patterns merit some horizontal and vertical comparisons — is the visual pattern itself easy to recognize among all visual cues consumers encounter every day? And is the underlying meaning easy to identify among all relevant brands and organizations? If the pattern itself is easily recognizable and the meaning appears clear as well, given the industry trends, the design could potentially be rather memorable and likable, all else being equal; if the pattern is easily recognizable but the meaning appears very ambiguous, the design will be more likely to be likable rather than memorable. For a graphic designer using CAD and AI for assistance, generating patterns based on readily available visual symbols tend to contribute to increased

likability, whereas generating patterns based on brand identity, industrial norms and symbols tend to contribute to increased likability.

For a marketing manager, the allocation of marketing resources could be better informed, given the results of the current study. On the one hand, some design features that positively affect both memorability and likability could help meeting marketing goals with regards to increased brand visibility and affinity. On the other hand, when some design features that change memorability and likability in opposite directions are in use — for instance, such as pattern complexity, use of distinct, saturated, and bright colors possibly lower memorability but help with likability — additional marketing resources should be allocated to counter the potential decrease in brand visibility. In addition, when composing marketing mix and strategies, design features that increase memorability but lower likability might be better coupled with marketing strategies more effective in boosting likability, whereas design features that help with both memorability and likability could potentially free marketing resources towards to other objectives than memorability or likability.

For an operations manager, the potentially opposing effects of visual features on memorability and likability could also affect scheduling decisions with regards to different marketing and operations efforts. When considering altering some visual designs that lead to further impacts on memorability and likability, some lead time might be taken into consideration if preventing potential interference with other marketing mechanisms targeted at increasing memorability and likability is desired. Scheduling different marketing strategies that target increased memorability and likability after the design alteration might be proper to prevent expected results from being compromised.

Lastly, we detail the scope of our results, the corresponding boundary conditions, and resulting limitations as follows:

Our results are based on datasets from public sources, which are gathered by automatic scrapers or crawlers and a comparatively negligible number of user uploads, and in aggregation are not particularly subject to selection biases.

The main methods used for measures of cognitive categorizations might be subject to adversarial perturbations as have been documented in various studies such as (Goodfellow, Shlens and Szegedy 2014b), (Kurakin, Goodfellow and Bengio 2016). However, since most demonstrated adversarial examples are synthetic artifacts applied to natural images, it is unclear how adversarial examples for design images, if any, can be constructed and prove detrimental to the classification systems in our case.

Our proposed methods to incorporate human biases are hinged upon the linearity assumption, albeit robust according to previous literature, could lack generality or flexibility compared to nonlinear approximation, masking more nuanced results.

The clustering methods we experimented with are for image clustering based only on pixel information, previously applied to large-scale natural images in literature. Potential integration with additional details of associated brands and companies in the case of logos might prove more informative and lead to more nuanced insights. Such leaves open a multitude of future research opportunities.

In a similar vein, the visualization methods are only based on pixel information. It would be potentially more interesting and informative if other information sources about associated brands and companies were baked into the visualization process that maps high-dimensional and multi-modal data onto a two-dimensional space intuitively and sensibly. We look forward to interesting future work along this line of research.

10. Conclusion

In this paper, we propose two new taxonomy of cognitive processes involved in the perception of logo designs. We rely on the cognitive psychology literature to develop constructs of logo design patterns; we rely on the computer vision and information theory research to develop methods that are automated and scalable.

We extract low-level image descriptors and the cognitive variables to analyze their absolute and relative effects on image memorability and likability, from which we identify novel and contrasting managerial insights. To further probe the landscape of logo designs, we cluster the entire dataset using different classes of image clustering algorithms, and identify clusters that (1) independently affect viewer memory, or liking; (2) simultaneously affect viewer memory and liking in the same direction; (3) simultaneously affect viewer memory and liking in opposite directions. Lastly, we use the classification image procedure to induce and measure categorical templates of human perception and incorporate these templates into our cognitive measures. We show the resulting differences and robustness of our analyses.

We acknowledge that there are at least three aspects where the current study could be improved. First, we elicit and annotate the dataset with memorability and likability scores by running online experiments on Amazon Mechanical Turk, which is by no means as well-

controlled as the conventional economic or psychological lab experiments, and therefore, could have introduced selection bias and weakened the external validity. Second, the linear observer model for understanding classification image methods, with which we constrain the proposed cognitive measures to test and validate theories, is certainly incomplete. For instance, it does not incorporate transduction nonlinearities, contrast normalization, spatial uncertainty, perceptual learning, or other known properties of human visual processing (Murray 2011). Some research on classification images has found better methods for characterizing linear observers. Other works have developed techniques that go beyond the linear observer model, taking into account nonlinearities in visual processing. We defer such technical improvement to future work. Third, multiple sections could have been improved if given more time, higher computing power, and more funding. For instance, the UMAP-powered visualization and the results from deep embedding clustering methods could be more precise and descriptive given more training time and iterations; the classification images procedure could been improved given more experimental samples; and some human-in-the-loop method could have been adopted to measure cognitive categorization.

We close by highlighting several opportunities for future research. First, our results may be replicated using other design image datasets such as book and album covers, movie posters, packaging, etc. Second, our approach may be tested with other important marketing schemes such as advertisements, and product placement in entertainment products, etc. Third, while we focused on aggregate outcomes, our approach may be applied to study individual or organizational recommendations about design, which brings us to the fourth point — academics and practitioners interested in incorporating the proposed method into predictive models would need to research the best way to combine it with various frameworks such as collaborative filtering or random forests. Fifth, given the focus in the perception, machine learning and AI research on improving learning, it would be worthwhile to study the link between design features and learning. Sixth, multi-modal recommendation engines may be developed based on the proposed approach. Such engines may be designed to increase not only consumption but also learning efficiency or time and cost utilization.

References

Abbey, Craig K and Miguel P Eckstein, "Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments," *Journal of vision*, 2002, *2* (1), 5–5.

Bainbridge, Wilma A, Phillip Isola, and Aude Oliva, "The intrinsic memorability of face photographs.," *Journal of Experimental Psychology: General*, 2013, *142* (4), 1323.

Baker, Michael J and John MT Balmer, "Visual identity: trappings or substance?," *European Journal of marketing*, 1997, *31* (5/6), 366–382.

Bhattacherjee, Anol and Clive Sanford, "Influence processes for information technology acceptance: An elaboration likelihood model," *MIS quarterly*, 2006, pp. 805–825.

Bornstein, Robert F and Paul R D'Agostino, "The attribution and discounting of perceptual fluency: Preliminary tests of a perceptual fluency/attributional model of the mere exposure effect," *Social Cognition*, 1994, *12* (2), 103–128.

Canny, John, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, 1986, (6), 679–698.

Chang, Jianlong, Lingfeng Wang, Gaofeng Meng, Shiming Xiang, and Chunhong Pan, "Deep adaptive image clustering," in "Proceedings of the IEEE International Conference on Computer Vision" 2017, pp. 5879–5887.

Cian, Luca, Aradhna Krishna, and Ryan S Elder, "This logo moves me: Dynamic imagery from static images," *Journal of Marketing Research*, 2014, *51* (2), 184–197.

Cui, Yin, Feng Zhou, Yuanqing Lin, and Serge Belongie, "Fine-grained Categorization and Dataset Bootstrapping using Deep Metric Learning with Human in the Loop," in "CVPR 2016" IEEE 2016.

Dalal, Navneet and Bill Triggs, "Histograms of oriented gradients for human detection," in "international Conference on computer vision & Pattern Recognition (CVPR'05)," Vol. 1 IEEE Computer Society 2005, pp. 886–893.

den Bosch, Annette LM Van, Menno DT De Jong, and Wim JL Elving, "How corporate visual identity supports reputation," *Corporate Communications: An International Journal*, 2005, *10* (2), 108–116.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in "Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on" IEEE 2009, pp. 248–255.

Deng, Li, "The MNIST database of handwritten digit images for machine learning research [best of the web]," *IEEE Signal Processing Magazine*, 2012, *29* (6), 141–142.

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

der Lans, Ralf Van, Joseph A Cote, Catherine A Cole, Siew Meng Leong, Ale Smidts, Pamela W Henderson, Christian Bluemelhuber, Paul A Bottomley, John R Doyle, Alexander Fedorikhin et al., "Cross-national logo evaluation analysis: An individual-level approach," *Marketing science*, 2009, *28* (5), 968–985.

Dew, Ryan, Asim Ansari, and Olivier Toubia, "Letting Logos Speak: A Machine Learning Approach for Data-Driven Logo Design," 2018.

____ , ____ , and ____ , "Letting logos speak: a machine learning approach to data-driven logo design," 2018.

Deza, Arturo and Devi Parikh, "Understanding image virality," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2015, pp. 1818–1826.

Donahue, Jeff, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," in "International conference on machine learning" 2014, pp. 647–655.

Doyle, John R and Paul A Bottomley, "Dressed for the occasion: Font-product congruity in the perception of logotype," *Journal of consumer psychology*, 2006, *16* (2), 112–123.

Dubey, Rachit, Joshua Peterson, Aditya Khosla, Ming-Hsuan Yang, and Bernard Ghanem, "What Makes an Object Memorable?," in "The IEEE International Conference on Computer Vision (ICCV)" December 2015.

Duda, Richard O and Peter E Hart, "Use of the Hough transformation to detect lines and curves in pictures," Technical Report, SRI INTERNATIONAL MENLO PARK CA ARTIFICIAL INTELLIGENCE CENTER 1971.

Dzyabura, Daria, Marat Ibragimov, and Siham El Kihal, 2018.

Eckstein, Miguel P and Albert J Ahumada, "Classification images: A tool to analyze visual strategies," *Journal of vision*, 2002, *2* (1), i–i.

Estes, William K, "Array models for category learning," *Cognitive psychology*, 1986, *18* (4), 500–549.

Felzenszwalb, Pedro F, Ross B Girshick, David McAllester, and Deva Ramanan, "Object detection with discriminatively trained part-based models," *IEEE transactions on pattern analysis and machine intelligence*, 2010, *32* (9), 1627–1645.

Flavell, John H, "Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry.," *American psychologist*, 1979, *34* (10), 906.

Girshick, Ross, "Fast r-cnn," in "Proceedings of the IEEE international conference on computer vision" 2015, pp. 1440–1448.

___ , Jeff Donahue, Trevor Darrell, and Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2014, pp. 580–587.

Goldstone, Robert L, "The role of similarity in categorization: Providing a groundwork," *Cognition*, 1994, *52* (2), 125–157.

Goodfellow, Ian J. et al., "Generative Adversarial Nets," in "NIPS" 2014.

Goodfellow, Ian J, Jonathon Shlens, and Christian Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

Gowda, K Chidananda and G Krishna, "Agglomerative clustering using the concept of mutual nearest neighbourhood," *Pattern recognition*, 1978, *10* (2), 105–112.

Greene, William H, "Econometric analysis," 2000.

Hagtvedt, Henrik, "The impact of incomplete typeface logos on perceptions of the firm," *Journal of Marketing*, 2011, *75* (4), 86–93.

Halberstadt, Jamin and Gillian Rhodes, "The attractiveness of nonface averages: Implications for an evolutionary explanation of the attractiveness of average faces," *Psychological Science*, 2000, *11* (4), 285–289.

Harris, Chris and Mike Stephens, "A combined corner and edge detector.," in "Alvey vision conference," Vol. 15 Citeseer 1988, pp. 10–5244.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in "Proceedings of the IEEE international conference on computer vision" 2015, pp. 1026–1034.

—— , —— , —— , and —— , "Deep residual learning for image recognition," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2016, pp. 770–778.

Henderson, Pamela W and Joseph A Cote, "Guidelines for selecting or modifying logos," *The Journal of Marketing*, 1998, pp. 14–30.

Hintzman, Douglas L, "Human learning and memory: Connections and dissociations," *Annual review of psychology*, 1990, *41* (1), 109–139.

Holland, John H, Keith J Holyoak, Richard E Nisbett, and Paul R Thagard, *Induction: Processes of inference, learning, and discovery*, MIT press, 1989.

Illingworth, John and Josef Kittler, "A survey of the Hough transform," *Computer vision, graphics, and image processing*, 1988, *44* (1), 87–116.

Isola, Phillip, Jianxiong Xiao, Antonio Torralba, and Aude Oliva, "What makes an image memorable?," 2011.

—— , —— , Devi Parikh, Antonio Torralba, and Aude Oliva, "What makes a photograph memorable?," *IEEE transactions on pattern analysis and machine intelligence*, 2014, *36* (7), 1469–1482.

Jiang, Yuwei, Gerald J Gorn, Maria Galli, and Amitava Chattopadhyay, "Does your company have the right logo? How and why circular-and angular-logo shapes influence brand attribute judgments," *Journal of Consumer Research*, 2015, *42* (5), 709–726.

___ , ___ , ___ , and ___ , "Does Your Company Have the Right Logo? How and Why Circular-and Angular-Logo Shapes Influence Brand Attribute Judgments," *Journal of Consumer Research*, 2016, *42* (5), 709–726.

Jones, Michael N, "Developing cognitive theory by mining large-scale naturalistic data," in "Big data in cognitive science," Psychology Press, 2016, pp. 10–21.

Jun, Jong Woo, Chang-Hoan Cho, and Hyuck Joon Kwon, "The role of affect and cognition in consumer evaluations of corporate visual identity: Perspectives from the United States and Korea," *Journal of Brand Management*, 2008, *15* (6), 382–398.

Kanopoulos, Nick, Nagesh Vasanthavada, and Robert L Baker, "Design of an image edge detection filter using the Sobel operator," *IEEE Journal of solid-state circuits*, 1988, *23* (2), 358–367.

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

Khan, Rahat, Joost Van de Weijer, Fahad Shahbaz Khan, Damien Muselet, Christophe Ducottet, and Cecile Barat, "Discriminative color descriptors," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2013, pp. 2866–2873.

Khosla, A, J Xiao, and A Torralba, "Memorability of image regions," *Advances in Neural ...*, 2012.

___ , WA Bainbridge, and A Torralba, "Modifying the memorability of face photographs," *Proceedings of the IEEE ...*, 2013.

Khosla, Aditya, Akhil S Raju, Antonio Torralba, and Aude Oliva, "Understanding and predicting image memorability at a large scale," in "Proceedings of the IEEE International Conference on Computer Vision" 2015, pp. 2390–2398.

___ , **Atish Das Sarma, and Raffay Hamid**, "What makes an image popular?," in "Proceedings of the 23rd international conference on World wide web" ACM 2014, pp. 867–876.

___ **et al.**, "Understanding and Predicting Image Memorability at a Large Scale," in "International Conference on Computer Vision (ICCV)" 2015.

Kingma, Diederik P and Max Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

Kiryati, Nahum, Yuval Eldar, and Alfred M Bruckstein, "A probabilistic Hough transform," *Pattern recognition*, 1991, *24* (4), 303–316.

Klink, Richard R, "Creating meaningful brands: The relationship between brand name and brand mark," *Marketing Letters*, 2003, *14* (3), 143–157.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds., *Advances in Neural Information Processing Systems 25*, Curran Associates, Inc., 2012, pp. 1097–1105.

___ , ___ , **and Geoffrey E. Hinton**, "ImageNet Classification with Deep Convolutional Neural Networks," in "NIPS" 2012.

___ , ___ , **and Geoffrey E Hinton**, "Imagenet classification with deep convolutional neural networks," in "Advances in neural information processing systems" 2012, pp. 1097–1105.

Kruskal, Joseph B, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, 1964, *29* (1), 1–27.

Kulis, Brian, Kate Saenko, and Trevor Darrell, "What you saw is not what you get: Domain adaptation using asymmetric kernel transforms," in "CVPR 2011" IEEE 2011, pp. 1785–1792.

Kurakin, Alexey, Ian Goodfellow, and Samy Bengio, "Adversarial examples in the physical world," *arXiv preprint arXiv:1607.02533*, 2016.

Landwehr, Jan R et al., "Gut liking for the ordinary: Incorporating design fluency improves automobile sales forecasts," *Marketing Science*, 2011, *30* (3), 416–429.

Langlois, Judith H and Lori A Roggman, "Attractive faces are only average," *Psychological science*, 1990, *1* (2), 115–121.

Lazebnik, Svetlana, Cordelia Schmid, and Jean Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in "null" IEEE 2006, pp. 2169–2178.

Lee, Angela Y, "Effects of implicit memory on memory-based versus stimulus-based brand choice," *Journal of Marketing Research*, 2002, *39* (4), 440–454.

____ **and Aparna A Labroo**, "The effect of conceptual and perceptual fluency on brand evaluation," *Journal of Marketing Research*, 2004, *41* (2), 151–165.

Lee, Dokyun, Kartik Hosanagar, and Harikesh S Nair, "Advertising content and consumer engagement on social media: Evidence from Facebook," *Management Science*, 2018, *64* (11), 5105–5131.

Liberman, Nira, Yaacov Trope, and Cheryl Wakslak, "Construal level theory and consumer behavior," *Journal of Consumer Psychology*, 2007, *17* (2), 113–117.

Lin, Min, Qiang Chen, and Shuicheng Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

Liu, Liu and Dina Mayzlin, "Visual Listening in: Extracting Brand Image Portrayed on Social Media," 2018.

Loy, Gareth and Jan-Olof Eklundh, "Detecting symmetry and symmetric constellations of features," in "European Conference on Computer Vision" Springer 2006, pp. 508–521.

Lu, Shasha, Li Xiao, and Min Ding, "A video-based automated recommender (VAR) system for garments," *Marketing Science*, 2016, *35* (3), 484–510.

Maaten, Laurens Van Der, "Accelerating t-SNE using tree-based algorithms," *The Journal of Machine Learning Research*, 2014, *15* (1), 3221–3245.

MacInnis, Deborah J, Stewart Shapiro, and Gayathri Mani, "Enhancing brand awareness through brand symbols," *ACR North American Advances*, 1999.

McInnes, Leland, John Healy, and James Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.

Mervis, Carolyn B and Eleanor Rosch, "Categorization of natural objects," *Annual review of psychology*, 1981, *32* (1), 89–115.

Müller, Brigitte, Bruno Kocher, and Antoine Crettaz, "The effects of visual rejuvenation through brand logos," *Journal of Business Research*, 2013, *66* (1), 82–88.

Murray, Richard F, "Classification images: A review," *Journal of vision*, 2011, *11* (5), 2–2.

Murray, Richard Frederick, "Perceptual organization and the efficiency of shape discrimination," *Unpublished doctoral dissertation, University of Toronto, Canada*, 2002.

Nedungadi, Prakash, "Recall and consumer consideration sets: Influencing choice without altering brand evaluations," *Journal of consumer research*, 1990, *17* (3), 263–276.

Ng, Andrew, "Sparse Auto-Encoder," *CS294A Lecture notes*, 2011, *72*, 1–19.

Ojala, Timo, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, 2002, *24* (7), 971–987.

Olins, Wally, *Corporate identity: Making business strategy visible through design*, Harvard Business School Pr, 1990.

Oliva, Aude and Antonio Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, 2001, *42* (3), 145–175.

Palmer, Stephen E, "Structural aspects of visual similarity," *Memory & Cognition*, 1978, *6* (2), 91–97.

Papatla, Purushottam, "Face, Body or Both? Effects of Partial and Full Visibility of People in VUGC on Consumer Response," 2018.

Petty, Richard E and John T Cacioppo, "The elaboration likelihood model of persuasion," in "Communication and persuasion," Springer, 1986, pp. 1–24.

___ , Duane T Wegener, and Leandre R Fabrigar, "Attitudes and attitude change," *Annual review of psychology*, 1997, *48* (1), 609–647.

___ , John T Cacioppo, and Rachel Goldman, "Personal involvement as a determinant of argument-based persuasion.," *Journal of personality and social psychology*, 1981, *41* (5), 847.

Porac, Joseph F and Howard Thomas, "Cognitive categorization and subjective rivalry among retailers in a small city.," *Journal of Applied Psychology*, 1994, *79* (1), 54.

Rahinel, Ryan and Noelle M Nelson, "When Brand Logos Describe the Environment: Design Instability and the Utility of Safety-Oriented Products," *Journal of Consumer Research*, 2016, *43* (3), 478–496.

Reber, Rolf, Norbert Schwarz, and Piotr Winkielman, "Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?," *Personality and social psychology review*, 2004, *8* (4), 364–382.

Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You only look once: Unified, real-time object detection," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2016, pp. 779–788.

Rosch, Eleanor, "Principles of categorization," *Concepts: core readings*, 1999, *189*.

___ and Barbara Bloom Lloyd, "Cognition and categorization," 1978.

___ , Carolyn B Mervis, Wayne D Gray, David M Johnson, and Penny Boyes-Braem, "Basic objects in natural categories," *Cognitive psychology*, 1976, *8* (3), 382–439.

Rosch, Eleanor H, "Natural categories," *Cognitive psychology*, 1973, *4* (3), 328–350.

Schetzen, Martin, "The Volterra and Wiener theories of nonlinear systems," 1980.

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

Shechtman, Eli and Michal Irani, "Matching local self-similarities across images and videos," in "Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on" IEEE 2007, pp. 1–8.

Simon, Herbert A, *The sciences of the artificial*, MIT press, 1996.

Simonyan, Karen and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

Stacey, M, "Psychological challenges for the analysis of style," *AIE EDAM: Artificial Intelligence for Engineering ...*, 2006.

Swartz, Teresa A, "Brand symbols and message differentiation.," *Journal of Advertising Research*, 1983.

Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2015, pp. 1–9.

Tkachenko, Yegor, Asim Ansari, and Olivier Toubia, "Computer-aided Exploration Of Product Designs in High-dimensional Visual Spaces," 2018.

Todorov, Alexander, "Modeling Visual Impressions of Faces," 2018.

Torralba, Antonio and Alexei A Efros, "Unbiased look at dataset bias," 2011.

Trope, Yaacov, Nira Liberman, and Cheryl Wakslak, "Construal levels and psychological distance: Effects on representation, prediction, evaluation, and behavior," *Journal of consumer psychology*, 2007, *17* (2), 83–95.

Ulyanov, Dmitry, "Multicore-TSNE," `https://github.com/DmitryUlyanov/Multicore-TSNE` 2016.

Valdez, Patricia and Albert Mehrabian, "Effects of color on emotions.," *Journal of experimental psychology: General*, 1994, *123* (4), 394.

van der Maaten, Laurens and Geoffrey Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, 2008, *9* (Nov), 2579–2605.

Veenman, Marcel VJ, Bernadette HAM Van Hout-Wolters, and Peter Afflerbach, "Metacognition and learning: Conceptual and methodological considerations," *Metacognition and learning*, 2006, *1* (1), 3–14.

Vincent, Pascal, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of machine learning research*, 2010, *11* (Dec), 3371–3408.

Volterra, Vito, *Theory of functionals and of integral and integro-differential equations*, Courier Corporation, 2005.

Vondrick, Carl, Aditya Khosla, Tomasz Malisiewicz, and Antonio Torralba, "Hoggles: Visualizing object detection features," in "Proceedings of the IEEE International Conference on Computer Vision" 2013, pp. 1–8.

___, Hamed Pirsiavash, Aude Oliva, and Antonio Torralba, "Learning visual biases from human imagination," in "Advances in neural information processing systems" 2015, pp. 289–297.

Walsh, Michael F, Karen Page Winterich, and Vikas Mittal, "Do logo redesigns help or hurt your brand? The role of brand commitment," *Journal of Product & Brand Management*, 2010, *19* (2), 76–84.

Wang, Jianfeng, Jingdong Wang, Jingkuan Song, Xin-Shun Xu, Heng Tao Shen, and Shipeng Li, "Optimized cartesian k-means," *IEEE Transactions on Knowledge and Data Engineering*, 2015, *27* (1), 180–192.

Wang, Margaret C, Geneva D Haertel, and Herbert J Walberg, "What influences learning? A content analysis of review literature," *The Journal of Educational Research*, 1990, *84* (1), 30–43.

Wattenberg, Martin, Fernanda Viégas, and Ian Johnson, "How to Use t-SNE Effectively," *Distill*, 2016.

Wiener, Norbert, "Nonlinear problems in random theory," *Nonlinear Problems in Random Theory, by Norbert Wiener, pp. 142. ISBN 0-262-73012-X. Cambridge, Massachusetts, USA: The MIT Press, August 1966.(Paper)*, 1966, p. 142.

Winkielman, Piotr and John T Cacioppo, "Mind at Ease Puts a Smile on the Face: Psychophysiological Evidence That Processing Facilitation Elicits Positive Affect," *Journal of Personality and Social Psychology*, 2001, *81* (6), 989–1000.

___ **, Jamin Halberstadt, Tedra Fazendeiro, and Steve Catty**, "Prototypes are attractive because they are easy on the mind," *Psychological science*, 2006, *17* (9), 799–806.

Xiao, Jianxiong, James Hays, Krista A Ehinger, Aude Oliva, and Antonio Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," in "Computer vision and pattern recognition (CVPR), 2010 IEEE conference on" IEEE 2010, pp. 3485–3492.

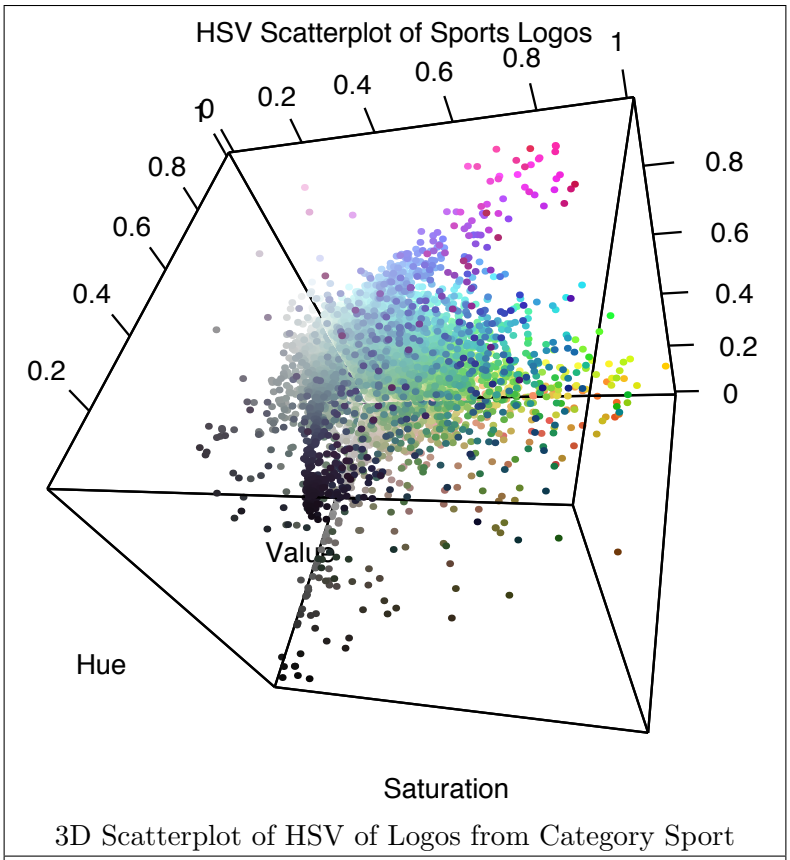
Xiao, Li and Min Ding, "Just the Faces: Exploring the Effects of Facial Features in Print Advertising," *Marketing Science*, 2014, *33*, 338–352.

Xie, Junyuan, Ross Girshick, and Ali Farhadi, "Unsupervised deep embedding for clustering analysis," in "International conference on machine learning" 2016, pp. 478–487.

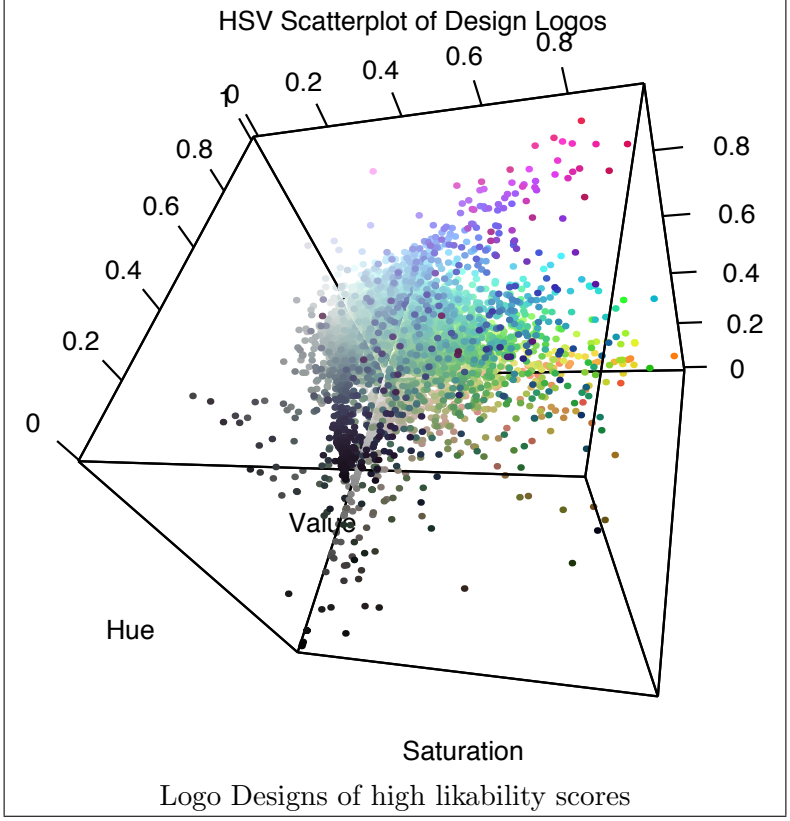
Yang, Jianwei, Devi Parikh, and Dhruv Batra, "Joint unsupervised learning of deep representations and image clusters," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition" 2016, pp. 5147–5156.

Zhang, Shunyuan, Dokyun Lee, Param Vir Singh, and Kannan Srinivasan, "How Much is an Image Worth? The Impact of Professional versus Amateur Airbnb Property Images on Property Demand," 2018.

Zhou, Xinyu, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, and Jiajun Liang, "EAST: an efficient and accurate scene text detector," in "Proc. CVPR" 2017, pp. 2642–2651.

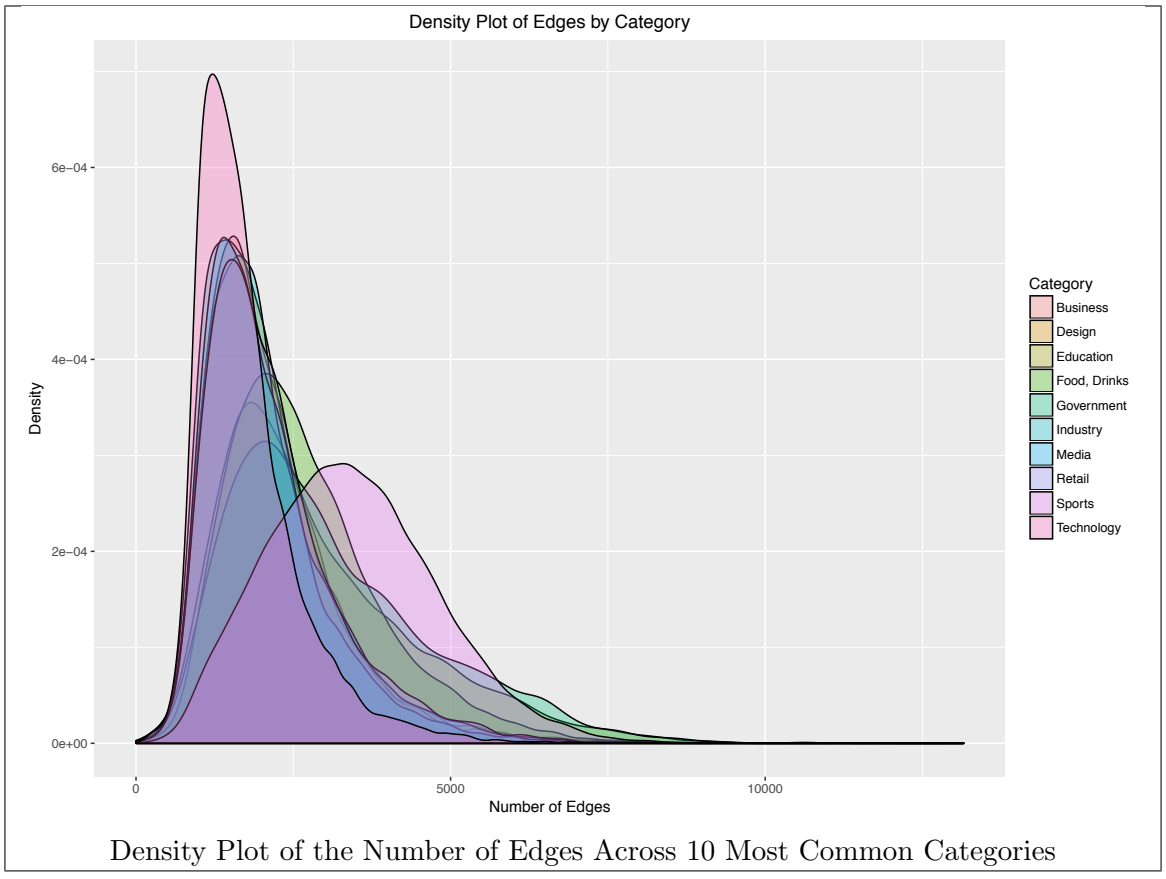


3D Scatterplot of HSV of Logos from Category Sport



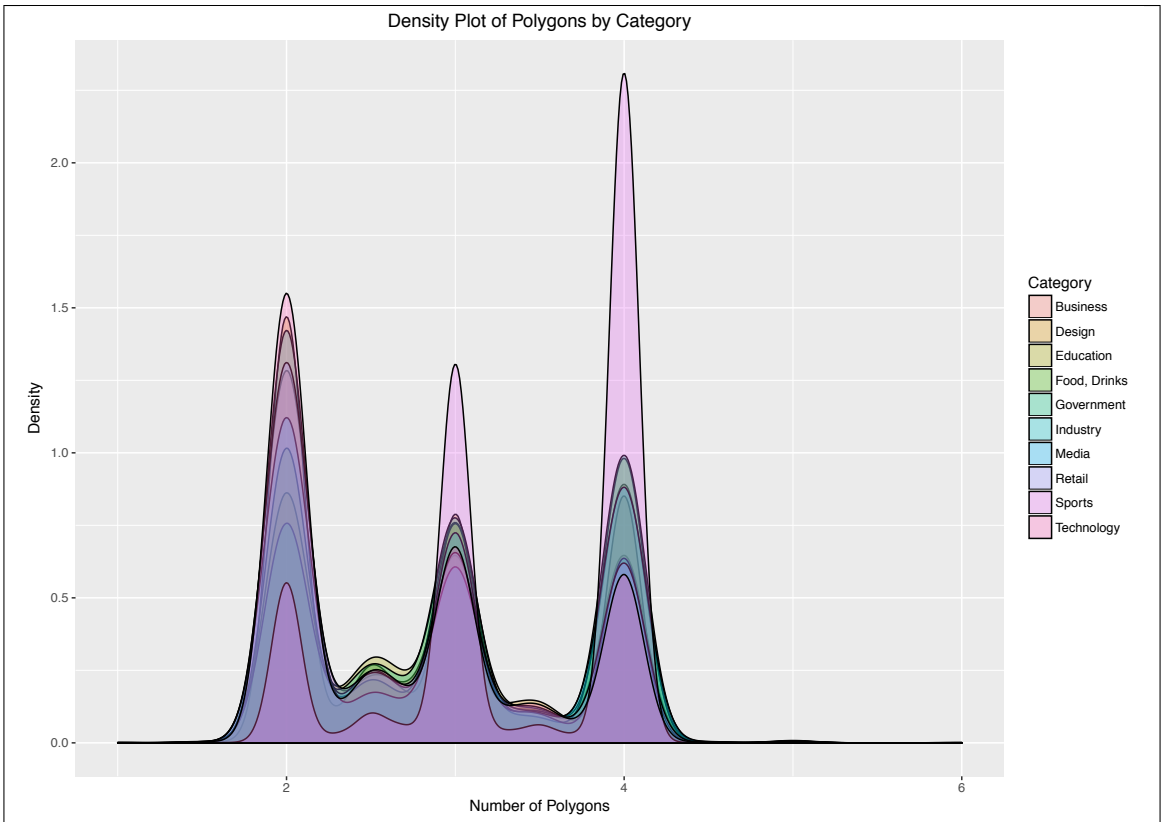
Logo Designs of high likability scores

Figure 2: 3D Scatterplots of Mean HSV Values Of Some Categories

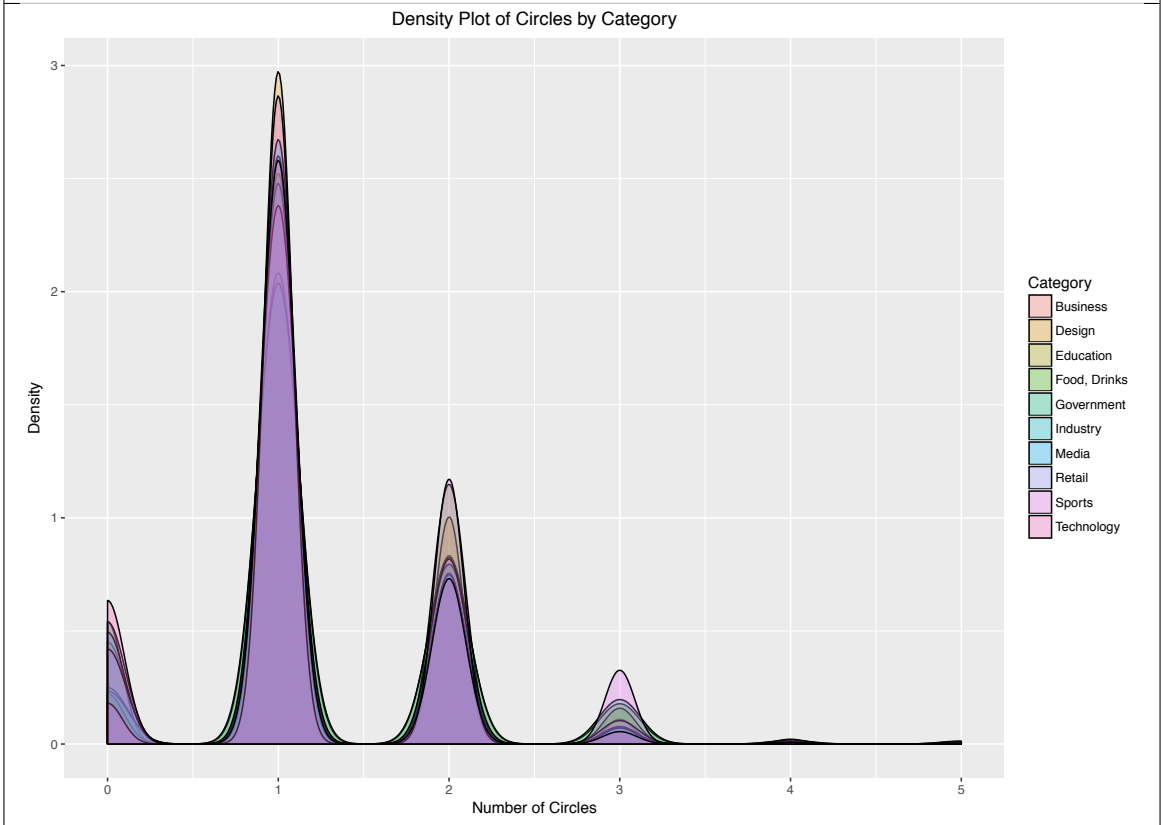


Density Plot of the Number of Edges Across 10 Most Common Categories

Figure 3: Density Plots of Low-level Features by Category



Density Plot of the Number of Polygons Across 10 Most Common Categories



Density Plot of the Number of Circles Across 10 Most Common Categories

Figure 4: Density Plots of Low-level Features by Category

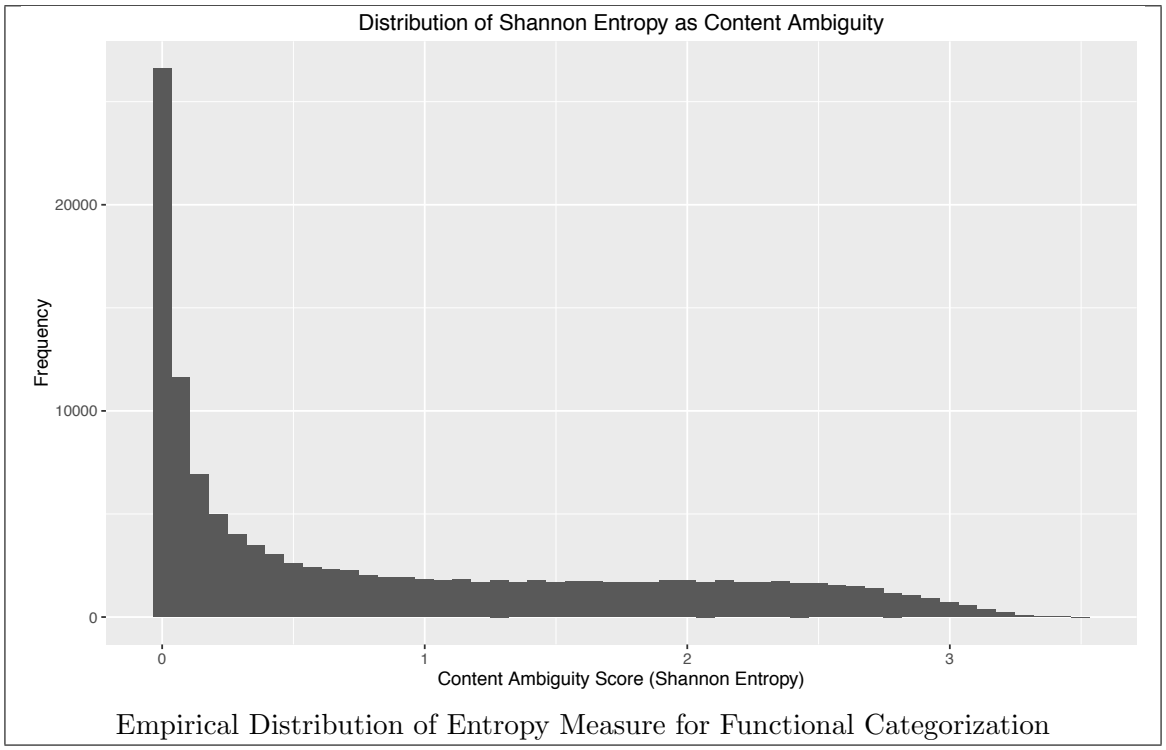


Figure 5: Empirical Distributions of Two Functional Categorization Measures (by intuition, they measure two dimensions of functional ambiguity of logos)

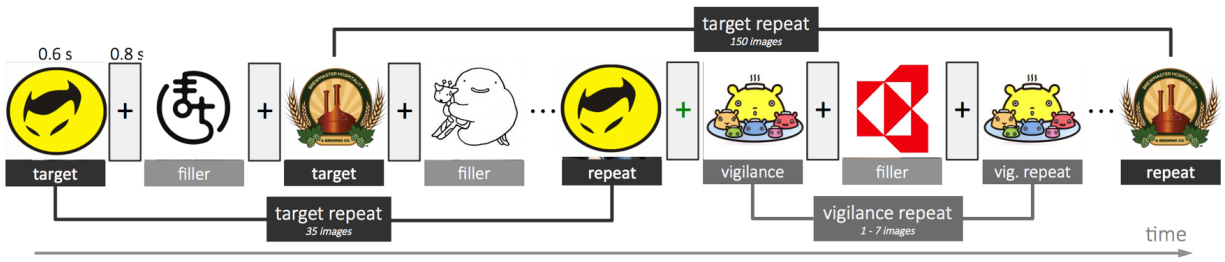


Figure 6: A Flow-chart of Experimental Procedure

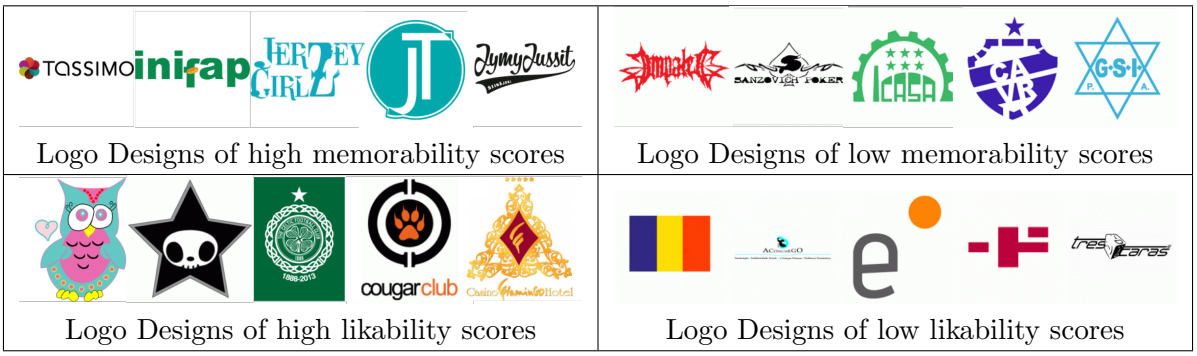


Figure 7: Sample images arranged by their memorability and likability scores: top left logos are of high memorability scores, top right of low memorability scores, bottom left of high likability scores, bottom right of low likability scores

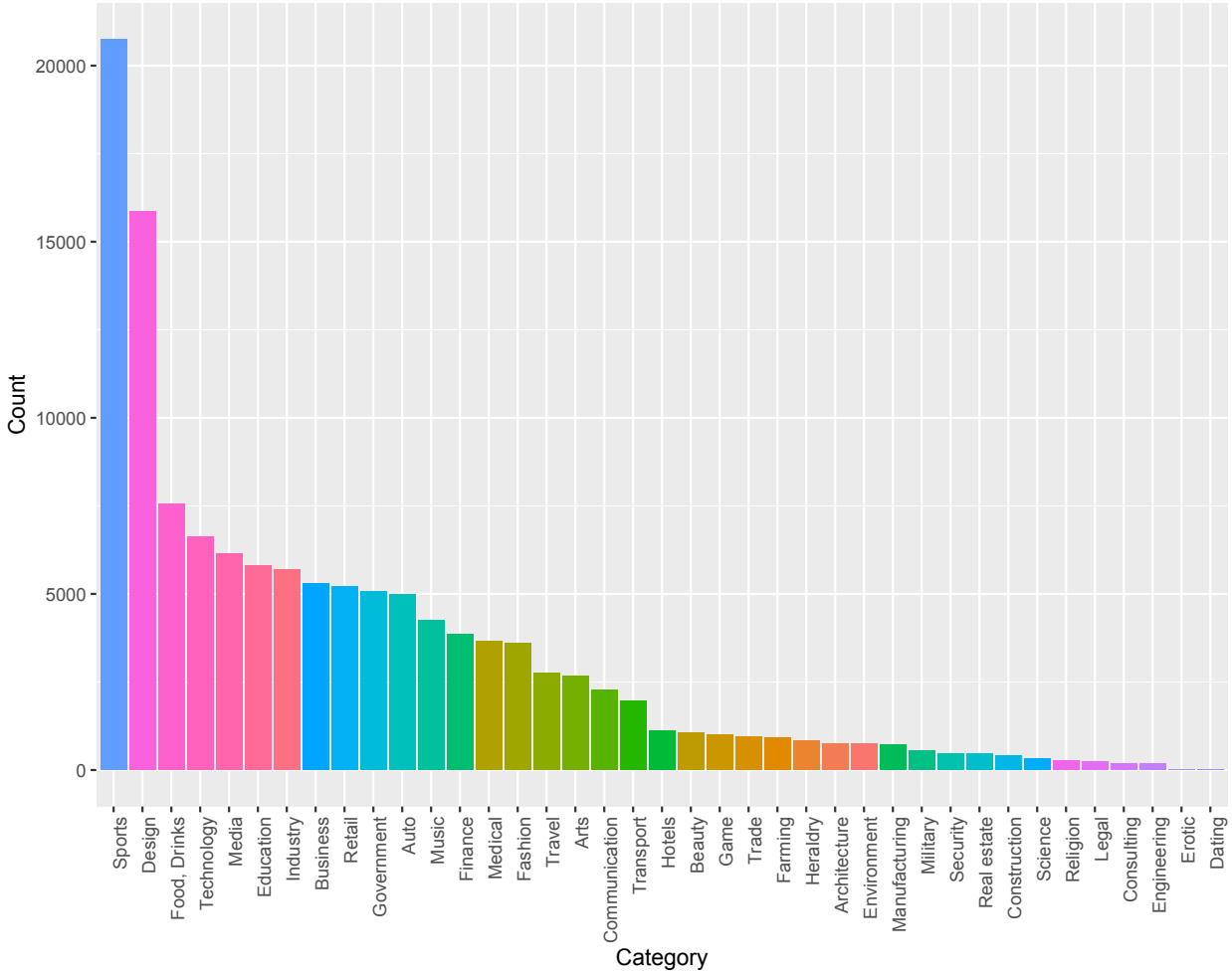


Figure 8: Histogram of Logo Industrial Category

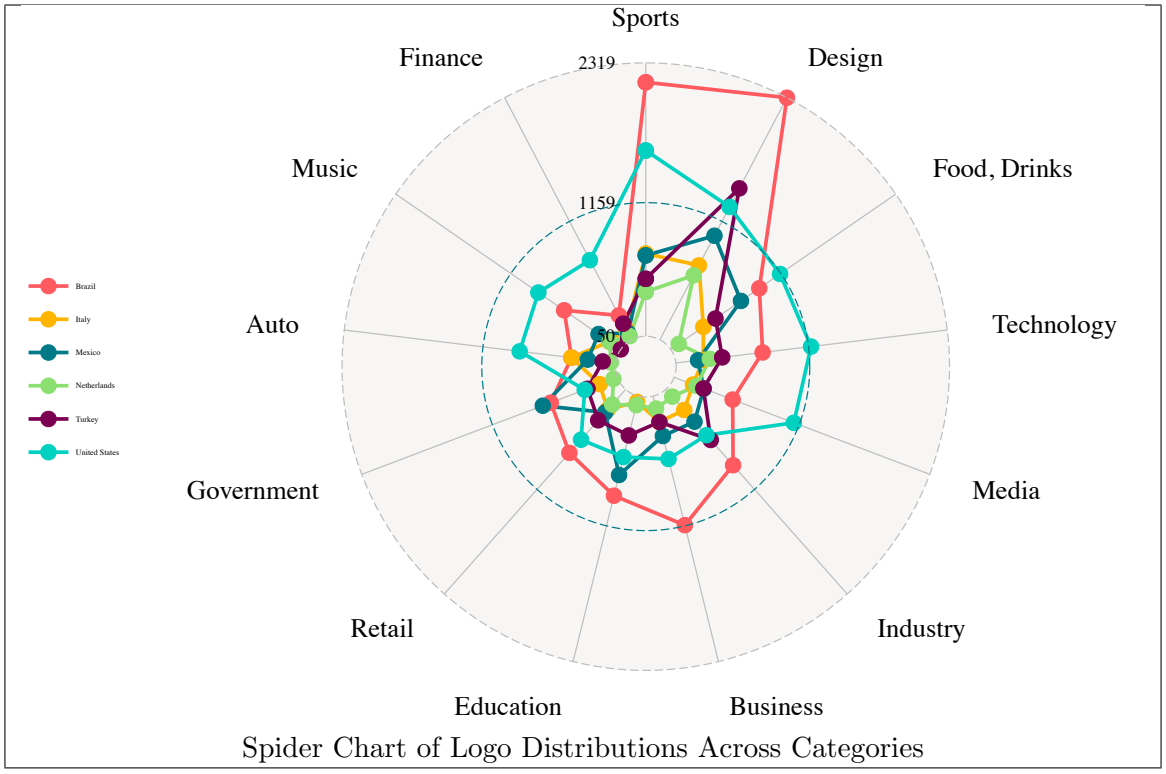


Figure 9: Spider Chart of Logo Distributions Across Categories

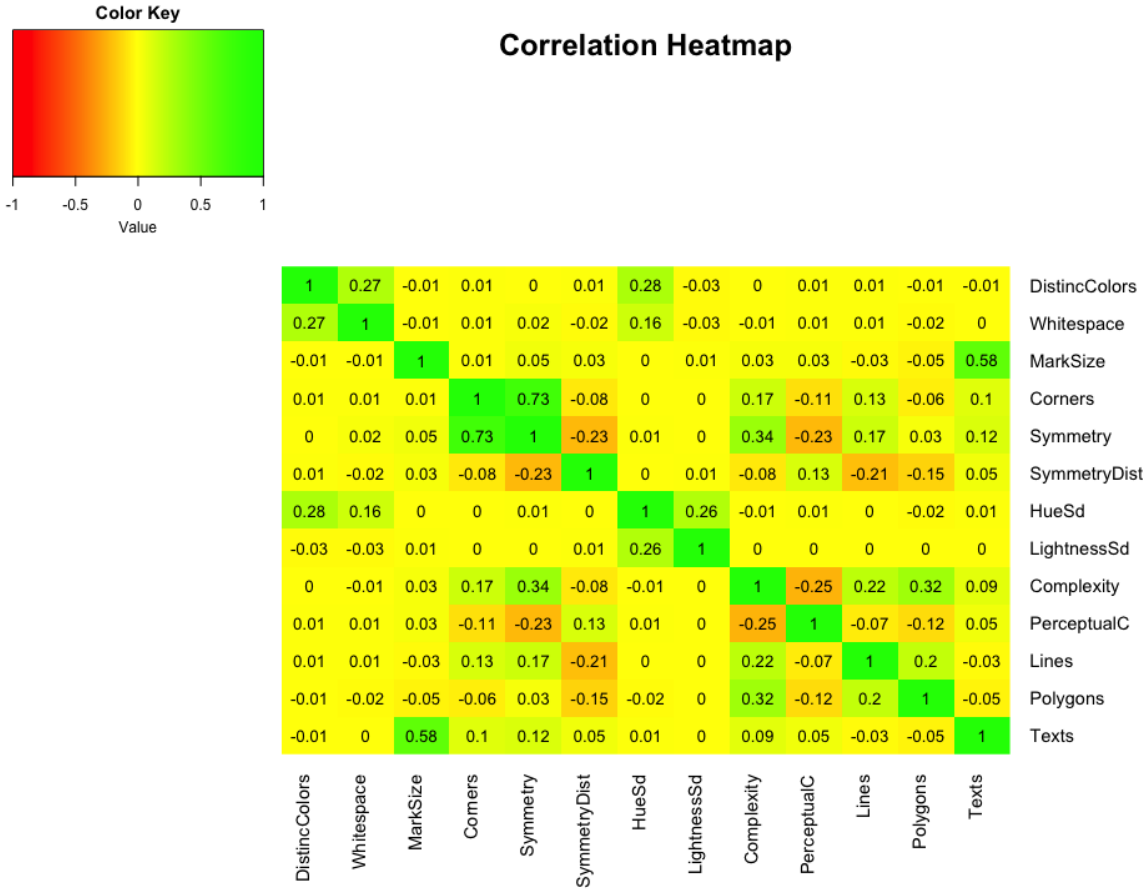


Figure 10: A Heatmap of Correlations between Variables

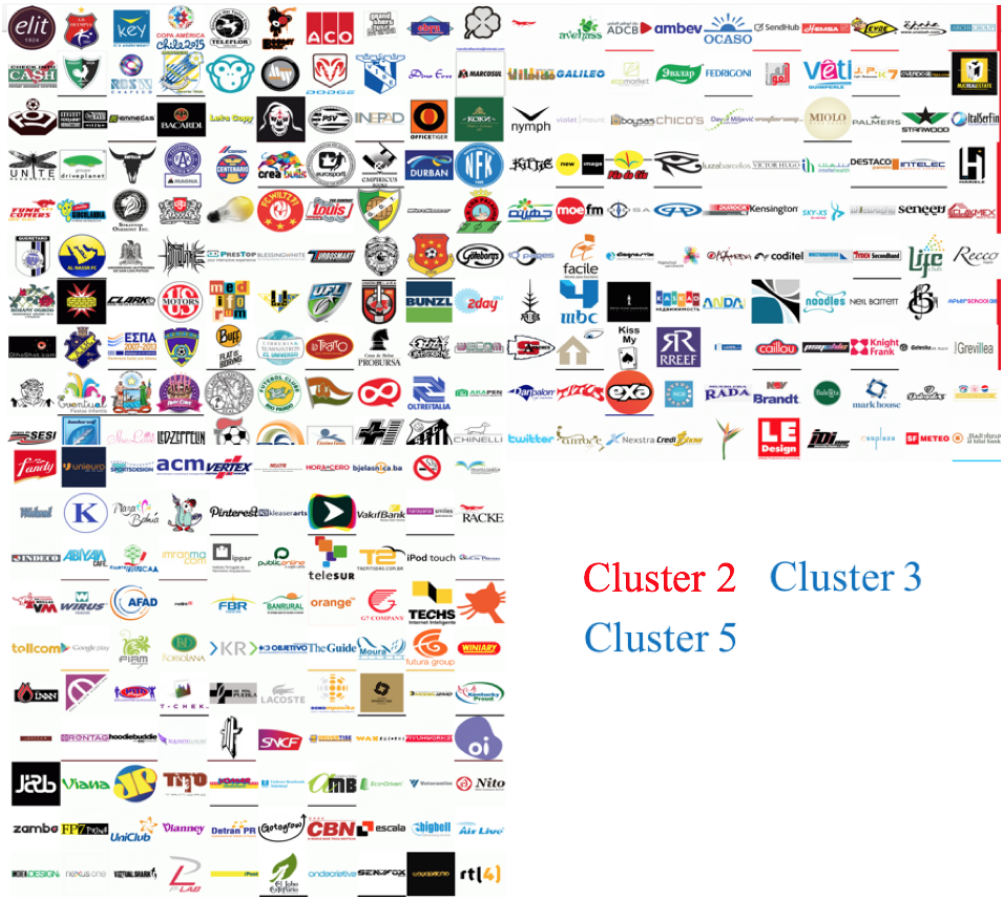


Figure 11: A Random Collage of Images in Cluster 2 (negative effects on memory but positive on liking), Cluster 3 (positive effects on memory but negative on liking), and Cluster 4 (positive effects on memory but negative on liking)

Figure 12: A Random Collage of Images in Cluster 2, cluster 4, Cluster 5, Cluster 7, and Cluster 9

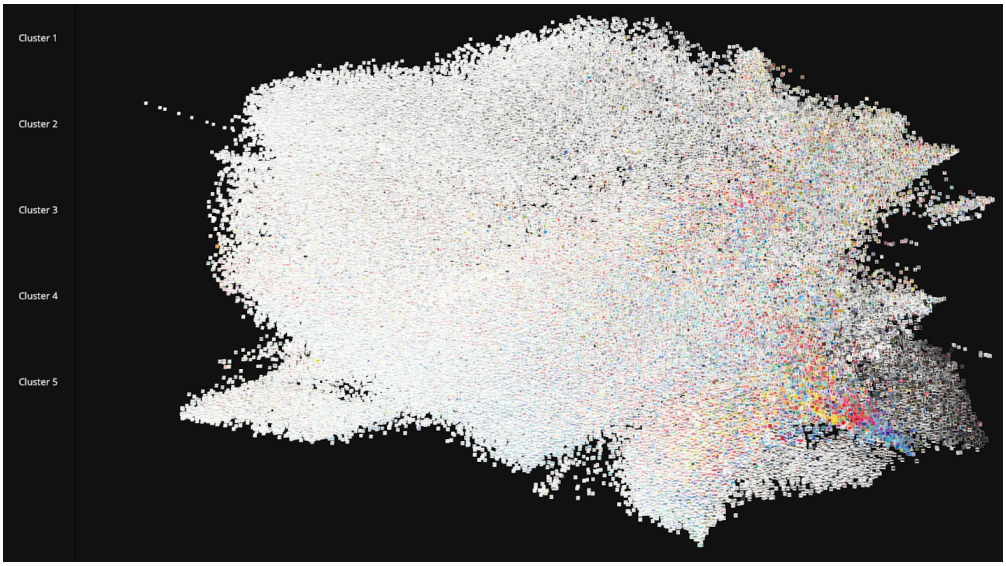


Figure 13: A Screen Shot of UMAP-empowered 2D Projection Visualization in 5 Clusters

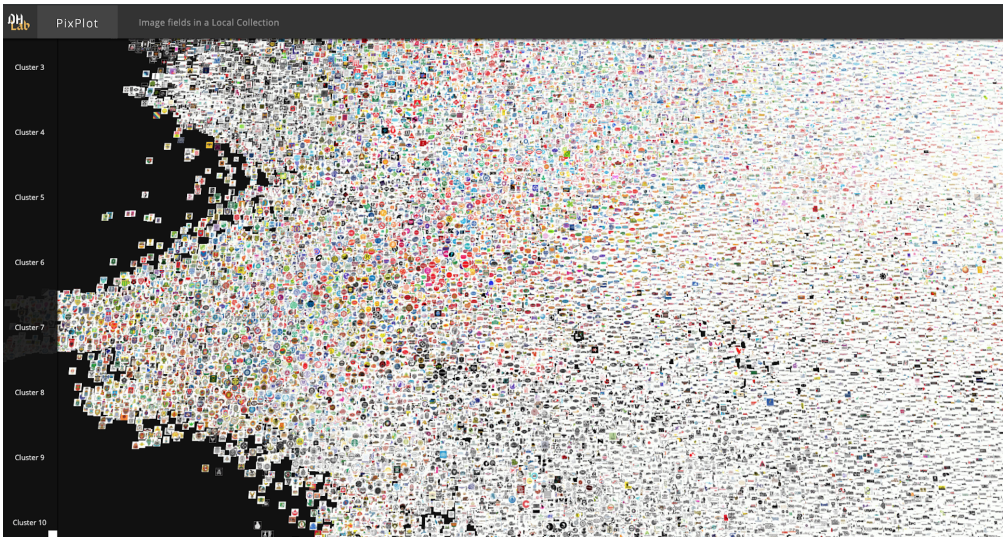


Figure 14: A Screen Shot of UMAP-empowered 2D Projection Visualization in 10 Clusters

Figure 15: A Screen Shot of UMAP-empowered 2D Projection Visualization in 20 Clusters

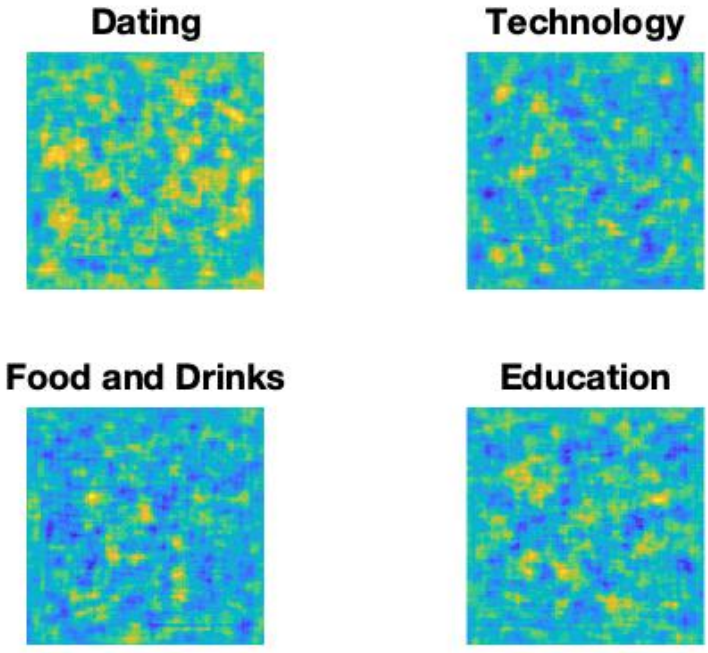


Figure 16: Visualizing Human Biases of Functional Categorization

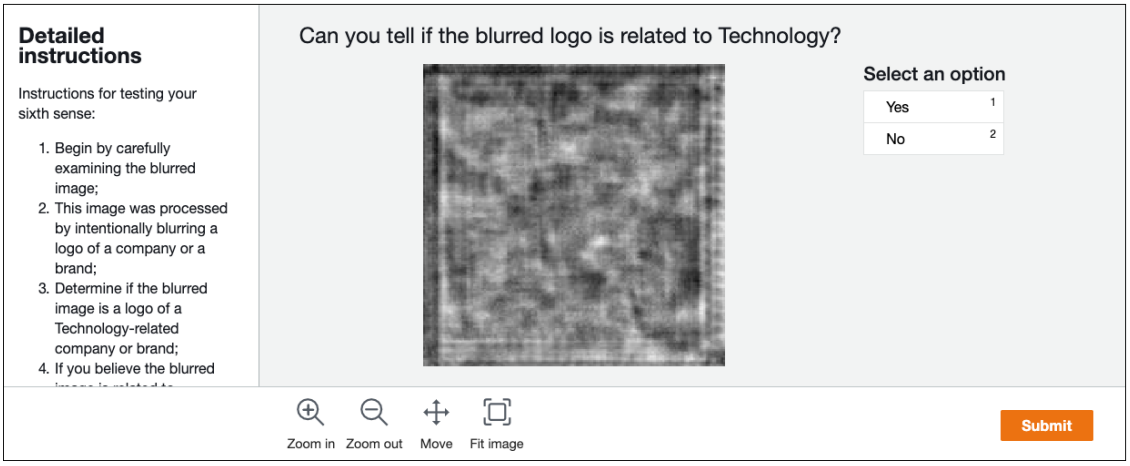


Figure 17: A Screenshot of the Hallucination Game for Measuring Human Biases

A Survey: Computer Vision for Visual Marketing

Shengli Hu

Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853

sh2264@cornell.edu

Since the revolution of big data and visual marketing began gaining prominence in the 2000s, a significant amount of research on both topics has emerged. This paper reviews the existing research on the intersection with a focus on visual image data and proposes a classification scheme for the diverse range of computer vision methods and constructs for visual marketing that has been developed in the literature. The classification criteria include the nature of research questions, application contexts, computational methods, and forms of big data. Additionally, the paper provides comparative evaluations of each criterion both horizontally and vertically, a normative guide on the use of these systems and results under different situations, and an agenda for future research.

Keywords: Marketing, Computer Vision, Business Application, Image Processing.

1. Introduction

Consumers today are inundated with visual marketing stimuli. From explicit daily advertisements on television, in newspapers, magazines, billboards, retail feature ads, and on the Internet, to the subtle product or brand placement in movies, shows, product packages, and public transportation boards, consumers are exposed to the omnipresent display of corporate marketing endeavor. As part of the corporate identity communication, such consists of ways companies employ to represent themselves to the public consistently visually. All this is part of what was termed visual marketing more than a decade ago: the important utilization

of commercial and non-commercial visual signs and symbols to deliver desirable and useful messages and experiences to consumers (Wedel and Pieters 2012). The importance of visual marketing has been increasingly recognized over time, as a search for visual marketing on Google produces over 640 million hits in March 2019 (as opposed to 46 million hits in November 2006). As previous marketing research has established the definite link between exposure and the probability of inclusion in the consumers' consideration set when shopping for a product. It is important to manage what consumers pay attention to to maximize profit.

Visual processes are a central component in the mental stream of consumers, consciously and unconsciously, and thus influence consumer behavior directly. With the exponential increase in the amount and diversity of visual stimuli in the daily marketplace, the needs of companies and professionals to better understand their impact on consumer behavior and how such insights could be used to improve visual marketing efforts also grew at an ever more rapid rate.

Despite the prevalence of visual marketing in practice, the vast amounts of funds poured in, and the proliferation of theoretical studies in academia based on small amounts of laboratory data samples gathered from university labs, efficient and scalable systems tailored for big data have long been lacking or were not synthesized in visual science. The body of both theoretical and methodological knowledge backing visual marketing efforts is still nascent.

Visual marketing lies at the intersection of vision science, cognitive psychology, and social psychology. Although vision science is interdisciplinary itself with roots in psychology, neuroscience, computer science, and optometry, and aesthetics, among others (Palmer 1999), there exists few well-structured efforts that illustrate how computer vision methods have been applied to big data for visual marketing applications. The objective of this paper is to provide a concise but comprehensive review of the academic literature on the intersection of computer vision and visual marketing, to propose a classification scheme for the diverse range of extant studies verging on both topics, and to offer a comparative evaluation of these studies in terms of theoretical, methodological, and practical contributions. Additionally, the paper provides a normative guide on the use of tailored computer vision methods under different visual marketing situations and an agenda for future research.

2. Why Computer Vision for Visual Marketing?

Marketing science has enjoyed a long tradition of embracing new challenges, new methods, and new disciplines (Chintagunta, Hanssens and Hauser 2016). The early 1980s had witnessed the data revolution experienced by consumer-packaged-goods firms, when the advent of scanner data enabled marketing scientists, to carefully observe the behavioral patterns of individual consumers purchasing groceries over many shopping trips, across many product categories, and for various purposes. Methods gradually evolved from initial descriptive analyses to discrete choice models, time series analyses, Bayesian methods, natural or field experiments, and beyond. They were first introduced into the discipline to solve marketing-science problems and then developed further to give back to disciplines from where they came.

The current big data revolution is to a great extent parallel to the scanner-data revolution. Roughly half a century later, what data scientists face today are extensive databases housing real-time unstructured data of consumers' online traces in the form of text, audio, image, and video.

To take advantage of big data for visual marketing, marketing science will need to embrace disciplines such as vision science and especially computer vision to make sense of the enormous volume of image and video files at a large scale. For instance, how can we efficiently parse out style information out of favored fashion images to infer consumers' preferences for fashion and art? What information can we glean from video data of movie trailers to make recommendations for new movie subscriptions?

Another reason why the marriage between visual marketing and computer vision is promising and called-for comes from the impetus for firms to improve operational efficiency by adopting machine learning and data science to assist decision making or automate essential marketing activities, many of which involve visual targeting. The relative scalability, automacy, and generalizability that come with computer vision methods in the era of big data prove to be essential for companies to gain competitive advantages. Such needs have been recognized from research communities in marketing and computer vision, evidenced by themed workshops at top vision conferences such as ECCV and CVPR, special issues on big data (Chintagunta et al. 2016) from top journal outlets of marketing science, and many more in the near future.

3. Review and Classification of Computer Vision for Visual Marketing Research

A steady stream of academic research on the interface of marketing and computer vision began gaining momentum in the 2010s. The Computer Vision Foundation (CVF) organized the first workshop on fashion, art, and design in conjunction with the European Conference on Computer Vision (ECCV) in 2018, and the first and second workshops on subjective attributes of data with applications in marketing and media in the conjunction to the annual conference of Computer Vision and Pattern Recognition (CVPR) in 2018 and 2019. The Marketing Science Institute (MSI) organized the first special issue of Marketing Science on Big Data in May 2016, which included some of the first few articles that apply computer vision methods to marketing problems (Lu, Xiao and Ding 2016).

While much of early research has focused on the designing computer vision systems for marketing applications, recent research has concentrated more on developing computational methods that connect marketing or consumer theories with large-scale unstructured datasets or machine learning systems. Alongside academic research, various AI startups have been actively deploying efficient computer vision systems for marketing applications at a large scale. For instance, GrokStyle[1] — Visual AI for Retail — develops an app that automatically identifies furniture and home decor from just about any picture or angle, and was recently acquired by Facebook.

We provide in Table 1 our classification scheme for the existing research on computer vision for visual marketing. We classify research according to the following criteria:

1. nature of research questions;
2. context of applications;
3. computational methods;
4. forms of data.

where each criterion consists of a number of categories that are hopefully comprehensive but by no means exhaustive.

[1] `https://www.grokstyle.com/`

Major Criteria	Sub-Criteria	Example Studies
Nature of Research Question	Causal Inference	Zhang et al. (2018), Xiao and Ding (2014)
	Prediction	Dzyabura et al. (2018)
	Descriptive Research	Liu and Mayzlin (2018)
	Recommender System	Chao et al. (2009), Lu et al. (2016), Veit et al. (2015)
	Representation Learning	Vittayakorn et al. (2015), Veit et al. (2015)
	Image Search	Liu et al. (2012), Bell and Bala (2014)
	Other System Designs	Mei et al. (2012), Zhou et al. (2016) Tkachenko et al. (2018)
Application Contexts	Fashion and Beauty	Malik et al. (2018), Todorov (2018)
	Advertisement	Joo et al. (2014), Hussain et al. (2017)
	Product Design	Zhang et al. (2018), Tkachenko et al. (2018)
	Subjective Attributes	Isola et al. (2014, 2016)
	Brand Image and Identity	Dew et al. (2018)
	User-generated Content	Papatla (2018), Liu and Mayzlin (2018)
	Storytelling	Rohrbach et al. (2015), Tapaswi et al. (2016)
	Style	Thomas and Kovashka (2016) Murillo et al. (2012)
Computational Method	Image Processing	Dew et al. (2018), Zhang et al. (2018)
	Machine Learning	Bossard et al. (2012), Dzyabura et al. (2018)
	Deep Learning	Liu and Mayzlin (2018), Joo et al. (2015)
	Combination	Shi et al. (2018)
Form of Data	Natural Image	McAuley et al. (2015), Isola et al. (2015)
	Synthetic Image	Wilber et al. (2017), Bylinski et al. (2017)
	Image and Text	Mei et al. (2012), Yamaguchi et al. (2012)
	Video	Hussain et al. (2017), Rohrbach et al. (2015)

Table 1: Overview of Classification Scheme for Computer Vision for Visual Marketing

Some observations are in order. First, all combinations of these criteria are meaningful. Second, these criteria are not independent nor exclusive. For instance, if one addresses a research question focused on causal inference (for criterion 1), econometric methods and random experiments (for criterion 3) might become more relevant.

Given the full range of research questions, approaches, datasets, and contexts that have

been adopted in this realm, the three criteria of Table 1 demonstrate a systematic way of identifying and grouping different kinds of research efforts regarding computer vision and visual marketing. Further, such a framework allows us to assess the strengths and weaknesses of the various methods, research angels, and datasets. Therefore, we detail the resulting comparative evaluations at the end of each subsection following the discussion for each criterion. Lastly, we apply these criteria for evaluating the practical relevance of each research project for different marketing management decisions. We detail such a normative guide on the use of computer vision for visual marketing in the following section.

3.1 Nature of Research Questions

One of the defining aspects of research efforts in different disciplines that characterizes and distinguishes from one another is the nature of the research questions and methods thereof, and it serves as the first criterion in our classification scheme.

3.1.1 Causal Inference

Causal inference has long been one of the foci of econometric analysis, and it is only after the successful establishment of causality can strategies and solutions be prescribed to optimize for the market outcomes of interest. Following this research paradigm, the effects of visuals on various market outcomes have been studied extensively in different contexts — sharing economy, fashion, social marketing, print advertising, among others.

Zhang, Lee, Singh and Srinivasan (2018) estimate the economic impact of images and low-level image features in sharing economy. More specifically, they explore the relevance and importance of aesthetic visual images of property with respect to demand in Airbnb. By classifying property photos based on different features derived from aesthetics and photography literature, they show that 48.9% of the effect of verified images boosting demand comes from the high image quality. They also identify and demonstrate twelve image attributes that have direct impacts on demand, thus prescribing optimal product image strategies to increase demand for housing and lodging managers.

Shi, Lee, Singh and Srinivasan (2018) study the substitutability between the value of brand and style in the fashion market. They quantify the style value by Xiao and Ding (2014) study the effect of non-celebrity faces in print advertising. Specifically, they propose the use of eigenface features to segment people based on their preferences towards

different faces. Significant and substantial effects of faces on viewers' attitudes towards the ad, the brand, and their purchase intentions were documented. Considerable consistency is identified within subjects, whereas substantial heterogeneity exists between subjects and among product categories: certain eigenfaces are more predictive of greater viewer affinity for specific product categories. They also find that the effect of faces interacts with product categories and is mediated by various facial traits such as attractiveness, trustworthiness, and competence.

Malik, Singh, Lee and Srinivasan (2018) investigate the dynamic effects of beauty over an individuals career. They score the attractiveness of every individual in a longitudinal sample on career milestones, with which they estimate a survival analysis where attractive men are found to progress faster in their career early on, and women are found to progress faster in their later career in comparison to their unattractive counterparts respectively. In other words, they find that men enjoy a beauty premium early in their career which disappears later in the career, whereas the opposite goes for women, even though the overall beauty premium is higher for women.

Papatla (2018) teases out the effect of individuals versus advertised products on viewers' attention in user-generated visual content on social media. More specifically, their study investigates whether the presence of faces in user-generated visual content could be less detrimental to brand image if they are less prominent relative to the product image.

Comparative Evaluation

Measuring the causal effects of visual images in marketing contexts provides valuable information on the effectiveness of visual marketing strategies and consumers' feelings and perceptions of the visuals. It also enables the counter-factual simulations of different compositions of visual cues and the resulting consumer response. Nevertheless, the primary weaknesses of this research paradigm lie in the measurements of visual stimuli — from the selection process of these measurements to the inherent weaknesses such as lack of interpretability, transparency, and fairness of the algorithms involved in these measurements. The common small sample sizes also cast doubts on the identifiability of causal inference.

For instance, the low-level image features identified in Zhang et al. (2018) were manually selected based on relevant literature, and used for classification tasks on property images. Such a two-stage pipe casts doubts on the validity of the image classifiers as data scientists

at Airbnb are adopting integrated systems where useful features are selected as the same time as the image classifiers are trained, with the entire pipeline's optimization objective aligned.

While Xiao and Ding (2014) is arguably the first study that has applied visual dimension reduction methods — eigenface decomposition — the interpretability is lacking. What does "eigenface 1 boosts the metric" mean in practice? Nor did the authors provide any prescriptive answers such as "how to generate the best face on the print advertising for beer?" Such could be quickly addressed by generative models based on their descriptive methods.

3.1.2 Prediction Research

The foundations of machine learning frameworks on which most contemporary computer vision methods rest make it unsurprising that most existing studies we review on the intersection of visual marketing and machine learning fall into this category — prediction, with the objective being notable increases in predictive powers. Marketing researchers and computer scientists alike have devised various predictive models for outcome variables of interest — consumer demand, consumer impression (click-through rate), political election results, perceived subjective attributes of images, among others.

Dzyabura, Ibragimov and Kihal (2018) use machine learning models to predict demand in online and offline retail channels, as well as returns for new products. They measure sale distributions across channels for each product category and find those predominantly sold online are more prone to customer returns. They also demonstrate the superior prediction performance when product image features — color histograms, texture, learned representation by training AlexNet (Krizhevsky, Sutskever and Hinton 2012) — are included.

Azimi, Zhang, Zhou, Navalpakkam, Mao and Fern (2012) study the relationship between the visual appearance and performance of creatives using large scale data in the worlds largest display ads exchange system, RightMedia. They design a set of 43 visual features, categorized into three different sets:

- Global features: grey level features, color distributions, model-based color harmony, color coherence, hue distribution, lightness features;
- Local features: segment size, segment hues, segment color harmony, segment lightness;
- Advanced features: saliency features, character counts, number of faces.

They extracted the visual features and conduct a series of experiments to evaluate the effectiveness of visual features to predict click-through rate, ranking and performance classification. Based on the evaluation results, they selected a subset of features that have the most important impact on click-through rates, useful for ads selection and developing visually appealing creatives.

Joo, Steen and Zhu (2015) infer the perceived traits of a person from his face — social dimensions, such as "intelligence", "honesty" and "competence" — and how those traits can be used to predict the outcomes of political elections, job hires, and marriage engagements. The authors propose a hierarchical model for enduring traits inferred from faces, incorporating high-level perceptions and intermediate-level attributes. Surprisingly, they show that the trained model can successfully classify the outcomes of two important political events, only using the photographs of politicians' faces. It classifies the winners of a series of U.S. elections with the accuracy of 67.9% (Governors) and 65.5% (Senators).

Huang and Kovashka (2016) extend Joo et al. (2015) by exploring a variety of features for predicting communicative intents. They study a number of facial expressions and body poses as cues for the implied nuances of the politician's personality, as well as how the environmental settings such as kitchen or hospital influence the audience's perception of the portrayed politician. They improve the performance by learning intermediate cues using convolutional neural networks and document state-of-the-art results on the *Visual Persuasion* dataset of Joo et al. (2015).

Comparative Evaluation

The predictive approaches are more scalable and rely more heavily on feature engineering, in the absence of meta-learning, towards the particular outcome. So the value of the research itself hinges upon the resulting predictive power over competitive alternative approaches, as well as the economic or social impact of the predictive outcomes. One common weakness of the reviewed articles is due to the subjective and heavy feature engineering, making the resulting predictive models context or application specific, and easily obsolete. For instance, with most recent advances in computer vision and deep learning, the feature extraction methods used in Dzyabura et al. (2018) and Azimi et al. (2012) are often replaced with more efficient models and architectures, and more adaptive learning methods have since been proposed to solving similar predictive problems to Huang and Kovashka (2016).

3.1.3 Descriptive Analysis

Descriptive studies could prove promising if presented with novel datasets, new research questions, and interesting or potentially impactful insights, spurring streams of future research to follow.

Liu and Mayzlin (2018) propose a "visual listening in" approach to measuring how brands are portrayed on social media (Instagram) by mining visual content posted by users. They first measure brand attributes (glamorous, rugged, healthy, fun) from images. Then they apply the classifiers to brand-related images posted on social media to measure what consumers are visually communicating about brands. By comparing the portrayals of 56 brands in the apparel and beverages categories in consumer-created images with images on the firms official Instagram account, further contrasted with consumer brand perceptions measured in a national brand survey, they find, despite convergent validity shown in all three measures, critical differences between how consumers and firms portray the brands on visual social media, and how the average consumer perceives the brands.

Comparative Evaluation

Given the descriptive nature of such studies, direct and practical applications would be lacking. However, the managerial insights and implications generated might be relevant for other downstream tasks such as prediction, causal inference, recommender system designs, among others.

3.1.4 Recommender System

Many major players in the consumer marketplace such as Amazon, Netflix, Spotify, Etsy, Pinterest, and Reddit have built their business models around powerful recommender systems. With the proliferation of visual content that defines information goods (movies, shows, aesthetics) and how consumers interact with products, it is not surprising research efforts that integrate image processing and computer vision into traditional recommender systems have been fruitful.

Chao, Huiskes, Gritti and Ciuhu (2009) present a clothing recommendation system called the Smart Mirror. They use computer vision to recognize classes and attributes of clothing for personal fashion recommendation, mimicking a real-time customer fashion assistant in a store's fitting room. Figure 1 showcases a prototype of the smart mirror. The end user stands

in front of a mirror-TV with an embedded web camera. First, a users' image is captured and the face region is detected automatically, based on which a rectangular region of interest (ROI) is selected to characterize the current clothing style. Using various image descriptors for ROI, similar clothes in style can be found in a database to be displayed to the user by the mirror-TV for inspiration. A slightly more marketing version of a garment recommendation system was published in Marketing Science *seven* years later (Lu et al. 2016), as reviewed below.

Figure 1: Prototype of the Smart Mirror

Lu et al. (2016) developed an automatic and scalable garment recommender system that integrates video analysis — real-time facial expression recognition and hand detection — at the individual customer level with existing marketing research methods to create useful managerial tools in retail. The garment recommender system:

1. uses a camera to capture a shopper's behavior in front of the mirror to make inferences about her preferences based on recognized facial expressions and region of interest (clothing tried on);

2. matches the customer with a database of individuals with known fashion tastes and preferences, within which those with similar tastes are identified as nearest neighbors;

3. makes fashion recommendations to the focal customer based on preferences of identified individuals in the database.

The significant difference between this marketing paper and the previous Smart Mirror is that Lu et al. (2016) added another feature of facial recognition to estimate customers' emotional responses to articles of clothing they are trying on. Such an additional design

element could backfire due to the heterogeneity of individual facial expressions. Neither studies addressed the garment compatibility recommendation problem — for instance, an article of clothing might be similar but not compatible to what the customer is wearing, and vice versa. Therefore it creates the problem that the recommender systems might not be able to help the focal customer find suitable clothing, which begs the next article we review below.

Veit, Kovacs, Bell, McAuley, Bala and Belongie (2015) propose a learning framework to recommend matching articles of clothing to consumers. The type of questions their system is trained to answer is along the lines of "What outfit goes well with this pair of shoes?" The idea of this framework is to learn a feature transformation from images of items into a latent space that expresses compatibility. The authors model compatibility based on large-scale user co-purchase data from Amazon.

Comparative Evaluation

Recommender systems developed for creative visuals appear still nascent in comparison to more mature recommender systems in other contexts such as music, product, and news. Therefore, one of the major weaknesses in this category stems from the difference between fashion or design products versus daily grocery products or movies. Not only are fashion and design products more difficult to categorize and objectively cluster, consumers purchasing decisions on fashion and design products are also more variable and context-dependent. Recommender systems for creative visual built upon traditional recommender systems do not provide satisfying solutions to such problems and therefore under-deliver. Music recommender systems that feature sequence modeling methods for precise description of hidden state transitions (Chen, Moore, Turnbull and Joachims 2012, Moore, Chen, Turnbull and Joachims 2013, Moore, Chen, Joachims and Turnbull 2012) could be of great value.

3.1.5 Feature Representation Learning

Representation learning could be particularly useful for visual marketing applications as it provides one way to perform unsupervised learning and semi-supervised learning, especially when large amounts of unlabeled training data are available (Goodfellow, Bengio and Courville 2016). Such is often the case with online user-generated visual content, which makes representation learning of image data appealing.

The idea of the learning framework proposed by Veit et al. (2015) to recommend matching articles of clothing to consumers is to learn a feature transformation from images of items into a latent space that expresses compatibility. For the feature transformation, a Siamese Convolutional Neural Network (CNN) architecture is used, where training examples are pairs of items that are either compatible or incompatible. The authors model compatibility based on large-scale user co-purchase data from Amazon. Pairs of heterogeneous dyads are constructed as training data in order to learn cross-category fit. They demonstrate that the proposed framework is capable of learning semantic information about visual style and can generate outfits of clothes, with items from different categories, that are aesthetically compatible.

Vittayakorn, Yamaguchi, Berg and Berg (2015) documents the first attempt to provide a quantitative analysis of fashion on the runway and in the streets. They develop a feature representation that can usefully capture the appearance of clothing items in outfits, through pose estimation, clothing parsing, and feature extraction. They collected human judgments of outfit similarity to train models for predicting the similarity between runway outfits and street outfits. They found their proposed representation boosts prediction performance of the season, year, and brand, outperforming humans.

Tkachenko, Ansari and Toubia (2018) apply deep learning for computer-aided exploration of visual product designs. In this work, they confirm properties of the lower-dimensional latent space that are desirable — they show (a) how distance between images in this latent space mimics human similarity judgments about the actual images, and (b) that essential characteristics of interest, such as product prices, can be predicted from latent image data alone.

Todorov (2018) model social perception of faces using data-driven approaches whose objective is to identify quantitative relationships between high-dimensional variables (e.g., visual images) and behaviors (e.g., perceptual decisions) with as little bias as possible. They conduct a series of studies using reverse correlation methods based on judgments of randomly generated faces from a statistical, multidimensional face model; a vector space where every face can be represented as a vector in the space. These methods are used to a) model evaluation of faces on any social dimension (e.g., trustworthiness), and b) to identify the perceptual basis of this evaluation, thus mapping configurations of face features to specific social inferences.

Comparative Evaluation

A good feature representation is one that makes a subsequent learning task easier (Goodfellow et al. 2016). The choice of representation usually depends on the choice of the subsequent learning task, which is defined by the research question itself. Therefore, this particular link between representation learning and the downstream learning tasks restricts the application of feature representation learning to a few subsets of research questions in our survey — prediction problems and recommender systems with less interpretability and transparency. Another concern stems from the lack of intuitive and interpretable interfaces for learned image representations.

3.1.6 Image Search

Content-based image retrieval has been spawned numerous research streams and industrial search engines. It covers and harmonizes with the image and text data as they flow from one computational component to another: query formulation, image feature extraction, representation learning, and indexing, similarity learning and search, visualization, among others (Gevers and Smeulders 2004). In this section, we review some research efforts in visual marketing that belong to this pipeline.

Garces, Agarwala, Gutierrez and Hertzmann (2014) present a method for measuring the similarity in style between two pieces of vector art, independent of content. The similarity is measured by the differences between four types of features: color, shading, texture, and stroke. Feature weightings are learned from crowdsourced experiments. This perceptual similarity enables style-based search, with which they demonstrate an application that allows users to create stylistically-coherent clip art mash-ups.

Liu, Song, Liu, Xu, Lu and Yan (2012b) develop a system that takes a daily street snapshot of individuals, and efficiently searches online for articles of clothing similar to the outfit in the street photo. It is a problem of cross-scenario clothing retrieval— the core of which lies in correctly identifying similarities between clothing against large discrepancies between street photos and online catalog photos due to distinct human posture and environmental background. The proposed solution leverages human pose estimation and offline structure detection. Bell and Bala (2015) tackle the same search problem but generalize to product images of various categories, the core algorithm and framework became the backbone of the AI startup, GrokStyle.

Yamaguchi, Hadi Kiapour and Berg (2013) introduce an effective retrieval-based clothing parsing method along with a large annotated dataset of fashion photos. For a query image, they find similar styles from a large database of tagged fashion images and use these examples to parse the query. Their approach combines parsing from: pre-trained global clothing models, local clothing models learned on the fly from retrieved examples and transferred parse masks (paper doll item transfer) from retrieved examples. They show that this approach significantly outperforms state-of-the-art in parsing accuracy.

Comparative Evaluation

As existing research that intersects image retrieval and visual marketing are sparse, one of the major drawbacks in this realm would be lack of comparison or standard benchmarks to establish the validity of methods introduced therein.

3.1.7 Other System Design

Many other machine learning systems for visual marketing show great promise for practical application, especially in the realm of advertisement generation and targeting, among others.

Chilton (2018) tackle the problem of creative ad generation given a central message to be conveyed to the target audience. For instance, given the message "Smoking kills you", their system outputs an image that blends two visual symbols — one represents the subject ("smoking"), and one represents the predicate ("kills you") — a handgun loaded with cigarettes. More details to be found in their draft yet to be released.

Mei, Li, Hua and Li (2012) presents a contextual advertising system driven by images, which automatically associates relevant ads with an image rather than the entire text in a Web page and seamlessly inserts the ads in the non-intrusive areas within each image. The proposed system, called ImageSense, supports scalable advertising of, from root to node, Web sites, pages, and images. In ImageSense, the ads are selected based on not only textual relevance but also visual similarity, so that the ads yield contextual relevance to both the text in the Web page and the image content. The ad insertion positions are detected based on image salience, as well as face and text detection, to minimize intrusiveness to the user.

Tkachenko et al. (2018) propose a deep-learning based exploratory system for visual product designs, where alternative design methods, such as conjoint or brute-force search, may not be applicable or may perform suboptimally. They demonstrate machine learning

techniques for exploration and ideation in the design space, such as image interpolation to generate product designs that are similar to competitors products or constrained Bayesian optimization to find novel designs that score high on quantitative characteristics of interest. They use images and attributes of products sold on Amazon.com as well as personal feedback from Amazon Turk workers as a basis for experiments. Their results imply that deep generative models offer a promising avenue for partial automation of the visual product design process.

Hussain, Zhang, Zhang, Ye, Thomas, Agha, Ong and Kovashka (2017) create two datasets of image ads and video ads, both of which contain rich annotations of the topics, the sentiments, questions, and answers describing the objectives, the reasoning the ad presents to persuade the viewer, as well as the symbolic references ads make. The authors develop a computer vision system to understand these ads and evaluate it on tasks such as symbolic question-answering, topic and sentiment recognition. The highest accuracy they achieved on symbolism prediction is 50% on the image dataset, and the accuracies for topic and sentiment predictions on video ads were 35.1% and 32.8%, respectively. They also predicted whether the videos ads were funny or exciting — the resulting accuracies were 78.6% and 78.2% accordingly.

Zhou, Lu and Ding (2016) propose a face anonymity-perceptibility framework to anonymize and distinguish facial images in online dating. It brings together face anonymity research from computer vision literature, and the perceptibility studies from the social and the neuropsychology literature for marketing applications such as online dating, hiring, sales, and security. They select a set of facial landmarks for local or global facial features depending on the abstraction method used to anonymize. Then they show users abbreviated profiles including facial abstractions that preserve anonymity and perceptibility at the same time for preference estimation. Finally, a smaller set of potential partners are selected for the user to the best of his or her liking. The major shortcoming of this study lies in the notorious issue of discrimination bias inherent in facial recognition algorithms. Such problems are especially relevant and essential in the online dating context. For instance, when most indicated user preferences are tilted towards facial features prominent in white males and Asian females, how would the algorithm perform, and how would a social planner like it to perform? When someone ethnically (facially) new appears, how would the algorithm cope compare to a social planner, as well as users themselves?

Comparative Evaluation

Due to the novelty of the machine learning systems reviewed above, relative to mature and clear-cut research streams such as recommender systems and search systems, most of these new systems are either based on small experimental data samples, or a collage of different and separate steps not yet optimized or integrated, and therefore not scalable for practical implementations. Greater concerns are cast to studies without performance comparisons between proposed systems and conventional systems in use in practice.

3.2 Context of Applications

In this section, we cluster existing research by the context of applications, compare and contrast within each cluster, and across clusters.

3.2.1 Fashion and Beauty

Fashion, beauty, art, and design have been the first few areas where marketing research and computer vision research intersect. The Computer Vision Foundation (CVF) has been organizing the Computer Vision for Fashion, Art, and Design workshop annually at top conferences of computer vision since 2018, accelerating and promoting research efforts in this realm.

Chen, Xu, Liu and Zhu (2006) use an And-Or representation to build a tree of composite clothing templates by gathering and segmenting artists' sketches and match those clothing templates to the image. This is one of the first studies on clothing in the computer vision community.

Malik et al. (2018) investigate the dynamic effects of beauty over an individuals career, using computer vision methods to score the attractiveness of every individual. With a survival model, they find that men enjoy a beauty premium early in their career which disappears later in the career, whereas the opposite for women, even though the overall beauty premium is higher for women. These results are biased until proven wrong, due to the inherent biases widely documented in computer vision classifiers applied to human faces. Discriminative biases against race, gender, and skin color are common pitfalls in such applications. Therefore more clarification would be useful in this aspect.

Todorov (2018) identify quantitative relationships between visual images and consumer perceptual decisions with as little bias as possible. They randomly generate faces from a statistical, multidimensional face model, from which a vector space where every face can be represented as a vector in the space is learned. Such methods are useful for model evaluation of faces on any social dimension such as trustworthiness, and identification of the perceptual basis of this evaluation, thus mapping configurations of face features to specific social inferences.

Veit et al. (2015) propose a learning framework based on feature transformation with a Siamese Convolutional Neural Network (CNN) architecture to recommend matching articles of clothing to consumers, addressing daily matching problems such as "what outfit goes well with this pair of shoes?" They demonstrate that the proposed framework is capable of learning semantic information about visual style and can generate outfits of clothes, with items from different categories, that are aesthetically compatible.

Vittayakorn et al. (2015) provide a first quantitative analysis of fashion on the runway and in the streets. With a large-scale dataset of runway fashion photos representing $9,328$ fashion shows over 15 years, complemented with human judgments of outfit similarity to train models for predicting the similarity between runway outfits and street outfits, they develop a feature representation that captures the appearance of clothing items in outfits. They found their proposed representation boosts prediction performance of the season, year, and brand, outperforming humans.

Bossard, Dantone, Leistner, Wengert, Quack and Van Gool (2012) build a pipeline for recognizing and classifying clothing in a natural setting, combining upper body detectors, style classification, attribute classification in a Random Forest.

Liu et al. (2012b) develop a system that takes a daily street snapshot of individuals, and efficiently searches online for articles of clothing similar to the outfit in the street photo. Bell and Bala (2015) tackle the same search problem but generalize to product images of various categories, the core algorithm and framework became the backbone of GrokStyle, the AI startup for product search acquired by Facebook.

Chao et al. (2009) and Lu et al. (2016) both build clothing recommendation systems that recognize classes and attributes of clothing, mimicking a real-time customer fashion assistant in a store's fitting room. The user stands in front of a mirror-TV with an embedded web camera, through which a users' image is captured, and the face region is detected automatically. Then a rectangular region of interest (ROI) is selected to characterize the

current clothing style.

Comparative Evaluation

Creative domains such as fashion, art, and design have sparked interests among marketing scholars and computer vision scientists alike. Many efforts have revolved around the creation, consumption, and analysis of creative visual content. Some of the recent trends in the vision community include (1) computer vision for fashion and (2) visual content generation for creative applications.

The large-scale analysis and understanding of fashion (or style) have growing interest with direct applications on advertising, product design, and other aspects of marketing. Some of the more fundamental studies address the design of unsupervised techniques to learn a visual embedding that is guided by the fashion style (Hsiao and Grauman 2017, Karayev, Trentacoste, Han, Agarwala, Darrell, Hertzmann and Winnemoeller 2013), or learn image similarity in the context of fashion (Veit et al. 2015) and design (Bell and Bala 2015). A closely related research stream involves learning visual representations for visual fashion search (Ak, Kassim, Lim and Tham 2018), especially in the context of social media (Li and Tang 2015, Gutierrez, Sondag, Butkovic, Lacy, Berges, Bertrand and Knudson 2018).

Research foci have been on tailoring both traditional and novel systems for the fashion, art, and design applications, which can be summarized as the following directions:

1. How to automatically generate creative artworks, fashion pieces, and visual designs?
2. How can algorithms help guide the creative processes of human?
3. How to automatically identify styles and measure consumer preferences for different styles?
4. How to tailor recommender systems, search systems, and knowledge graphs for creative domains?

Even though existing research is sparse on these topics, each of the above directions has been investigated by computer scientists and marketing scholars separately, almost in parallel. For instance, Sbai, Elhoseiny, Bordes, LeCun and Couprie (2018) and Huynh, Ciptadi, Tyagi and Agrawal (2018) introduce generative models that facilitate design inspirations and matching garment recommendation, whereas Tkachenko et al. (2018) and Lu et al. (2016)

tackle computer-aided design and garment recommendation problems with a greater focus on psychological and consumer theory; Hsiao and Grauman (2017) and Veit et al. (2015) use fashion images to create virtual wardrobe with special attention paid to matching compatibility, whereas Yoganarasimhan (2017) and Shi et al. (2018) use traditional econometric methods to study fashion trends and to separate fashion from brand.

3.2.2 Advertisement

Advertising has been one of the most popular and mature topics in research communities of both marketing and computer science. We review vision research on advertising with emphases on the overlap, as well as differentiating methodological and substantive innovations.

Xiao and Ding (2014) study the effect of non-celebrity faces in print advertising with established face recognition methods from computer vision. Specifically, they propose the use of eigenface features to segment people based on their preferences towards different faces. Significant and substantial effects of faces on viewers' attitudes towards the ad, the brand, and their purchase intentions were documented, differentiating their work from computer science research with an algorithmic focus.

Considerable consistency is identified within subjects, whereas substantial heterogeneity exists between subjects and among product categories: certain eigenfaces are more predictive of greater viewer affinity for specific product categories. They also find that the effect of faces interacts with product categories and is mediated by various facial traits such as attractiveness, trustworthiness, and competence.

Joo, Li, Steen and Zhu (2014) first introduce the problem of understanding visual persuasion in computer vision. A compelling image has an underlying intention to persuade the viewer by its visuals and is widely used in mass media, such as TV news, advertisements, and political campaigns. Joo et al. (2014) focus on understanding the underlying intents of such persuasive images. They identify twelve syntactical features as predictors and nine dimensions of communicative intents as labels. The communicative intents are grouped into three buckets:

1. Emotional traits: happy, angry, fearful;

2. Personality traits and values: competent, energetic, comforting, trustworthy, socially dominant;

3. Overall favorability: favorable.

As are syntactical features:

1. Facial display: smile, look down, eye open, mouth Open;
2. Body cues — gestures: hand-wave, hand-shake, finger-point, touch-head, hug;
3. Scene context: large crowd, dark-background, indoor.

Figure 6 from the original study is reproduced in Figure 2. For all dimensions, the full approach that exploits all three types of syntactical features yields the best result. Besides, the facial display type outperforms the other cues on the emotional dimensions while the gesture type is more discriminative for three among five dimensions of personality traits and values. The heavy feature engineering and lack of connection to market outcomes

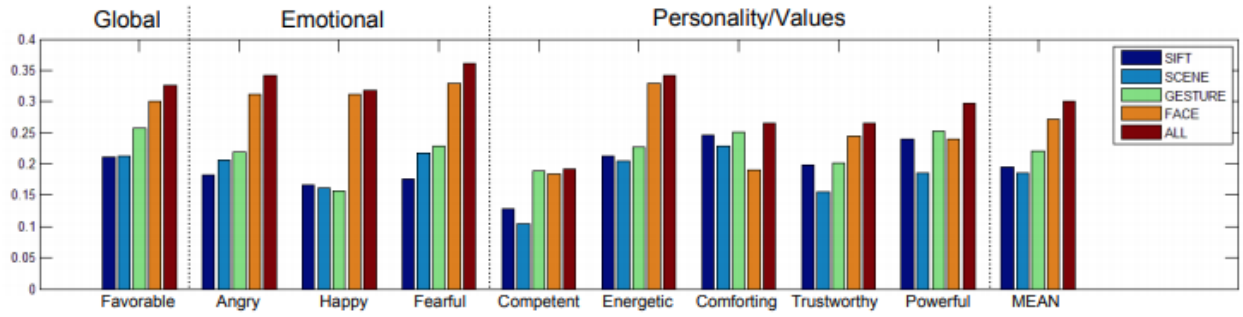


Figure 2: Figure 6 in Joo et al. (2014): Intents prediction performance evluation[2]

distinguishes this work from marketing research of similar objectives.

Hussain et al. (2017) create two datasets of image ads and video ads, both of which contain rich annotations of the topics, the sentiments, questions, and answers describing the objectives, the reasoning the ad presents to persuade the viewer, as well as the symbolic references ads make. The authors develop a computer vision system to understand these ads and evaluate it on tasks such as symbolic question-answering, topic and sentiment recognition. More interesting applications could build on the result from Hussain et al. (2017). For instance,

- methods could be developed to predict how effective a certain ad will be given a target audience;
- on the part of viewers, the automatic ad-understanding system could help prevent being tricked into buying certain products;

- by decoding the messages of ads, better ad-targeting strategies could be tailored according to user interests;
- it could be extended to generate automatic ad descriptions and summarizations.

In the same vein, but reverse order, Chilton (2018) tackle the problem of creative ad generation given a central message to be conveyed to the target audience. For instance, given the message "Smoking kills you", their system outputs an image that blends two visual symbols — one represents the subject ("smoking"), and one represents the predicate ("kills you") — a handgun loaded with cigarettes.

State-of-the-art prediction algorithms for click-through rates on non-guaranteed display advertising rely heavily on historical information collected for advertisers, users and publishers, which makes it challenging when new advertisements are encountered due to the lack of historical data. Cheng, Zwol, Azimi, Manavoglu, Zhang, Zhou and Navalpakkam (2012) propose to mitigate this problem by integrating multimedia features extracted from display ads into the click prediction models because multimedia features can help capture the attractiveness of the ads with similar contents or aesthetics. The authors evaluate the use of numerous multimedia features, in addition to commonly used user, advertiser and publisher features, and demonstrate that adding multimedia features can significantly improve the accuracy of click prediction for new ads, compared to a baseline model.

McDuff, El Kaliouby, Cohn and Picard (2015) collected over $12,0000$ facial responses from $1,223$ people to 170 ads from a range of markets and product categories. The facial responses were automatically coded frame by frame at a large scale. Detected expressions were found, but aggregate responses revealed rich emotion trajectories. They found that ad liking is driven by mostly positive expressions, whereas determinants of purchase intent are more complicated: peak positive responses that are immediately preceded by a brand appearance are more likely to be effective.

Mei et al. (2012) presents a contextual advertising system driven by images, which automatically associates relevant ads with an image rather than the entire text in a Web page and seamlessly inserts the ads in the non-intrusive areas within each image. The proposed system, called ImageSense, supports scalable advertising of, from root to node, Web sites, pages, and images. In ImageSense, the ads are selected based on not only textual relevance but also visual similarity, so that the ads yield contextual relevance to both the text in the Web page and the image content. The ad insertion positions are detected based on

image salience, as well as face and text detection, to minimize intrusiveness to the user. They collected a unique advertisement-image dataset and demonstrated the effectiveness of ImageSense for online image advertising.

Likewise, Yadati, Katti and Kankanhalli (2014) propose an in-stream video advertising strategy for advertising platforms such as Youtube, which they term as Computational Affective Video-in-Video Advertising (CAVVA). They focus on the role emotions play in influencing the buying behavior of users and therefore factor in the emotional impact of the videos as well as advertisements. Given a video and a set of advertisements, their method first identifies candidate advertisement insertion points, as well as the suitable advertisements based on marketing and consumer psychology theories. The two-stage problem is integrated as a single optimization function in a non-linear 01 integer programming framework and a solution identified based on genetic algorithm. They evaluate CAVVA with a subjective user-study and eye-tracking experiment and were able to demonstrate that CAVVA strikes a balance between the conflicting goals of (a) minimizing the user disturbance because of advertisement insertion while (b) enhancing the user engagement with the advertising content. Their method was benchmarked against existing advertising strategies, and they show that CAVVA can enhance the users experience and increase the monetization potential of the advertising content.

Sánchez, Binefa and Vitrià (2002) approach the problem of automatic TV commercials recognition, and introduce an algorithm for scene break detection. The structure of each commercial is represented by the set of its key-frames, which are automatically extracted from the video stream. The commonly used shot boundary detection techniques are based on individual image features or visual cues, which show significant performance lacks when they are applied to complex video content domains like commercials. The authors present a scene break detection algorithm based on the combined analysis of edge and color features. Local motion estimation is applied to each edge in a frame, and the continuity of the color around them is checked in the following frame, and thus the continuous presence of the objects and the background of the scene during each shot is essential for the proposed algorithm. They show that this approach outperforms single feature algorithms in terms of precision and recall.

Lastly, Zhao, Yuan, Xu and Wu (2011) propose a data-mining method for thematic object discovery in commercials by finding spatially collocated visual features, such as logos, human facial reactions, ad placement, and recognition.

3.2.3 Product Design

The process of engineering products, so they are more likely to be shared among peers, purchased by loyal customers, appealing to potential new customers, among other marketing goals, has been extensively researched on since decades ago. We review below most recent efforts focused on design issues in visual marketing.

Zhang et al. (2018) estimate the economic impact of images and low-level image features — in other words, visual product packing — on property demand in Airbnb. By classifying property photos based on different features derived from aesthetics and photography literature, they show that 48.9% of the effect of verified images boosting demand comes from the high image quality.

Dzyabura et al. (2018) convey a similar message of the importance of visually representing products online by predicting demand in online and offline retail channels, using product image features. They demonstrate the superior prediction performance when product image features — color histograms, texture, learned representation by training AlexNet (Krizhevsky et al. 2012) — are included.

Tkachenko et al. (2018) propose computer-aided exploration of visual product designs-Their results imply that deep generative models offer a promising avenue for partial automation of the visual product design process.

Chan, Mihm and Sosa (2017) look into the ebbs and flows of styles in product design and how it relates to the evolution of product functionality over time, based on a large-scale US design patent dataset. They identify the styles in design using clustering and provide empirical evidence for the long-lived architectural and design mantra "Form follows function".

Comparative Evaluation

Similar to fashion, recent trends in both the marketing and the vision community for visual designs include (1) computer vision metrics for visual designs and (2) visual content generation for design applications.

Given the similarity of fashion and design in the feature space, methods applied to learn a visual embedding that is guided by the fashion style (Hsiao and Grauman 2017, Karayev et al. 2013), or learn image similarity in the context of fashion (Veit et al. 2015) can be readily

applied to similar design problems, for instance, Bell and Bala (2015). Therefore, the same major challenges and questions remain for design:

1. How to automatically generate visual designs?
2. How can algorithms help guide the creative processes of human designers?
3. How to automatically identify design patterns and measure consumer preferences for different designs?
4. How to tailor recommender systems, search systems, and knowledge graphs for creative designs?

3.2.4 Subjective Attributes of Images

Traditionally, the recognition of *tangible* properties of data, such as objects and scenes, have been the focus of applications in computer vision. In the recent years, the understanding of *subjective attributes* of images has attracted attention of many vision researchers. Examples of subjective attributes of image data include interestingness (Dhar, Ordonez and Berg 2011, Gygli, Grabner, Riemenschneider, Nater and Van Gool 2013, Fu, Hospedales, Xiang, Gong and Yao 2014), evoked emotions and sentiment (Mohammad and Turney 2010), memorability (Isola, Xiao, Parikh, Torralba and Oliva 2014, Dubey, Peterson, Khosla, Yang and Ghanem 2015), creativity (Khosla, Xiao and Torralba 2012, Lowrey 2006, Khosla, Bainbridge and Torralba 2013, Isola et al. 2014), and aesthetics (Repp 1997, Reber, Schwarz and Winkielman 2004, Dhar et al. 2011, Quercia, O'Hare and Cramer 2014, Eisenman 2013), which have been the centerpieces of content marketing, advertising, product design, innovation, and consumer behavior research in marketing. For instance, marketing scholars have researched extensively on creativity (Toubia and Netzer 2016, Amabile, Hennessey and Grossman 1986), the effect of evoked emotions on consumer experience (Aurîer 1994, Novak, Hoffman and Yung 2000, Mano and Oliver 1993), aesthetics (Repp 1997, Krishna, Elder and Caldara 2010), interestingness (Alexandrov and Pollack 2013), etc.

Given the abstract nature of such constructs, many challenges arise regarding various aspects of relevant tasks:

1. Unbiased data collection and annotation of subjective assessments, or creative natural experiments that generate desirable annotations;

2. Tailored visual representations for subjective measures, and how the visual representations help predict market outcomes;

3. Sound measures of accuracy regarding subjective measures, where debiasing methods tailored for subjective assessments would be key;

4. Incorporating psychological or consumer behavior theories into machine learning approaches to automatically understand human perception at a large scale.

3.2.5 Brand Image, and User-generated Visual Content on Social Media

With the ubiquity of companies and brands on social media, coupled with the common practice of content marketing and brand storytelling, marketing academics and specialists recognize the importance of maintaining active social media presence, positive consumer engagement on social media, and monitoring online consumer conversations about brands. Most of the earlier research efforts were focused on user-generated content in the form of text. However, images are on their way to surpassing text as the medium of choice for social conversations, which necessitates the emergence of computer vision methods for information extraction in the wild.

Liu and Mayzlin (2018) propose a "visual listening in" approach to measuring how brands are portrayed on social media by mining visual content posted by users. By comparing the portrayals of brands in consumer-created images with images on the firms official Instagram account, they find critical differences between how consumers and firms portray the brands on visual social media, and how the average consumer perceives the brands.

Dew, Ansari and Toubia (2018) explore the visual elements in logos that express brand personality traits, based on which they introduce a logo tokenization algorithm that decomposes logos into theory-based and human-meaningful visual features. Applied to a small dataset of logos, matched with textual data from firms' websites, consumer evaluations of brands, third-party descriptions of the companies, they uncover uncovers links that exist between a brands logo, description, and personality, and thereby facilitate a better understanding of the underpinnings of good design, and inform the design of new logos.

Papatla (2018) teases out the effect of individuals versus advertised products on viewers' attention paid to user-generated visual content on social media. Their study provides answers to whether the presence of faces in user-generated visual content could be less detrimental

to brand image if they are less prominent relative to the product image. Based on findings that faces and bodies of humans in the visual field are processed holistically even if they are seen as distinct stimuli, they analyze consumer response to about 12,000 photos of 800 different products in six categories displayed by 35 online retailers.

Comparative Evaluation

Some of the major drawbacks of studies reviewed above include the heavy feature engineering based on marketing assumptions, relatively small sample sizes and therefore relatively high risk of dataset bias and the disconnect between best methodological practices and what was adopted in the studies.

3.2.6 Storytelling

Brand storytelling has been gaining momentum, as stories are one of the most effective media to connect with consumers, with a focus on sharing value and emotional empathy. Storytelling topics such as humor, suspense, surprise, and character identification have been extensively researched in marketing, media, and social psychology. Therefore, the visual aspects of storytelling could be particularly fruitful in facilitating consumer engagement.

Humor is a highly-valued personal skill — arguably a sign of intelligence and creativity. Chandrasekaran, Vijayakumar, Antol, Bansal, Batra, Lawrence Zitnick and Parikh (2016) document progress in understanding the subtleties of human expressions such as humor. They collected two datasets of abstract scenes that facilitate the study of humor at both the scene-level and the object-level. They annotated the funny scenes and explored the different types of humor depicted in them. By designing computational models that predict the funniness and alter the funniness of a scene, they were able to provide answers to questions such as "what content in a scene causes it to be funny?" They show that their models perform well quantitatively, and qualitatively through human studies. In the same vein, Chilton, Landay and Weld (2018) surveyed professional comedians and found evidence that the humor-generation process can be described. Based on this survey, they performed an analysis of news satire from *The Onion* and decomposed the process of humor creation into seven microtasks — aspect, expected reactions, expected reasons, associations, expectations violation mechanisms, beliefs, and evaluation.

Rohrbach, Rohrbach, Tandon and Schiele (2015) introduce a dataset of transcribed Descriptive video service (DVS) that is temporally aligned to full-length HD movies. DVS provides linguistic descriptions of movies and is by design mainly visual. Comparing DVS to scripts, they find that DVS is far more visual and describes precisely what is shown rather than what should happen according to the scripts created before movie production. Building on Rohrbach et al. (2015), Tapaswi, Zhu, Stiefelhagen, Torralba, Urtasun and Fidler (2016) evaluate automatic story comprehension from both video and text. Their dataset consists of 14,944 questions about 408 movies with high semantic diversity, with which the authors extended existing Question Answering techniques to show that question-answering with such open-ended semantics is hard and much future work awaits in this challenging domain.

Comparative Evaluation

One of the major difficulties in research on storytelling lies in designing and training machine learning models capable of understanding narratives and reasoning with common sense, which coincide with some of the state-of-the-art reasoning and visual-semantic frameworks in the language and vision research community. Similar to the interdisciplinary language and vision, many new challenges arise:

1. Unbiased data collection and annotation of brand storytelling, or creative natural experiments that generate desirable annotations;

2. Tailored visual representations for brand storylines, and how the visual representations help predict consumer engagement;

3. Incorporating psychological or consumer behavior theories into machine learning approaches to automatically understand and generate visual narratives at a large scale.

3.2.7 Style

There is a fast growing body of research that aims at learning a notion of styles from images, whether it be art (Karayev et al. 2013, Liu, Yan, Ricci, Yang, Han, Winkler and Sebe 2015) (including the explosion of studies on style transfer, following the seminal work of Gatys, Ecker and Bethge (2015) or Gatys, Ecker and Bethge (2016), which we review in Section 5.1.3), vehicle (Jae Lee, Efros and Hebert 2013), scenic spots (Doersch, Singh, Gupta,

Sivic and Efros 2012, Quercia et al. 2014), photograph (Thomas and Kovashka 2016), and clothing (Bossard et al. 2012, Veit et al. 2015, Kiapour, Yamaguchi, Berg and Berg 2014), which we review below.

Learning styles of art. Garces et al. (2014) present a method for measuring the similarity in style between two pieces of vector art, independent of content. The similarity is measured by the differences between four types of features: color, shading, texture, and stroke. Feature weightings are learned from crowdsourced experiments. This perceptual similarity enables style-based search, with which they demonstrate an application that allows users to create stylistically-coherent clip art mash-ups.

Learning city icons. Doersch et al. (2012) seek to automatically find visual elements that are most distinctive for a specific geo-spatial area, for example, the city of Paris, given a vast repository of geotagged imagery. A discriminative clustering approach is proposed to show that geographically representative image elements can be discovered automatically from Google Street View imagery. They demonstrate that these elements are visually interpretable and perceptually geo-informative, for tourism marketing of Paris. The discovered visual elements can also support a variety of computational geography tasks, such as mapping architectural correspondences and influences within and across cities, finding representative elements at different geo-spatial scales, and geographically-informed image retrieval.

Quercia et al. (2014) present a crowdsourcing project that aims to investigate, at scale, which visual aspects of a city neighborhood (e.g., London) make them appear beautiful, quiet, and happy. They collected votes from over 3.3K individuals and translate them into quantitative measures of urban perception, thereby quantifying each neighborhood's beautiful capital. By then using state-of-the-art image processing techniques, the authors were able to determine visual cues that may cause a street to be perceived as being beautiful, quiet, or happy. Effects of color, texture and visual words were identified. For example, the amount of greenery is the most positively associated visual cue with each of three qualities; by contrast, broad streets, fortress-like buildings, and council houses tend to be associated with the opposite qualities (ugly, noisy, and unhappy). Such insights are especially useful for marketing purposes in the travel industry, as well as the visualization of city identities and the personification of cities.

Learning styles of photos. (Thomas and Kovashka 2016) introduce the novel problem of identifying the photographer behind a photograph. They created a dataset of over 180,000 images taken by 41 well-known photographers, and examined the effectiveness of a variety

of features (low and high-level, including CNN features,) at identifying the photographer. They also trained a deep convolutional neural network tailored for this task. The authors show that high-level features significantly outperform low-level features.

Learning vehicle style. Jae Lee et al. (2013) present a weakly-supervised visual data mining approach that discovers connections between recurring midlevel visual elements in historical (temporal) and geographic (spatial) image collections, and attempts to capture the underlying visual style. In contrast to existing discovery methods that mine for patterns that remain visually consistent throughout the dataset, they discover visual elements whose appearance changes due to change in time or location; i.e., exhibit consistent stylistic variations across the label space (date or geo-location). Their approach first identifies groups of patches that are style sensitive; it then incrementally builds correspondences to find the same element across the entire dataset. Finally, they train style-aware regressors that model each element's range of stylistic differences. They apply it to date and geo-location prediction and show substantial improvement over several baselines that do not model visual style. The method's effectiveness is also demonstrated on the related task of fine-grained classification.

Learning clothing style. Murillo, Kwak, Bourdev, Kriegman and Belongie (2012) consider photos of groups of people to learn which groups are more likely to socialize with one another. This problem implies learning a distance metric between images. However, they require manually specified styles, called "urban tribes". Similarly, Bossard et al. (2012) who use a random forest approach to classify the style of clothing images, require pre-specified classes of style. Vittayakorn et al. (2015) learn outfit similarity, based on specific descriptors for color, texture, and shape. Therefore their system can retrieve similar outfits to a query image. McAuley, Targett, Shi and Van Den Hengel (2015) collect a large scale co-purchase dataset from Amazon and learn a notion of style and retrieve products from different categories that are supposed to be of similar style, using the image features from AlexNet (Krizhevsky et al. 2012) that was trained for object classification to learn their distance metric. Veit et al. (2015) address the same problem. Rather than using logistic regression, Veit et al. (2015) advance the method by fine-tuning the entire network with a Siamese architecture with a different sampling strategy, which they claim to be less prone to the cold-start problem in the previous paper. More specifically, Veit et al. (2015) propose a framework to learn a feature transformation from images of items into a latent space that expresses compatibility. For the feature transformation, a Siamese Convolutional Neural Network (CNN) architecture is used, where training examples are pairs of items that are either compatible or incompatible.

The authors model compatibility based on large-scale user co-purchase data from Amazon, as do McAuley et al. (2015). Veit et al. (2015) construct pairs of heterogeneous dyads as training data in order to learn cross-category fit. They demonstrate that the proposed framework is capable of learning semantic information about visual style and can generate outfits of clothes, with items from different categories, that are aesthetically compatible. Kiapour et al. (2014) further extend the identified style of clothing to personal wealth, occupation, and socio-identity, based on their assumption or observation that the clothing we wear and our identities are closely tied, revealing to the world clues other aspects of our lives. They designed an online competitive Style Rating Game called *Hipster Wars* to crowdsource reliable human judgments of style, with which they collected a dataset of clothing outfits with associated style ratings for 5 style categories: hipster, bohemian, pinup, preppy, and goth. They train models for between-class and within-class classification of styles and thus identifying clothing elements that are generally discriminative for a style, as well as items in a particular outfit that may indicate a style. Yamaguchi et al. (2013) introduce an effective retrieval-based clothing parsing method wherefor a query image, they find similar styles from a large database of tagged fashion images and use these examples to parse the query.

Comparative Evaluation

Major drawbacks of research on styles in various contexts include:

- the lack of generalizability of methods proposed across different contexts;
- the disconnect between style identification and the most probable downstream applications and predictions, such as using detected styles for consumer profiling, thus limiting the practical relevance and potential;
- the lack of psychological or consumer behavior theories that could have potentially guided the design of machine learning approaches.

3.3 Computational Methods

We further classify marketing studies by methodology, into image processing studies, machine learning studies, and deep learning studies. Since machine learning is a superset of deep

learning, the studies classified as those using deep learning technically belong to the category of machine learning. We also exclude computer vision literature in this section because almost all computer vision studies belong to deep learning or machine learning if earlier.

3.3.1 Image Processing

Before the takeoff of machine or deep learning, computer vision and image processing consist of low-level feature extraction and dimension reduction. We review the few studies that first introduced these methods into marketing research.

Zhang et al. (2018) classified property photos with computer vision models, even though the features they use are manually selected based on relevant literature, with which they approach the research question with low-level features for interpretability. Xiao and Ding (2014) propose the use of eigenface features to segment people based on their preferences towards different faces. Considerable consistency is identified within subjects, whereas substantial heterogeneity exists between subjects and among product categories: certain eigenfaces are more predictive of greater viewer affinity for specific product categories. Dew et al. (2018) introduce a logo tokenization algorithm that decomposes logos into theory-based and human-meaningful visual features to uncover uncovers links that exist between a brands logo, description, and personality, and thereby facilitate a better understanding of the underpinnings of good design, and inform the design of new logos.

3.3.2 Machine Learning

Dzyabura et al. (2018) use traditional machine learning models to predict demand in online and offline retail channels, as well as returns for new products. Bossard et al. (2012) build a machine learning pipeline for recognizing and classifying clothing in a natural setting, combining upper body detectors, style classification, attribute classification in a Random Forest.

3.3.3 Deep Learning

Liu and Mayzlin (2018) use supervised machine learning methods, traditional support vector machine classifiers, and deep convolutional neural networks, to measure brand attributes from images. Then they apply the classifiers to brand-related images posted on social media

to measure what consumers are visually communicating about brands. Joo et al. (2015) design a fully automated system that can infer the perceived traits of a person from his face and propose a hierarchical model for enduring traits inferred from faces, incorporating high-level perceptions and intermediate-level attributes. Huang and Kovashka (2016) improve the performance by learning intermediate cues using convolutional neural networks and document state-of-the-art results on the *Visual Persuasion* dataset of Joo et al. (2015). Tkachenko et al. (2018) apply deep learning techniques for computer-aided exploration of visual product designs, where alternative design methods, such as conjoint or brute-force search, may not be applicable or may perform suboptimally. Their results imply that deep generative models offer a promising avenue for partial automation of the visual product design process.

3.3.4 Combinations

Shi et al. (2018) quantify the style value by employing deep learning based computer vision techniques to create style features, including clothing style (for instance, compatibility between clothing items, creativity), model style (for instance, facial and body attractiveness), and photo style. These style features are incorporated in a dynamic structural model to estimate a dynamic structural model to analyze the content creation and consumption behavior of influencers in a fashion social network community.

Comparative Evaluation

The choice of computational methods almost always depends on the research problem being solved. The studies that adopt machine learning and deep learning methods exclusively are benchmarked against state-of-the-art methods and results in the computer vision community, as they are proposed to solve similar problems by nature. Therefore, major drawbacks of such studies include:

- the lack of scalability, parallelism, and flexibility for production in practice;
- the relatively small size of datasets from which the results are based on, and the absence of comparisons of model performance when applied to benchmark datasets in the vision community;
- the lack of important steps during training or post-training to ensure interpretability, transparency, fairness, and accountability of models and algorithms;

- the lack of large-scale dataset to benchmark against in the visual marketing community.

3.4 Nature and Structure of Data

In this section, we single out and classify studies whose contributions include the collection of large-scale datasets that could benefit the entire research community. We review them below by the type of data they introduce.

3.4.1 Natural Image Data

McAuley et al. (2015) introduced a large scale co-purchase dataset gathered from Amazon for the learning task of a notion of style. The dataset contains over 180 million relationships between a pool of almost 6 million objects. These relationships are a result of visiting Amazon and recording the product recommendations that it provides given our (apparent) interest in the subject of a particular web page. The statistics of the dataset shown in Table 1 of the paper is reproduced in Table 2. An image and a category label are available for each object, as is the set of users who reviewed it. Potential research questions involving this dataset include the interaction between product images and consumer co-purchase behavior with respect to different product categories, the effect of images on new product adoption and competition, among others.

3.4.2 Synthetic Image Data

Wilber, Fang, Jin, Hertzmann, Collomosse and Belongie (2017) collected a large-scale dataset of contemporary artwork from Behance, a website containing millions of portfolios from professional and commercial artists. The resulting dataset which they termed "the Behance Artistic Media Dataset", containing almost 65 million images and quality assurance thresholds, is available at `https://bam-dataset.org/`. They also create an expert defined vocabulary of binary artistic attributes that spans the broad spectrum of artistic styles and content represented in the dataset:

- Media attributes: they label images created in 3D computer graphics, comics, oil painting, pen ink, pencil sketches, vector art, and watercolor;
- Emotion attributes: they label images that are likely to make the viewer feel calm/peaceful, happy/cheerful, sad/gloomy, and scary/fearful;

Category	Users	Items	Ratings	Edges
Books	8,201,127	1,606,219	25,875,237	51,276,522
Cell Phones & Accessories	2,296,534	223,680	5,929,668	4,485,570
Clothing, Shoes & Jewelry	3,260,278	773,465	25,361,968	16,508,162
Digital Music	490,058	91,236	950,621	1,615,473
Electronics	4,248,431	305,029	11,355,142	7,500,100
Grocery & Gourmet Food	774,095	120,774	1,997,599	4,452,989
Home & Kitchen	2,541,693	282,779	6,543,736	9,240,125
Movies & TV	2,114,748	150,334	6,174,098	5,474,976
Musical Instruments	353,983	65,588	596,095	1,719,204
Office Products	919,512	94,820	1,514,235	3,257,651
Toys & Games	1,352,110	259,290	2,386,102	13,921,925
Total	20,980,320	5,933,184	143,663,229	180,827,502

Table 2: The types of objects from a few categories in the dataset and the number of relationships between them.

- Entry-level object category attributes: they label images containing bicycles, birds, buildings, cars, cats, dogs, flowers, people, and trees.

Furthermore, with computational experiments, the authors show the value of this dataset for artistic style prediction, for improving the generality of existing object classifiers, and for the study of visual domain adaptation. Other research problems that can be addressed with this massive dataset include quantifying perceptual biases, estimating visual templates, approximating the mapping between concrete and abstract concepts, among others.

Thomas and Kovashka (2016) created a dataset of over 180,000 images taken by 41 well-known photographers, exhibiting various artistic photographic styles.

Bylinskii, Kim, Donovan, Alsheikh, Madan, Pfister, Durand, Russell and Hertzmann (2017) introduced a curated dataset of 29K large infographic images sampled across 26 categories and 391 tags for training deep learning models for visual summarization, which they call "visual hashtags".

3.4.3 Image and Text

One important archival source containing hundreds of thousands of designs is the US design patent database. The US design patent chronicles the creation of new product forms patented in the US since 1842. The United States Patent and Trademark Office (USPTO) makes data on patents publicly accessible. Chan et al. (2017) maintain this large-scale public US design patent dataset at `http://www.stylesinproductdesign.com/data`. This could prove useful for detecting fashion cycles from design images over time, testing cognitive theories of creative processes and activities, among others.

Mei et al. (2012) released a unique advertisement-image dataset that consists of $7,285$ unique ad product logos with annotations of $32,480$ unique ad words done by 20 subjects, as well as $382,371$ images from `http://www.tango.msra` and $200,000$ images from Flickr, based on which they evaluate their proposed ad targeting system ImageSense with $1,100$ ad triggering pages (100 web pages from major news sites, $1,000$ images searched by top 100 image queries).

Yamaguchi, Kiapour, Ortiz and Berg (2012) introduced *the Fashionista dataset*, consisting of $158,235$ fashion photos with associated text annotations, and web-based tools for labeling. The dataset was collected from `Chictopia.com`, a social networking website for fashion bloggers. On this website, fashionistas upload "outfit of the day" type pictures, designed to draw attention to their fashion choices or as a form of social interaction with peers. They tend to display a wide range of styles, accessories, and garments. Besides, the pictures are also often depicted in relatively simple poses (mostly standing), against relatively clean backgrounds, and without many other people in the picture. Fashionista could be particularly illuminating when it comes to detecting fashion cycles over time, identifying style diffusion, testing competition and imitation theories in the context of the design and fashion industry, among others.

Yamaguchi et al. (2013) introduced *The Paper Doll dataset*, which is a large, complex, real-world collection of tagged outfit pictures from the aforementioned social network focused on fashion, `chictopia.com`. It consists of over 1 million pictures from Chictopia with associated metadata tags denoting characteristics such as color, clothing item, or occasion. Since *the Fashionista dataset* also uses Chictopia, *the Paper Doll dataset* exclude any duplicate pictures from *the Fashionista dataset*. From the remaining, they selected pictures tagged with at least one clothing item and ran a full-body pose detector (Yang and Ramanan 2011), keep-

ing those that have a person detection. This resulted in 339,797 pictures weakly annotated with clothing items and estimated pose.

Vittayakorn et al. (2015) released a large-scale dataset containing 348,598 runway fashion photos representing 9,328 fashion shows over 15 years, combined with the Paper Doll dataset released by Yamaguchi et al. (2013).

Kiapour et al. (2014) crowd-sourced reliable human judgments of clothing styles, and therefore introduced a dataset of clothing outfits with associated style ratings for 5 style categories: hipster, bohemian, pinup, preppy, and goth.

3.4.4 Video

Hussain et al. (2017) released two valuable datasets of advertisements:

1. an image dataset of 64,832 image ads;

2. a video dataset of 3,477 ads.

Both contain rich annotations of the topics, the sentiments, questions and answers describing the objectives, the reasoning the ad presents to persuade the viewer, as well as the symbolic references ads make. A rich set of research questions that can be potentially answered with this Visual Ad Dataset include identifying effective visual or visual-semantic marketing strategy combinations of persuasion.

Rohrbach et al. (2015) released a dataset of transcribed Descriptive video service (DVS), temporally aligned to full-length HD movies, and supplemented with the aligned movie scripts. DVS provides linguistic descriptions of movies and is by design mainly visual. In total the Movie Description dataset contains a parallel corpus of over 54,000 sentences and video snippets from 72 HD movies, building on which, Tapaswi et al. (2016) introduce the MovieQA dataset consisting of 14,944 questions about 408 movies with high semantic diversity. Each question comes with a set of five possible answers; a correct one and four deceiving answers provided by human annotators. Multiple sources of information video clips, plots, subtitles, scripts, and DVS (Rohrbach et al. 2015) were all included.

4. Normative Guide to the use of Computer Vision for Visual Marketing

4.1 Brand Management on Social Media

Shi et al. (2018) identify significant effects of brands and style features on the trendiness of a fashion look, as well as substitutability patterns between style features and brand levels. For managers and influencers in the fashion market, their results provide guidelines on how to engineer a fashion "look" that can attract the most attention.

4.2 Product Design

According to Zhang et al. (2018), high quality of verified images of rental property explains half of the boosted demand. And there exist twelve image attributes categorized in 3 components — composition, color, and figure-ground relationship — that have direct impacts on demand — diagonal dominance, rule of thirds, visual balance intensity, hue, saturation, brightness, contrast, clarity, area difference, color difference, and texture difference. Therefore, for housing and lodging managers, the optimal product image strategies to increase demand include optimizing for each of the twelve image attributes.

Similarly, when presenting the focal product or brand using visual images, marketing managers that are meticulous with image quality could be rewarded with higher online and offline demand, according to Dzyabura et al. (2018).

When designing for new products, brainstorming for new ideas, or testing for new product lines, the computer-aided design pipeline introduced by Tkachenko et al. (2018) could be potentially helpful in suggesting innovative feature combinations and component selections.

4.3 Advertising

Various facial traits such as perceived attractiveness, trustworthiness, and competence were found to significantly influence viewers' attitudes towards the print ad, the brand, their purchasing intentions and willingness to pay, barring substantial heterogeneity among viewer demographics and product categories (Xiao and Ding 2014). Advertising managers could leverage such results to increase the visibility, acceptance, and pricing of the focal product

and the brand by the careful design and choice of faces on magazines or newspapers tailored for every consumer segment and product category.

For the design of video advertisements, a consumer response simulator based on Joo et al. (2014) or Hussain et al. (2017) could be efficient during the iterations of ad engineering, facilitating fast ad prototyping and product shipping. In addition, systems such as Chilton (2018) could aid human designers in their creative processes to craft the visuals, the messages, and the combination at the same time. In particular, systems such as Mei et al. (2012) could be potentially fruitful in the brainstorming stage of ad creation, whereas those similar to Zhao et al. (2011) and Sánchez et al. (2002) could be helpful in the synthesizing and editing phase of ad creation.

4.4 Consumer profiling

Marketing analysts that segment, analyze, and profile customers could leverage various tools and results in sections of subjective attributes of data (Section 3.2.4), style detection, and style classification (Section 3.2.7), to learn highly valuable semantic user and product embeddings for a number of downstream applications such as recommender systems, search engines, network analysis, among others.

4.5 Brand positioning

Social media marketers could potentially leverage the results from Liu and Mayzlin (2018) by first understanding the disparity between the brand image crafted by the company and that perceived by consumers, identifying the link between such disparity and desirable market outcomes, and therefore, optimizing social marketing strategies for the desirable market outcome. When in execution, the particular choice of social influencers and the visual product placement could be informed by the results in Papatla (2018) to maximize influence and acceptance on social media.

Similarly, for corporate logo designers, an intricate and fine-grained correspondence between image features and perceived brand image or personality documented by Dew et al. (2018) could be of valuable assistance in the process of logo creation mindful of audience response.

Brand ambassadors and marketers looking for the best brand storytelling practices could

also benefit from the insights from Chandrasekaran et al. (2016) and Chilton et al. (2018) to craft the most engaging stories in anticipation of consumer emotional responses.

4.6 Financial decisions of marketing surveys

Instead of circulating expensive yet highly attritive consumer census, a large-scale consumer profiling system based on computer vision (and natural language processing) methods, similar to Gebru, Krause, Wang, Chen, Deng and Fei-Fei (2017) could be favored due to its time and cost efficiency, and the advantage of potentially eliminating selection bias, and inaccurate reporting.

Computer vision methods could also be readily integrated into existing dynamic pricing systems that assume multi-armed bandit or dynamic auction frameworks, as was mentioned in Sutton, Barto et al. (1998).

5. Future Research

In this section, we first review a few trending topics in computer vision, and suggest potential marketing applications. In the following subsection, we cover visual marketing or vision science research that could potentially benefit from computer vision methods.

5.1 Computer Vision Research for Marketing Applications

5.1.1 Language and Vision

There appears a growing body of research at the intersection of natural language processing and computer vision that provides ample opportunities for marketing applications. For instance, Frome, Corrado, Shlens, Bengio, Dean, Mikolov et al. (2013) and Karpathy and Fei-Fei (2015) propose deep visual-semantic embedding learning methods for tasks such as generating textual descriptions of images and constituents. Such multi-modal learning frameworks could be especially valuable to marketing research because visual marketing and content marketing go hand in hand. Therefore, an integrated framework borrowing from the emerging body of research on language and vision could pay high dividends, by

complementing previous research efforts that have focused on separate effects of linguistic subtleties and visual stimuli.

5.1.2 Fine-grained Classification

In the past few decades, information technology and the internet markets have substantially increased the collective share of niche products, thereby creating a longer tail in the distribution of sales, due to the lower search cost and the increase in product selection on the Internet (Brynjolfsson, Hu and Simester 2011). Such a long tail phenomenon has led to challenges in estimating consumer preferences with regards to the vast number of product categories and sparse consumer traces within each category. The emerging subfield of *fine-grained classification* in the computer vision community aims to overcome the same challenge albeit in a broader sense. Previous research has addressed fine-grained classification problems such as named entities (Fleischman and Hovy 2002), cars (Yang, Luo, Change Loy and Tang 2015, Gebru et al. 2017), aircrafts (Maji, Rahtu, Kannala, Blaschko and Vedaldi 2013), cooking activities (Rohrbach, Amin, Andriluka and Schiele 2012), dog breeds (Liu, Kanazawa, Jacobs and Belhumeur 2012a), and so forth. For instance, Gebru et al. (2017) showcase how fine-grained classification could be especially cost-effective and time-saving for large-scale socio-economic demographic estimation and therefore facilitate public economic policy implementations. In the same vein, such fine-grained classification methods could be valuable to identify subtle differences in shopping environments — brand images, product designs, logo designs, and so on — automatically at a large scale, which could enable marketers to segment consumers into even finer categories and devise better targeting strategies.

5.1.3 Style Transfer

The body of literature on style transfer grew rapidly since the seminal paper of Gatys et al. (2015), or, Gatys et al. (2016), first emerged in 2015.

Gatys et al. (2015) (or Gatys et al. (2016)) introduce an artificial system based on a Deep Neural Network that creates artistic images of high perceptual quality. The system uses neural representations to separate and recombine *content* and *style* of arbitrary images, providing a neural algorithm for the creation of artistic images. Moreover, the study offers

a path forward to an algorithmic understanding of how humans create and perceive artistic imagery.

Such scientific advancements appear especially promising for marketing applications concerning creative content generation. For instance, algorithmic marketing content generation systems could be adopted to balance content and style in visual designs to tailor for particular brand images, corporate logos, and product packaging.

The proliferation of extensions including Johnson, Alahi and Fei-Fei (2016) that solves the underlying optimization problem in real time, Selim, Elgharib and Doyle (2016) and Li and Wand (2016) that generalize the original work to a broader class of applications, and Luan, Paris, Shechtman and Bala (2017) that incorporates a deep-learning approach to photographic style transfer, further boosts the feasibility and practical relevance of potential marketing applications of style transfer methods.

5.1.4 3D Reconstruction and Virtual Reality

As marketing applications of virtual reality and augmented reality gain momentum for consumer engagement, integrating 3D reconstruction methods into marketing research and applications could potentially introduce new questions and propose new solutions. For instance, which aspects of virtual reality could lead to higher consumer engagement, result in customer satisfaction, and are more effective in persuading consumers into purchasing? Which dimensions of augmented reality applications might help or defeat which marketing objectives and market outcomes? Existing research in 3D reconstruction that might provide fruitful directions include tourist spots and landmarks reconstruction (Agarwal, Snavely, Simon, Seitz and Szeliski 2009, Frahm, Fite-Georgel, Gallup, Johnson, Raguram, Wu, Jen, Dunn, Clipp, Lazebnik et al. 2010, Snavely, Seitz and Szeliski 2006), reconstruction of social scenes (Snavely, Seitz and Szeliski 2008) and buildings such as museums, galleries (Xiao and Furukawa 2014), and commercial apartment (Zou, Colburn, Shan and Hoiem 2018).

5.1.5 Multi-task Learning, Multimodal Learning, and Multimedia

Marketing efforts typically comprise creative visuals, music and content generation and consumption at the same time. Therefore, multi-modal and multi-task learning methods could be particularly promising when applied to marketing applications and enable us to build

end-to-end systems for consumer behavior or preference profiling and analyses. For example, how could we jointly predict sales, consumer engagement, and consumer willingness to buy given numerical, textual, auditory, pictorial, and temporal data?

5.2 Marketing Research with Computer Vision Applications

5.2.1 Advertising

Advertising studies in computer vision appear to be more focused on automatic understanding and reasoning about existing advertisements at the moment, whereas advertising in marketing appears more concerned with the effect of various aspects of advertising on market outcomes, whether it be sales, consumer engagement, revenue, market share, and so on, with exceptions in both cases. Automatic advertisement understanding, for instance, in the form of Visual Question Answering (VQA) (Hussain et al. 2017), could be integrated with discrete choice models and datasets associated with consumer response and market outcomes, to create end-to-end advertisement generation systems that tailor advertisements are shown to consumer preferences and personalized design advertisements for each consumer, in order to maximize marketing return.

5.2.2 Movies

Similarly, movie studies in computer vision appear more concerned about understanding movie plots based on scripts and frames (Rohrbach et al. 2015, Tapaswi et al. 2016), whereas a plethora of work in marketing on motion pictures (Eliashberg, Elberse and Leenders 2006) appear more about market outcomes. If combined, a deeper understanding could be achieved about which aspects of movies drive which aspects of audience response in terms of box office, or consumer sentiment.

5.2.3 Fashion

Studies about fashion style and related recommendation systems abound in multiple research fields. There appears to be more machine learning and deep learning in fashion studies in computer vision (Chao et al. 2009, Vittayakorn et al. 2015), and more time-series analysis and discrete choice models in fashion studies in marketing (Yoganarasimhan 2017, Shi et al. 2018). If combined, a more holistic analyses of fashion phenomena might be possible. In

addition, future work might benefit from more exploration from the source of the fashion industry — designers. For instance, generative models that inherit the idea of perfectly balancing style and substance from the body of literature on artistic style transfer (Gatys et al. 2015) reviewed in Section 5.1.3 could be applied to the realms of fashion design, logo design, architecture design, interior design, and so on. Furthermore, when combined with market outcome data collected from end consumers, such automatic design generation systems could be promising and valuable for marketing practitioners and consumers alike. Tkachenko et al. (2018) appears to be one notable step towards this direction.

5.2.4 Faces, emotions, postures, and more

Face, emotion, and posture detection studies abound in computer vision, whereas studies revolving these topics in marketing appear restricted to small sample size, controlled laboratory experiments using college students or AMT workers as subjects. Due to the almost mutually exclusive research objectives — detection accuracy for computer vision studies and market outcome for marketing studies — there appears to be little overlap. An end-to-end automatic system starting from large-scale unstructured data to the ultimate measures of market outcomes, albeit challenging, could be designed by integrating studies on both fronts. Studies in computer vision could serve to support the the front end and those in marketing the intermediate pipelines and connections to the ultimate market outcome variables for research questions such as:

- "which aspects of reality TV shows invoke the greatest consumer engagement?"
- "what kinds of election campaign video clips gather the greatest support from the audience?"
- "how to automatically generate the most persuasive business pitch that gets the largest amount of investment given the product idea?"
- "could we automatically generate enticing advertisements that lead to the greatest purchase intention given a particular consumer profile?"

For clarity, we tabulate the sub-categories in normative guides and future directions in Table 3 below.

Section	Sub-category	Relevant Research
Normative Guide to Marketing Managers	Brand Management Social Media Marketing	Shi et al. (2018), Liu and Mayzlin (2018)
	Product Design	Zhang et al. (2018) Tkachenko et al. (2018) Dzyabura et al. (2018)
	Advertising	Xiao and Ding (2014) Chilton (2018) Mei et al. (2012) Zhao et al. (2011), Sanchez et al. (2002)
	Consumer Profiling	Murillo et al. (2012), Lu et al. (2016)
	Brand Positioning	Liu and Mayzlin (2018) Papatla (2018), Dew et al. (2018)
	Financial Decisions	Gebru et al. (2017)
Future Directions	Language and Vision	Frome et al. (2013), Karpathy and Fei-fei (2015)
	Fine-grained Classification	Yang et al. (2015), Gebru et al. (2017)
	Style Transfer	Gatys et al. (2015, 2016)
	3D Reconstruction, VR	Agarwal et al. (2009), Seitz and Szeliski (2006)
	Multimodal Learning	Covington et al. (2016)
	Advertising VQA	Hussain et al. (2017)
	Faces, emotions, postures	Murillo et al. (2012)
	Movies	Rohrbach et al. (2015), Tapaswi et al. (2016)
	Fashion	Chao et al. (2009), Yoganarasimhan (2017)

Table 3: Overview of Normative Guides and Future Directions

6. Conclusions

"A picture is worth a thousand words." Memorable visual presentations could be surprisingly effective in communication, persuasion, and customer engagement, and thus have piqued the interest of marketing professional and scholars alike. Computer vision research for business applications has come a long way in the last two decades, but even more awaits to be explored. In the current review, we classify existing research by four primary criteria and provide comparative evaluations of each category and subcategory. We close by predicting future directions and identifying disparities and opportunities that define the interdisciplinary research front.

References

Agarwal, Sameer, Noah Snavely, Ian Simon, Steven M Seitz, and Richard Szeliski, "Building rome in a day," in "Computer Vision, 2009 IEEE 12th International Conference on" IEEE 2009, pp. 72–79.

Ak, Kenan E, Ashraf A Kassim, Joo Hwee Lim, and Jo Yew Tham, "Fashion-SearchNet: Fashion Search with Attribute Manipulation," in "European Conference on Computer Vision" Springer 2018, pp. 45–53.

Alexandrov, Aliosha and Birgit Leisen Pollack, "Exploring and Conceptualizing Brand Interestingness," *Annals of the Society for Marketing Advances Volume 2*, 2013, p. 70.

Amabile, Teresa M., Beth Ann Hennessey, and Barbara S Grossman, "Social Influence on Creativity: The Effects of Contracted-for Reward," *Journal of Personality and Social Psychology*, 1986, *81* (4), 813–846.

Aurîer, P, "The influence of emotions on satisfaction with movie consumption," 1994.

Azimi, Javad, Ruofei Zhang, Yang Zhou, Vidhya Navalpakkam, Jianchang Mao, and Xiaoli Fern, "Visual appearance of display ads and its effect on click through rate," in "Proceedings of the 21st ACM international conference on Information and knowledge management" ACM 2012, pp. 495–504.

Bell, Sean and Kavita Bala, "Learning visual similarity for product design with convolutional neural networks," *ACM Transactions on Graphics (TOG)*, 2015, *34* (4), 98.

Bossard, Lukas, Matthias Dantone, Christian Leistner, Christian Wengert, Till Quack, and Luc Van Gool, "Apparel classification with style," in "Asian conference on computer vision" Springer 2012, pp. 321–335.

Brynjolfsson, Erik, Yu Hu, and Duncan Simester, "Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales," *Management Science*, 2011, *57* (8), 1373–1386.

Bylinskii, Zoya, Nam Wook Kim, Peter O Donovan, Sami Alsheikh, Spandan Madan, Hanspeter Pfister, Fredo Durand, Bryan Russell, and Aaron Hertzmann, "Learning Visual Importance for Graphic Designs and Data Visualizations,"

in "Proceedings of the 30th Annual ACM Symposium on User Interface Software & Technology" 2017.

Chan, Tian Heong, Jürgen Mihm, and Manuel E Sosa, "On Styles in Product Design: An Analysis of US Design Patents," *Management Science*, 2017, *64* (3), 1230–1249.

Chandrasekaran, Arjun, Ashwin K Vijayakumar, Stanislaw Antol, Mohit Bansal, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh, "We are humor beings: Understanding and predicting visual humor," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition" 2016, pp. 4603–4612.

Chao, Xiaofei, Mark J Huiskes, Tommaso Gritti, and Calina Ciuhu, "A framework for robust feature selection for real-time fashion style recommendation," in "Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics" ACM 2009, pp. 35–42.

Chen, Hong, Zi Jian Xu, Zi Qiang Liu, and Song Chun Zhu, "Composite Templates for Cloth Modeling and Sketching," in "Computer Vision and Pattern Recognition (CVPR), 2006 IEEE Conference on" IEEE 2006.

Chen, Shuo, Josh L Moore, Douglas Turnbull, and Thorsten Joachims, "Playlist prediction via metric embedding," in "Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining" ACM 2012, pp. 714–722.

Cheng, Haibin, Roelof van Zwol, Javad Azimi, Eren Manavoglu, Ruofei Zhang, Yang Zhou, and Vidhya Navalpakkam, "Multimedia features for click prediction of new ads in display advertising," in "Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining" ACM 2012, pp. 777–785.

Chilton, Lydia B, "Constructing Visual Metaphors for Creative Ads," 2018.

___ , James A Landay, and Daniel S Weld, "HumorTools: A Microtask Workflow for Writing News Satire," 2018.

Chintagunta, Pradeep, Dominique M Hanssens, and John R Hauser, "Marketing science and big data," 2016.

Dew, Ryan, Asim Ansari, and Olivier Toubia, "Letting Logos Speak: A Machine Learning Approach for Data-Driven Logo Design," 2018.

Dhar, Sagnik, Vicente Ordonez, and Tamara L Berg, "High level describable attributes for predicting aesthetics and interestingness," in "CVPR 2011" IEEE 2011, pp. 1657–1664.

Doersch, Carl, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros, "What Makes Paris Look like Paris?," *ACM Transactions on Graphics (SIGGRAPH)*, 2012, *31* (4), 101:1–101:9.

Dubey, Rachit, Joshua Peterson, Aditya Khosla, Ming-Hsuan Yang, and Bernard Ghanem, "What Makes an Object Memorable?," in "The IEEE International Conference on Computer Vision (ICCV)" December 2015.

Dzyabura, Daria, Marat Ibragimov, and Siham El Kihal, 2018.

Eisenman, Micki, "Understanding aesthetic innovation in the context of technological evolution," *Academy of Management Review*, 2013, *38* (3), 332–351.

Eliashberg, Jehoshua, Anita Elberse, and Mark AAM Leenders, "The motion picture industry: Critical issues in practice, current research, and new research directions," *Marketing science*, 2006, *25* (6), 638–661.

Fleischman, Michael and Eduard Hovy, "Fine grained classification of named entities," in "Proceedings of the 19th international conference on Computational linguistics-Volume 1" Association for Computational Linguistics 2002, pp. 1–7.

Frahm, Jan-Michael, Pierre Fite-Georgel, David Gallup, Tim Johnson, Rahul Raguram, Changchang Wu, Yi-Hung Jen, Enrique Dunn, Brian Clipp, Svetlana Lazebnik et al., "Building rome on a cloudless day," in "European Conference on Computer Vision" Springer 2010, pp. 368–381.

Frome, Andrea, Greg S Corrado, Jon Shlens, Samy Bengio, Jeff Dean, Tomas Mikolov et al., "Devise: A deep visual-semantic embedding model," in "Advances in neural information processing systems" 2013, pp. 2121–2129.

Fu, Yanwei, Timothy M Hospedales, Tao Xiang, Shaogang Gong, and Yuan Yao, "Interestingness prediction by robust learning to rank," in "European conference on computer vision" Springer 2014, pp. 488–503.

Garces, Elena, Aseem Agarwala, Diego Gutierrez, and Aaron Hertzmann, "A similarity measure for illustration style," *ACM Trans. Graph.*, 2014, *33*, 93:1–93:9.

Gatys, Leon A, Alexander S Ecker, and Matthias Bethge, "A neural algorithm of artistic style," *arXiv preprint arXiv:1508.06576*, 2015.

___ , ___ , and ___ , "Image style transfer using convolutional neural networks," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition" 2016, pp. 2414–2423.

Gebru, Timnit, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, and Li Fei-Fei, "Fine-Grained Car Detection for Visual Census Estimation.," in "AAAI," Vol. 2 2017, p. 6.

Gevers, Th and AWM Smeulders, "Image search engines: An overview," *Emerging Topics in Computer Vision*, 2004, pp. 1–54.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville, *Deep learning*, MIT press, 2016.

Gutierrez, Patricia, Pierre-Antoine Sondag, Petar Butkovic, Mauro Lacy, Jordi Berges, Felipe Bertrand, and Arne Knudson, "Deep Learning for Automated Tagging of Fashion Images," in "European Conference on Computer Vision" Springer 2018, pp. 3–11.

Gygli, Michael, Helmut Grabner, Hayko Riemenschneider, Fabian Nater, and Luc Van Gool, "The interestingness of images," in "Proceedings of the IEEE International Conference on Computer Vision" 2013, pp. 1633–1640.

Hsiao, Wei-Lin and Kristen Grauman, "Learning the latent look: Unsupervised discovery of a style-coherent embedding from fashion images," in "2017 IEEE International Conference on Computer Vision (ICCV)" IEEE 2017, pp. 4213–4222.

Huang, Xinyue and Adriana Kovashka, "Inferring visual persuasion via body language, setting, and deep features," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops" 2016, pp. 73–79.

Hussain, Zaeem, Mingda Zhang, Xiaozhong Zhang, Keren Ye, Christopher Thomas, Zuha Agha, Nathan Ong, and Adriana Kovashka, "Automatic understanding of image and video advertisements," in "2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)" IEEE 2017, pp. 1100–1110.

Huynh, Cong Phuoc, Arridhana Ciptadi, Ambrish Tyagi, and Amit Agrawal, "CRAFT: Complementary Recommendation by Adversarial Feature Transform," in "European Conference on Computer Vision" Springer 2018, pp. 54–66.

Isola, Phillip, Jianxiong Xiao, Devi Parikh, Antonio Torralba, and Aude Oliva, "What makes a photograph memorable?," *IEEE transactions on pattern analysis and machine intelligence*, 2014, *36* (7), 1469–1482.

Johnson, Justin, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," *arXiv preprint arXiv:1603.08155*, 2016.

Joo, Jungseock, Francis F Steen, and Song-Chun Zhu, "Automated facial trait judgment and election outcome prediction: Social dimensions of face," in "Proceedings of the IEEE international conference on computer vision" 2015, pp. 3712–3720.

____ , Weixin Li, Francis F Steen, and Song-Chun Zhu, "Visual persuasion: Inferring communicative intents of images," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2014, pp. 216–223.

Karayev, Sergey, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller, "Recognizing image style," *arXiv preprint arXiv:1311.3715*, 2013.

Karpathy, Andrej and Li Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2015, pp. 3128–3137.

Khosla, A, J Xiao, and A Torralba, "Memorability of image regions," *Advances in Neural* , 2012.

___ , WA Bainbridge, and A Torralba, "Modifying the memorability of face photographs," *Proceedings of the IEEE* , 2013.

Kiapour, M Hadi, Kota Yamaguchi, Alexander C Berg, and Tamara L Berg, "Hipster wars: Discovering elements of fashion styles," in "European conference on computer vision" Springer 2014, pp. 472–488.

Krishna, Aradhna, Ryan S Elder, and Cindy Caldara, "Feminine to smell but masculine to touch? Multisensory congruence and its effect on the aesthetic experience," *Journal of Consumer Psychology*, 2010, *20* (4), 410–418.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds., *Advances in Neural Information Processing Systems 25*, Curran Associates, Inc., 2012, pp. 1097–1105.

Lee, Yong Jae, Alexei A. Efros, and Martial Hebert, "Style-Aware Mid-level Representation for Discovering Visual Connections in Space and Time," in "The IEEE International Conference on Computer Vision (ICCV)" December 2013.

Li, Chuan and Michael Wand, "Combining markov random fields and convolutional neural networks for image synthesis," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition" 2016, pp. 2479–2486.

Li, Zechao and Jinhui Tang, "Weakly supervised deep metric learning for community-contributed image retrieval," *IEEE Transactions on Multimedia*, 2015, *17* (11), 1989–1999.

Liu, Gaowen, Yan Yan, Elisa Ricci, Yi Yang, Yahong Han, Stefan Winkler, and Nicu Sebe, "Inferring Painting Style with Multi-Task Dictionary Learning.," in "IJCAI" 2015, pp. 2162–2168.

Liu, Jiongxin, Angjoo Kanazawa, David Jacobs, and Peter Belhumeur, "Dog breed classification using part localization," in "European Conference on Computer Vision" Springer 2012, pp. 172–185.

Liu, Liu and Dina Mayzlin, "Visual Listening in: Extracting Brand Image Portrayed on Social Media," 2018.

Liu, Si, Zheng Song, Guangcan Liu, Changsheng Xu, Hanqing Lu, and Shuicheng Yan, "Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set," in "Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on" IEEE 2012, pp. 3330–3337.

Lowrey, Tina M, "The relation between script complexity and commercial memorability," *Journal of Advertising*, 2006, *35* (3), 7–15.

Lu, Shasha, Li Xiao, and Min Ding, "A video-based automated recommender (VAR) system for garments," *Marketing Science*, 2016, *35* (3), 484–510.

Luan, Fujun, Sylvain Paris, Eli Shechtman, and Kavita Bala, "Deep Photo Style Transfer," *arXiv preprint arXiv:1703.07511*, 2017.

Maji, Subhransu, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi, "Fine-grained visual classification of aircraft," *arXiv preprint arXiv:1306.5151*, 2013.

Malik, Nikhil, Param Vir Singh, Dokyun Lee, and Kannan Srinivasan, "When Does Beauty Pay. A Large Scale Image Based Appearance Analysis on Career Transitions," 2018.

Mano, Haim and Richard L Oliver, "Assessing the dimensionality and structure of the consumption experience: evaluation, feeling, and satisfaction," *Journal of Consumer research*, 1993, *20* (3), 451–466.

McAuley, Julian, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel, "Image-based recommendations on styles and substitutes," in "Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval" ACM 2015, pp. 43–52.

McDuff, Daniel, Rana El Kaliouby, Jeffrey F Cohn, and Rosalind W Picard, "Predicting ad liking and purchase intent: Large-scale analysis of facial responses to ads," *IEEE Transactions on Affective Computing*, 2015, *6* (3), 223–235.

Mei, Tao, Lusong Li, Xian-Sheng Hua, and Shipeng Li, "ImageSense: Towards contextual image advertising," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2012, *8* (1), 6.

Mohammad, Saif M and Peter D Turney, "Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon," 2010, pp. 26–34.

Moore, Joshua L, Shuo Chen, Douglas Turnbull, and Thorsten Joachims, "Taste Over Time: The Temporal Dynamics of User Preferences.," in "ISMIR" 2013, pp. 401–406.

—— , —— , Thorsten Joachims, and Douglas Turnbull, "Learning to Embed Songs and Tags for Playlist Prediction.," in "ISMIR," Vol. 12 2012, pp. 349–354.

Murillo, Ana C, Iljung S Kwak, Lubomir Bourdev, David Kriegman, and Serge Belongie, "Urban tribes: Analyzing group photos from a social perspective," in "Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on" IEEE 2012, pp. 28–35.

Novak, Thomas P, Donna L Hoffman, and Yiu-Fai Yung, "Measuring the customer experience in online environments: A structural modeling approach," *Marketing science*, 2000, *19* (1), 22–42.

Palmer, Stephen E, *Vision science: Photons to phenomenology*, MIT press, 1999.

Papatla, Purushottam, "Face, Body or Both? Effects of Partial and Full Visibility of People in VUGC on Consumer Response," 2018.

Quercia, Daniele, Neil Keith O'Hare, and Henriette Cramer, "Aesthetic capital: what makes london look beautiful, quiet, and happy?," in "CSCW" 2014.

Reber, Rolf, Norbert Schwarz, and Piotr Winkielman, "Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?," *Personality and social psychology review*, 2004, *8* (4), 364–382.

Repp, Bruno H, "The aesthetic quality of a quantitatively average music performance: Two preliminary experiments," *Music Perception: An Interdisciplinary Journal*, 1997, *14* (4), 419–444.

Rohrbach, Anna, Marcus Rohrbach, Niket Tandon, and Bernt Schiele, "A dataset for movie description," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2015, pp. 3202–3212.

Rohrbach, Marcus, Sikandar Amin, Mykhaylo Andriluka, and Bernt Schiele, "A database for fine grained activity detection of cooking activities," in "Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on" IEEE 2012, pp. 1194–1201.

Sánchez, Juan M, Xavier Binefa, and Jordi Vitrià, "Shot partitioning based recognition of tv commercials," *Multimedia Tools and Applications*, 2002, *18* (3), 233–247.

Sbai, Othman, Mohamed Elhoseiny, Antoine Bordes, Yann LeCun, and Camille Couprie, "Design: Design inspiration from generative networks," in "European Conference on Computer Vision" Springer 2018, pp. 37–44.

Selim, Ahmed, Mohamed Elgharib, and Linda Doyle, "Painting style transfer for head portraits using convolutional neural networks," *ACM Transactions on Graphics (ToG)*, 2016, *35* (4), 129.

Shi, Zijun (June), Dokyun Lee, Param Vir Singh, and Kannan Srinivasan, "Design of Fashion: Can Brand Value be Separated from Style Value?," 2018.

Snavely, Noah, Steven M Seitz, and Richard Szeliski, "Photo tourism: exploring photo collections in 3D," in "ACM transactions on graphics (TOG)," Vol. 25 ACM 2006, pp. 835–846.

___ , ___ , and ___ , "Modeling the world from internet photo collections," *International journal of computer vision*, 2008, *80* (2), 189–210.

Sutton, Richard S, Andrew G Barto et al., *Introduction to reinforcement learning*, Vol. 135, MIT press Cambridge, 1998.

Tapaswi, Makarand, Yukun Zhu, Rainer Stiefelhagen, Antonio Torralba, Raquel Urtasun, and Sanja Fidler, "Movieqa: Understanding stories in movies through question-answering," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2016, pp. 4631–4640.

Thomas, Christopher and Adriana Kovashka, "Seeing behind the camera: Identifying the authorship of a photograph," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition" 2016, pp. 3494–3502.

Tkachenko, Yegor, Asim Ansari, and Olivier Toubia, "Computer-aided Exploration Of Product Designs in High-dimensional Visual Spaces," 2018.

Todorov, Alexander, "Modeling Visual Impressions of Faces," 2018.

Toubia, Olivier and Oded Netzer, "Idea Generation, Creativity, and Prototypicality," *Marketing Science*, 2016.

Veit, Andreas, Balazs Kovacs, Sean Bell, Julian McAuley, Kavita Bala, and Serge Belongie, "Learning visual clothing style with heterogeneous dyadic co-occurrences," in "Proceedings of the IEEE International Conference on Computer Vision" 2015, pp. 4642–4650.

Vittayakorn, Sirion, Kota Yamaguchi, Alexander C Berg, and Tamara L Berg, "Runway to realway: Visual analysis of fashion," in "Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on" IEEE 2015, pp. 951–958.

Wedel, Michel and Rik Pieters, *Visual marketing: From attention to action*, Psychology Press, 2012.

Wilber, Michael J, Chen Fang, Hailin Jin, Aaron Hertzmann, John Collomosse, and Serge J Belongie, "BAM! The Behance Artistic Media Dataset for Recognition Beyond Photography.," in "ICCV" 2017, pp. 1211–1220.

Xiao, Jianxiong and Yasutaka Furukawa, "Reconstructing the worlds museums," *International journal of computer vision*, 2014, *110* (3), 243–258.

Xiao, Li and Min Ding, "Just the Faces: Exploring the Effects of Facial Features in Print Advertising," *Marketing Science*, 2014, *33*, 338–352.

Yadati, Karthik, Harish Katti, and Mohan Kankanhalli, "CAVVA: Computational affective video-in-video advertising," *IEEE Transactions on Multimedia*, 2014, *16* (1), 15–23.

Yamaguchi, Kota, M Hadi Kiapour, and Tamara L Berg, "Paper doll parsing: Retrieving similar styles to parse clothing items," in "Proceedings of the IEEE international conference on computer vision" 2013, pp. 3519–3526.

___ , **M Hadi Kiapour, Luis E Ortiz, and Tamara L Berg**, "Parsing clothing in fashion photographs," in "Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on" IEEE 2012, pp. 3570–3577.

Yang, Linjie, Ping Luo, Chen Change Loy, and Xiaoou Tang, "A large-scale car dataset for fine-grained categorization and verification," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition" 2015, pp. 3973–3981.

Yang, Yi and Deva Ramanan, "Articulated pose estimation with flexible mixtures-of-parts," in "Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on" IEEE 2011, pp. 1385–1392.

Yoganarasimhan, Hema, "Identifying the presence and cause of fashion cycles in data," *Journal of Marketing Research*, 2017, *54* (1), 5–26.

Zhang, Shunyuan, Dokyun Lee, Param Vir Singh, and Kannan Srinivasan, "How Much is an Image Worth? The Impact of Professional versus Amateur Airbnb Property Images on Property Demand," 2018.

Zhao, Gangqiang, Junsong Yuan, Jiang Xu, and Ying Wu, "Discovering the thematic object in commercial videos," *IEEE MultiMedia*, 2011, *18* (3), 56–65.

Zhou, YingHui, Shasha Lu, and Min Ding, "A Face Anonymity-Perceptibility Paradigm and an Application in the Online Dating Industry," 2016.

Zou, Chuhang, Alex Colburn, Qi Shan, and Derek Hoiem, "LayoutNet: Reconstructing the 3D Room Layout from a Single RGB Image," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition" 2018, pp. 2051–2059.