A

Colors –Messengers of Concepts: Visual Design Mining for Learning Color Semantics

Ali Jahanian, MIT
Shaiyan Keshvari, MIT
S.V.N. Vishwanathan, Purdue University
Jan P. Allebach, Purdue University

We study the concept of color semantics by modeling a dataset of magazine cover designs, evaluating the model via crowdsourcing, and demonstrating several prototypes that facilitate color-related design tasks. We investigate a probabilistic generative modeling framework that expresses semantic concepts as a combination of color and word distributions –color-word topics. We adopt an extension to Latent Dirichlet Allocation (LDA) topic modeling, called LDA-dual, to infer a set of color-word topics over a corpus of 2,654 magazine covers spanning 71 distinct titles and 12 genres. While LDA models text documents as distributions over word topics, we model magazine covers as distributions over color-word topics. The results of our crowdsourcing experiments confirm that the model is able to successfully discover the associations between colors and linguistic concepts. Finally, we demonstrate several prototype applications that use the learned model to enable more meaningful interactions in color palette recommendation, design example retrieval, pattern recoloring, image retrieval, and image color selection.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems –human factors; H.5.2 [**Information Interfaces and Presentation**]: UI

General Terms: Human Factors, Theory

Additional Key Words and Phrases: Color semantics, topic modeling, generative models, visual design mining, visual design language, interaction design, aesthetics, color palette recommendation, design example retrieval, image retrieval, image color selection, pattern recoloring.

ACM Reference Format:
Ali Jahanian, Shaiyan Keshvari, S.V.N. Vishwanathan, and Jan P. Allebach, 2016. Colors –Messengers of Concepts: Visual Design Mining for Learning Color Semantics *ACM Trans. Comput.-Hum. Interact.* V, N, Article A (January YYYY), 37 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Color conveys meaning. Beyond basic visual perception of color itself, humans classify colors at higher levels of abstraction into verbal and nonverbal semantic categories [Humphreys and Bruce 1989; Barsalou 1999; Derefeldt et al. 2004]. In practice, designers carefully choose color combinations not only to be appealing, but also to communicate specific concepts, moods, and styles [Eisemann 2000; Frascara 2004; Newark 2007; Samara 2007].

Previous work has attempted to understand how colors map onto color names and onto semantic concepts. For instance, a particular range of hue is called “blue”, and may semantically relate to

Author’s addresses: A. Jahanian (current address) Computer Science and Artificial Intelligence Laboratory (CSAIL), MIT, ali-design@csail.mit.edu. This work was partially performed while A. Jahanian was with Purdue University; S. Keshvari (current address) Department of Brain & Cognitive Sciences, MIT, shaiyan@mit.edu; S.V.N. Vishwanathan, (current address) Jack Baskin School of Engineering, Computer Science Department, University of California Santa Cruz, vishy@ucsc.edu. This work was partially performed while S.V.N. Vishwanathan was with Purdue University; J. P. Allebach, School of Electrical and Computer Engineering, Purdue University, allebach@purdue.edu.

DOI: http://dx.doi.org/10.1145/0000000.0000000

Fig. 1: Application of color semantics in color palette selection, and design example retrieval. See Sec. 8.

concepts of “coolness” or “potency” [Berlin 1969; Osgood 1971; Ou et al. 2004a; 2004b; 2004c]. Kobayashi’s *Color Image Scale* is a notable attempt to understand the implications of color semantics in design [Kobayashi 1981; 1991]. Kobayashi used crowdsourcing experiments to collect ratings of colors and 3-color palettes along 180 meaningful qualities, e.g. “modern” vs “conservative” or “stylish” vs “rustic”. Using the chromaticity and values of colors as well as the “warm” vs “cool” and “soft” vs “hard” ratings, he organized the color palettes on a $2D$ space. He then used the remaining ratings and factor analysis to define groups corresponding to fashion, product design, and textile in this space. This color scale, however, suffers several fundamental limitations. Importantly, there is no rigorous mapping function for adding new concepts. Furthermore, the discriminative nature of the space precludes combinations of non-adjacent concepts, for instance, “both casual and modern”.

More recent work addresses some of these shortcomings by using data mining and discriminative models to automatically classify colors and color palettes into categories practical for product design [Csurka et al. 2010; Murray et al. 2012]. While certainly a major advance, this approach has two major shortcomings with respect to design. First, it is highly dependent on context-free, human-labelled color palettes. People may associate different labels to the same color or color palette depending on the context, or even use an arbitrary name such as “my-theme” [O’Donovan et al. 2011]. This is particulary relevant for design. For instance, magazine covers must compete with other magazines on a newsstand, so designers spend many days conceptualizing and creating covers that attract customers at a glance [Foges 1999]. This requires the designers to carefully choose a color palette for the cover based on the magazine’s general topic and the specific stories in the issue. Second, the discriminative approach alone does not allow for generation of novel palettes or palette combinations for design applications; research has shown that suggesting designs or elements of design can help users be more creative and productive [Herring et al. 2009].

We bring together probabilistic models and a novel dataset to address these challenges. Specifically, we adapt LDA-dual, an extension of Latent Dirichlet Allocation (LDA) topic modeling [Shu

et al. 2009], as a way to *discover* meaningful color-word combinations, which we call *color-word topics*. We simultaneously infer novel color-word topics from the distributions of colors and words occurring within our corpus of 2,654 magazine covers, which spans 71 distinct titles and 12 genres. Furthermore, our framework harnesses the LDA model's generated color topics to interactively create original color combinations, and select perceptually similar 5-color palettes from those color combinations for design. This link between color, language, and semantic concepts opens the door to many possible applications in design. The user could, for example, choose color palettes based on topic words, and use those color palettes to retrieve design examples (Fig. 1).

To verify whether or not users agree with the associations between color combinations and linguistic concepts produced by the model, we conducted a crowdsourcing experiment. We used the model to generate pairs of word clouds and discretized color palettes. Users viewed the color palettes and chose the most appropriate corresponding word clouds from 4 alternatives (one of which came from the model). To complement this evaluation, we conducted a second experiment with the same setup, but instead showing a word cloud and asking users to match it with color palettes. Based on the user feedback, we inferred the strength of the association between each color palette and word cloud in the experiment. This allowed us to test whether the model produced intuitive pairs of colors and word clouds. This crowdsourcing strategy is a superior way to evaluate the model when compared to held-out likelihood methods (see [Wallach et al. 2009]), which are suboptimal when applied to data from semantically meaningful topics [Chang et al. 2009].

Given a verified model of color-word topics that is both inferential and generative, how can we use it for design? Our rigorous model of color semantics enables many applications, including image retrieval [Solli and Lenz 2010], recommending design alternatives [Jahanian et al. 2013], editing graphics, and creating color palettes [Heer and Stone 2012]. There are several online communities for color palette design (e.g. [Adobe Kuler 2016; ColourLovers 2016]), each with thousands to millions of user-created, named, and rated palettes. Despite the expansive scope, however, it is quite difficult for users to navigate them to find useful examples. These online services commonly use sparse, user-labeled keywords to aid search; users are at the mercy of whether a previous user labelled a color palette with the concept desired. Furthermore, the open-ended labelling procedure leads to little agreement between labels, which makes search more noisy. Color semantics, on the other hand, can provide a meaningful and tractable way to find palettes. We show how we apply our model's discovered color-word topics to recommend palettes based on both perceptual similarity and semantic concepts. The user can then automatically discover the palettes that match their application. Importantly, we can retrieve design examples by mapping from recommended palettes to a pool of magazine covers (Fig. 1). These applications are particularly relevant in light of the emergence of design-by-example, a concept in HCI that enables more creative design by users through display of related examples [Herring et al. 2009].

The overarching contribution of our work is to provide a novel solution to the "gap" of automatically connecting media to semantic information. This gap has been a major point of discussion for over a decade [Smeulders et al. 2000; Sethi et al. 2001; Mojsilovic and Rogowitz 2001; Liu et al. 2007] and is considered to be "a major challenge to solve in the multimedia community" ([Lindner and Süsstrunk 2015]). Our framework tackles the gap directly, taking media in the form of magazine cover designs, and extracting semantic information in the form of color-word histograms. Furthermore, the flexibility and richness of the model provides a way to traverse the gap in the other direction, going from semantic information to media. Our work makes the following specific contributions: First, rather than use the typical approach of extracting hand-crafted features or utilizing supervised learning, we train an unsupervised topic model to discover the inherent relationships between sets of multiple words and colors found in designs. This approach better reflects the underlying rich associations between colors and words. Second, we provide an intuitive and useful technique to address the challenge of visualizing compound topics discovered by topic models. Third, we use crowdsourcing to validate our modeling, unlike typical approaches that instead use crowdsourcing to drive the modeling. Furthermore, our crowdsourcing study covers a large wide range of demographics, allowing us to test known variations in color semantics between cultures.

Finally, we demonstrate how to integrate our approach with typical design applications to support intuitive interactions. Because our model implements the notion that color is understood different levels of abstraction, we support the user to select a set of arbitrary words, anything from "red" to "science" to "dancer", that describes the purpose or context of their media. Our model's associations then link the words to relevant design examples, images, or color palettes.

The flow of this paper is as follows. In Sec. 2, we discuss prior work on both theoretical and practical aspects of color semantics. In Sec. 3, we introduce the dataset we collected. We then discuss the inference and generative mechanisms in the LDA-dual modeling framework in Sec. 4. In Sec. 5, we illustrate how to visualize the discovered semantic topics. We then explain our design of the crowdsourcing experiment in Sec. 6, and analyze the crowd responses in Sec. 7. In Sec. 8, we demonstrate a number of applications for color semantics, specifically color palette selection, design example recommendation, pattern recoloring, image retrieval, and color region selection in images. We conclude by discussing remaining limitations of our approach, and suggest a number of avenues for future work in Sec. 9.

2. PRIOR WORK

2.1. Color Cognition

There is more to our experience with color than low level perception; humans classify colors into multiple progressively higher levels of abstraction. The study of the verbal and semantic categories associated with colors is called *color cognition* [Humphreys and Bruce 1989; Barsalou 1999; Derefeldt et al. 2004]. These verbal and semantic categories enable us to communicate about colors. For instance, not only can we identify a color as "red", but we can further describe it as "warm", or even more abstractly, as "romantic". The extent of the linkage between color and meaning, and its cross-cultural variation, has spurred an entire field of research in color naming, emotional meanings of colors, and visual communication design.

Color naming refers to associating colors with names like, "blue" or "red". The early work of Berlin and Kay [1969] introduced the study of the consistency of color naming between cultures. They studied many different languages, and concluded that there exists a set of universal 11 basic color categories, and that any given language always draws its basic color terms from these categories. Later studies reformulated each of the basic terms as continuous functions of a fuzzy set to account for evolving terms [Kay and McDaniel 1978]. Other studies, however, have challenged these universal terms, for example finding two terms for "blue" in Russian language [Winawer et al. 2007]. Cultural semiotics also appear to influence the basic terms [Paramei 2005]. Others have shown how proposing a list of predefined basic terms in an experiment can influence color category judgements [Roberson et al. 2000]. This approach mathematically and computationally limits models of color categorization [Chuang et al. 2008].

Complementary to color naming, research on color semantics aims to discover the "meaning" of colors. The first systematic approach to quantifying meanings of linguistic concepts came from *measurement of meaning* [Osgood 1952]. Osgood [1952] proposed an affective space based on 12 pairs of bipolar terms (such as *happy-sad* or *kind-cruel*). In a later study, Adam and Osgood [1973] found that while there are differences across cultures between the affective meanings attributed to the colors, there are also consistencies. For instance, among all the cultures, *red* is *strong* and *active*. The ability of such bipolar scales to capture semantics continues to be an active line of research [Ou et al. 2004a; 2004b; Ou et al. 2012]. Among these bipolar scales, Kobayashi's Color Image Scale [1981; 1991] is relevant to the current study, since it contains multi-color combinations with associated linguistic concepts. The Color Image Scale is a semantic space of bipolar terms, augmented with terms from fashion and textile products such as "chic" and "dandy". This scale comprises of two dimensions, *warm-cool* and *soft-hard*, 180 adjectives (e.g. "festive", "romantic", etc.), and 15 high clusters (e.g. "modern", "natural", etc.). Using the chromaticity and values of colors as well as the "warm" vs "cool" and "soft" vs "hard" ratings, he organized the color palettes on a $2D$ space. He then used the remaining ratings and factor analysis to define groups corresponding to

fashion, product design, and textile in this space. He conducted several crowdsourcing experiments where participants rated the similarity between color palettes and descriptive adjectives in order to map color combinations onto this space. Later cross-cultural studies examined the universality of the Color Image Scale [Ou et al. 2004a; 2004b; Ou et al. 2012].

2.2. Data Mining Approaches

Our work focuses on mining the association between colors and linguistic concepts in the context of design. On the other hand, existing machine learning models of color and language have largely been restricted to the domain of color naming. As color semantics builds on color naming, however, it is important to examine existing data-driven models of color naming.

Prior work in modeling color naming attempts to fit statistical models to databases of color names. Specifically, it links a set of labels to the Berlin and Kay basic colors (see Heer and Stone [2012] for a review of these models). The main limitation is that these labels are combinations of basic color terms (e.g. "greenish-blue"), and do not necessarily map to real-world objects. Lin et al. [2013a], and more recently Setlur and Stone [2016], extended this work in the domain of data visualization. Both approaches aim to improve user interactions with color by decreasing *Stroop* Interference (see [MacLeod 1991] for a review), or the difficulty observers have when there is a mismatch between a color-word combination. For example, coloring the word "apple" with blue in a visualization can lead to confusion. Setlur and Stone's main contribution was to mine Google n-grams (see [NgramViewer 2016; Michel et al. 2011]) and discover more word context to incorporate word context with respect to objects and brands.

Researchers in computer vision have approached color naming from a more image-based perspective. Importantly, they have used large datasets of images and captions from internet search engines and topic modeling to ascertain the associations between words and basic colors. Weijer et al. [2009] use Probabilistic Latent Semantic Analysis (PLSA), and Schauerte and Stiefelhagen [2012] use Latent Dirichlet Allocation (LDA), to learn these associations. In the case of PLSA, the authors adapt and extend the model by defining prior Dirichlet distributions for color labels as well as a regularization term to control the shape of the model. Schauerte and Stiefelhagen use a supervised version of LDA ([Mcauliffe and Blei 2008; Wang et al. 2009]) to learn word-basic color associations. It is important to note the similarity and two key differences between this LDA model and the one we present: While both approaches simultaneously learn the co-occurrences of visual features and words, our model is unsupervised and uses a different graphical model (a mixture of color-word proportions) to describe the topics. Furthermore, one key aspect of both of these previous modelling approaches is that they were evaluated by using cross-validation to maximize the likelihood of the data given the model. This method of cross-validation poses problems when capturing semantically meaningful topics [Chang et al. 2009]. As we will show, our approach circumvents cross-validation by using crowdsourcing to validate the inferred topics.

Building towards richer color semantics, researchers have recently modeled more abstract linguistic concepts. Csurka et al. [2010] discuss color moods, while Solli and Lenz [2010] algebraically implement Kobayashi's Color Image Scale. Csurka and colleagues selected 15 linguistic concepts with associated color combinations from [Eisemann 2000] and an online community called ColourLovers [ColourLovers 2016] to create a vocabulary of labels. They trained a classifier to associate these linguistic concepts with colors. Furthermore, Murray et al. [2012] utilized this framework for transferring color moods to images. In image retrieval, Solli and Lenz define a mathematical framework for Kobayashi's Color Image Scale. Their goal was to index any given image based on the proportions of Kobayahsi's 3-color combinations that it contains. Given its effectiveness, we previously utilized this framework in a system for designing alternative and customized magazine covers (see [Jahanian et al. 2013]). A notable difference between Csurka and colleagues' approach and the current study is that our inferred clusters take into account the proportions of the colors and not simply their presence. Importantly, as mentioned earlier, the online color palettes used by Csurka and colleagues are potentially noisy and do not necessarily indicate context [O'Donovan et al. 2011].

Attempting to outperform the typical method of finding palettes by querying words on Adobe Kuler [Adobe Kuler 2016], Lindner and Süsstrunk [2013] suggested a method to automatically generate color palettes based on users' input words. First, they constructed a database of the 100,000 most frequent words in a subset of Google n-gram text [Google n-grams 2016]. Next, for each of these words, they found the top 60 images returned by Google image search. For each image, they extracted a set of 5-color palettes using four different "harmonious templates" (Adobe Kuler, Matsuda [1995]), and designed a tool that delivered the best palette from each template at a user's request. Finally, they tested their results using a small set (30) of color palettes and found that average users preferred their palettes better (but not statistically significant) than those retrieved from Adobe Kuler. Our work is different in three important ways. First, our approach discovers the color-word associations made by designers in a corpus of magazine covers. Using our corpus is beneficial because it gives us a large diversity of data, while avoiding the potential pitfalls of using the top Google search results. Namely, search results can be influenced by particulars of the search algorithm, e.g. time and location. Second, in contrast with the single word to multiple palettes mapping discovered by Lindner and Süsstrunk [2013], we find the relationships between sets of multiple words and sets of multiple colors. These many-to-many associations underlie the true semantic nature of media, but can be complex to analyze; modeling them effectively requires the use of flexible models. LDA-dual is particularly well suited for this task. Finally, since our model internally uses a richer representation of colors than a simple 5-color palette, we enable ranking color palettes both extracted from our dataset as well as those from any existing database of palettes, like Adobe Kuler or ColourLovers.

3. DATA COLLECTION

Our dataset of magazine covers includes 2,654 covers from 71 magazine titles and 12 genres, spanning 14 years, from 2000 to 2013 (and one cover from 1998). We collected approximately 1,500 of these covers by scanning them from magazines held by libraries and newsstands in our university. The rest of the cover images were downloaded from the Internet. Although we developed a web crawler tool to collect magazine covers, because many magazine publishers do not provide archives with high quality images, in half of the cases we had to collect online images by hand[1]. We attempted to collect roughly 12 different genres of magazines to capture different contexts of design. These genres include *Art*, *Business*, *Education*, *Entertainment*, *Family*, *Fashion*, *Health*, *Nature*, *Politics*, *Science*, *Sports*, and *Technology*. To this end, we obtained category labels from the Dewey Classification method [OCLC 2016a], the WorldCat indexing system [OCLC 2016b], suggestions from our librarians, as well as the description of the magazine by the publishers. We used overlapping methods to disambiguate categories like "general", which were sometimes assigned to titles by the Dewey method. Table III in the appendix contains a summary of our dataset. Note that the genres are fluid and could change depending on use.

3.1. Preprocessing of Images

The preprocessing of cover images was performed using the Matlab *Image Processing* toolbox[2]. For the scanned images, gamma correction was applied. We use 512 basic colors obtained by quantizing the sRGB color space with 8 bins in each channel. Given this color basis, each magazine cover (image) is then a histogram of these colors. We chose sRGB mostly for processing convenience, as it has a cubic space and is thus readily divided into bins. Conceivably, we could use the CIELab color space, which is considered more perceptually uniform but creates computational challenges. Importantly, whenever we compare colors (for finding palettes close to color histograms, etc.) we do convert to the CIELab color space and subjects see color palettes that are closest to their respective

[1]The preprocessed data is available at https://github.com/ali-design/ColorSemantics.git. Note however that copyright concerns prevent us from distributing the raw data freely in many locales. Please contact us if you would like to make use of the raw data in your own work.
[2]The MathWorks, Inc., Natick, MA.

color word topics in CIELab space rather than sRGB. This dual approach is common practice in color applications; for example, Soli and Lenz [2010] use sRGB for quantization prior to implementing applications in the CIELab color space. To feed the images to LDA-dual, we scale them to 300×200 pixels using bicubic interpolation. The down-sizing was done to reduce the computation without affecting the distribution of the colors in the images.

3.2. Creating the Word Vocabulary

To capture the words to be associated with color distributions of the magazine covers, the words on the covers were transcribed by hand. To create a word vocabulary, we first prune the transcribed words (as described below) and then create a histogram of words. Because a more meaningful vocabulary results in more meaningful topics, we filter out special characters, numbers, common stop words[3] (e.g. articles and lexical words), and an additional handcrafted list of stop words (see Table IV in the appendix). Compound words formed with a hyphen or dash are decomposed; both the separated words and the original compound word are included. In this fashion, we defined a vocabulary of 9,929 words. A version of the Porter Stemming algorithm [Porter 1980] is used to equate different forms of a word, for instance "elegant" and "elegance." Finally, a mapping from month to season is applied. In order to include the context and classes of the magazines with the associated words, the periodical category to which each magazine title belongs was added to the set of words. We collected these periodical categories from the WorldCat indexing system, which is the largest international network of library content and services [OCLC 2016b].

(a) (b) (c) (d)

Fig. 2: Pink is used in all of these designs, despite the fact that each of these designs belongs to a different context and genre of magazines.

4. STATISTICAL MODEL

When ideating about visual design, the designer takes into account the topic or the context within which he or she is asked to convey his or her message. For instance, when the context is *politics*, the designer may tend to use darker, "heavier" and more "formal" colors. However this is not the only factor, the words in the design also influence the designer's choice of colors. Figure 2 illustrates that *pink* –which may be stereotypically associated with feminity– has been used in a variety of magazines from different genres. This observation suggests that each design's theme might be a combination of words and color distributions; and each design may include a proportion of various themes. Our goal is to model these combinations of words and colors, and infer proportions of these combinations in magazine cover designs. A similar intuition has been argued in statistical topic modeling, specifically LDA [Blei et al. 2003], for modeling word distributions in documents as proportions of different word topics.

LDA (Latent Dirichlet Allocation) is an intuitive approach to infer topics from text data. As Blei et al. [2003; 2012] describe, instead of categorizing and exploring documents using tools such as keywords, we may first categorize documents based on topics. This allows us to explore topics of

[3]Provided by MySQL database, available at https://dev.mysql.com/doc/refman/5.1/en/fulltext-stopwords.html.

interest and find related documents. For example, a document about sociology may include different topics, such as biology, evolution, history, and statistics, with different proportions. Each of these individual topics can be viewed as a multinomial distribution over a fixed vocabulary of words. Accordingly, each document, which can be viewed as a bag of words, is a combination of these topics with some proportions. Typically, a value for the number of topics is chosen by hand. The latent topics, as well as the topic proportions of each document, are inferred by LDA using the observed data, which are the words in the documents.

Just as word topics are distributions over words, one may think of color topics as distributions over colors. This way, we can model the associations between the color topics and word topics and infer combined color-word topics, as we show in the next section. Jointly inferring topics between two different domains requires the LDA framework to be extended. Such an extension was recently proposed by Shu et al. [2009], for identifying unique authors in bibliography databases.

4.1. LDA-dual Model for Color Semantics

In this section, we explain how to adapt the LDA-dual model proposed by Shu et al. [2009] for color semantics. Our implementation[4] of the model is an adaptation of the Matlab *Topic Modeling* toolbox [Steyvers and Griffiths 2014; Griffiths and Steyvers 2004] for use in LDA (see [Jahanian 2014] for relevant derivations).

Assume that there are K color-word topics denoted by $k_1, k_2, ..., k_K$ and D magazine covers denoted by $d_1, d_2, ..., d_D$. Let W denote the number of words in the vocabulary and C denote the number of color swatches, where each swatch is a patch of color defined by using its sRGB values[5]. Moreover, let M_d denote the number of words and N_d denote the number of color swatches in magazine cover d_d. Let $w_{d,m}$ denote the m-th word in the d-th document and $c_{d,n}$ denote the n-th color swatch in the d-th document. Each magazine cover includes some proportion of each word topic, as well as each color topic. Let $y_{d,m}$ denote the word topic assignment to the word $w_{d,m}$ and $z_{d,n}$ denote the color topic assignment to the color swatch $c_{d,n}$. Note that these assignments are latent. Also let $\psi_{y_{d,m}}$ and $\phi_{z_{d,n}}$ denote the multinomial distributions of the word topics and the color topics, respectively.

Each magazine cover includes some proportion of the color-word topics. These proportions are latent, and one may use the K dimensional probability vector θ_d to denote the corresponding multinomial distribution for a document d_d.

Let β, γ, and α be the hyper-parameters of the three Dirichlet distributions for the color topics, word topics, and the proportions θ_d, respectively. Let $\texttt{Dirichlet}(\cdot)$ denote the Dirichlet distribution, and $\texttt{Discrete}(\texttt{Dirichlet}(\cdot))$ denote the discrete distribution that is drawn from a Dirichlet distribution.

Given the above notation, the generative model for LDA-dual can be written as follows:

(1) Draw K word topics $\psi_k \sim \texttt{Dirichlet}(\gamma)$.
(2) Draw K color topics $\phi_k \sim \texttt{Dirichlet}(\beta)$.
(3) For each document $d_d \in \{d_1, d_2, \ldots, d_D\}$:
  — Draw $\theta_d \sim \texttt{Dirichlet}(\alpha)$.
  — For each word $w_{d,m}$ with $m = 1, \ldots, M_d$
    — Draw $y_{d,m} \sim \texttt{Discrete}(\theta_d)$
    — Draw $w_{d,m} \sim \texttt{Discrete}(\psi_{y_{d,m}})$
  — For each color $c_{d,n}$ with $n = 1, \ldots, N_d$
    — Draw $z_{d,n} \sim \texttt{Discrete}(\theta_d)$
    — Draw $c_{d,n} \sim \texttt{Discrete}(\phi_{z_{d,n}})$

A graphical model for this generative process is illustrated in Fig. 3, where the shaded nodes denote observed random variables and the unshaded nodes are latent random variables.

[4]Available at https://github.com/ali-design/ColorSemantics.git.
[5]Recall that we discretize and use 8 values for each of the three sRGB color channels. Therefore, $C = 512$.

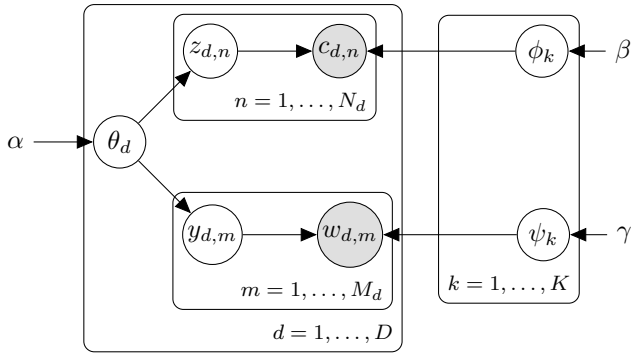


Fig. 3: Hierarchical Bayesian plate model for the LDA-dual model, which combines color and word topics. Here, D is the number of magazine covers; K is the number of color-word topics; and N_d and M_d are the number of color swatches and words, respectively, in the d-th magazine cover.

If we let $\boldsymbol{\phi} = \{\phi_1, \ldots, \phi_K\}$, $\boldsymbol{\psi} = \{\psi_1, \ldots, \psi_K\}$, $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_D\}$, $z_d = \{z_{d,1}, \ldots, z_{d,N_d}\}$, $y_d = \{y_{d,1}, \ldots, y_{d,M_d}\}$, $\boldsymbol{z} = \{z_1, \ldots, z_D\}$, $\boldsymbol{y} = \{y_1, \ldots, y_D\}$, $\boldsymbol{w} = \{w_1, \ldots, w_d\}$, and $\boldsymbol{c} = \{c_1, \ldots, c_d\}$, then the joint distribution corresponding to the LDA-dual model above can be written as

$$p(\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{c}, \boldsymbol{y}, \boldsymbol{w}) = \prod_{i=1}^{K} p(\phi_i|\beta) \cdot p(\psi_i|\gamma) \cdot \prod_{d=1}^{D} p(\theta_d|\alpha) \cdot \left(\prod_{n=1}^{N} p(z_{d,n}|\theta_d)\, p(c_{d,n}|\boldsymbol{\phi}, z_{d,n}) \right) \cdot \left(\prod_{m=1}^{M} p(y_{d,m}|\theta_d)\, p(w_{d,m}|\boldsymbol{\psi}, y_{d,m}) \right). \quad (1)$$

Figure 4 provides a graphical illustration of the generative mechanism and the inference procedure described below. This figure is a symbolic representation of the model. In each sub-figure, a cylinder represents a color-word topic. Each arrow represents the probability of each cover being drawn from a given color-word topic. Each cover includes a histogram of colors and a list of words (each word is superscripted by its corresponding color-word topic). In the generative process, we know the distribution of the color-word topics, and can produce the distribution of the colors and words on the magazine covers. For instance, "Cover 1" is completely (with probability 1.0) generated by color-word topic 1. "Cover 2" is generated by equal distributions of both "color-word topic 1" and "color-word topic 2". In the statistical inference mechanism, we only know the distribution of the colors and the words for each cover. We do not know (represented by question marks) the color-word topics, their proportions, and the assignments of the colors and words of each cover to these color-word topics.

4.2. Inference

Since $\boldsymbol{c}$ and $\boldsymbol{w}$ are observed, inference entails computing

$$p(\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{y}|\boldsymbol{c}, \boldsymbol{w}) = \frac{p(\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{c}, \boldsymbol{w})}{p(\boldsymbol{c}, \boldsymbol{w})}. \quad (2)$$

Theoretically, the above distribution can be obtained by computing the joint probability distribution of the latent and the observed variables, and then computing the marginal probability of the observations. In practice, however, topic modeling algorithms approximate the result to bypass the computational complexity of the solution. There are often two approaches for this approximation [Blei 2012]: variational inference [Jordan et al. 1999; Teh et al. 2006] and Markov chain Monte

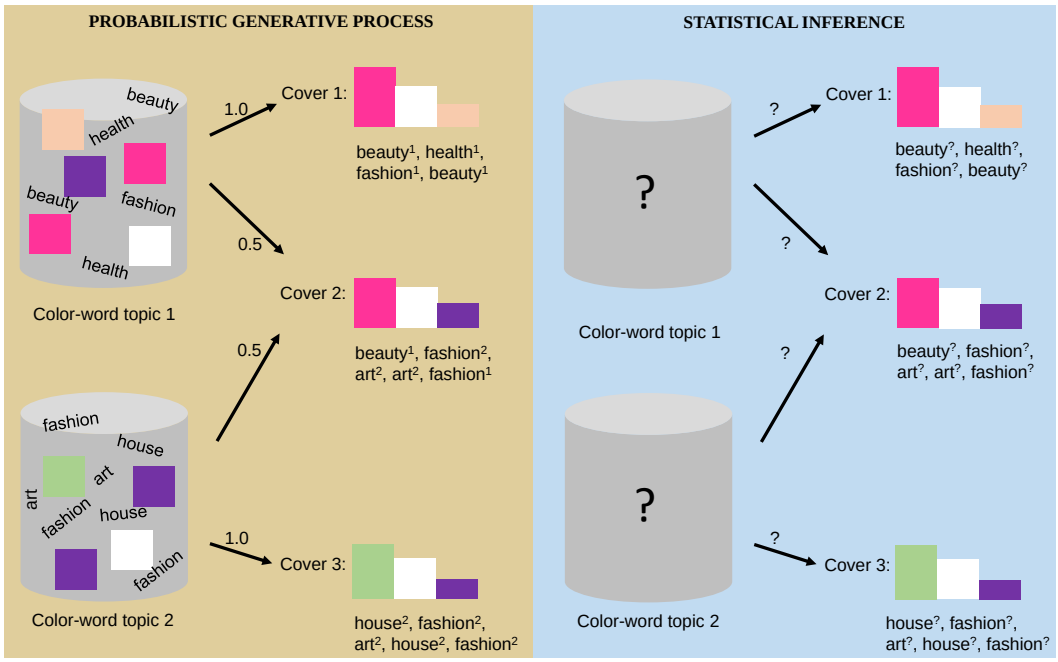


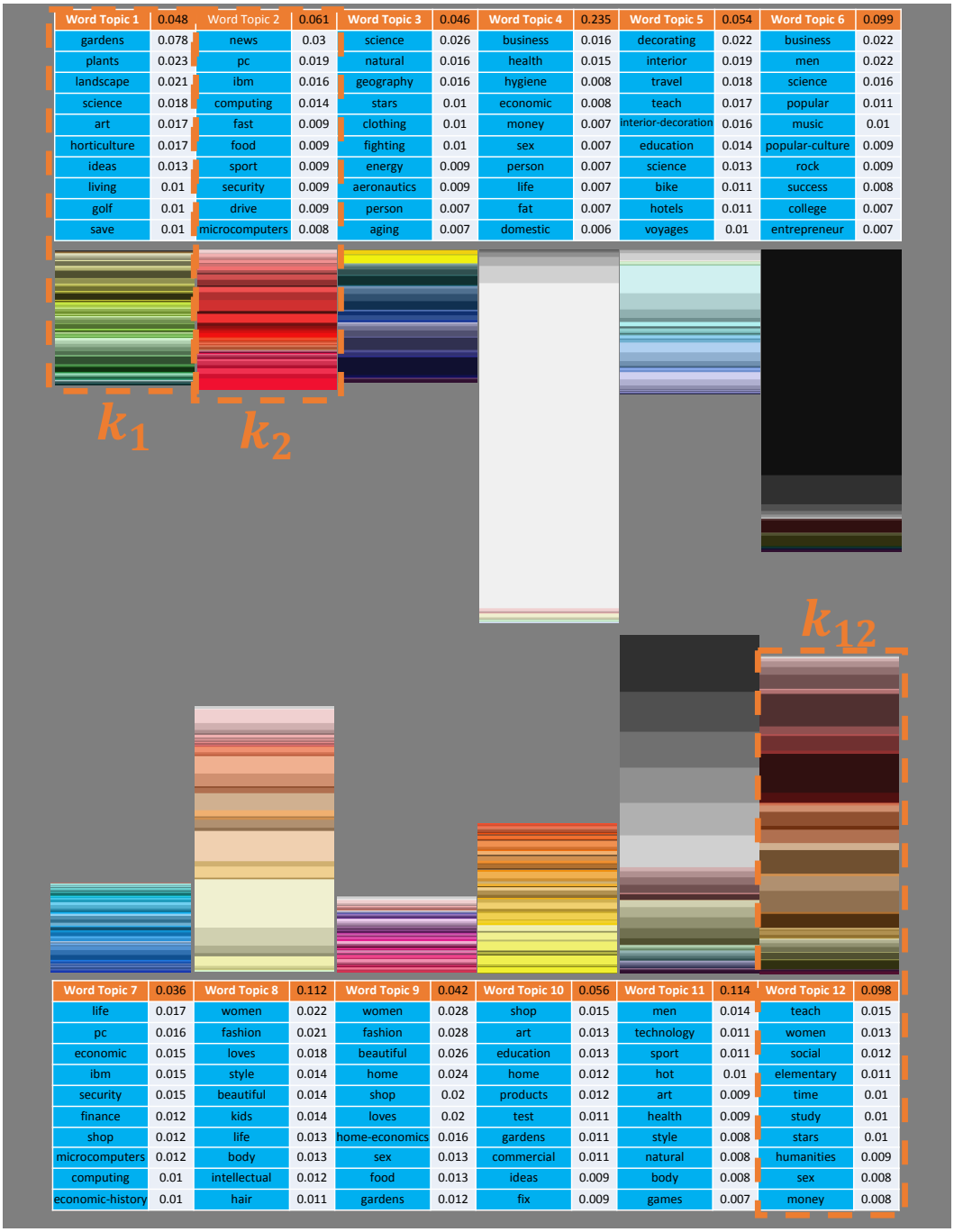
Fig. 4: LDA is both a generative and inference model. This image is inspired by [Steyvers and Griffiths 2007].

Carlo (MCMC) sampling [Andrieu et al. 2003; Griffiths 2002]. We adapted MCMC collapsed Gibbs sampling from the Matlab Topic Modeling Toolbox [Griffiths and Steyvers 2004; Steyvers and Griffiths 2014].

In this paper, we set the number of topics to $K = 12$, with hyper-parameters $\alpha = 0.8$, and $\beta = \gamma = 0.1$. We chose β and γ to match the values used in the original version of LDA applied to text documents [Griffiths and Steyvers 2004]. Also, we assume a symmetrical distribution for the topics, and thus a higher value for α means each cover is a mixture of most of the topics. Note that the 12 general categories (genres) of magazines and the 12 topics ($K = 12$) are independent. There is no a priori relationship between the number of genres (determined by the methods in Sec. 3) and the topics produced by LDA-dual. Furthermore, note that the choice of hyper-parameters in probabilistic models, such as the hyper-parameters of a Dirichlet Process Model, affect the resulting model. It is possible to select parameters that maximize data likelihood [Wallach et al. 2009]. However, other studies have shown this does not necessarily discover topics that correspond to human intuitions [Chang et al. 2009]. In this paper, we present our approach for validating the results according to human judgement. Importantly, given our parameter choices, a reader should readily be able to reproduce our findings.

Figure 5 illustrates the 12 color-word topics inferred by the model for the given parameters. Note that because each color-word topic includes proportions of the color basis and the vocabulary words, in this figure, we visualize a topic as a pair of colors and words histograms. The visualized histograms just illustrate the principal components.

Note that in this figure, next to each word topic (e.g. “Word Topic 1”), there is a distribution over some colors representing the associated color topic. Each word topic and color topic has some proportion in the entire dataset (e.g. 0.0483 for “Word Topic 1”). The summation of all the 12 word topics proportions is 1. Similarly, all the 12 color topic proportions add up to 1. Here, just for visualization, we show the length of each color topic based on its ratio to the “Color Topic 4” (in color-word topic k_4), which has the largest proportion.

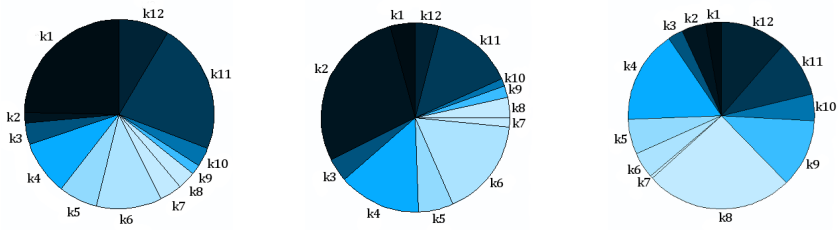


Word Topic 1	0.048	Word Topic 2	0.061	Word Topic 3	0.046	Word Topic 4	0.235	Word Topic 5	0.054	Word Topic 6	0.099
gardens	0.078	news	0.03	science	0.026	business	0.016	decorating	0.022	business	0.022
plants	0.023	pc	0.019	natural	0.016	health	0.015	interior	0.019	men	0.022
landscape	0.021	ibm	0.016	geography	0.016	hygiene	0.008	travel	0.018	science	0.016
science	0.018	computing	0.014	stars	0.01	economic	0.008	teach	0.017	popular	0.011
art	0.017	fast	0.009	clothing	0.01	money	0.007	interior-decoration	0.016	music	0.01
horticulture	0.017	food	0.009	fighting	0.01	sex	0.007	education	0.014	popular-culture	0.009
ideas	0.013	sport	0.009	energy	0.009	person	0.007	science	0.013	rock	0.009
living	0.01	security	0.009	aeronautics	0.009	life	0.007	bike	0.011	success	0.008
golf	0.01	drive	0.009	person	0.007	fat	0.007	hotels	0.011	college	0.007
save	0.01	microcomputers	0.008	aging	0.007	domestic	0.006	voyages	0.01	entrepreneur	0.007

Word Topic 7	0.036	Word Topic 8	0.112	Word Topic 9	0.042	Word Topic 10	0.056	Word Topic 11	0.114	Word Topic 12	0.098
life	0.017	women	0.022	women	0.028	shop	0.015	men	0.014	teach	0.015
pc	0.016	fashion	0.021	fashion	0.028	art	0.013	technology	0.011	women	0.013
economic	0.015	loves	0.018	beautiful	0.026	education	0.013	sport	0.011	social	0.012
ibm	0.015	style	0.014	home	0.024	home	0.012	hot	0.01	elementary	0.011
security	0.015	beautiful	0.014	shop	0.02	products	0.012	art	0.009	time	0.01
finance	0.012	kids	0.014	loves	0.02	test	0.011	health	0.009	study	0.01
shop	0.012	life	0.013	home-economics	0.016	gardens	0.011	style	0.008	stars	0.01
microcomputers	0.012	body	0.013	sex	0.013	commercial	0.011	natural	0.008	humanities	0.009
computing	0.01	intellectual	0.012	food	0.013	ideas	0.009	body	0.008	sex	0.008
economic-history	0.01	hair	0.011	gardens	0.012	fix	0.009	games	0.007	money	0.008

Fig. 5: Color-word topics inferred by the LDA-dual model. Illustration of the 12 color topics in the middle, and their corresponding 12 word topics; 6 on top for the first 6 color histograms from left, and the other 6 on the bottom. Note that for visualization, only the principal elements in the histograms are shown. Also note that the numerical weight of each word topic is shown next to heading of each word topic histogram.

Figures 6 (a), (b), and (c) illustrate the proportions of each of the inferred color-word topics for three magazine title designs in the dataset. For instance, note that *Vogue* as a fashion magazine has k_8 and k_9 as two of the dominant color-word topics. As can be seen, k_8 and k_9 contain words such as "women", "fashion", "love", and "beauty", while the corresponding color histograms contain pastel and pink colors, which are often associated with fashion magazines. On the other hand, *Horticulture*, which is a nature magazine, has the highest proportion of k_1, which pre-dominantly contains shades of green. The words in k_1 include gardening-related words such as "gardens", "landscapes", and "plants". See Fig. 16 for all 71 magazine titles. Additionally, Table II in the appendix illustrates the proportions of the top 10 magazine titles in the color-word topics.



(a) *Horticulture* (b) *Time* (c) *Vogue*

Fig. 6: Proportions of each of the inferred color-word topics for three sample magazine title designs in the dataset. (a) *Horticulture*, (b) *Time*, and (c) *Vogue* magazines (including all the issues in the dataset) are shown. Note that k's are the same as in Fig. 5 (and Fig. 8). See Fig. 16 for all of the magazines.

5. INTERPRETING THE MODEL OUTPUT

Visualizing the results of LDA is a topic of research [Chaney and Blei 2012; Chuang et al. 2012]. Chaney and Blei [2012], for example, suggest a visualization mechanism for exploring and navigating through inferred topics from LDA and their corresponding documents. Although their work does not completely address the usability evaluation of this mechanism, it inspired our visualization mechanism for our user study. In order to evaluate the color semantics hypothesis, we need to display both the color histogram and the word histogram to the participants in our user study in a comprehensive, yet unbiased fashion. We address this via a two-step process. The word histograms are converted to word clouds, while the color histograms are converted to 5-color palettes. Figure 7 illustrates the visualization process. We discuss our choices for colors and words, and describe our implementions for each decision in the two following sections.

5.1. From Color Histograms to Color Palettes

We use 5-color palettes as proxies to represent each of the color histograms in Fig. 5. That is, we chose to match 5-swatch color palettes to the 512-bin color histogram returned by the model. This was for several key reasons. First, a 512-bin color histogram is perceptually hard to be shown to the participants, given that many colors might be invisible at each level of visualization. In other words, such a color histogram encapsulates too much of information to be comprehended. On the other hand, it is not practical to show the entire histogram to a user for his/her usage in an application. Therefore, we need to downsample the histogram. Second, 5-color palettes are standard in the design industry and prior computer science work, and thus can be easily obtained ([Murray et al. 2012; Lin and Hanrahan 2013; Adobe Kuler 2016; ColourLovers 2016], etc.). Designers argue that using more than 5 can lead to clutter (e.g. see [Samara 2007]), and most magazines use 2-3 additional colors (e.g. for typography and bells-and-whistles) beyond those contained in the images (from interviews with designers [Jahanian 2011]).

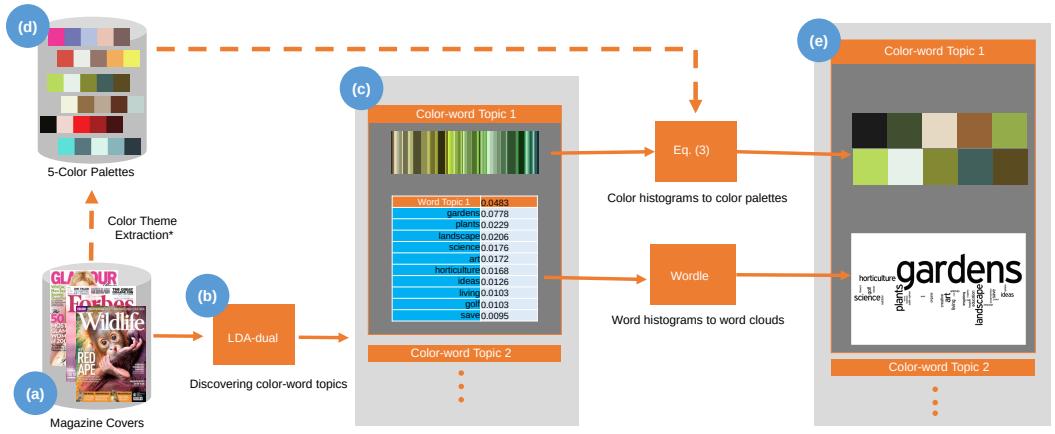


Fig. 7: Visualization process for the inferred color-word topics. To visualize the color-word topic histograms inferred by the model (see Fig. 5), we use 5-color palettes and word clouds as proxies to color histograms and word histograms, respectively. (a) The magazine cover dataset. (b) Applying the LDA-dual model on the dataset. (c) Output of the LDA-dual –color-word topics. (d) Extracting color palettes from the dataset to be used in Eq. (3) for finding the closest color palettes to each color topic histogram. *See Sec. 5.1 for the color theme extraction details. (e) The closest color palettes (here only two), to the color histogram, and the word cloud for the word histogram.

The corresponding color palettes of the color histograms are drawn from a pool of 5-color palettes, one for each magazine cover in the dataset. To extract color palettes from the images, we used the color theme extraction code provided by [Lin and Hanrahan 2013]. In their implementation, the algorithm requires a saliency map of the given image (they used the code from [Judd et al. 2009]), as well as the segmentation of the given image (they used the code from [Felzenszwalb and Huttenlocher 2004]). In our work, we however used the saliency map code from [Harel et al. 2007], since it was easily accessible. Note that these color palettes are not the input to the model; they are only used to visualize the inferred color histograms.

In order to find the 5-color palettes that are closest to the color topic histograms, we define a similarity metric as follows: Let S^{512} denote a color topic histogram with the 512 color basis defined earlier, and S^5 denote a 5-color palette. An intuitive similarity metric is the Euclidean distance between color swatches of S^{512} and S^5. Among the possible color spaces, we choose the CIELab color space with a D65 reference white point. It is considered to be a perceptually uniform space, where ΔE around 2.3 (the distance between two colors) corresponds to one JND (Just Noticeable Difference) [Sharma 2002].

Defining the color similarity distance problem as a bipartite graph matching between S^{512} and S^5 with 512 and 5 nodes, respectively, we find the minimum distance cost of this graph using the Hungarian method [Kuhn 1955]. Equation 3 defines the weighted Euclidean distances d_{WED} between the nodes of these two graphs. Here, the weight w_i corresponds to the weight of the i-th color in the color topic histogram S^{512}, and $\left\|S_i^{512} - S_j^5\right\|_2$ denotes the distance in CIELab between the i-th color from S^{512} and the j-th color from S^5. This metric can be thought as a version of The Earth Mover's distance suggested by Rubner et al. [Rubner et al. 2000] for image retrieval, with the weight vector representing color importance.

$$d_{WED} = \sum_{i=1}^{512} \frac{1}{w_i} \sum_{j=1}^{5} \left\|S_i^{512} - S_j^5\right\|_2. \tag{3}$$

Fig. 8: Alternative visualization of the 12 color-word topics shown in Fig. 5. In each color-word topic (e.g. k_1), the top panel shows the color histogram and the second and third color panels show the top two color palettes we extracted from this histogram. The word topics are visualized in the bottom panel as word clouds, with the size of a word being proportional to its weight.

Computing d_{WED} for a given color topic histogram and all 5-color palettes, we can choose the closest of them as proxies to the histogram (see Fig. 7). In the user study, we present two series of questions for the first and the second closest color palettes, because just one color palette may not provide an adequate visualization of the entire topic histogram. See Fig. 8 for the entire color-word topics.

Note that we could have simply chosen the top 5 colors represented in the histogram, but this has a few shortcomings. First, this would not take into account the variability in color distributions. One could use k-means clustering to get a representative set of "average" colors, but simply averaging

color values often results in perceptually different (and thus non-representative) colors [Lin and Hanrahan 2013]. Another approach might be to choose 5 peaks or modes in the histogram, according to some measure. This suffers the same problem of picking the top 5, since there is no guarantee of capturing the natural perceptual variability in the color distributions.

5.2. From Weighted Bag of Words to Word Clouds

Figure 7 illustrates how we visualize each word topic histogram with a word cloud. The word cloud (or tag cloud) is a visualization technique used to show the relative weights of words through different font sizes. The weights resemble frequency of occurrence or importance of the words in a word dataset. A suite of word cloud algorithms and their usabilities are discussed in [Seifert et al. 2008]. Because of the popularity of word clouds in visualizing categories, and the fact that words are randomly scattered over a layout, we used this technique in our user study. Using wordle[6], we generated black and white word clouds to avoid introducing any color bias. Note that whereas we chose two color palettes for each color topic, we developed only one word cloud for each word topic. This is because we are downsampling the color histograms a lot more than the word histograms, and it makes sense to test the colors with a stronger test (two palettes per color histogram).

6. USER STUDY

The main aim of this section is to validate the output of the probabilistic topic model[7]. In particular, we want to understand if casual users (who are not necessarily designers) agree with the association between color combinations and linguistic concepts produced by our model. We conducted two experiments to study how users match a given color palette with the 12 word clouds –Experiment I (color palette to word cloud direction)– and vise versa –Experiment II (word cloud to color palette direction). Since Experiment I was our main survey with a larger number of participants, we discuss our evaluation framework through this experiment. We then use the same framework and notation to report Experiment II. We conducted the second experiment to complement the first experiment; however, we note that it has a relatively smaller number of participants.

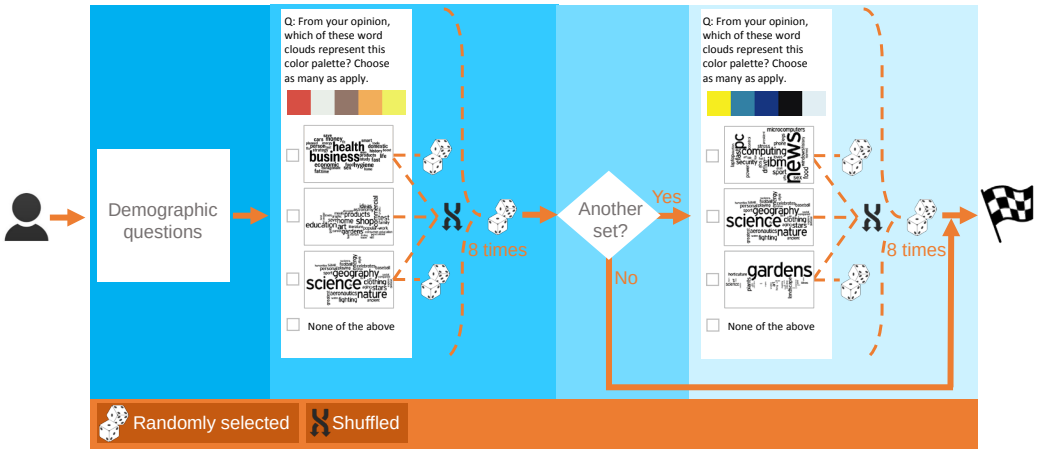


Fig. 9: Flow of the user study.

[6]http://www.wordle.net
[7]The data collected for the experiments is available at https://github.com/ali-design/ColorSemantics.git.

6.1. Experiment I

6.1.1. Stimuli and Procedure. Figure 9 illustrates the flow of the survey. In order to simulate a matching experiment between pairs of color and word topics, we designed a question as follows: one 5-color palette was shown in the left side of the screen, and three shuffled and randomly chosen word clouds, as well as a "None of the above" option were shown on the right side arranged in vertical order. Each question was a multiple choice (represented by multiple checkboxes). For each question, we asked the participant to choose as many word clouds, as in his/her opinion applied to the 5-color palette shown. If the participant felt that none of the word clouds applied to the 5-color palette, he/she could choose "None of the above". Among the three randomly drawn word clouds, one was the word cloud inferred from the model.

The survey was divided into two subsets of questions. The reason for this is that, if otherwise, some participants may lose interest in finishing the experiment. This observation was made in the pilot experiment[8]. More specifically, we created 24 questions for the first and second closest 5-color palettes corresponding to the 12 inferred color topics. However to avoid exhausting the participants, we randomly drew 8 questions from the 12 questions of the closest color palettes and asked the participants to answer them. Then we asked the participants if they would like to continue by taking another set of 8 questions (this time drawn from the 12 questions of the second closest color palettes). Of all participants, 61.35% of the users chose to continue, and answered all 16 questions.

6.1.2. Participants. Our survey[9] was advertised through social networks and universities (Purdue and MIT) email networks. Because trials were randomly ordered, we included all completed trials from all participants (except the few exceptions listed below), regardless of how many trials a participant completed. This survey attracted 1091 participants. The data from 8 participants were removed because they left comments that claimed they had trouble viewing the images on their display. In the early stages of the survey, for the first 177 participants, the "correct" word cloud was not always shown as a possible answer (the word cloud derived from the same topic as the color palette on that trial). We fixed this error for the rest of the participants, and removed 167 trials in which this condition happened. This resulted in 846 participants who completed at least one trial. Over all of these participants, a total of 9,098 trials were completed. Of those trials, 5,523 were in the first subset of questions, and 3,575 in the second.

We collected 846 responses from 481 (56.86%) females, 361 (42.67%) males, and 4 others (0.47%), in the range of 18 to 80 years (with mean = 31.04 and standard deviation = 12.10). The participants are from 70 countries and natively speak 66 different languages, with the majority from the U.S. (60.05%). There are 340 (40.19%) participants who have lived in more than one country. There are 346 (40.90%) participants with college degrees, 445 (52.6%) with graduate degrees (graduate school, PhD, postdoctoral), and 55 others (pre-high school, high school, and professional degree). The majority of the participants, 705 (83.33%) are non-designers. In contrast, there are 129 (15.25%) participants with three or more years of experience in visual design (including graphic design, interior design, and textiles.) Participants spent on average 6.27 (standard deviation = 3.00) hours per day on the Internet.

6.2. Experiment II

6.2.1. Stimuli and Procedure. We used the same procedure as we did for Experiment I. The only difference was that we showed a word cloud in each question and examined the color palettes in the multiple options for the answers. That is, one word cloud was shown in the left side of the screen, and three shuffled and randomly chosen 5-color palettes (including the correct one), as well as a "None of the above" option were shown on the right side arranged in vertical order. Each question was a multiple choice (represented by multiple checkboxes). For each question, we asked the participant to choose as many 5-color palettes, as in his/her opinion applied to the word cloud

[8]This pilot experiment is hosted at https://purdue.qualtrics.com/jfe/form/SV_7WmxYF575nFx7DL.
[9]https://purdue.qualtrics.com/jfe/form/SV_1AqhT38FJKZ5Vrf.

shown. If the participant felt that none of the word clouds applied to the world cloud, he/she could choose "None of the above".

6.2.2. Participants. This survey[10] attracted 447 participants. The data from 4 participants were excluded based on comments claiming they either had trouble viewing the images or were severely visually impaired. Of the remaining participants, 378 completed at least one trial. In the end, 4,447 trials were kept. Of those, 2,772 trials in subset 1 and 1,675 in subset 2 were completed by all participants.

We collected 378 responses, with 39.15% male, 60.05% female, and 0.79% others, in the age range of 18 to 70 years (with average 24.53, and standard deviation 8.22). The participants are from 35 countries and natively speak 41 different languages, with the majority from the U.S. (75.13%). There are 118 (31.22%) participants who have lived in more than one country. There are 179 (47.35%) participants with college degrees, 166 (43.92%) with graduate degrees (graduate school, PhD, postdoctoral), and 33 others (pre-high school, high school, and professional degree). The majority of the participants, 332 (87.83%) are non-designers. In contrast, there are 36 (9.52%) participants with three or more years of experience in visual design (including graphic design, interior design, and textiles.) Participants spent on average 5.80 (standard deviation = 2.66) hours per day on the Internet.

7. INTERPRETING THE USER STUDY

In this section we explain the statistical inference mechanism that we used to understand the user responses for Experiment I. We then use the same framework and notation to report the responses of Experiment II.

Table I: Permutation test results.

Color-word topic	Ratio of observed "correct" responses to prediction from random guessing	
	Experiment I	Experiment II
k_1	2.66*	3.26*
k_2	0.69	0.91
k_3	1.59*	1.40*
k_4	1.49*	1.33*
k_5	1.63*	1.51*
k_6	2.03*	1.69*
k_7	1.74*	1.80*
k_8	1.49*	1.48*
k_9	2.58*	2.40*
k_{10}	1.63*	1.96*
k_{11}	1.61*	1.63*
k_{12}	1.01	1.00
average	1.68*	1.70*

* indicates significantly greater than random guessing at $p < 5 \times 10^{-4}$, two-sided permutation test.

7.1. Summary of Basic Statistics on Correct Responses

Before accounting for detailed statistical results, we first show the basic strength of the associations between palettes and their corresponding word clouds in the participant data. In this section, by "correct" response, we mean the participant's choice agreeing with the model's output.

Participants in Experiment I, pooled over both subsets, were 1.68 times more likely than random guessing to choose the "correct" word cloud for a given color palette. This is the ratio of how many times participants chose the correct word cloud to how many would be expected if each participant's

[10] https://jfe.qualtrics.com/form/SV_8B9qc7nqnUKPuEl.

per-trial responses were randomized (using the same number, but not order, of choices per trial). This effect is highly significant ($p < 5 \times 10^{-4}$, two-sided permutation test) for all but two color-word topics (topics k_2 and k_{12} were not significantly above chance). The results are summarized in Table I.

Analogously to Experiment I, participants in Experiment II were 1.70 times more likely to choose the "correct" color palette for a given word cloud than chance. This effect is significant for all topics ($p < 5 \times 10^{-4}$, two-sided permutation test) except topics k_2 and k_{12}. These results are also available in Table I.

7.2. Statistical Model

In order to interpret the results in more detail, we present a statistical analysis. First, we define some notation. Let c_i denote the event that the i-th color palette was displayed. Also, let w_j denote the event that the user selected (clicked on) the j-th word cloud, and u_{ij} denote the probability that the j-th word cloud was selected by the user in response to the i-th color palette. In order to compute u_{ij} we note that

$$u_{ij} = \Pr(w_j|c_i). \tag{4}$$

There are three possible positions $p \in \{1, 2, 3\}$ at which a word cloud can be displayed. Let d_{jp} denote the event that the j-th word cloud was displayed at position p, and let w_{jp} denote the event that the user selected the j-th word cloud which was displayed at the p-th position. Then

$$u_{ij} = \sum_{p\in\{1,2,3\}} \Pr(w_{jp}|d_{jp}, c_i) \cdot \Pr(d_{jp}|c_i). \tag{5}$$

If d_j denotes the event that the j-th word cloud was selected for display and d_j^p the event that it was displayed at position p, then

$$\Pr(d_{jp}|c_i) = \Pr(d_j|c_i) \cdot \Pr(d_j^p|c_i). \tag{6}$$

According to our experimental design, each word cloud has an equal probability of appearing in any one of the three positions. Therefore

$$\Pr(d_j^p|c_i) = \frac{1}{3}. \tag{7}$$

On the other hand, we always select the i-th word cloud (the true word cloud according to our model) for the i-th color palette. The other two slots are filled by selecting any two of the remaining 11 word clouds uniformly at random. Therefore

$$\Pr(d_j|c_i) = \begin{cases} 1 & \text{if } i = j \\ \frac{2}{11} & \text{otherwise.} \end{cases} \tag{8}$$

All that remains is to estimate $\Pr(w_{jp}|d_{jp}, c_i)$. To estimate this quantity, we utilize the technique known as "cascade click modeling" [Govindaraj et al. 2014]. Cascade click modeling was originally introduced to model a user's back and forth clicks on a list of URLs (resulting from an online search query), regardless of their content. This model allows us to simultaneously estimate two quantities: the first is position bias b_p, the probability that the p-th position is examined by a user; the second is r_{ij}, the intrinsic relevance of the word cloud j to the color palette i. In other words,

$$\Pr(w_{jp}|d_{jp}, c_i) = r_{ij} \cdot b_p, \tag{9}$$

and by using (5) and letting $q_{ij} = \sum_{p\in\{1,2,3\}} \Pr(d_{jp}|c_i) \cdot b_p$, we can write

$$u_{ij} = r_{ij} \cdot \sum_{p\in\{1,2,3\}} \Pr(d_{jp}|c_i) \cdot b_p = r_{ij} \cdot q_{ij}. \tag{10}$$

Note that b_p can be pre-computed as follows:

$$b_p = \frac{m_p}{m}, \tag{11}$$

where m denotes the total number of trials (each question in our survey is equivalent to one trial), and m_p denotes the number of times the word cloud at position p was selected in any of the trials. In Experiment I, we found the position bias of the options (in vertical order) for the first set of the questions to be $0.3308, 0.3797, 0.3473, 0.147$, and $0.4254, 0.3932, 0.3564, 0.1269$ for the second set of the questions. Note that for each set, these numbers do not sum up to 1, because of the fact that the participant could choose more than one word cloud. The position bias for Experiment II is $0.337, 0.4173, 0.3465, 0.1083$ for the first set of the questions, and $0.4641, 0.3654, 0.3434, 0.1319$ for the second set of the questions.

These numbers indicate that the position bias for each option is not equal, and even though we shuffled the three choices of word clouds in the first three vertical positions, we need to account for the position bias. We note that the fourth option –"None of the above"– is clicked less than the other options. This indicates that our participants wished to provide an answer, as well as they may have not thought that the associations between the word clouds and the colors were too abstract. We also note that in the second set of the questions, the fourth number is lower than the one in the first set of questions. This perhaps means that the participants who chose to participate in one more set of the questions in the survey were more confident with their conclusions.

As the last step, let m_i denote the number of trials in which the i-th color palette was displayed, and m_{ij} denote the number of trials in which the i-th color palette was displayed and the j-th word cloud was selected. We can assume that the trials are independent, and therefore the probability of observing this data under model (10) can be written as

$$\Pr(m_i, m_{ij}) = (r_{ij} \cdot q_{ij})^{m_{ij}} (1 - r_{ij} \cdot q_{ij})^{m_i - m_{ij}}. \tag{12}$$

The maximum likelihood estimate for $r_{ij} \cdot q_{ij}$ is simply $\frac{m_{ij}}{m_i}$, from which we can infer $\hat{r}_{ij}$, the maximum likelihood estimate for r_{ij}, as

$$\hat{r}_{ij} = \frac{m_{ij}}{m_i \cdot q_{ij}}. \tag{13}$$

7.3. Analyzing the Results for Experiment I

Figure 10 illustrates two relevance matrices. The matrices corresponds to the inferred relevance of the first and second closest color palette, respectively, to the word cloud produced by LDA-dual. The rows correspond to color palettes, as proxies to color topics histograms, and the columns correspond to word clouds. The (i, j)-th elements of these matrices are the intrinsic relevance values $\hat{r}_{ij}$, computed from the observed responses of the participants using the model described in the previous section. Higher values of $\hat{r}_{ij}$ mean that the users found a high correlation between the i-th color palette and the j-th word cloud. If the participants find the word cloud produced by our model to be the most relevant for a given color palette, then the diagonal entries, marked in blue, should contain the highest values. Whenever an off-diagonal entry is larger than the corresponding diagonal entry in its row, it is marked yellow in the figure.

Note that for the first set of 12 color palettes, participants selected the "None of the above" option on average for 14.70% of the trials. The minimum was 6.21%, which occurred for k_1; and the maximum was 27.54%, which occurred for k_2. For the second set of 12 color palettes, the average was 12.69%. The minimum was 1.34%, which occurred for k_1; and the maximum was 21.81%, which occurred for k_2. Frequent selection of the "None of the above" option for a given color palette suggests that participants had more difficulty associating this palette with the word clouds that were shown. In the relevance matrices, we do not compute the numbers for each color palette against the "None of the above" option. Thus the matrices do not contain the 13-th column. However, if the

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.14	0.33	1.38	0.39	0.49	0.38	0.26	0.39	0.28	0.81	0.40	0.30
c_2	0.58	0.27	0.56	0.63	1.18	0.38	0.36	1.80	1.86	0.70	0.68	1.45
c_3	0.14	1.25	1.53	0.84	0.87	0.89	1.18	0.52	0.28	0.68	1.04	0.34
c_4	0.03	1.62	1.16	1.29	0.71	1.18	1.65	0.12	0.26	0.64	1.39	0.27
c_5	0.39	1.23	1.46	1.09	1.45	1.05	1.61	0.43	0.36	0.93	1.13	0.53
c_6	0.74	0.99	1.09	0.74	0.54	1.84	0.98	0.27	0.10	0.61	1.01	0.41
c_7	0.13	1.31	1.23	1.00	1.11	1.01	1.63	0.29	0.48	0.39	1.05	0.48
c_8	0.51	0.34	0.60	0.41	1.49	0.40	0.34	1.79	1.31	1.03	0.48	1.07
c_9	0.44	0.03	0.07	0.30	1.30	0.07	0.11	1.96	2.06	0.63	0.34	1.20
c_{10}	1.06	0.25	0.73	0.58	1.82	0.34	0.18	1.54	1.00	1.43	0.61	1.18
c_{11}	0.30	1.68	0.91	1.41	0.48	1.23	1.52	0.23	0.23	0.74	1.37	0.37
c_{12}	0.99	0.50	1.58	0.94	0.87	1.53	0.64	0.57	0.63	1.09	1.09	0.58

(a)

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.23	0.24	1.45	0.48	1.06	0.47	0.56	0.51	0.29	1.01	0.49	0.45
c_2	0.05	0.97	0.77	0.81	0.74	0.45	0.41	1.10	1.57	0.58	0.89	1.07
c_3	0.31	1.17	1.18	1.04	1.40	0.95	1.31	0.60	0.66	0.85	1.38	0.48
c_4	0.20	1.19	0.91	1.31	0.96	1.16	1.52	0.29	0.43	0.80	0.92	0.63
c_5	0.65	0.93	1.56	0.94	1.31	0.84	0.78	0.49	0.56	1.10	1.40	0.74
c_6	0.19	1.32	1.16	0.79	0.46	1.81	1.29	0.18	0.24	0.54	1.40	0.37
c_7	0.05	1.19	0.97	1.24	1.03	1.38	1.38	0.25	0.70	0.73	1.20	0.36
c_8	1.21	0.84	1.79	0.90	1.42	0.92	0.75	0.62	0.65	1.25	0.83	0.88
c_9	0.26	0.09	0.13	0.32	0.92	0.09	0.13	1.65	2.42	0.47	0.18	1.20
c_{10}	0.74	0.36	0.60	0.70	1.98	0.35	0.45	0.95	0.82	1.29	0.61	1.00
c_{11}	0.09	1.62	0.99	1.11	0.57	1.20	1.40	0.05	0.00	0.63	1.43	0.10
c_{12}	0.53	0.42	1.21	0.96	0.87	0.66	0.62	0.89	0.98	0.91	0.49	1.11

(b)

Fig. 10: Relevance matrices $\hat{R}^1$ and $\hat{R}^2$ for the first and second set of questions, respectively. For the first set of questions, the participants were shown the closest palettes identified by LDA-dual. For the second set of questions, the participants were shown the second closest palettes identified by LDA-dual. The elements of these matrices are the estimated intrinsic relevance of associations between colors and words, calculated from the participants' responses. The higher the value, the greater the intrinsic relevance associated by the users. Ideally, the diagonals should contain the highest values. Whenever an off-diagonal entry is larger than the corresponding diagonal entry in its row, it is marked yellow in the figure.

participant has selected "None of the above" as well as other options, we take into account those options, effectively treating the "None of the above" response as a vote of lower confidence.

The relevance matrices in Fig. 10 show that most diagonal elements are larger than their corresponding off-diagonal ones. This indicates a strong correlation between the results of our application of LDA-dual and participants' opinions (also see the next section for an aggregate measure). It is interesting to note that not all diagonal elements have the same value. $\hat{r}^1_{11}$, for example, has the largest diagonal value, suggesting that "green" and "garden" are closely associated by most participants.

There are a few color palettes such as c_2 in $\hat{R}^1$ where the users assign higher relevance to word clouds other than the one produced by the LDA-dual model. To understand this, note that in Fig. 8 (b), the first 5-color palette which is c_2 predominantly contains shades of red and black. Users assign higher relevance to word clouds w_8 and w_9, which are about "sex" and "beauty." In our dataset however, the red and black color combinations are often used by news magazines such as *Time* and *The Economist*, and computer magazines such as *PC Magazine*. One can compare other color palettes such as c_{12} in $\hat{R}^1$ and c_8 in $\hat{R}^2$ to infer why there is a mismatch between the model output and the relevance values assigned by the users. Moreover, these results illustrate the importance of conducting a user study; domain specific color palettes and their associated linguistic concepts may not always transfer to a general context.

To understand differences between female versus male and non-U.S. versus U.S. participants, we computed the corresponding relevance matrices (see Fig. 19). Comparing these matrices with Fig. 10, we do not observe any striking differences. This suggests that our results do not depend strongly on the gender or cultural background of the users. When we compare designers versus non-designers, however, we find that there are more zero values in the off-diagonals for designers. This indicates that designers are more consistent with each other in color-word associations, perhaps because of their training [Whitfield and Wiltshire 1982]. We directly compare the populations (male vs female, U.S. vs non-U.S., designer vs non-designer) in Fig. 21 in the appendix.

7.4. Aggregate Measures

We have defined two different measures to assess the strength of the relationships of elements in a matrix as indicated by the magnitude of the diagonal elements relative to the off-diagonal elements in each row.

For each N by N relevance matrix $\hat{R}^1$ and $\hat{R}^2$, let r_{ij} be the (i,j)-th element. The first measure is the *diagonal dominance* D. For each row, it is simply the ratio of the diagonal element in that row to the average value of the off-diagonal elements in that row:

$$D_i = \frac{r_{ii}}{\bar{r_i'}}, \tag{14}$$

where r_{ii} is a diagonal element, and $\bar{r_i'}$ is the mean of off-diagonal elements:

$$\bar{r_i'} = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} r_{ij}. \tag{15}$$

To get a summary assessment, we can average the diagonal dominance over all the rows of the matrix:

$$\bar{D} = \frac{1}{N} \sum_{i=1}^{N} D_i. \tag{16}$$

For the two matrices $\hat{R}^1$ and $\hat{R}^2$ in Fig.10, the average diagonal dominances are 2.06 and 1.99, respectively (see Fig. 11). So on average, the diagonal element is about twice as large as the other elements in the row. This suggests that despite the wide variability of the data in the matrices, the diagonal elements tend to dominate.

	$\bar{D}$	$\bar{S}$
$\hat{R}^1$	2.06	1.71
$\hat{R}^2$	1.99	1.52

Fig. 11: Aggregated measures for relevance matrices $\hat{R}^1$ and $\hat{R}^2$. Column $\bar{D}$ is the diagonal dominance measure computed according to (16). Column $\bar{S}$ is the diagonal separation computed according to (19). Overall, these two separate measures support the conclusion that on average, the diagonal elements in the relevance matrices are stronger than the off-diagonal elements.

The second measurement is the *diagonal separation* S. It is also defined for each row, and is also a ratio:

$$S_i = \frac{r_{ii} - \bar{r_i'}}{{}_r\acute{\sigma}_i}. \tag{17}$$

The numerator is the difference between the diagonal element and the mean of the off-diagonal elements in that row. The denominator is the standard deviation of the off-diagonal elements in that row:

$${}_r\acute{\sigma}_i = \left(\frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(r_{ij} - \bar{r_i'}\right)^2 \right)^{\frac{1}{2}}. \tag{18}$$

To get a summary assessment, we can again average the diagonal separation over all the rows of the matrix:

$$\bar{S} = \frac{1}{N} \sum_{i=1}^{N} S_i. \tag{19}$$

As Fig. 11 summarizes, for matrix $\hat{R}^1$ (Fig. 10 (a)), the average diagonal separation is 1.71. Thus, on average, the diagonal element is more than one and a half standard deviations away from the mean of the off-diagonal elements. For the second matrix, $\hat{R}^2$ (Fig. 10 (b)), the average diagonal separation is 1.52. While this is smaller than it is for the first matrix, it still indicates good separation.

Overall, these two separate measures support the conclusion that on average, the diagonal elements are indeed stronger than the off-diagonal elements.

7.5. Analyzing the Results for Experiment II

Figure 12 illustrates the relevance matrices of the responses in Experiment II. For the first set of 12 color palettes, participants selected the "None of the above" option on average for 10.81% of the trials. The minimum was 4.27%, which occurred for k_1; and the maximum was 17.57%, which occurred for k_{12}. For the second set of 12 color palettes, the average was 13.18%. The minimum was 1.44%, which occurred for k_1; and the maximum was 29.79%, which occurred for k_8. For the two matrices $\hat{R}^1$ and $\hat{R}^2$ in Fig.12, the average diagonal dominances are 2.36 and 2.61, respectively. This indicates that on average, the diagonal element is more than twice as large as the other elements in the row; the diagonal elements do tend to dominate. For matrix $\hat{R}^1$ (Fig. 12 (a)), the average diagonal separation is 2.38. For the second matrix, $\hat{R}^2$ (Fig. 12 (b)), the average diagonal separation is 2.17, suggesting a high level of separation between diagonal and off-diagonal elements for both of the matrices. Overall, these two separate measures support the conclusion that on average, the diagonal elements are stronger than the off-diagonal elements. Therefore, these aggregate statistics show that for both Experiments I and II, participants favored the color-word associations discovered by our model.

$\hat{R}^1$	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}
w_1	2.08	0.48	0.14	0.20	0.00	0.34	0.07	0.41	0.14	0.56	0.07	0.34
w_2	0.30	0.64	1.46	1.81	1.75	1.16	1.07	0.82	0.43	0.43	1.63	0.65
w_3	2.00	0.30	1.49	0.56	0.83	0.86	1.12	0.70	0.19	0.68	0.50	0.80
w_4	1.04	0.57	1.10	1.24	1.16	0.70	1.02	0.82	0.42	0.72	0.94	0.56
w_5	1.13	1.01	0.73	0.96	1.28	0.37	1.06	1.32	1.08	1.78	0.41	0.49
w_6	1.29	0.55	1.11	1.08	1.15	1.48	1.26	0.54	0.13	0.68	1.21	0.96
w_7	0.72	0.43	0.95	1.30	1.20	0.78	1.61	0.37	0.13	0.58	1.57	0.79
w_8	0.31	1.44	0.43	0.44	0.68	0.13	0.37	1.72	1.80	1.49	0.13	0.25
w_9	0.07	1.22	0.77	0.54	0.35	0.07	0.61	0.81	1.78	1.19	0.33	0.47
w_{10}	0.85	0.48	1.06	0.52	0.84	0.35	0.54	0.68	0.87	1.76	0.25	0.50
w_{11}	1.36	0.57	1.30	0.71	1.01	0.98	1.28	0.43	0.19	0.61	1.60	0.95
w_{12}	0.63	1.27	0.96	0.60	1.15	0.37	0.49	1.52	1.74	1.51	0.55	0.71

(a)

$\hat{R}^2$	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}
w_1	2.09	0.11	0.46	0.00	0.33	0.23	0.23	0.34	0.11	0.43	0.24	0.34
w_2	0.28	1.05	1.16	1.32	0.86	0.86	1.59	0.55	0.19	0.68	1.16	0.55
w_3	2.06	0.39	0.94	0.88	1.09	0.98	1.23	1.03	0.40	1.12	0.83	0.91
w_4	1.58	0.76	0.87	1.15	1.52	1.25	1.06	0.77	0.20	0.94	1.07	0.98
w_5	1.52	0.44	0.34	1.02	1.19	0.54	0.86	0.76	1.31	1.85	0.47	1.09
w_6	0.93	0.49	1.45	0.88	1.18	1.78	0.95	0.79	0.09	0.62	1.26	0.47
w_7	0.68	0.57	1.17	1.01	1.09	1.08	1.26	0.76	0.23	0.32	1.41	0.32
w_8	0.60	1.44	0.71	0.56	0.73	0.10	0.92	0.64	2.08	1.07	0.00	1.42
w_9	1.07	1.38	0.76	0.55	1.01	0.39	0.09	0.10	2.29	0.94	0.20	0.55
w_{10}	2.02	1.01	0.43	0.65	0.97	0.44	0.76	1.29	0.76	1.40	0.11	0.79
w_{11}	1.18	0.93	1.00	0.68	1.17	0.90	0.96	0.67	0.42	1.31	1.38	0.67
w_{12}	0.48	0.56	1.30	0.29	1.11	0.85	0.50	0.47	1.51	1.71	0.59	1.09

(b)

Fig. 12: Relevance matrices $\hat{R}^1$ and $\hat{R}^2$ for the first and second set of questions, respectively, in Experiment II. For the first set of questions, in each question, the participants were shown a word cloud versus its associated closest palettes identified by LDA-dual. For the second set of questions, for each question, the participants were shown a word cloud versus its second closest palettes identified by LDA-dual. The elements of these matrices are estimated intrinsic relevance of associations between words (rows) and colors (columns), calculated from the participants' responses, using the same statistical model described in 7.2.

It is interesting to note that some participants in Experiment II mentioned preferring certain palettes over others, based purely on aesthetics and not the task. Any effects of such preferences are not readily observed in the data, however. Furthermore, just as for Experiment I, we compared the data from subpopulations (male vs female, non-U.S. vs U.S., designer vs non-designer) of participants in Experiment II (Fig 21). We did not find any striking differences between the subgroups.

8. APPLICATIONS

Understanding of color semantics as represented by color-word topic modeling could be useful for a number of real-world applications. Here we give some examples. Note that we have not yet conducted formal user studies with these applications; our informal use, however, has shown them to be useful. Although they are by no means "finished", we argue that they demonstrate the value of color semantics and motivate future development of tools. We have implemented these applications (except pattern recoloring, see Sec. 8.1.2) as prototypes in Matlab.

8.1. From Semantics to Color Palettes

Effectively selecting color palettes is important for many domains, including product design, image recoloring based on color mood, image retrieval, and visual feature-based recommendation systems. In visual design, for instance, the user needs a color combination that is both appealing and aligned with the purpose of design [Samara 2007]. This is particularly important for a non-designer, who may have little training in how to choose a good color palette. Nonetheless designers may also prefer to use automatically generated examples as inspiration (e.g. see [Starmer 2005]). There exist several online communities for color palette design (e.g. [Adobe Kuler 2016; ColourLovers 2016]), each with millions of user-created, named, and rated palettes. As we mention in the second-to-last paragraph of the Introduction section, current online palette design communities only have sparse, user-created labels, that usually lack semantic information [O'Donovan et al. 2011]. Understanding the associations between colors and linguistic concepts provides a more meaningful and tractable way to find palettes. Since LDA-dual relates sets of colors to sets of words, we can map user-input words to retrieve color palettes for use in design.

8.1.1. Recommending Color Palettes. Our model's inferred color semantics can be applied directly to color palette recommendation. Consider a scenario in which the user wishes to find a color palette for a tech magazine's edition on how developments in material engineering change what manufacturers are using to make dresses. Figure 1 illustrates such a scenario, where the user queries for "technology" and "fashion". The user can also provide weights to each word, e.g. 80% to "technology" and 20% to "fashion". This enables a richer and more customized way to find semantically appropriate palettes. As discussed in Sec. 2, unlike the current state of the art in color palette recommendation, we incorporate the knowledge embedded in designers' work and rank palettes both extracted from our dataset as well as those from any existing database of palettes, like Adobe Kuler or ColourLovers.

The text-input query is mapped to the word topics, and then to their corresponding color topics. We weight and map these color topics to a ranked set of color palettes in a pool of color palettes (see Fig. 1) created from the magazine dataset (see Sec. 3.1). The user can then choose his/her preferred color palettes from the recommended set.

In short, the tool works by creating a color histogram which is the weighted sum of the color histograms from color-word topics discovered by LDA-dual. There are two sets of weights. The first is a weighting based on how often the input word occurs in the word histogram of each color-word topic. For example, if the word "technology" occurs twice as often in topic 1 than topic 2, the color histogram from topic 1 will be weighted twice relative to topic 2. These weighted histograms are combined to create a histogram for each user-input word. The resulting histograms are then themselves weighted based on the user-input weight per word. If the user input a weight 80% for "technology" and 20% for "fashion", the "technology" color histogram would be weighted four times that of the "fashion" histogram, and they would then be combined into one color histogram. The tool then computes the Euclidean distance of the weighted color histogram to each color in the 5-color palettes from our database, using [Lin and Hanrahan 2013], just as described in Sec. 5.1. The weighted color histogram has 512 elements, most of which are negligible. For this reason, prior to finding the nearest palettes to the weighted color histogram, we truncate the weighted color histogram to include only the 10% largest color bins, as this gives us nearly identical results using the whole histogram, but is an order of magnitude faster to compute.

Furthermore, because the color palettes are derived from magazine covers, we can easily retrieve the covers that correspond to a set of recommended palettes determined from the above method (see Fig. 1). Such design examples can be utilized in creativity support tools [Shneiderman 2009] and to facilitate design prototyping [Dow et al. 2010].

8.1.2. Recoloring Patterns. Designers regularly modify color themes of existing designs to impart new meanings. Part of this color modification may involve recoloring an image [Lin et al. 2013b], transferring a color theme to an image [Murray et al. 2012], or enhancing the color theme of an image [Wang et al. 2010]. We suggest an application of color semantics for pattern recoloring (Fig. 13), based on techniques introduced by Lin et al. [Lin et al. 2013b]. Figure 13 (a) illustrates an original magazine cover from the *Science* magazine. Using the associations embedded in our color-word topics, we are able to accept a user query, map it to the word topics, and use a 5-color palette as a representative of this color topic to recolor the original pattern. Figure 13 illustrates the results of recoloring the original pattern using "shop" (b) and (c), and "sport" (d) and (e) queries, respectively.

(a) (b) (c) (d) (e)

Fig. 13: Pattern recoloring using color semantics. We removed the cover lines of a *Science* magazine cover in (a) and recolored it with color palettes derived from the terms "shop" (b) and (c), "sport" (d) and (e), and added thematic titles. The recoloring method is from [Lin et al. 2013b].

8.2. Image Retrieval

As discussed in Sec. 1, our work targets the "gap" of automatically connecting media to semantic information. Image retrieval is a way to showcase how we traverse the gap in the opposite direction: going from semantic information in the form of text to media in the form of images. Our approach provides a natural way to incorporate high level image features into current image retrieval algorithms. Figure 12 illustrates an application of our inferred color semantics in image retrieval. Consider a scenario in which the user makes a query in an image search community, e.g Pinterest.com [Pinterest 2016], Flickr.com [Flickr 2016], etc. about "interior design". The site will almost invariably suggest an overwhelming number of images. In order to explore and navigate through the retrieved images, however, the user can request a subset of "gardens"-inspired images; using color semantics, we are able to map this query to the color-word topics. We can then map the combined color histograms to rank the already retrieved images based on how well they represent "gardens", similarly to how we retrieve palettes and design examples in the previous section. We show the results of this application in Fig 14 for "beach" and "gardens" images related to interior design.

8.3. Color Selection

Users could utilize color semantics to more intuitively select regions of color within an image. As Heer and Stone [2012] note, the image editing community often uses tools to select subsets pixels of a certain color or color distribution. They suggest tools for selecting colors in images using color name queries. Building upon this concept, we extend the idea one step further and demonstrate the usage of our inferred color semantics for this type of color selection tool. In our case, we use the user's query and map it to the set of pixels that are relevant to not just the queried word, but

(a)

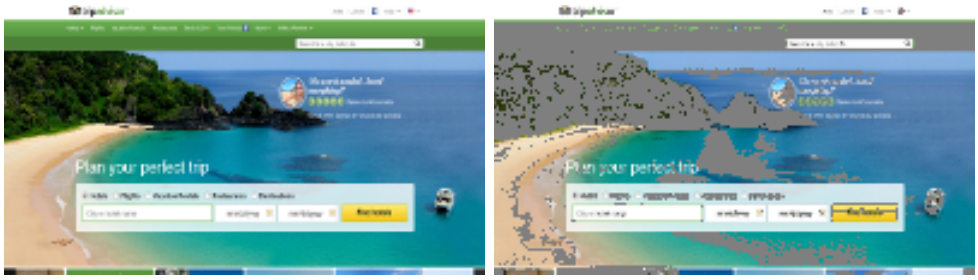
(b)

Fig. 14: Image retrieval using color semantics. (a) “beach” inspired images in a dataset of interior design images. (b) “gardens” inspired images in the same dataset. The interior design dataset is retrieved from scraping the House & Garden website [House&Garden 2016].

also other semantically related words. Figure 15 illustrates this kind of interaction. Figure 15 (a) is the original image, a screenshot of a travel agency website [TripAdvisor 2014]. A user may be interested to know what color regions have contributed to the concepts of “travel” and “trip” in this image. Figure 15 (b) represents the pixels selected by our algorithm for these regions, while turning the other regions to grayscale. Note that in order to find these color regions, we map the users queries to the word topics, and preserve their associated color topics in the image, similar to the previous applications.

9. CONCLUSION AND FUTURE WORK

The goal of visual design is both to convey a message and to be aesthetically appealing. We used data mining to investigate how designers associate colors with linguistic concepts. We collected high quality examples of professional designs, resulting in the largest dataset to date of magazine covers with associated text transcription. We then adapted LDA-dual, an extension of the popular LDA topic model, to simultaneously model designers’ choice of both colors and words for the magazine covers. We used a crowdsourcing experiment to verify the model’s color-word topics. The results



(a) (b)

Fig. 15: Image color selection using color semantics. (a) The original image, (b) colors that contribute to "travel" and "trip" in the original image. Image from the home page of Tripadvisor.com [TripAdvisor 2014].

confirm that our model is able to successfully discover the association between colors and linguistic concepts. This closes the loop of our design mining system, from data to inference to validation.

Our work demonstrates a new methodological approach to color semantics and design. As this is a first pass at using probabilistic models to formalize designers' intuitions about color, we made several assumptions in order to get tangible results; future work should study in more detail and optimize these assumptions. For the purposes of this study, we consider semantics to be the associations that designers create between and within groups of colors and words. Our instantiation of LDA-dual does not explicitly model higher-order semantic relationships, such as the word order in a cover line or the spatial distributions of various colors, that exist in design examples. Future work extending the model to cover more variables is required to uncover these structures. For instance, our method for generating 5-color palettes from the magazine covers explicitly uses a measure of color saliency [Harel et al. 2007], which in turn affects the color-proportion weighting (Eq. 3) used by the model. This method, however, does not capture the spatial layout of color or how color is used differently for foreground and background elements. One extension of our model could be to explicitly represent the saliency of colors or words as an independent input to the model.

Furthermore, since LDA-dual can combine words and topics into an arbitrary number of clusters, one must choose the number of clusters. It is not obvious a priori how many clusters to choose, nor how to optimize this number. Larger numbers of clusters generally result in higher levels of granularity. We found that for our data, 12 topics produced an intuitively parsimonious set of clusters. In the appendix we show the results of finding 6 topics (Fig. 18 in the appendix), and 24 topics online[11]. Ultimately one could create interactive tools for changing the number of topics. Another assumption we made was our choice of hyper-parameters used in training the model. They were chosen to be in line with previous modeling work [Griffiths and Steyvers 2004]. However, even given these assumptions, our validation study strongly suggests that the intuitions discovered by the model are shared by designers and non-designers alike.

A key feature of our work is the development of an extensive database of magazine colors spanning 14 years of publication. Because the present study is meant to be a demonstration of the concept of color semantics, we evaluated all of the covers together, not taking into account how designs may have changed over the years. There clearly exist trends in design, and such trends might be interesting and important to study. One might want to know, for instance, how the association between pink hues and terms like "women" and "fashion" has developed over time, or one might want to know what color-word associations exist in designs from the last year. Extensions to our present work could tease out the formation and evolution of such trends. A potential approach might be to use a predictive model that represents time, e.g. logistic regression, in conjunction with LDA-dual.

[11] https://github.com/ali-design/ColorSemantics.git.

Our validation study shows that users across different countries, genders, and age groups largely agree with the colors and linguistic concept associations discovered by our model. This is not to say that the results are completely general across all subsets of the population; participants all read English, were mostly college-educated, and had access to the Internet. An important scientific extension would be to study color semantics in different cultures (similar to how [Reinecke and Gajos 2014] study aesthetics of low level color features). Our methodology could easily be generalized to uncover how different communities might agree or disagree on the meanings of colors. There is already extensive and comprehensive research on color naming across cultures [Kay et al. 2009], and we argue that investigating agreement on color semantics is a natural next step. Interestingly, we were contacted by participants who had color vision deficiency but still wished to perform the task. While we did not include those participants in this study, various color blind communities might have different conceptions of color semantics, and there are clearly applications for design accessibility (e.g. [Flatla et al. 2013]).

Finally, we presented a number of applications for color semantics to illustrate how it can enable more meaningful user interactions, and perhaps help non-designers generate more creative and appealing designs. We specifically demonstrated color palette selection, design example recommendation, image retrieval, color region selection in images, and pattern recoloring. We demonstrated an initial pass at instantiating these applications, which we hope to develop and test in terms of user experience and performance. Future work will hopefully incorporate these applications into current design creation tools and extend this first step towards a broader goal of making design accessible to the general public.

ACKNOWLEDGMENTS

We would like to sincerely thank Editor Jeffrey Nichols for his invaluable guidance and the anonymous reviewers for their feedback; this help has significantly improved the work and its presentation. We would also like to thank Sergey Kirshner for help with deriving the model; Ruth Rosenholtz for providing her expert advice and extensive review; Kalyan Veeramachaneni and Una-May O'Reilly for their support; Phillip Isola, Jeffrey Heer, Jodi Forlizzi, Jean-Daniel Fekete, and Petronio Bendito for helpful review and discussion on both the technical and presentation aspects of the project; and Sharon Lin for sharing code and recoloring some of the images.

APPENDIX

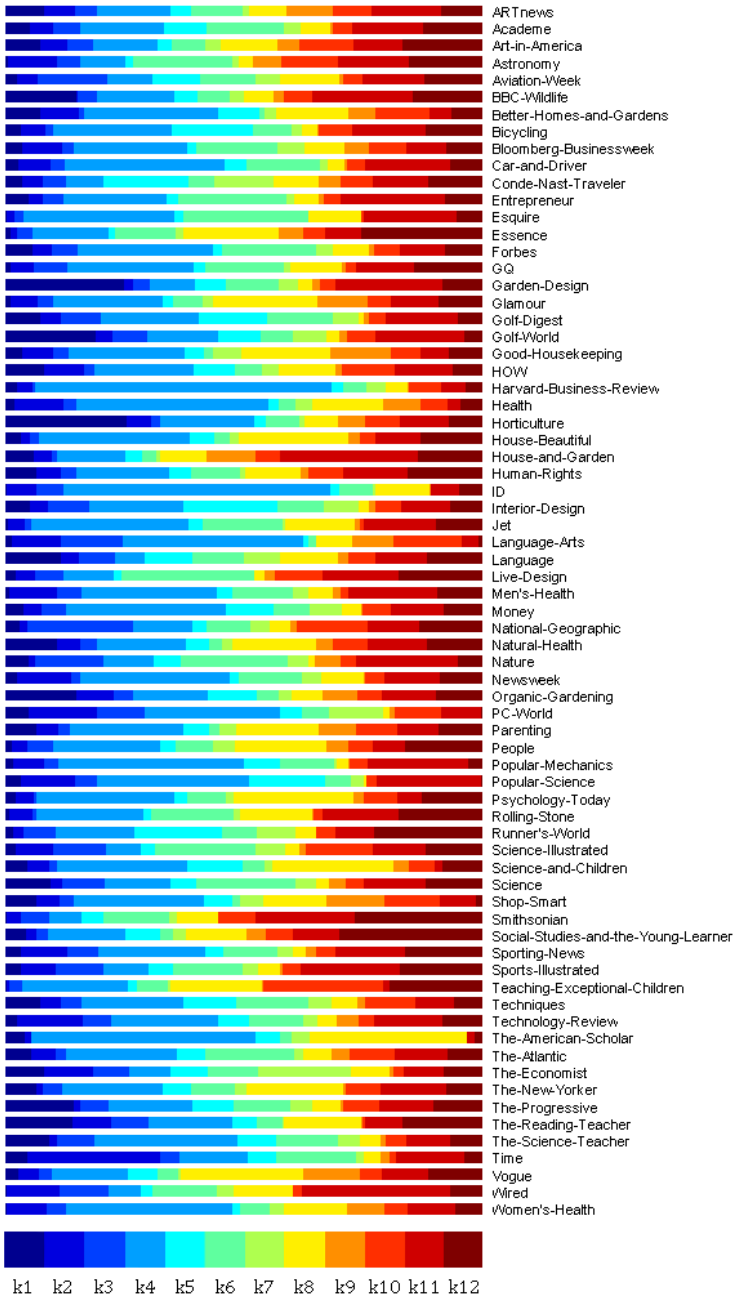


Fig. 16: Topics vs. titles. The proportion of each of the 12 color-word topics, k_1 to k_{12} (see Fig. 5 and Fig. 8) for each magazine title including all the issues in the dataset (see Table III) is illustrated. Note that the colors here are just legends for the purpose of visualization, and are not related to the color-word topics.

Table II: Magazine titles' proportions in the color-work topics.

k_1	*prop*	k_2	*prop*	k_3	*prop*
Horticulture	0.0684	Time	0.0693	National Geographic	0.0609
Garden Design	0.0670	PC World	0.0361	Aviation Week	0.0404
Golf World	0.0513	Technology Review	0.0341	Nature	0.0395
BBC Wildlife	0.0401	Newsweek	0.0283	Language Arts	0.0355
Organic Gardening	0.0399	Popular Science	0.0280	Science Illustrated	0.0302
The Progressive	0.0386	Wired	0.0277	Wired	0.0282
The Reading Teacher	0.0380	Language Arts	0.0259	Sports Illustrated	0.0280
Language	0.0310	The Economist	0.0258	PC World	0.0268
Natural Health	0.0293	Astronomy	0.0258	Science	0.0249
Science	0.0254	Health	0.0251	Interior Design	0.0236
k_4	*prop*	k_5	*prop*	k_6	*prop*
Harvard Business Review	0.0392	Interior Design	0.0465	Live Design	0.0376
ID	0.0353	Runner's World	0.0431	Esquire	0.0351
The American Scholar	0.0297	Conde Nast Traveler	0.0419	Entrepreneur	0.0306
Health	0.0253	Bicycling	0.0403	Nature	0.0304
Popular Mechanics	0.0245	Popular Science	0.0374	Science Illustrated	0.0283
Language Arts	0.0239	Golf Digest	0.0344	Astronomy	0.0279
Women's Health	0.0220	Science and Children	0.0275	Science	0.0279
Car and Driver	0.0211	Techniques	0.0261	Forbes	0.0265
Money	0.0211	Organic Gardening	0.0244	The Atlantic	0.0251
Jet	0.0208	Aviation Week	0.0240	Rolling Stone	0.0236
k_7	*prop*	k_8	*prop*	k_9	*prop*
The Economist	0.0780	The American Scholar	0.0468	Good Housekeeping	0.0589
Conde Nast Traveler	0.0493	Vogue	0.0362	Shop Smart	0.0555
PC World	0.0463	Science and Children	0.0361	Vogue	0.0544
Runner's World	0.0323	Psychology Today	0.0356	Glamour	0.0484
Language	0.0302	House Beautiful	0.0325	House and Garden	0.0470
Interior Design	0.0296	Glamour	0.0311	ARTnews	0.0440
Money	0.0275	Essence	0.0286	Language Arts	0.0398
Golf World	0.0274	The New Yorker	0.0284	Health	0.0363
Science Illustrated	0.0250	Teaching Exceptional Children	0.0272	Women's Health	0.0357
Good Housekeeping	0.0232	People	0.0272	Parenting	0.0349
k_{10}	*prop*	k_{11}	*prop*	k_{12}	*prop*
Teaching Exceptional Children	0.0588	Wired	0.0355	Social Studies and the Young Le	0.0409
National Geographic	0.0345	House and Garden	0.0332	Smithsonian	0.0368
Science Illustrated	0.0331	Garden Design	0.0256	Essence	0.0348
Language Arts	0.0331	Entrepreneur	0.0250	Runner's World	0.0310
Astronomy	0.0281	Popular Science	0.0249	Teaching Exceptional Children	0.0268
Shop Smart	0.0273	Nature	0.0243	Rolling Stone	0.0242
Better Homes and Gardens	0.0266	Popular Mechanics	0.0240	Live Design	0.0240
Art in America	0.0264	BBC Wildlife	0.0239	Sports Illustrated	0.0237
HOW	0.0261	Sports Illustrated	0.0237	Art in America	0.0231
Techniques	0.0246	Smithsonian	0.0235	The Reading Teacher	0.0229

Proportions (denoted by *prop*) of magazine titles in the color-work topics ($k_1, k_2, ..., k_{12}$). Only the top 10 titles in each color-word topic are shown.

Table III: Summary of Our Magazine Covers Dataset.

Art		Business		Education	
Magazine Title	**# collected**	**Magazine Title**	**# collected**	**Magazine Title**	**# collected**
ARTNews	50	Entrepreneur	52	Social Studies and the	50
Interior Design	50	Bloomberg Businessweek	50	The Science Teacher	50
The New Yorker	50	Forbes	50	Techniques	48
Art in America	49	Harvard Business Review	50	Academe	40
HOW	41	Money	48	Language	36
Live Design	9	The Economist	29	The American scholar	3
ID	1			The Reading Teacher	3
				Language Arts	2
				Teaching Exceptional Children	2
Art Total:	250	Business Total:	279	Education Total:	234
Entertainment		**Family**		**Fashion**	
Magazine Title	**# collected**	**Magazine Title**	**# collected**	**Magazine Title**	**# collected**
Conde Nast Traveler	50	Good Housekeeping	61	Essence	50
Jet	50	Parenting	51	Glamour	50
People	50	House Beautiful	50	GQ	50
Rolling Stone	50	ShopSmart	40	Vogue	50
National Geographic	44	Better Homes and Gardens	30	Esquire	11
Entertainment Total:	244	Family Total:	232	Fashion Total:	211
Health		**Nature**		**Politics**	
Magazine Title	**# collected**	**Magazine Title**	**# collected**	**Magazine Title**	**# collected**
Men's Health	50	Garden Design	35	Newsweek	50
Women's Health	50	BBC Wildlife	25	Time	50
Health	17	Organic Gardening	17	Human Rights	42
Natural Health	17	House & Garden	15	The Atlantic	25
				The Progressive	22
Health Total:	134	Nature Total:	92	Politics Total:	189
Science		**Sports**		**Technology**	
Magazine Title	**# collected**	**Magazine Title**	**# collected**	**Magazine Title**	**# collected**
Science	119	Sports Illustrated	50	PC World	50
Nature	50	Sporting News	44	Aviation Week	45
Smithsonian	38	Car and Driver	29	Wired	45
Science Illustrated	30	Golf Digest	26	Popular Mechanics	34
Popular Science	29	Golf World	24	Technology Review	17
Astronomy	10	Bicycling	22		
Science and Children	5	Runner's World	22		
Science Total:	281	Sports Total:	217	Technology Total:	191

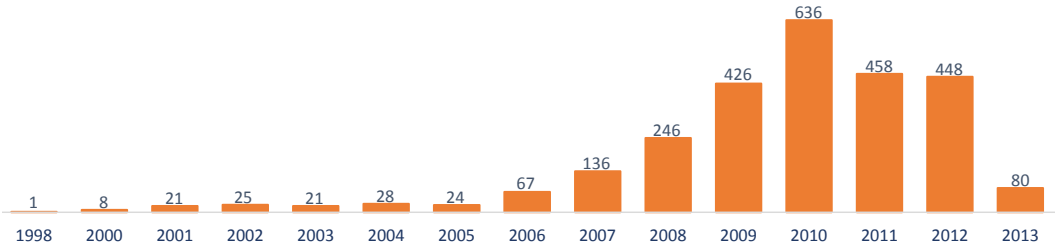
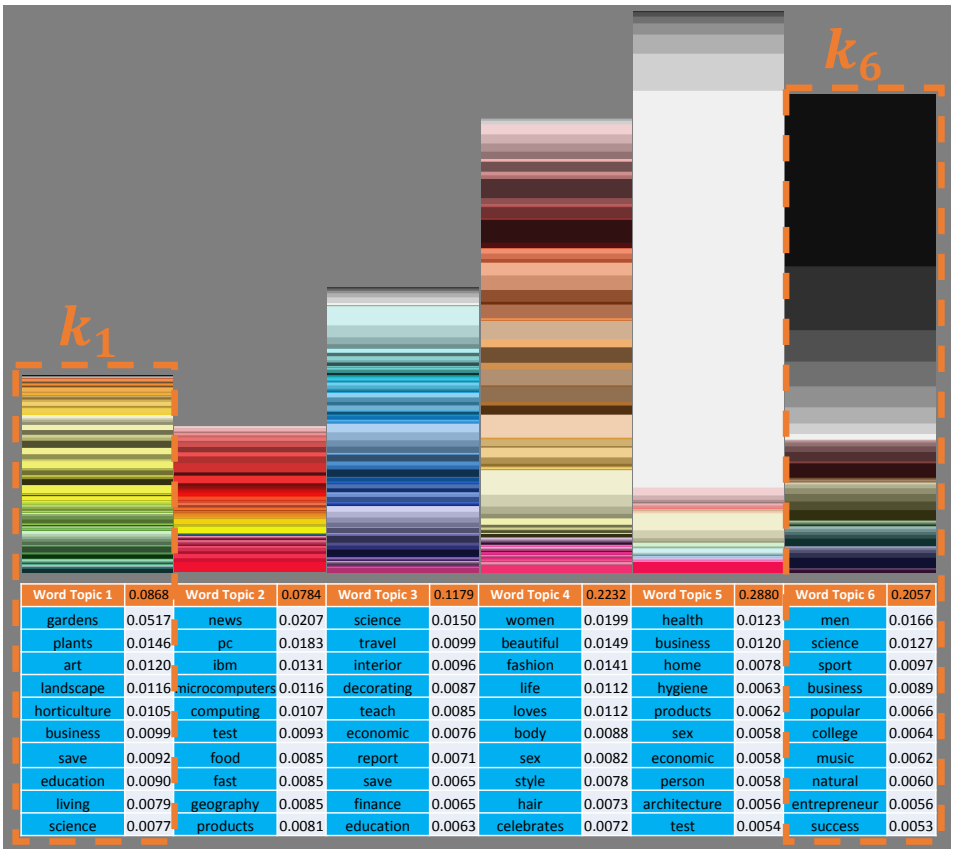


Fig. 17: Histogram of the number of collected magazine covers per year.

Table IV: Handcrafted Stop Word List.

amazing	britain	easy	great-britain	johns	minutes	rated	special	tells	ups
america	canada	essence	green	julie	mitt	real	spring	things	wanted
american	change	exclusive	grows	kate	month	red	states	tips	ways
annual	china	eye	guide	klein	nation	reveals	steve	today	week
autumn	colors	faces	i	lost	needed	romney	stop	top	white
avoiding	cte	falling	inside	makeover	obama	ryan	stories	trick	winning
awards	cutting	frances	issue	making	ons	secret	stuff	u.s.	work
back	day	free	italy	matter	pages	share	summer	ultimate	world
bad	design	good	japan	meaning	perfect	shows	super	undos	year
big	dos	great	jennifer	meet	picks	simple	takes	united	
black	double	great	joe	minute	preview	small	talking	united-states	

Note that these words are inspected and excluded by manually visiting the first 30 words in the word topics inferred by the model.



Word Topic 1	0.0868	Word Topic 2	0.0784	Word Topic 3	0.1179	Word Topic 4	0.2232	Word Topic 5	0.2880	Word Topic 6	0.2057
gardens	0.0517	news	0.0207	science	0.0150	women	0.0199	health	0.0123	men	0.0166
plants	0.0146	pc	0.0183	travel	0.0099	beautiful	0.0149	business	0.0120	science	0.0127
art	0.0120	ibm	0.0131	interior	0.0096	fashion	0.0141	home	0.0078	sport	0.0097
landscape	0.0116	microcomputers	0.0116	decorating	0.0087	life	0.0112	hygiene	0.0063	business	0.0089
horticulture	0.0105	computing	0.0107	teach	0.0085	loves	0.0112	products	0.0062	popular	0.0066
business	0.0099	test	0.0093	economic	0.0076	body	0.0088	sex	0.0058	college	0.0064
save	0.0092	food	0.0085	report	0.0071	sex	0.0082	economic	0.0058	music	0.0062
education	0.0090	fast	0.0085	save	0.0065	style	0.0078	person	0.0058	natural	0.0060
living	0.0079	geography	0.0085	finance	0.0065	hair	0.0073	architecture	0.0056	entrepreneur	0.0056
science	0.0077	products	0.0081	education	0.0063	celebrates	0.0072	test	0.0054	success	0.0053

Fig. 18: Color-word topics inferred by the LDA-dual model. Illustration of the 6 color topics. Note that for visualization, only the principal elements in the histograms are shown. Also note that the numerical weight of each word topic is shown next to heading of each word topic histogram.

(a)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.21	0.12	1.69	0.29	0.32	0.48	0.28	0.35	0.23	0.77	0.50	0.19
c_2	0.44	0.28	0.40	0.65	1.21	0.38	0.31	1.90	1.87	0.90	0.62	1.50
c_3	0.13	1.31	1.57	0.92	0.94	1.18	1.41	0.56	0.34	0.74	0.94	0.32
c_4	0.05	1.80	1.12	1.36	0.87	1.32	1.59	0.20	0.16	0.50	1.41	0.18
c_5	0.27	1.11	1.41	1.20	1.47	0.93	1.69	0.48	0.48	0.88	1.21	0.30
c_6	0.60	1.00	1.01	0.78	0.51	1.92	1.08	0.24	0.10	0.67	1.12	0.27
c_7	0.10	1.01	1.29	0.84	1.36	0.94	1.54	0.25	0.54	0.42	0.80	0.43
c_8	0.72	0.28	0.56	0.51	1.01	0.18	0.21	1.76	1.27	1.14	0.38	1.21
c_9	0.36	0.06	0.00	0.29	1.22	0.13	0.06	2.09	2.11	0.63	0.19	1.03
c_{10}	1.04	0.28	0.67	0.52	1.80	0.33	0.16	1.33	0.95	1.44	0.66	1.25
c_{11}	0.27	1.71	1.01	1.48	0.55	1.11	1.80	0.11	0.33	0.76	1.49	0.45
c_{12}	0.85	0.43	1.73	1.02	0.86	1.34	0.61	0.22	0.63	1.17	1.04	0.52

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.29	0.31	1.52	0.41	1.04	0.57	0.70	0.32	0.17	0.96	0.38	0.55
c_2	0.06	1.16	0.67	0.95	0.67	0.38	0.47	0.81	1.38	0.67	0.74	1.04
c_3	0.32	1.23	1.17	1.19	1.35	1.03	1.40	0.63	0.72	0.92	1.33	0.43
c_4	0.16	1.31	0.94	1.37	1.18	0.91	1.52	0.27	0.53	0.86	1.07	0.45
c_5	0.66	0.55	1.74	1.08	1.23	0.95	0.96	0.56	0.43	1.26	1.25	0.70
c_6	0.18	1.68	1.13	0.85	0.42	1.82	1.27	0.13	0.28	0.42	1.26	0.22
c_7	0.00	1.11	1.02	1.29	1.08	1.19	1.34	0.24	0.72	0.81	1.28	0.26
c_8	1.21	0.92	2.04	0.93	1.23	0.82	0.96	0.56	0.70	1.29	0.95	0.76
c_9	0.36	0.07	0.18	0.31	1.00	0.07	0.08	1.53	2.41	0.44	0.28	0.90
c_{10}	0.78	0.45	0.62	0.66	1.92	0.52	0.50	0.79	0.79	1.29	0.72	0.94
c_{11}	0.07	1.81	0.81	1.12	0.54	1.17	1.38	0.07	0.00	0.62	1.52	0.00
c_{12}	0.59	0.49	1.17	0.77	0.77	0.58	0.76	0.78	1.10	0.76	0.41	1.10

(b)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.07	0.60	0.95	0.45	0.82	0.28	0.24	0.38	0.36	0.86	0.26	0.43
c_2	0.77	0.27	0.74	0.59	1.14	0.39	0.42	1.67	1.84	0.37	0.70	1.43
c_3	0.15	1.15	1.48	0.73	0.78	0.56	0.99	0.47	0.18	0.61	1.16	0.37
c_4	0.00	1.48	1.20	1.21	0.50	0.99	1.72	0.00	0.43	0.80	1.36	0.37
c_5	0.58	1.39	1.50	0.97	1.42	1.25	1.54	0.35	0.10	1.05	1.05	0.80
c_6	0.88	0.97	1.25	0.69	0.58	1.73	0.84	0.30	0.09	0.53	0.92	0.61
c_7	0.18	1.69	1.09	1.25	0.80	1.10	1.75	0.37	0.42	0.36	1.40	0.54
c_8	0.23	0.39	0.66	0.29	2.09	0.66	0.54	1.82	1.37	0.89	0.58	0.83
c_9	0.54	0.00	0.19	0.32	1.42	0.00	0.17	1.71	2.00	0.64	0.53	1.51
c_{10}	1.06	0.21	0.80	0.67	1.90	0.35	0.22	1.81	1.13	1.41	0.55	1.06
c_{11}	0.33	1.61	0.74	1.30	0.28	1.33	1.26	0.39	0.08	0.71	1.21	0.26
c_{12}	1.19	0.56	1.28	0.84	0.88	1.82	0.67	0.81	0.62	0.97	1.20	0.67

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.09	0.00	1.23	0.63	1.12	0.22	0.30	1.00	0.44	1.11	0.73	0.19
c_2	0.00	0.59	0.93	0.47	0.92	0.65	0.27	1.65	2.13	0.34	1.13	1.12
c_3	0.31	0.95	1.23	0.62	1.50	0.77	1.16	0.54	0.55	0.53	1.58	0.59
c_4	0.27	1.01	0.86	1.18	0.46	1.52	1.50	0.33	0.26	0.51	0.63	1.08
c_5	0.63	1.44	1.33	0.57	1.49	0.53	0.35	0.33	0.95	0.85	1.69	0.80
c_6	0.21	0.81	1.24	0.70	0.53	1.78	1.30	0.30	0.15	0.79	1.91	0.64
c_7	0.21	1.43	0.78	1.07	0.83	1.61	1.46	0.27	0.66	0.59	1.09	0.48
c_8	1.24	0.67	1.34	0.83	1.93	1.14	0.44	0.73	0.52	1.22	0.50	1.14
c_9	0.00	0.16	0.00	0.35	0.74	0.12	0.21	1.85	2.44	0.58	0.00	1.72
c_{10}	0.66	0.00	0.56	0.80	2.10	0.00	0.35	1.41	0.89	1.29	0.42	1.15
c_{11}	0.13	1.18	1.21	1.08	0.73	1.24	1.44	0.00	0.00	0.66	1.25	0.32
c_{12}	0.36	0.28	1.29	1.31	1.03	0.88	0.30	1.13	0.50	1.39	0.69	1.12

(c)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	1.97	0.70	1.30	0.50	0.30	0.35	0.24	0.44	0.40	0.97	0.34	0.17
c_2	0.65	0.28	0.76	0.49	1.27	0.20	0.44	1.77	1.73	0.83	0.58	1.74
c_3	0.12	1.17	1.46	0.79	0.91	0.74	1.19	0.74	0.39	0.94	0.87	0.52
c_4	0.00	1.69	1.25	1.15	0.95	1.09	1.54	0.08	0.35	1.04	1.34	0.22
c_5	0.47	0.84	1.55	1.15	1.43	1.15	1.46	0.36	0.24	1.25	1.23	0.75
c_6	0.81	1.15	0.93	0.70	0.65	1.74	0.89	0.29	0.12	0.66	0.74	0.79
c_7	0.15	1.27	1.05	1.11	1.20	0.79	1.84	0.43	0.40	0.35	1.32	0.53
c_8	0.09	0.27	0.92	0.43	1.54	0.72	0.39	1.93	0.98	1.03	0.53	1.28
c_9	0.58	0.00	0.10	0.20	1.43	0.11	0.20	1.88	1.93	0.51	0.47	1.24
c_{10}	0.60	0.42	0.81	0.66	1.71	0.49	0.00	1.46	0.97	1.40	0.53	0.92
c_{11}	0.45	1.71	0.65	1.31	0.53	1.41	1.02	0.38	0.43	0.65	1.24	0.33
c_{12}	0.90	0.51	1.26	1.02	0.55	1.50	0.73	0.66	0.39	1.24	0.94	0.59

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.43	0.48	1.44	0.38	1.10	0.26	0.50	0.57	0.18	0.90	0.30	0.48
c_2	0.00	0.80	0.91	1.13	0.92	0.50	0.33	1.28	1.96	0.45	0.67	1.34
c_3	0.25	0.91	1.43	0.60	1.33	0.85	1.74	0.16	0.42	0.87	1.64	0.29
c_4	0.49	1.47	0.73	1.26	0.75	0.95	1.60	0.18	0.32	1.10	0.61	0.83
c_5	0.41	1.15	1.69	0.72	1.27	1.42	0.44	0.57	0.44	0.72	2.14	0.77
c_6	0.24	1.16	1.20	0.29	0.77	1.93	1.54	0.21	0.00	0.57	1.35	0.29
c_7	0.15	1.18	0.44	1.53	0.84	1.66	1.50	0.19	0.95	0.80	1.18	0.64
c_8	1.03	0.89	1.42	0.83	1.44	1.38	0.42	0.70	0.00	1.37	0.75	0.85
c_9	0.00	0.15	0.41	0.42	1.29	0.14	0.57	1.92	2.34	0.48	0.16	1.08
c_{10}	0.90	0.29	0.51	0.20	2.39	0.15	0.36	1.25	1.53	1.40	0.28	0.41
c_{11}	0.00	1.27	1.14	1.02	0.41	1.14	1.37	0.00	0.00	0.86	1.31	0.00
c_{12}	0.44	0.40	1.28	1.21	0.32	0.45	1.14	0.43	1.36	0.46	0.78	1.00

(d)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.25	0.15	1.44	0.34	0.53	0.40	0.28	0.35	0.22	0.71	0.45	0.38
c_2	0.53	0.27	0.45	0.69	1.14	0.45	0.32	1.81	1.94	0.64	0.73	1.28
c_3	0.15	1.29	1.58	0.88	0.85	0.98	1.17	0.40	0.21	0.56	1.17	0.22
c_4	0.05	1.57	1.11	1.35	0.58	1.23	1.70	0.15	0.20	0.35	1.44	0.29
c_5	0.34	1.48	1.40	1.06	1.47	0.99	1.67	0.46	0.45	0.77	1.09	0.41
c_6	0.68	0.92	1.19	0.75	0.47	1.89	1.03	0.25	0.09	0.58	1.22	0.23
c_7	0.12	1.32	1.31	0.94	1.07	1.13	1.52	0.20	0.52	0.41	0.88	0.46
c_8	0.75	0.38	0.37	0.40	1.46	0.25	0.31	1.71	1.50	1.03	0.45	0.95
c_9	0.37	0.05	0.05	0.36	1.22	0.05	0.05	2.04	2.15	0.69	0.24	1.19
c_{10}	1.33	0.15	0.68	0.53	1.90	0.22	0.24	1.57	1.01	1.45	0.64	1.36
c_{11}	0.21	1.66	1.03	1.47	0.46	1.14	1.78	0.17	0.10	0.79	1.45	0.40
c_{12}	1.03	0.49	1.74	0.91	1.09	1.55	0.58	0.49	0.74	1.01	1.19	0.57

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.16	0.18	1.45	0.50	1.04	0.52	0.58	0.49	0.33	1.11	0.60	0.44
c_2	0.06	1.02	0.73	0.74	0.69	0.44	0.43	1.04	1.46	0.64	0.94	0.97
c_3	0.32	1.25	1.09	1.21	1.44	0.98	1.17	0.79	0.79	0.85	1.31	0.56
c_4	0.12	1.11	0.97	1.32	1.04	1.27	1.49	0.33	0.47	0.64	1.06	0.54
c_5	0.72	0.83	1.52	1.03	1.32	0.70	0.88	0.47	0.59	1.19	1.13	0.73
c_6	0.18	1.40	1.14	1.04	0.33	1.76	1.21	0.17	0.29	0.53	1.42	0.42
c_7	0.00	1.20	1.12	1.13	1.12	1.31	1.34	0.27	0.62	0.71	1.20	0.30
c_8	1.28	0.83	2.02	0.90	1.41	0.79	0.92	0.59	0.80	1.21	0.84	0.89
c_9	0.34	0.07	0.06	0.28	0.74	0.07	0.00	1.57	2.45	0.47	0.19	1.24
c_{10}	0.68	0.37	0.64	0.85	1.81	0.45	0.49	0.84	0.58	1.26	0.77	1.19
c_{11}	0.13	1.73	0.93	1.14	0.62	1.21	1.42	0.06	0.00	0.52	1.47	0.14
c_{12}	0.55	0.43	1.18	0.87	1.10	0.71	0.44	1.12	0.86	1.02	0.39	1.13

(e)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.11	0.40	1.40	0.31	0.53	0.34	0.29	0.41	0.27	0.84	0.41	0.30
c_2	0.61	0.30	0.49	0.71	1.21	0.34	0.39	1.80	1.97	0.75	0.62	1.41
c_3	0.16	1.24	1.51	0.82	0.97	0.97	1.05	0.50	0.28	0.58	1.04	0.33
c_4	0.04	1.66	1.14	1.26	0.65	1.20	1.70	0.12	0.28	0.65	1.47	0.35
c_5	0.43	1.29	1.52	1.11	1.45	1.10	1.67	0.42	0.38	0.93	1.07	0.52
c_6	0.74	1.02	1.06	0.66	0.57	1.87	0.92	0.23	0.12	0.60	1.11	0.43
c_7	0.13	1.34	1.24	1.00	1.04	0.97	1.67	0.30	0.44	0.45	0.99	0.41
c_8	0.49	0.30	0.53	0.47	1.35	0.42	0.34	1.80	1.39	1.03	0.47	1.00
c_9	0.48	0.04	0.09	0.28	1.29	0.08	0.12	1.99	2.04	0.46	0.41	1.25
c_{10}	1.00	0.28	0.69	0.53	1.81	0.35	0.22	1.53	1.09	1.42	0.59	1.15
c_{11}	0.28	1.66	0.96	1.33	0.39	1.17	1.58	0.26	0.27	0.69	1.40	0.42
c_{12}	0.96	0.55	1.53	0.99	0.82	1.61	0.52	0.61	0.65	1.11	1.10	0.57

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.23	0.26	1.52	0.44	1.08	0.43	0.36	0.47	0.23	0.99	0.46	0.38
c_2	0.00	0.99	0.83	0.76	0.80	0.40	0.44	1.13	1.55	0.54	0.95	0.94
c_3	0.25	1.17	1.18	1.01	1.42	0.82	1.34	0.64	0.71	0.82	1.43	0.46
c_4	0.12	1.16	0.98	1.28	0.97	1.15	1.48	0.30	0.40	0.67	0.88	0.58
c_5	0.70	0.85	1.60	1.00	1.33	0.95	0.81	0.64	0.56	0.94	1.55	0.73
c_6	0.12	1.44	1.15	0.87	0.46	1.83	1.39	0.15	0.25	0.54	1.54	0.45
c_7	0.06	1.15	0.93	1.13	1.07	1.46	1.36	0.29	0.83	0.89	1.12	0.45
c_8	1.32	0.78	1.75	0.92	1.52	1.02	0.69	0.64	0.69	1.25	0.79	1.00
c_9	0.30	0.11	0.10	0.36	0.82	0.11	0.17	1.62	2.40	0.47	0.10	1.17
c_{10}	0.78	0.44	0.54	0.67	1.99	0.43	0.38	0.95	0.74	1.23	0.61	0.92
c_{11}	0.12	1.56	0.86	1.08	0.52	1.29	1.27	0.00	0.00	0.57	1.43	0.00
c_{12}	0.54	0.45	1.13	1.06	0.95	0.62	0.63	1.01	1.07	0.90	0.50	1.11

(f)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.37	0.00	1.18	0.71	0.23	0.65	0.00	0.26	0.46	0.63	0.20	0.35
c_2	0.42	0.18	1.56	0.34	0.91	0.45	0.00	1.74	1.43	0.30	1.13	1.74
c_3	0.00	1.31	1.60	0.96	0.25	0.35	1.91	0.69	0.23	1.20	1.15	0.40
c_4	0.00	1.36	1.28	1.47	0.96	1.20	1.45	0.17	0.19	0.62	0.83	0.00
c_5	0.20	1.18	0.94	0.81	1.47	0.69	1.52	0.46	0.24	0.67	1.51	0.65
c_6	0.72	0.82	1.28	1.13	0.43	1.70	1.04	0.46	0.00	0.68	0.85	0.36
c_7	0.14	1.25	1.13	0.89	1.47	1.13	1.39	0.23	0.60	0.17	1.21	0.91
c_8	0.55	0.52	0.89	0.17	2.77	0.28	0.37	1.73	0.93	0.70	0.40	1.39
c_9	0.26	0.00	0.00	0.43	1.31	0.00	0.00	1.71	2.25	1.21	0.00	0.81
c_{10}	1.36	0.00	0.81	0.90	1.87	0.30	0.00	1.60	0.75	1.48	0.77	1.55
c_{11}	0.42	1.80	0.39	1.82	0.93	1.57	1.37	0.00	0.00	0.86	1.24	0.20
c_{12}	1.06	0.00	1.91	0.88	1.22	0.83	1.08	0.28	0.60	1.09	1.22	0.61

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.21	0.19	1.00	0.69	0.98	0.64	1.95	0.73	0.67	1.10	0.45	0.68
c_2	0.24	0.89	0.55	0.94	0.56	0.66	0.27	0.87	1.78	0.90	0.42	1.53
c_3	0.81	1.13	1.17	1.24	1.31	1.32	1.18	0.54	0.40	0.99	1.20	0.62
c_4	0.57	1.37	0.67	1.39	0.94	1.41	1.89	0.00	0.54	1.19	1.29	0.90
c_5	0.39	1.20	1.18	0.00	1.21	0.41	0.81	0.00	0.34	1.83	0.58	1.10
c_6	0.60	0.81	1.19	0.60	0.26	1.77	0.37	0.31	0.22	0.57	0.78	0.00
c_7	0.00	1.37	1.10	1.81	0.93	0.45	1.47	0.00	0.24	0.24	1.54	0.00
c_8	0.88	1.17	1.90	0.81	0.95	0.00	1.19	0.53	0.40	1.24	0.98	0.61
c_9	0.16	0.00	0.35	0.20	1.36	0.00	0.00	1.71	2.50	0.55	0.82	1.35
c_{10}	0.33	0.00	0.63	0.88	1.65	0.00	0.80	1.13	0.94	1.57	0.42	1.25
c_{11}	0.00	1.79	1.94	1.29	0.63	0.90	1.88	0.19	0.00	0.77	1.42	0.31
c_{12}	0.50	0.27	1.59	0.48	0.53	0.68	0.55	0.31	0.67	1.01	0.42	1.13

Fig. 19: Relevance matrices $\hat{R}^1$ and $\hat{R}^2$ for Experiment I, for the first (left) and second (right) set of questions, respectively, computed for: (a) females, (b) males, (c) non-U.S. participants, (d) U.S. participants, (e) non-designers, and (f) designers.

(a)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.04	0.40	0.15	0.28	0.00	0.26	0.12	0.31	0.23	0.70	0.13	0.11
c_2	0.30	0.58	1.52	1.93	1.88	1.07	1.04	0.64	0.36	0.59	1.47	0.67
c_3	2.02	0.21	1.37	0.76	0.75	0.76	0.67	0.77	0.28	0.70	0.45	0.58
c_4	0.97	0.61	1.49	1.27	1.33	0.60	0.79	0.68	0.50	0.67	0.80	0.55
c_5	1.16	0.79	0.74	0.93	1.40	0.32	1.00	1.19	0.91	1.71	0.44	0.32
c_6	1.38	0.60	1.35	1.14	1.02	1.39	1.34	0.35	0.21	0.70	1.15	0.56
c_7	1.04	0.37	0.85	1.44	1.22	1.03	1.68	0.29	0.27	0.11	1.51	0.72
c_8	0.30	1.51	0.42	0.56	0.73	0.10	0.33	1.84	1.55	1.54	0.20	0.23
c_9	0.09	1.07	0.82	0.68	0.46	0.00	0.83	1.05	1.69	1.11	0.34	0.73
c_{10}	0.77	0.42	1.00	0.38	1.09	0.11	0.42	0.84	0.84	1.96	0.17	0.45
c_{11}	1.29	0.57	1.36	0.70	1.02	0.77	1.11	0.34	0.32	0.75	1.57	0.76
c_{12}	0.67	1.16	0.76	0.38	1.46	0.24	0.57	1.89	1.81	1.84	0.77	0.69

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.03	0.14	0.37	0.00	0.35	0.16	0.20	0.50	0.14	0.74	0.36	0.37
c_2	0.35	1.12	1.12	1.07	0.80	1.10	1.60	0.57	0.14	0.94	0.95	0.65
c_3	1.99	0.28	0.79	0.98	1.09	0.63	1.48	1.42	0.31	1.17	0.72	1.40
c_4	1.72	0.64	0.78	1.24	1.46	1.32	1.02	1.31	0.14	0.83	1.11	0.93
c_5	1.25	0.43	0.51	0.97	1.14	0.70	0.90	0.47	1.01	1.90	0.36	1.36
c_6	0.85	0.26	1.22	1.04	1.01	1.86	0.98	0.95	0.16	0.27	1.16	0.52
c_7	0.68	0.37	1.51	0.96	1.16	1.29	1.19	0.84	0.19	0.35	1.52	0.38
c_8	0.54	1.17	0.79	0.76	0.86	0.00	0.87	0.65	1.88	1.20	0.00	1.60
c_9	1.07	1.22	0.85	0.50	1.06	0.66	0.00	0.15	2.31	0.93	0.13	0.49
c_{10}	2.01	1.10	0.54	0.51	1.21	0.29	0.42	1.81	0.76	1.41	0.00	0.75
c_{11}	1.04	0.88	1.15	0.69	1.15	1.29	0.83	0.45	0.22	1.43	1.28	0.78
c_{12}	0.52	0.54	1.41	0.28	1.31	0.78	0.61	0.40	1.52	1.56	0.40	1.10

(b)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.13	0.58	0.13	0.14	0.00	0.49	0.00	0.59	0.00	0.34	0.00	0.74
c_2	0.29	0.70	1.36	1.60	1.53	1.28	1.11	1.08	0.51	0.25	1.78	0.66
c_3	1.95	0.45	1.69	0.18	0.90	1.00	1.50	0.61	0.00	0.63	0.60	1.07
c_4	1.15	0.46	0.67	1.21	0.56	0.82	1.26	0.99	0.21	0.80	1.13	0.57
c_5	1.01	1.28	0.72	0.97	1.14	0.47	1.15	1.56	1.23	1.88	0.43	0.73
c_6	1.13	0.50	0.80	0.94	1.27	1.60	1.04	0.75	0.00	0.52	1.32	1.47
c_7	0.40	0.48	1.14	0.94	1.17	0.33	1.57	0.52	0.00	1.19	1.64	0.89
c_8	0.33	1.35	0.44	0.31	0.62	0.20	0.50	1.46	2.34	1.43	0.00	0.26
c_9	0.00	1.48	0.69	0.22	0.00	0.18	0.35	0.49	1.94	1.25	0.32	0.16
c_{10}	0.92	0.59	1.06	0.69	0.21	0.66	0.73	0.38	0.91	1.45	0.46	0.57
c_{11}	1.48	0.58	1.24	0.73	0.91	1.25	1.59	0.58	0.00	0.36	1.65	1.22
c_{12}	0.61	1.61	1.30	0.97	0.60	0.50	0.37	1.09	1.65	0.99	0.28	0.72

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.26	0.00	0.69	0.00	0.27	0.40	0.25	0.00	0.00	0.00	0.00	0.32
c_2	0.00	0.98	1.26	2.25	1.00	0.00	1.63	0.45	0.32	0.00	1.46	0.00
c_3	2.26	0.67	1.21	0.76	1.06	1.96	0.40	0.00	0.64	1.05	1.04	0.00
c_4	1.45	1.25	1.05	0.86	1.70	1.20	1.10	0.00	0.36	1.11	1.02	1.20
c_5	1.84	0.45	0.00	1.13	1.32	0.43	0.74	1.00	1.80	2.03	0.69	0.55
c_6	0.89	0.97	1.57	0.59	1.47	1.66	1.11	0.35	0.00	1.98	1.27	0.24
c_7	0.69	0.68	0.62	1.43	0.99	0.71	1.44	0.61	0.31	0.30	0.92	0.25
c_8	0.78	2.16	0.51	0.24	0.35	0.27	1.04	0.67	2.79	0.91	0.00	0.92
c_9	1.01	1.75	0.00	0.68	0.92	0.00	0.27	0.00	2.28	1.20	0.38	0.51
c_{10}	2.18	0.78	0.32	0.91	0.68	0.94	1.43	0.54	0.82	1.40	0.47	0.85
c_{11}	1.89	1.60	0.90	0.63	1.21	0.40	0.95	0.92	0.61	0.97	1.57	0.35
c_{12}	0.41	0.60	1.09	0.30	0.81	0.97	0.28	0.68	1.39	2.21	1.08	1.07

(c)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	1.88	0.67	0.00	0.25	0.00	0.76	0.00	0.27	0.30	0.65	0.00	0.23
c_2	0.42	1.04	1.03	1.15	2.21	1.53	0.53	0.63	0.23	0.59	1.37	0.26
c_3	2.55	0.57	1.63	0.46	0.22	0.00	2.00	0.00	0.18	0.62	0.16	0.60
c_4	0.86	0.91	1.03	0.96	1.51	0.00	1.58	0.96	0.63	0.77	1.03	0.30
c_5	1.48	0.89	0.60	0.25	1.35	0.43	0.97	0.72	1.00	1.98	1.33	0.40
c_6	1.08	0.82	0.58	0.75	1.40	1.39	2.59	0.99	0.22	0.46	1.18	1.42
c_7	0.93	0.63	0.00	0.32	0.88	0.60	1.75	0.59	0.41	0.45	1.84	0.70
c_8	0.27	1.26	0.82	0.24	0.52	0.20	0.00	1.67	2.26	1.65	0.00	0.46
c_9	0.00	1.00	0.87	0.00	0.50	0.00	0.65	0.27	2.00	1.37	0.70	0.33
c_{10}	0.53	0.82	1.13	0.74	1.27	0.21	0.00	0.66	0.84	1.71	0.00	1.21
c_{11}	1.69	0.49	1.36	0.71	0.72	1.58	1.21	0.00	0.22	0.76	1.45	1.48
c_{12}	1.03	1.48	1.08	0.42	0.41	0.00	0.00	0.99	1.86	2.00	1.16	0.82

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.08	0.00	0.54	0.00	1.07	0.00	0.00	0.00	0.00	0.00	0.77	0.00
c_2	0.34	1.23	0.91	0.84	1.18	0.93	1.49	0.75	0.00	0.50	1.15	0.00
c_3	2.43	0.81	1.15	0.00	1.53	1.27	1.58	0.34	0.00	1.25	0.34	0.73
c_4	2.07	0.31	1.22	1.01	2.53	1.69	1.22	1.04	0.31	0.89	1.29	2.53
c_5	1.08	0.00	0.45	2.24	1.12	0.89	0.00	0.54	1.19	1.96	0.90	0.39
c_6	0.49	0.83	1.84	1.58	0.96	1.81	0.84	1.19	0.00	0.86	0.41	0.82
c_7	0.77	0.45	0.90	1.34	0.67	0.90	1.12	0.67	0.54	0.00	2.01	0.67
c_8	0.50	1.91	0.86	1.03	0.60	0.00	0.61	0.40	1.62	1.44	0.00	1.75
c_9	0.89	0.81	0.65	0.75	1.53	0.00	0.00	0.00	1.95	1.62	0.00	0.79
c_{10}	1.67	1.15	0.45	0.67	1.20	0.90	1.34	0.67	1.53	1.30	0.00	0.54
c_{11}	1.29	0.79	1.44	0.64	1.17	0.40	0.73	1.22	0.71	2.00	1.56	0.00
c_{12}	0.00	0.64	1.32	0.35	1.75	0.61	0.00	0.96	1.77	2.37	1.20	0.76

(d)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.15	0.35	0.17	0.19	0.00	0.25	0.09	0.45	0.09	0.52	0.08	0.39
c_2	0.25	0.51	1.59	1.93	1.55	1.10	1.24	0.93	0.50	0.39	1.72	0.76
c_3	1.88	0.23	1.45	0.57	1.06	1.12	0.90	0.83	0.20	0.69	0.71	0.90
c_4	1.11	0.48	1.12	1.32	1.03	0.84	0.84	0.78	0.37	0.71	0.89	0.63
c_5	0.98	1.06	0.79	1.21	1.26	0.35	1.10	1.65	1.11	1.73	0.32	0.51
c_6	1.36	0.47	1.25	1.18	1.10	1.51	0.98	0.40	0.10	0.76	1.21	0.78
c_7	0.65	0.38	1.14	1.54	1.33	0.82	1.57	0.31	0.08	0.63	1.49	0.84
c_8	0.32	1.49	0.31	0.51	0.73	0.09	0.49	1.74	1.64	1.43	0.17	0.17
c_9	0.09	1.25	0.74	0.77	0.29	0.10	0.60	0.99	1.71	1.12	0.19	0.51
c_{10}	0.94	0.29	1.03	0.48	0.68	0.41	0.63	0.69	0.87	1.77	0.32	0.18
c_{11}	1.23	0.60	1.28	0.71	1.15	0.72	1.30	0.53	0.17	0.56	1.66	0.72
c_{12}	0.50	1.19	0.91	0.67	1.45	0.43	0.59	1.69	1.70	1.39	0.43	0.68

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.09	0.14	0.43	0.00	0.14	0.29	0.27	0.40	0.13	0.54	0.00	0.48
c_2	0.25	0.99	1.28	1.57	0.75	0.83	1.61	0.53	0.22	0.73	1.17	0.67
c_3	1.95	0.26	0.86	1.05	0.96	0.84	1.18	1.28	0.45	1.07	1.02	1.02
c_4	1.40	0.95	0.80	1.18	1.39	1.19	1.01	0.67	0.15	0.97	1.02	0.78
c_5	1.64	0.57	0.30	0.61	1.22	0.43	0.98	0.90	1.36	1.73	0.41	1.36
c_6	1.19	0.30	1.38	0.78	1.25	1.77	0.97	0.66	0.13	0.48	1.56	0.36
c_7	0.64	0.61	1.26	0.91	1.18	1.12	1.29	0.78	0.14	0.45	1.31	0.26
c_8	0.67	1.32	0.65	0.45	0.79	0.12	1.05	0.72	2.22	0.99	0.00	1.36
c_9	1.08	1.54	0.80	0.52	0.85	0.51	0.11	0.14	2.41	0.73	0.32	0.47
c_{10}	2.20	0.95	0.42	0.64	0.84	0.29	0.57	1.43	0.45	1.43	0.15	0.87
c_{11}	1.16	0.98	0.88	0.69	1.18	1.05	1.04	0.57	0.30	1.14	1.33	0.72
c_{12}	0.59	0.54	1.25	0.26	0.99	0.95	0.63	0.35	1.41	1.32	0.29	1.21

(e)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.04	0.49	0.16	0.15	0.00	0.37	0.08	0.48	0.17	0.62	0.07	0.32
c_2	0.26	0.64	1.44	1.68	1.92	1.19	1.08	0.87	0.35	0.38	1.65	0.68
c_3	1.98	0.26	1.52	0.52	0.83	1.00	1.09	0.62	0.24	0.72	0.41	0.78
c_4	1.11	0.43	1.02	1.23	1.27	0.70	1.02	0.79	0.48	0.70	1.02	0.58
c_5	1.24	0.99	0.80	0.88	1.26	0.38	1.09	1.25	1.06	1.80	0.41	0.48
c_6	1.37	0.34	1.05	1.18	0.96	1.52	1.25	0.52	0.16	0.74	1.22	0.97
c_7	0.79	0.43	1.03	1.33	1.12	0.72	1.58	0.35	0.08	0.60	1.59	0.89
c_8	0.31	1.41	0.45	0.40	0.70	0.13	0.35	1.73	1.78	1.48	0.15	0.28
c_9	0.07	1.26	0.73	0.48	0.41	0.09	0.68	0.62	1.82	1.08	0.25	0.55
c_{10}	0.84	0.53	0.97	0.45	0.92	0.39	0.51	0.54	0.76	1.79	0.30	0.48
c_{11}	1.34	0.60	1.27	0.75	1.07	0.99	1.35	0.48	0.14	0.66	1.56	0.93
c_{12}	0.73	1.20	0.97	0.58	1.13	0.42	0.60	1.43	1.71	1.47	0.55	0.70

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.08	0.13	0.50	0.00	0.25	0.26	0.25	0.42	0.12	0.39	0.13	0.42
c_2	0.30	1.03	1.22	1.22	0.92	0.66	1.60	0.68	0.20	0.78	1.22	0.57
c_3	2.08	0.41	0.96	0.84	0.91	0.96	1.23	1.07	0.26	0.98	0.91	0.92
c_4	1.62	0.78	0.87	1.17	1.52	1.25	1.05	0.78	0.13	1.07	0.97	0.97
c_5	1.42	0.54	0.43	1.07	1.12	0.42	1.12	0.77	1.42	2.08	0.28	0.99
c_6	1.00	0.48	1.49	0.90	1.15	1.70	1.18	0.90	0.11	0.56	1.43	0.46
c_7	0.74	0.57	1.13	1.03	0.91	0.99	1.32	0.81	0.14	0.41	1.13	0.25
c_8	0.61	1.69	0.50	0.61	0.79	0.12	0.82	0.60	1.95	1.15	0.00	1.35
c_9	1.16	1.46	0.85	0.55	1.13	0.45	0.10	0.00	2.23	0.90	0.24	0.64
c_{10}	2.09	1.03	0.47	0.68	0.90	0.39	0.74	1.29	0.68	1.38	0.13	0.63
c_{11}	1.08	0.93	0.97	0.89	1.07	0.78	0.94	0.67	0.39	1.17	1.47	0.58
c_{12}	0.47	0.59	1.46	0.24	1.25	0.93	0.61	0.55	1.68	1.59	0.51	1.04

(f)

$\hat{R}^1$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.40	0.00	0.00	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
c_2	0.00	0.64	1.78	2.71	1.28	0.62	1.20	0.55	0.68	2.84	1.36	0.51
c_3	2.93	1.32	1.20	2.93	0.85	0.35	0.98	1.20	0.00	0.43	1.47	1.47
c_4	0.51	2.04	1.20	1.47	0.55	0.89	0.89	1.20	0.00	0.84	0.51	0.53
c_5	0.00	0.94	0.50	2.64	1.55	0.00	0.69	2.64	1.70	1.32	0.56	0.00
c_6	0.65	2.05	1.30	0.00	2.74	1.47	1.30	0.00	0.00	0.42	1.06	0.65
c_7	0.43	0.55	0.00	0.89	1.75	1.30	2.11	0.55	0.43	0.00	1.78	0.00
c_8	0.43	1.98	0.00	0.47	0.00	NaN	0.00	1.70	1.20	1.39	0.00	0.00
c_9	0.00	1.16	NaN	1.15	0.00	0.00	0.00	2.89	1.68	1.47	0.48	0.00
c_{10}	0.53	0.29	1.59	1.02	0.00	0.00	2.84	1.22	1.89	1.47	0.00	0.89
c_{11}	1.26	0.00	1.70	0.00	0.66	0.91	0.66	0.34	0.91	0.00	1.89	0.00
c_{12}	0.00	1.89	0.91	0.66	1.08	0.00	0.00	1.98	2.16	2.64	0.54	0.66

$\hat{R}^2$	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
c_1	2.06	0.00	0.00	0.00	0.89	0.00	0.00	0.00	0.00	0.00	1.23	1.30
c_2	0.00	1.26	0.00	2.47	0.00	2.20	2.33	0.00	0.00	0.00	0.73	0.00
c_3	1.80	NaN	0.92	1.23	1.80	NaN	0.76	0.00	0.99	2.26	0.00	1.16
c_4	0.00	0.61	1.23	0.99	2.76	0.79	1.38	0.00	0.00	0.00	1.84	1.23
c_5	2.47	0.00	0.00	0.00	1.68	0.00	0.00	0.69	1.03	0.85	0.89	1.77
c_6	0.00	0.00	NaN	0.73	2.45	2.29	0.00	0.00	0.00	0.00	0.00	0.58
c_7	0.00	NaN	1.71	1.30	2.47	1.71	1.04	0.54	0.52	0.00	2.16	0.85
c_8	0.73	0.00	1.45	NaN	0.00	0.00	1.58	0.74	2.41	0.76	0.00	2.20
c_9	0.76	0.00	0.48	0.00	0.00	0.00	0.00	0.46	2.61	1.23	0.00	0.00
c_{10}	1.30	0.85	0.00	NaN	0.89	0.89	1.23	1.30	1.27	1.47	0.00	1.14
c_{11}	2.20	0.92	0.00	0.00	2.20	1.38	2.20	2.76	0.00	1.84	0.99	1.23
c_{12}	NaN	0.00	0.53	0.79	0.00	0.48	0.00	0.00	1.23	2.32	1.57	1.02

Fig. 20: Relevance matrices $\hat{R}^1$ and $\hat{R}^2$ for Experiment II, for the first (left) and second (right) set of questions, respectively, computed for: (a) females, (b) males, (c) non-U.S. participants, (d) U.S. participants, (e) non-designers, and (f) designers.

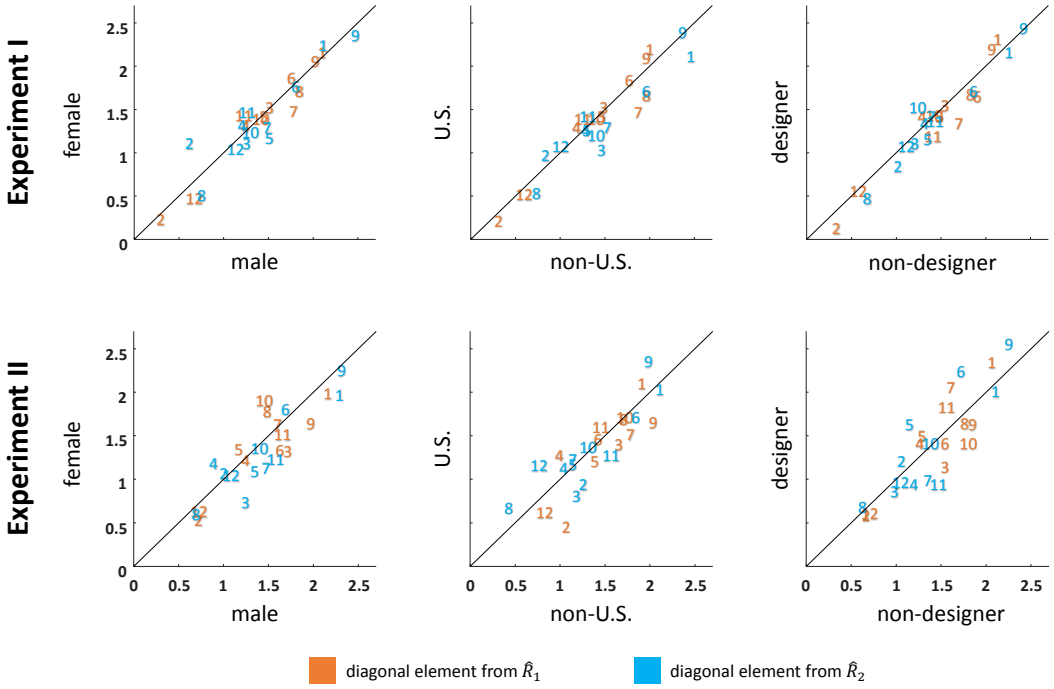


Fig. 21: Comparison of diagonal elements of relevance matrices between demographic subsets. The first row of plots is data from color palette to word clouds (Experiment I), and the second row is word cloud to color palettes (Experiment II). Orange indicates elements compared between diagonals of $\hat{R}^1$ matrices while blue indicates $\hat{R}^2$ matrices. Each number refers to a color-word topic. Notice that the values are quite close to the diagonal, indicating high similarity between the color palette-word cloud associations between demographic groups. Over all comparisons in Experiment I (considering all data points in the first row), the R^2 of the identity line ($y = x$) is 0.89, and for Experiment II (all data points in second row) it is 0.71. There is possibly one point, the blue "2" in the top left (male vs female), where males made a slightly weaker association than women between the colors and words in topic k_2. Given the relatively small number of comparisons in this plot, however, this finding is unlikely to be statistically significant. The overall variance in the second row is larger, due to the smaller amount of data collected in Experiment II.

REFERENCES

Francis M. Adams and Charles E. Osgood. 1973. A cross-cultural study of the affective meanings of color. *Journal of Cross-Cultural Psychology* 4, 2 (1973), 135–156.

Adobe Kuler. 2016. Adobe System Incorporated. (2016). Retrieved May 20, 2016 from https://color.adobe.com

Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I. Jordan. 2003. An introduction to MCMC for machine learning. *Machine Learning* 50, 1-2 (2003), 5–43.

Lawrence W Barsalou. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22, 04 (1999), 577–660.

Brent Berlin. 1969. *Basic color terms: their universality and evolution*. University of California Press.

David M. Blei. 2012. Probabilistic topic models. *Commun. ACM* 55, 4 (April 2012), 77–84.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *The Journal of Machine Learning Research* 3 (2003), 993–1022.

Allison June-Barlow Chaney and David M. Blei. 2012. Visualizing Topic Models.. In *The International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L Boyd-Graber, and David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*. 288–296.

Jason Chuang, Christopher D. Manning, and Jeffrey Heer. 2012. Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 74–77.

Jason Chuang, Maureen Stone, and Pat Hanrahan. 2008. A probabilistic model of the categorical association between colors. In *Color and Imaging Conference*, Vol. 2008. Society for Imaging Science and Technology, 6–11.

ColourLovers. 2016. ColourLovers Creative Market Labs Inc. (2016). Retrieved May 20, 2016 from http://www.colourlovers.com

Gabriela Csurka, Sandra Skaff, Luca Marchesotti, and Craig Saunders. 2010. Learning moods and emotions from color combinations. In *Proceedings of the Seventh Indian Conference on Computer Vision, Graphics and Image Processing*. ACM, 298–305.

Gunilla Derefeldt, Tiina Swartling, Ulf Berggrund, and Peter Bodrogi. 2004. Cognitive color. *Color Research & Application* 29, 1 (2004), 7–19.

Steven P. Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L. Schwartz, and Scott R. Klemmer. 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 4 (2010), 18.

Leatrice Eisemann. 2000. *Pantones Guide to Communicating with Color*. Grafix Press, Ltd.

Pedro F. Felzenszwalb and Daniel P. Huttenlocher. 2004. Efficient graph-based image segmentation. *International Journal of Computer Vision* 59, 2 (2004), 167–181.

David R Flatla, Katharina Reinecke, Carl Gutwin, and Krzysztof Z Gajos. 2013. SPRWeb: preserving subjective responses to website colour schemes through automatic recolouring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2069–2078.

Flickr. 2016. Yahoo! Inc. (2016). Retrieved May 20, 2016 from https://www.flickr.com/

Chris Foges. 1999. *Magazine Design*. RotoVision Press, Switzerland.

Jorge Frascara. 2004. *Communication design: principles, methods, and practice*. Allworth Communications, Inc.

Google n-grams. 2016. Google Inc. (2016). Retrieved May 20, 2016 from https://books.google.com/ngrams/

Dinesh Govindaraj, Tao Wang, and S. V. N. Vishwanathan. 2014. Modeling Attractiveness and Multiple Clicks in Sponsored Search Results. *arXiv preprint arXiv:1401.0255* (2014).

Thomas Griffiths. 2002. *Gibbs sampling in the generative model of latent Dirichlet allocation*. Technical Report. Standford University.

Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America* 101, Suppl 1 (2004), 5228–5235.

Jonathan Harel, Christof Koch, and Pietro Perona. 2007. Graph-based visual saliency. In *Advances in Neural Information Processing Systems 19*. MIT Press, 545–552.

Jeffrey Heer and Maureen Stone. 2012. Color naming models for color selection, image editing and palette design. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*. ACM, 1007–1016.

Scarlett R. Herring, Chia-Chen Chang, Jesse Krantzler, and Brian P. Bailey. 2009. Getting inspired!: understanding how and why examples are used in creative design practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 87–96.

House&Garden. 2016. Cond Nast Publications LTD. (2016). Retrieved May 20, 2016 from http://www.houseandgarden.co.uk/

Glyn W. Humphreys and Vicki Bruce. 1989. *Visual cognition: Computational, experimental, and neuropsychological perspectives.* Lawrence Erlbaum Associates, Inc.

Ali Jahanian. 2011. *Automatic Magazine Cover Design*. Master's thesis. Purdue University.

Ali Jahanian. 2014. *Quantifying aesthetics of visual design applied to automatic design*. Ph.D. Dissertation. Purdue University.

Ali Jahanian, Jerry Liu, Daniel R. Tretter, Qian Lin, Eamonn O'Brien-Strain, Seungyon Lee, Nic Lyons, and Jan P. Allebach. 2013. Recommendation System for Automatic Design of Magazine Covers. In *Proceedings of the 2013 ACM international conference on Intelligent User Interfaces*. ACM.

Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. 1999. An introduction to variational methods for graphical models. *Machine Learning* 37, 2 (1999), 183–233.

Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. 2009. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on*. IEEE, 2106–2113.

Paul Kay, Brent Berlin, Luisa Maffi, William R Merrifield, and Richard Cook. 2009. *The world color survey*. Springer.

Paul Kay and Chad K. McDaniel. 1978. The linguistic significance of the meanings of basic color terms. *Language* (1978), 610–646.

Shigenobu Kobayashi. 1981. The aim and method of the color image scale. *Color Research & Application* 6, 2 (1981), 93–107.

Shigenobu Kobayashi. 1991. *Color Image Scale*. Kodansha Intern.

Harold W. Kuhn. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 2, 1-2 (1955), 83–97.

Sharon Lin, Julie Fortuna, Chinmay Kulkarni, Maureen Stone, and Jeffrey Heer. 2013a. Selecting Semantically-Resonant Colors for Data Visualization. In *Eurographics Conference on Visualization (EuroVis), 2013 15th Annual Visualization Conference on*, Vol. 32. EuroVis.

Sharon Lin and Pat Hanrahan. 2013. Modeling How People Extract Color Themes from Images. In *ACM Human Factors in Computing Systems (CHI)*.

Sharon Lin, Daniel Ritchie, Matthew Fisher, and Pat Hanrahan. 2013b. Probabilistic Color-by-Numbers: Suggesting Pattern Colorizations Using Factor Graphs. In *ACM SIGGRAPH 2013 Papers*.

Albrecht Lindner and Sabine Süsstrunk. 2013. Automatic color palette creation from words. In *Color and Imaging Conference*, Vol. 2013. Society for Imaging Science and Technology, 69–74.

Albrecht Lindner and Sabine Süsstrunk. 2015. Semantic-Improved Color Imaging Applications: It Is All About Context. *IEEE Transactions on Multimedia* 17, 5 (2015), 700–710.

Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. 2007. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition* 40, 1 (2007), 262–282.

Colin M MacLeod. 1991. Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin* 109, 2 (1991), 163.

Y Matsuda. 1995. *Color Design*. Asakura Shoten.

Jon D Mcauliffe and David M Blei. 2008. Supervised topic models. In *Advances in neural information processing systems*. 121–128.

Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K Gray, Joseph P Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, and others. 2011. Quantitative analysis of culture using millions of digitized books. *science* 331, 6014 (2011), 176–182.

Aleksandra Mojsilovic and Bernice Rogowitz. 2001. Capturing image semantics with low-level descriptors. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, Vol. 1. IEEE, 18–21.

Naila Murray, Sandra Skaff, Luca Marchesotti, and Florent Perronnin. 2012. Toward automatic and flexible concept transfer. *Computers & Graphics* 36, 6 (2012), 622–634.

Quentin Newark. 2007. *What is graphic design?* Rockport Publishers.

NgramViewer. 2016. Google Inc. (2016). Retrieved May 20, 2016 from https://books.google.com/ngrams

OCLC. 2016a. Dewey Decimal Classification. (2016). Retrieved May 20, 2016 from http://www.oclc.org/dewey.en.html

OCLC. 2016b. WorldCat. (2016). Retrieved May 20, 2016 from http://www.worldcat.org

Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. 2011. Color compatibility from large datasets. *ACM Transactions on Graphics (TOG)* 30, 4 (2011), 63.

Charles E. Osgood. 1952. The nature and measurement of meaning. *Psychological bulletin* 49, 3 (1952), 197.

Charles E. Osgood. 1971. Exploration in Semantic Space: A Personal Diary1. *Journal of Social Issues* 27, 4 (1971), 5–64.

Li-Chen Ou, M. Ronnier Luo, Andrée Woodcock, and Angela Wright. 2004a. A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application* 29, 3 (2004), 232–240.

Li-Chen Ou, M. Ronnier Luo, Andree Woodcock, and Angela Wright. 2004b. A study of colour emotion and colour preference. Part II: Colour emotions for two-colour combinations. *Color Research & Application* 29, 4 (2004), 292–298.

Li-Chen Ou, M. Ronnier Luo, Andrée Woodcock, and Angela Wright. 2004c. A study of colour emotion and colour preference. Part III: Colour preference modeling. *Color Research & Application* 29, 5 (2004), 381–389.

Li-Chen Ou, M. Ronnier Luo, Pei-Li Sun, Neng-Chung Hu, Hung-Shing Chen, Shing-Sheng Guan, Andree Woodcock, José Luis Caivano, Rafael Huertas, Alain Treméau, and others. 2012. A cross-cultural comparison of colour emotion for two-colour combinations. *Color Research & Application* 37, 1 (2012), 23–43.

Galina V Paramei. 2005. Singing the Russian blues: An argument for culturally basic color terms. *Cross-Cultural Research* 39, 1 (2005), 10–38.

Pinterest. 2016. Pinterest. (2016). Retrieved May 20, 2016 from https://www.pinterest.com/

Martin F. Porter. 1980. An algorithm for suffix stripping. *Program: Electronic Library and Information Systems* 14, 3 (1980), 130–137.

Katharina Reinecke and Krzysztof Z. Gajos. 2014. Quantifying Visual Preferences Around the World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM.

Debi Roberson, Ian Davies, and Jules Davidoff. 2000. Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General* 129, 3 (2000), 369.

Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. 2000. The Earth Mover's distance as a metric for image retrieval. *International Journal of Computer Vision* 40, 2 (2000), 99–121.

Tim Samara. 2007. *Design elements: A graphic style manual*. Rockport Publishers.

Boris Schauerte and Rainer Stiefelhagen. 2012. Learning robust color name models from web images. In *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 3598–3601.

Christin Seifert, Barbara Kump, Wolfgang Kienreich, Gisela Granitzer, and Michael Granitzer. 2008. On the beauty and usability of tag clouds. In *Information Visualisation, 2008. IV'08. 12th International Conference*. IEEE, 17–25.

Ishwar K. Sethi, Ioana L. Coman, and Daniela Stan. 2001. Mining association rules between low-level image features and high-level concepts. *Proceedings of the SPIE Data Mining and Knowledge Discovery* 3 (2001), 279–290.

Vidya Setlur and Maureen C Stone. 2016. A Linguistic Approach to Categorical Color Assignment for Data Visualization. *Visualization and Computer Graphics, IEEE Transactions on* 22, 1 (2016), 698–707.

Gaurav Sharma. 2002. Color fundamentals for digital imaging. In *Digital Color Imaging Handbook*, Gaurav Sharma (Ed.). CRC Press.

Ben Shneiderman. 2009. Creativity support tools: A grand challenge for HCI researchers. *Engineering the User Interface* (2009), 1–9.

Liangcai Shu, Bo Long, and Weiyi Meng. 2009. A latent topic model for complete entity resolution. In *Data Engineering, 2009. ICDE'09. IEEE 25th International Conference on*. IEEE, 880–891.

Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. 2000. Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22, 12 (2000), 1349–1380.

Martin Solli and Reiner Lenz. 2010. Color semantics for image indexing. In *CGIV 2010: 5th European Conference on Colour in Graphics, Imaging, and Vision*. IS&T, 353–358.

Anna Starmer. 2005. *The Color Scheme Bible: Inspirational Palettes for Designing Home Interiors*. Firefly Books.

Mark Steyvers and Tom Griffiths. 2007. Probabilistic topic models. In *Handbook of Latent Semantic Analysis*, Simon Dennis Thomas K. Landauer, Danielle S. McNamara and Walter Kintsch (Eds.). Lawrence Erlbaum.

Mark Steyvers and Tom Griffiths. 2014. Matlab Topic Modeling Toolbox. (2014). Retrieved November 14, 2014 from http://psiexp.ss.uci.edu/research/programs_data/toolbox.htm

Yee Whye Teh, David Newman, and Max Welling. 2006. A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In *The Neural Information Processing Systems (NIPS)*, Vol. 6. 1378–1385.

TripAdvisor. 2014. TripAdvisor LCC. (2014). Retrieved November 14, 2014 from http://www.tripadvisor.com/

Joost Van De Weijer, Cordelia Schmid, Jakob Verbeek, and Diane Larlus. 2009. Learning color names for real-world applications. *Image Processing, IEEE Transactions on* 18, 7 (2009), 1512–1523.

Hanna M Wallach, David M Mimno, and Andrew McCallum. 2009. Rethinking LDA: Why priors matter. In *Advances in neural information processing systems*. 1973–1981.

Baoyuan Wang, Yizhou Yu, Tien-Tsin Wong, Chun Chen, and Ying-Qing Xu. 2010. Data-driven image color theme enhancement. *ACM Transactions on Graphics (TOG)* 29, 6 (2010), 146.

Chong Wang, David Blei, and Fei-Fei Li. 2009. Simultaneous image classification and annotation. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 1903–1910.

Allan Whitfield and John Wiltshire. 1982. Design training and aesthetic evaluation: An intergroup comparison. *journal of Environmental Psychology* 2, 2 (1982), 109–117.

Jonathan Winawer, Nathan Witthoft, Michael C Frank, Lisa Wu, Alex R Wade, and Lera Boroditsky. 2007. Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences* 104, 19 (2007), 7780–7785.