

Affect-Aware Word Clouds

TUGBA KULAHCIOGLU, Rutgers University, Piscataway, New Jersey

GERARD DE MELO, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

Word clouds are widely used for non-analytic purposes, such as introducing a topic to students, or creating a gift with personally meaningful text. Surveys show that users prefer tools that yield word clouds with a stronger emotional impact. Fonts and color palettes are powerful typographical signals that may determine this impact. Typically, these signals are assigned randomly, or expected to be chosen by the users. We present an affect-aware font and color palette selection methodology that aims to facilitate more informed choices. We infer associations of fonts with a set of eight affects, and evaluate the resulting data in a series of user studies both on individual words as well as in word clouds. Relying on a recent study to procure affective color palettes, we carry out a similar user study to understand the impact of color choices on word clouds. Our findings suggest that both fonts and color palettes are powerful tools contributing to the affects evoked by a word cloud. The experiments further confirm that the novel datasets we propose are successful in enabling this. We also find that, for the majority of the affects, both signals need to be congruent to create a stronger impact. Based on this data, we implement a prototype that allows users to specify a desired affect and recommends congruent fonts and color palettes for the word.

CCS Concepts: • **Human-centered computing** → **Information visualization**; **Visualization toolkits**;

Additional Key Words and Phrases: Affective interfaces, word clouds, typography, color palettes

ACM Reference format:

Tugba Kulahcioglu and Gerard de Melo. 2020. Affect-Aware Word Clouds. *ACM Trans. Interact. Intell. Syst.* 10, 4, Article 34 (November 2020), 25 pages.

<https://doi.org/10.1145/3370928>

1 INTRODUCTION

While word clouds have certain drawbacks as an analytical tool [1, 12], they are widely adopted by people of all ages and occupations for non-analytic purposes, such as preparing a gift, or introducing a topic [48]. Viegas et al. [48] focus on a specific type of word cloud, namely those generated by Wordle, finding that a key reason for the tool's greater success in comparison with alternative word cloud tools is the "emotional impact" it creates through the tight layout it applies, and the

The reviewing of this article was managed by special issue associate editors Oliver Brdiczka, Polo Chau, Gaelle Calvary, and Minsuk Kahng.

Research conducted while Gerard de Melo was a faculty member at Rutgers University.

Authors' addresses: T. Kulahcioglu, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ, 08854-8019; email: tugba.kulahcioglu@rutgers.edu; G. de Melo, Hasso Plattner Institute, University of Potsdam, Prof.-Dr.-Helmert-Str. 2-3, 14482, Potsdam, Germany; email: gdm@demelo.org.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2160-6455/2020/11-ART34 \$15.00

<https://doi.org/10.1145/3370928>



Fig. 1. Word cloud examples using fonts (first row) and color palettes (second row) that are congruent with their message. Instead of randomly picking these paralinguistic signals as current tools do, in this study we determine their congruency with a set of eight affects, and propose a word cloud tool that helps users make congruent choices.

font and color palette options it provides. This is exemplified by user comments from the study that emphasize the importance of these signals, e.g.:

“Wordles have more emotional emphasis, colors, and layouts to enhance the meaning.” [48]

“Wordles are colorful, more visually interesting, more of an emotional response and connection with the viewer than the tag clouds.” [48]

For some use cases, the primary purpose of the cloud is to convey the sentiment of its input, e.g., when visualizing restaurant reviews. In such cases, the affective perception of the cloud plays a critical role. Affect, thus, is a key factor to be considered in the design of word clouds. Typography and color are two important aspects that may influence the affect evoked by a word cloud. Figure 1 demonstrates how different choices of fonts and colors may give rise to quite diverse affective associations.

The congruency between the textual content and the selected font has shown to have a pronounced impact on the affective interpretation of the text [35, 40]. For instance, different typefaces may result in distinct ratings for the same textual content with respect to its perceived *excitingness*. It has been observed that the response times of users decrease when fonts that are congruent with the message are used [14, 26]. Similar ties have also been observed for individual colors [4, 29, 52] and color palettes [2].

Designers routinely expend significant effort in making typographical choices that accord with the message that visual material is intended to convey. Thus, a word cloud produced by a designer would typically incorporate colors and fonts that are congruent with the message of the cloud. On the contrary, this may not be the case for non-professional users, for a variety of reasons, including a lack of such awareness, or simply an unwillingness to put in the time and effort. Despite the established relationship between typographic signals and affect, and the known impact of their congruency, word cloud tools thus far neglect to support the users in making semantically informed typographical choices. Hence, the impact of automatic means of improving the congruency of fonts and colors in word cloud tools has remained under-explored. Towards the aim of enabling such support, our contributions in this study are as follows:

- (1) Previous studies use human annotations to solicit font–attribute ratings. We propose computational methods to obtain affective relationships based on other ratings (Section 3).

- (2) Through three user studies,¹ in Sections 4, 5, and 6, we show that congruent fonts and color palettes, respectively, better reflect word clouds with affective content, for a range of different affects.²
- (3) We find that fonts and color palettes have complementary strengths in conveying affects, and for the majority of the considered affects, it is important to include congruent signals using both the color palette and the font.
- (4) Existing word cloud generators choose fonts and colors based solely on aesthetics or even randomly, neglecting their congruency with the intended affect. We instead show how they can be utilized to make typographical recommendations based on user-specified affects.
- (5) We further discuss in detail a number of promising new directions for future research.

This article extends a previous conference publication [24] with the following novel contributions: (1) In Section 6, we present a user study that investigates the relative affective power of fonts and color palettes on word clouds and user preferences to combine these signals to create an intended affect in a word cloud; (2) In Section 7.1.5, we present new functionality of the prototype enabling users to keep track of different word cloud designs for the same input; and (3) Throughout this article, we have added additional details and references.

2 RELATED WORK

Before describing the details of our method, we review related work on word clouds, fonts, and color.

2.1 Word Clouds

Word clouds have been of substantial interest to the academic community, especially with regard to the employed layout algorithms [10, 49]. Several studies [27, 38, 53] focus on semantic relationships of the words, e.g., placing semantically similar words closer to each other. There are also studies aiming to create comparison clouds by combining visualizations for multiple texts [7, 28].

Wordle [45] is a widely appreciated word cloud generation tool that aims to create more pleasing word clouds (“wordles”) by adopting tighter layouts, e.g., allowing a tiny word to appear within a character from a larger word, with several font and color palette options. Through a user survey and analysis of resources on the Web, Viegas et al. [48] analyze the users and common use cases of wordles. The results show that it attracts people of all ages, most of whom are educators or students (29%), with no other occupation accounting for more than 6%. The main reason for the tool’s broad appeal is found to be the its power to create an emotional impact by means of fonts, colors, and layouts. Two other important factors are the attention-grabbing nature and the organic non-linear layout of the generated word clouds. The use cases vary from education (e.g., using Wordle to aid in introducing a topic to students) to gift-giving (e.g., creating a wordle from wedding vows), and even *guess the cloud* games.³ Another insightful result of this analysis was that 88% of the users reported that they feel creative when using Wordle.

ManiWordle [19] enables custom manipulations in a word cloud, based on changes with regard to fonts, colors, and the layout, both at the cloud and word level. For example, one may alter the color of a selected word so as to emphasize it. Although this provides more control, users did not report feeling a strong difference in terms of creativity compared to the regular Wordle

¹All studies in this article received IRB approval.

²Specifically, we consider the affective attributes *calm*, *exciting*, *positive*, *negative*, *playful*, *serious*, *disturbing*, and *trustworthy*.

³<http://guessthewordle.weebly.com/>.

tool. WordlePlus [16] further extends ManiWordle to allow natural interaction on pen and touch-enabled tablets. EdWordle [50] facilitates multi-word editing, while preserving the neighborhood, i.e., keeping non-edited words close to their original locations. It applies a local re-wordle algorithm that re-arranges the words to close gaps.

We are not aware of detailed studies on the use of fonts or colors in text visualization from a semantic congruency perspective. A related study [5] uses font attributes, such as underlining or small caps, to distinguish set membership in set visualizations, including for emotions. Wecker et al. [51] propose using font properties such as size and color to highlight the sentiment polarity of text passages. However, they do not consider typefaces or their semantic connotations.

2.2 Fonts

In the following, we consider related research on affective attributes of fonts, and studies revealing their impact.

Affective Attributes of Fonts. Through a crowdsourced study, O'Donovan et al. [35] associate 200 diverse fonts with 37 semantic attributes (e.g., *happy*). Specifically, they request participants to pick one of two presented fonts for a given attribute, and then aggregate these choices to assign fonts a series of scores for each attribute. Another online survey [41] assesses the characteristics of 20 fonts with respect to 15 adjective-based scales (e.g., *stable-unstable*). Similarly, many font-focused websites⁴ allow for tagging fonts with various attributes, some of which are semantic in nature. Further studies [46, 47] explored the relationship between visual font characteristics and taste attributes (sweet, sour, etc.) through user studies, concluding, for instance, that round-shaped fonts are associated with a sweet taste.

Kulahcioglu and de Melo [23] extend the dataset from O'Donovan et al. [35] to a larger number of fonts using a k-Nearest Neighbors (k-NN) approach with deep Convolutional Neural Network (CNN) embeddings as a similarity metric between fonts. FontLex [21, 22] is a lexicon providing word–font associations for arbitrary English words, automatically induced using affective associations of words and fonts, the latter based on word embedding similarities. In this study, we propose new computational methods that are based on affect relationships to obtain font–affect scores, instead of using word similarities.

Psychological Effects of Font Selection. We next survey studies that investigate the effect of font characteristics on perception.

Several Stroop-style studies have been conducted in this regard. Hazlett et al. [14] asked users to judge whether a displayed word is positive or negative, comparing 5 fonts and 25 words that are all strongly associated with positive or negative emotions. The results indicate that congruent typefaces yield faster responses. Lewis and Walker [26] ask users to press a left-hand key if the words *slow* or *heavy* appear, versus a right-hand one if *fast* or *light* appears. In a second experiment, they display related words (e.g., *fox*) instead of the original words (e.g., *fast*) to ensure that the user needs to grasp the meaning of the displayed word. In both experiments, they repeat the tasks with congruent and incongruent fonts, finding that the former significantly reduce the response time.

There are also studies that use survey-style approaches. Juni and Gross [17] present newspaper articles using two different fonts, finding that the same text may be perceived as more humorous or angry depending on the chosen font. Shaikh [40] collects ratings to understand the perception of the document and the perceived personality of the author. The results reveal that fonts with different congruencies heavily alter the perception of a document, while congruent and neutral fonts appear to evoke similar perceptions of an author's personality.

⁴For instance, <https://fontsinuse.com/>, <http://www.dafont.com>, and <http://www.1001fonts.com>.

Another study by Shaikh et al. [42] implements a related test for the perception of emails. The results suggest that incongruent fonts may give rise to a different perception of an email compared to the use of a congruent font. A similar study on the perception of a company website [39] demonstrates that neutral and low congruency fonts can negatively affect a company's perception in terms of professionalism, believability, trust, and intent to act on the site.

Marketing research has a strong interest in the impact of font choices in packaging design. Fligner [13] shows that *natural* looking fonts in a packaging design increase the perceived *healthfulness* of products, especially if the products' intrinsic (e.g., being fat-free) and extrinsic cues (e.g., being sold at Whole Foods Market) also support this perception. Childers and Jass [8] show that font choices have an impact on product perception for both high and low user engagement levels.

Font Recommendation. Traditional font recommendation tools have focused on font-to-font recommendation topics, such as font pairing [3, 32, 36] and font similarity [35]), neglecting the automated discovery of semantic connections between the data and the fonts. Recently, FontMatcher [9] studied font recommendation for image captions, using a color–font mapping in the soft–strong and warm–cool dimensions, which, similar to our study, goes beyond the traditional font recommendation scenarios.

2.3 Color Palettes

The impact of color has been studied extensively [4, 52]. Certain colors have been tied to specific forms of impact on cognitive tasks, although their effect may vary between tasks [31]. For instance, a choice of red has been found to be beneficial for detail-oriented tasks, but it does not have the same effect on creative tasks. Similar to fonts, colors have also been of special interest in marketing research [25, 44].

In contrast, the use of multiple colors together as a palette has not been studied to the same extent, as explained by Bartram et al. [2]. To address this, they conducted a study of color palette choices for a set of eight affects, which is also the source of the color palettes in our work. The study establishes several relationships between affects and perceptual color properties (hue, chroma, and lightness). As an example, the attributes *calm*, *playful*, and *positive* are found to be associated with the lightest colors. Among the core affects, *calm* is found to evoke the strongest color preferences. Other notable findings include that highly saturated light colors are not a good fit for *serious*, *trustworthy*, or *calm*, and that light colors, in general, are not very successful in conveying a *negative* sentiment.

3 AFFECTIVE FONTS

In this section, we present our method to computationally obtain font attribute associations using crowdsourced seed data, and we evaluate that method through a user study.

3.1 Method

Our goal is to compute font attribute vectors $\mathbf{v}_a \in \mathbb{R}^{|\mathbb{F}|}$ which, for a given affective attribute $a \in \mathbb{A}$, reflect the perception of fonts $f \in \mathbb{F}$ with respect to the affect. The set of affective attributes $\mathbb{A}$ is given in Table 1 and was chosen because it includes the core affects in the PAD emotional state model [30] and is used in a previous study on color [2], allowing us to compare the impact of fonts and color palettes. Given the indices i from Table 1, we denote each affective attribute as a_i and also use the notation v_i as a shorthand for v_{a_i} , e.g., v_1 as the vector for the affective attribute *calm*. Each dimension of a particular $\mathbf{v}_a$ reflects the congruency of attribute a with respect to a different font $f \in \mathbb{F}$.

Table 1. Affective Attributes Associated with Fonts

Index	Attribute Name	Index	Attribute Name
1	calm	5	playful
2	exciting	6	serious
3	positive	7	trustworthy
4	negative	8	disturbing

In our study, the set $\mathbb{F}$ consists of 200 different fonts, taken from the aforementioned study by O'Donovan et al. [35], due to it being the largest of its sort available for download. Their data provides us with $|\mathbb{F}|$ -dimensional vectors $\mathbf{x}_t$ for a set of 31 font traits $t \in \mathbb{T}$, providing the scores for that trait over all fonts in $\mathbb{F}$. The set $\mathbb{T}$ of 31 traits they consider includes visual ones such as *thin*, *angular*, but also more subjective ones. We rely on the following multi-pronged procedure to obtain the desired vectors $\mathbf{v}_a$ for affective attributes $a \in \mathbb{A}$ from this data.

Scores for “calm”, “playful”. Fortunately, the crowdsourced data already includes two of our eight considered affective attributes. Thus, we directly obtain $\mathbf{v}_1$ and $\mathbf{v}_5$ by selecting the relevant $\mathbf{x}_t$.

Scores for “positive”. We obtain scores for the attribute *positive* by clustering all emotional traits included in the data crowdsourced by O'Donovan et al. [35]. For this, we manually filter the set of font traits $t \in \mathbb{T}$ so as to retain only the ones of a strong emotional nature. Then, we apply k-means clustering and obtain the following three clusters:

- C_1 : *bad, boring*
- C_2 : *happy, playful, attractive*
- C_3 : *calm, charming, fresh, friendly, gentle, graceful, soft, warm*.

The first observation about these clusters is that C_1 contains traits with a negative connotation, while the other two each contain positive traits. A detailed analysis reveals that C_2 includes high-arousal positive emotions, whereas C_3 includes low-arousal ones. We compute score vectors $\mathbf{x}_C$ for each positive cluster $C \in \mathbb{C} = \{C_2, C_3\}$ as

$$\mathbf{x}_C = \sum_{t \in C} \sigma(C, t) \mathbf{x}_t, \quad (1)$$

where the weight of each trait is calculated as

$$\sigma(C, t) = \frac{1}{|C| - 1} \frac{\sum_{t' \in C, t' \neq t} d(C, \mathbf{x}_{t'})}{\sum_{t' \in C} d(C, \mathbf{x}_{t'})}. \quad (2)$$

with $d(C, \mathbf{x})$ corresponding to the distance between the cluster centroid and a trait-specific vector $\mathbf{x}$.

Using these two cluster centers, the scores for the target attribute *positive* is obtained as:

$$\mathbf{v}_3 = \frac{1}{|\mathbb{C}|} \sum_{C \in \mathbb{C}} \mathbf{x}_C. \quad (3)$$

Scores for “exciting”, “serious”, and “negative”. For the attributes *exciting*, *serious*, and *negative*, we take advantage of the antonymy relationships and use scores for *calm*, *playful*, and *positive*, respectively, as:

$$\mathbf{v}_i = \mathbf{1} - \mathbf{v}_{\alpha(i)}, \quad (4)$$

calm	exciting	positive	negative	playful	serious	disturbing	trustworthy
calm	exciting	<i>positive</i>	negative	playful	serious	disturbing	<i>trustworthy</i>
<i>calm</i>	EXCITING	<i>positive</i>	negative	<i>playful</i>	serious	DISTURBING	trustworthy
calm	exciting	<i>positive</i>	negative	<i>playful</i>	serious	disturbing	trustworthy
calm	exciting	<i>positive</i>	NEGATIVE	<i>playful</i>	serious	disturbing	trustworthy
calm	exciting	<i>positive</i>	<i>negative</i>	<i>playful</i>	serious	DISTURBING	<i>trustworthy</i>

Fig. 2. Top five congruent fonts obtained for each of the eight affects used in this study.

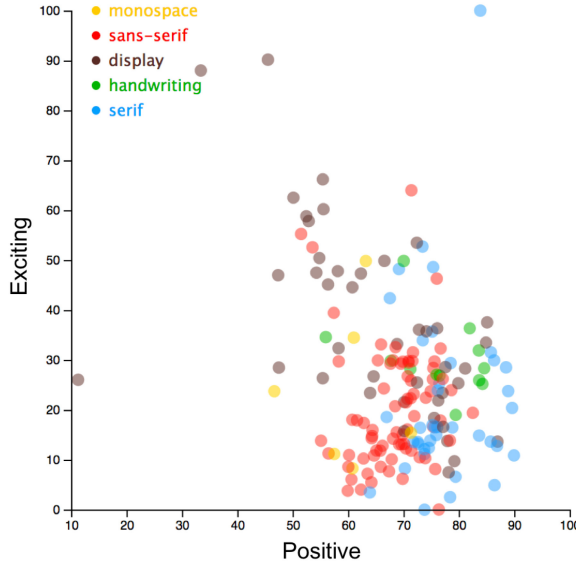


Fig. 3. Scatter plot of the resulting font scores based on the *positive* and *exciting* attributes (scaled to [0, 100] range).

where $\mathbf{1} \in \mathbb{R}^{|\mathbb{F}|}$ is a vector of ones, $i = 2, 4, 6$ are indices from Table 1, and $\alpha(i)$ denotes the respective antonym of affective attribute i . We follow this method based on the observation that a font that is least representative of an attribute is a candidate to best represent the opposite attribute. However, this method assumes that a font cannot be a good representative for each of the two opposing attributes, which may not always hold, given different contexts the fonts could be used in.

Scores for “trustworthy” and “disturbing”. The PAD model [30] posits that complex emotions are composed of more basic ones. Following Bartram et al. [2], *trustworthy* can be defined as *positive* + *calm*. To obtain font-specific scores, we accordingly average the values for *positive* and *calm* as $\mathbf{v}_7 = \frac{1}{2}(\mathbf{v}_3 + \mathbf{v}_1)$. We performed an analogous derivation to obtain values for *disturbing* using *negative* and *exciting*, i.e., $\mathbf{v}_8 = \frac{1}{2}(\mathbf{v}_4 + \mathbf{v}_2)$.

3.2 Results

The top five congruent fonts for each affect in Table 1, as determined by the above process, are depicted in Figure 2. The top fonts for a particular affect appear to have similar visual characteristics, except for *exciting* and *negative*. To investigate this, Figure 3 plots the distribution of fonts based on the *positive* and *exciting* attribute scores. These attributes correspond to the opposites of

Option 1	Option 2	Option 3	Option 4	Option 5
serious	serious	serious	serious	serious

Fig. 4. A sample task from User Study I. For this sample task, the third and fifth images are generated using incongruent fonts, the second one uses a neutral font, and the first and fourth images use congruent fonts.

negative and *calm*. For instance, the fonts that correspond to the lower end of the *positive* scale are the fonts assessed as highly *negative*. We find that the top fonts for the *exciting* and *negative* attributes have a wide range of values in the respective other scale. For example, fonts with the highest scores for *exciting* exhibit a large degree of variance with regard to their scores for the attribute *positive*. The same applies for *negative*, i.e., fonts with the lowest scores for *positive* have a wide range of scores for *exciting*. This can explain why for these two attributes, fonts with the highest congruency exhibit different visual characteristics. The top fonts for other attributes, on the other hand, are found to reside in a smaller area in the chart, i.e., have a narrower range of scores in the other scale. As an example, fonts with the highest scores for *positive* are concentrated in a region with low scores for *exciting* (between 10 and 25). This analysis is further expanded and validated in Section 7, where we group the top *negative* and *positive* fonts based on their scores for *calm* and *exciting*.

Overall, the majority of fonts considered are deemed highly *positive* and *calm*. Figure 3 also provides the categories to which the fonts belong. *Handwriting* typefaces are designed to give the impression of being hand-rendered. The characters of *monospace* typefaces occupy equal horizontal space. *Serif* typefaces have small lines attached to the end of the strokes in their characters, whereas *sans-serif* ones lack such attached lines. *Display* typefaces do not share typical typographic properties other than a lower degree of legibility when used for body text, so they are reserved mostly for headings and other kinds of display purposes.

When we analyze the score distributions in conjunction with their font categories, we find that *Serif* fonts appear to have higher *positive* scores compared to *sans-serif* fonts. *Display* fonts are found to be *exciting*, which accords with their decorative nature. With significantly fewer instances in the dataset, *monospace* and *handwriting* fonts are scattered along a wide range of different values.

3.3 Evaluation (User Study I)

We evaluate the above font score computations through a user study, in which we collect user preferences among congruent, incongruent, and neutral fonts for affective word clouds.

Hypothesis. Fonts with higher congruency scores for a given attribute are assessed as better representing that attribute than fonts with lower scores.

Participants and Method. We recruited 40 participants via Mechanical Turk, all from the United States, with at least 50 approved HITs and an overall approval rating of 90% or more. Participants were paid \$0.01 for each task. We rely on a within-subject design, and perform counterbalancing by reversing the order of the tasks for half of the participants. The study involves 50 tasks for each participant, consisting of 6 tasks for each of the 8 affective attributes, and 2 additional validation tasks. Each task presents the name of the affective attribute using five different fonts, as exemplified in Figures 4 and 5. To allow for comparison with the color palette study by Bartram et al. [2], we select two congruent, two incongruent, and a neutral font. The congruent ones are selected randomly amongst the six most congruent fonts for the corresponding affect, while the two incongruent fonts are selected randomly among the least congruent six fonts. The fifth font is selected randomly among the three fonts that are in the middle of the ranked font list for

Option 1	Option 2	Option 3	Option 4	Option 5
<i>negative</i>	negative	NEGATIVE	negative	<i>negative</i>

Fig. 5. A sample task from User Study I. For this sample task, the third and fifth images are generated using incongruent fonts, the second one uses a neutral font, and the first and fourth images use congruent fonts.

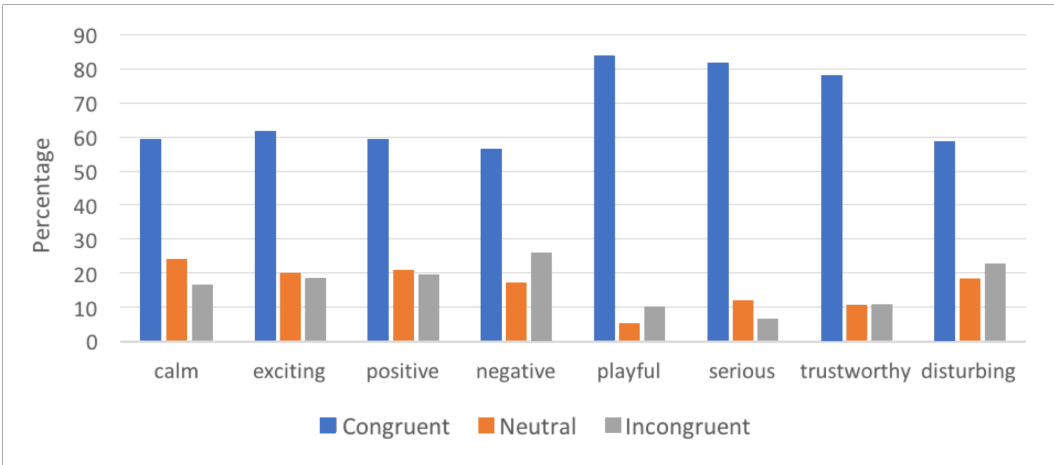


Fig. 6. Results of User Study I, evaluating the font scores with respect to renderings of single words. The user preferences for congruent, neutral, and incongruent font choices are compared for each of the eight affective attributes. With uniformly random selections, these, respectively, have a 40%, 20%, or 40% chance of being selected, since two out of five options are congruent or incongruent, while one option out of five uses a neutral font.

each affect. The order of fonts with different congruencies are selected randomly for each task. The validation tasks include words written with one congruent font and four incongruent fonts. Participants were instructed to “Pick the image that best represents the word”, with an additional detailed version given as “Select the image that you think best reflects the meaning of the word shown in the images.”

Results and Analysis. The results of our study are summarized in Figure 6. Across all attributes, the options with fonts determined to be congruent are frequently selected by the participants, according with our hypothesis, while the options with fonts determined to be incongruent are less frequently selected. We conducted chi-square goodness of fit tests of user preferences for each affect based on the three font category choices (congruent, neutral, and incongruent fonts). Table 2 provides results of these analyses, which are found to be statistically significant for each of the eight affects.

Given that the scores for *calm* and *playful* were obtained via crowdsourcing, in our analysis, they may serve as ground truth benchmarks as to what range of scores we are to expect from high-quality human-provided ratings. Fonts rated strongly as *calm* appear to be less preferred than those for *playful*, possibly owing to the fact that even regular fonts may also have a tendency to be perceived as calm. Indeed, the median congruency score in our data for *calm* was 76.3%, while for *playful*, it was 34.5%, confirming that neutral fonts are more likely perceived as calm.

Both *exciting* and *serious* acquire similar results to the baselines, which suggests that our method of computing their scores as reversed opposites suffices to select fonts perceived as congruent. The

Table 2. Chi-Square Goodness of Fit Test Results for the Four User Studies Presented in this Article

<i>i</i>	Attribute	Study I (Section 3.3)		Study II (Section 4.1)		Study III (Section 5.2)		Study IV (Section 6.1)	
		$\bar{X}^2(2)$	$p <$	$\bar{X}^2(2)$	$p <$	$\bar{X}^2(2)$	$p <$	$\bar{X}^2(2)$	p
1	calm	56.79	0.001	54.14	0.001	121.84	0.001	48.2	< 0.001
2	exciting	56.33	0.001	16.80	0.001	44.96	0.001	46.25	< 0.001
3	positive	48.10	0.001	43.78	0.001	43.13	0.001	48.65	< 0.001
4	negative	29.03	0.001	51.17	0.001	89.54	0.001	1.8	0.4066
5	playful	195.07	0.001	70.15	0.001	115.02	0.001	54.6	< 0.001
6	serious	180.03	0.001	142.06	0.001	43.09	0.001	0.6	0.7408
7	trustworthy	149.29	0.001	86.68	0.001	20.79	0.001	14.6	< 0.001
8	disturbing	38.94	0.001	126.69	0.001	76.75	0.001	27.95	< 0.001

Each affective attribute is represented by the corresponding index (*i*) as defined in Table 1. For the first three experiments, expected values are specified as 0.4, 0.2, and 0.4 for the categories *congruent*, *neutral*, and *incongruent*, respectively. For the last experiment expected values are specified as 1/3 for each of the three options. The significance level is set as 0.05 for all experiments. Results for each affective attribute in the first three user studies are found to be highly statistically significant, as for each analysis $p < 0.001$. For the last user study, all affects except *serious* and *negative* received statistically significant results. Discussions on the results can be found in the respective sections.

clustering approach used for *positive*, from which in turn scores for *negative* are derived as well, appears to yield reasonable but not overly strong ratings. While this might stem from inaccuracies in the automatic clustering, it may also be the case that it is less trivial to convey positive and negative sentiment than to convey attributes such *playful* and *exciting*. The most peculiar finding is that fonts with high scores for the attribute *trustworthy* manifest stronger preferences than those for the attributes used to compute its values (namely, *positive* and *calm*). Despite being computed analogously, the ratings for *disturbing* do not exceed those for *negative* and *exciting*.

4 AFFECTIVE FONTS IN WORD CLOUDS

Using the font scores obtained in the preceding section, we seek to understand the impact of the affective nature of fonts on affective word clouds. In particular, we shall determine to what extent users may prefer fonts that accord with the content of a word cloud with regard to affect.

4.1 User Study II

Through a user study on Mechanical Turk, we evaluate the impact of fonts on word clouds with affective content.

Hypothesis. Word clouds using fonts determined to be congruent are assessed as being more representative of pertinent affect-evoking words than word clouds using fonts determined as neutral or incongruent.

Participants and Method. The details of the participants and method of evaluation are as in Study I (Section 3.3), except that the displayed renderings here include word clouds instead of single words. Sample tasks are given in Figures 7 and 8. For each affective attribute, we created word clouds of 10 words, one of which is the affective attribute name itself, coupled with 9 further semantically related words to avoid confounding effects potentially caused by irrelevant words. For the same reason as earlier, each task includes five word clouds with the same content and layout, just using a different font. As earlier, we randomly select two congruent fonts, two incongruent ones, and a

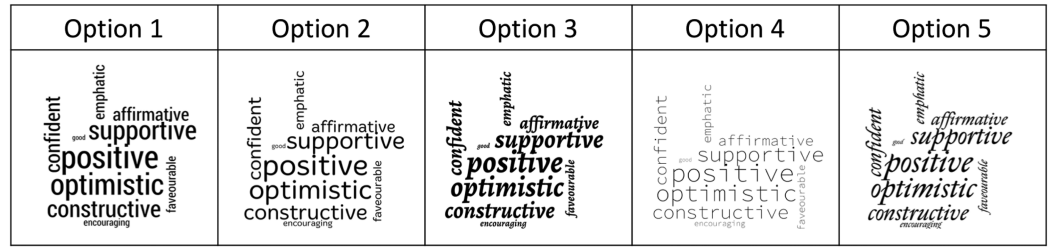


Fig. 7. A sample task from User Study II for the affect *positive*. The third and fifth images use congruent fonts, the first uses a neutral font, and the second and fourth ones use incongruent fonts. The participants are asked to pick the image that best represents the words in the word cloud.

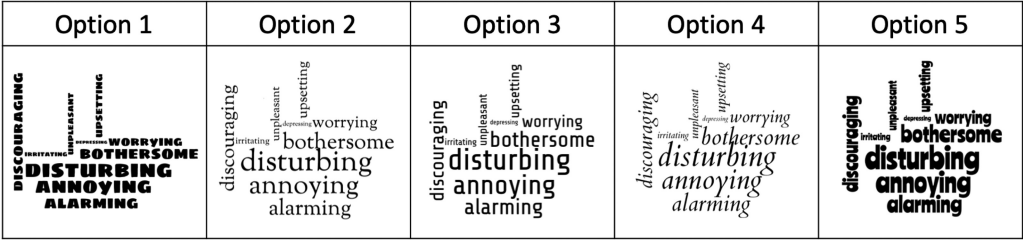


Fig. 8. A sample task from User Study II for the affect *disturbing*. The first and fifth images use congruent fonts, the third uses a neutral font, and the second and fourth ones use incongruent fonts. The participants are asked to pick the image that best represents the words in the word cloud.

neutral one, following the procedure for Study I. Figure 1 provides samples of word clouds with congruent fonts from this user study.

4.2 Results and Analysis

The frequencies of participant responses are visually presented in Figure 9, and chi-square goodness of fit test results are given in Table 2. Similar to Study I, results for each of the eight affective attributes are found to be statistically significant.

The results are in general consistent with those from the first study, as both show strong support for congruency with *serious* and *trustworthy*. Although the differences are less pronounced than earlier, across all attributes, the options with congruent fonts were chosen notably more frequently than incongruent ones, and the options with incongruent fonts were chosen less frequently than chance would predict, i.e., $\frac{2}{5} = 40\%$. It is observed that the congruent font options are more frequently selected for complex affects compared to core affects.

For *calm* and *positive*, neutral fonts were also chosen in many cases. As explained for Study I, and shown in Figure 3, larger numbers of fonts might appear somewhat calm or positive, and hence fonts in the middle of the ranked lists, which we assumed as *neutral*, may be more congruent.

Interestingly, *disturbing* received higher scores in this experiment compared to Study I. This may result from the layout of the word cloud, which mixes horizontal and vertical orientations, different font sizes, and different alignments, and, hence, in itself may already embody an appearance congruent with the notion of being *disturbing*. This circumstance may also explain the comparably lower scores for congruence with *calm* in comparison with Study I. We conclude that, in addition to the layout, font congruency as well merits significant consideration when designing word clouds.

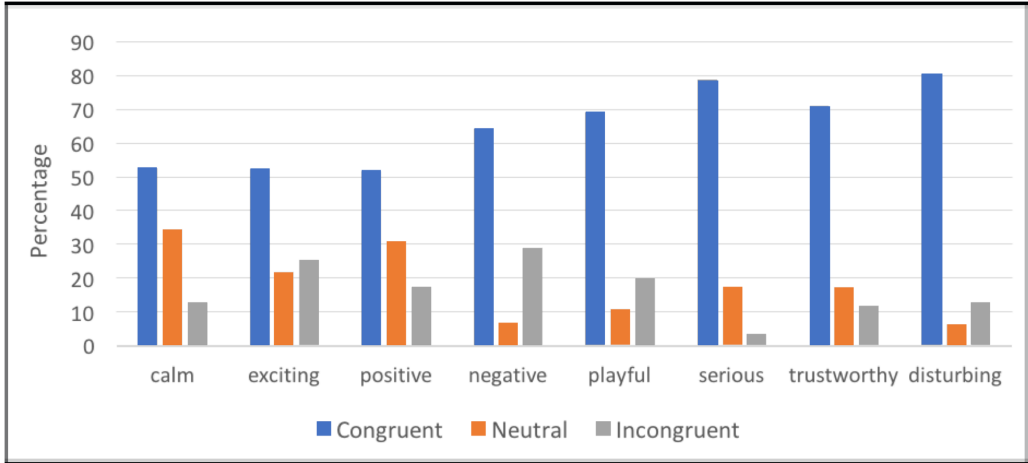


Fig. 9. Results of User Study II, providing percentages of user responses for congruent, incongruent, and neutral font choices in the word clouds. With uniformly random selections, the expected values for the congruent and incongruent options are 40%, while for neutral it is 20%.

5 AFFECTIVE COLOR PALETTES IN WORD CLOUDS

Apart from the choice of font, we further wish to understand the impact of affective color palettes on affective word clouds, specifically, to what extent users prefer color palettes that match the content of the word cloud affectively. To achieve this, we rely on the data from Bartram et al. [2] and carry out a user study using these palettes to create affective word clouds.

5.1 Data

The Bartram et al. study considers the same affect categories as this study, and their methodology to obtain the color palettes is as follows: First, 8,608 images tagged with one of the 8 affective categories are analyzed to find the most common colors for each category. Then, with the support from a visualization color expert, a set of 41 colors is determined, which combines the most representative colors for all 8 affects.

Using this set of colors, a user study is carried out requesting participants to design color palettes for one of the affect categories to be used in either a bar chart or a map visualization. The frequencies from this user study reveal the preferred colors for each affect, which are then used to create weights to come up with a palette weight concept. A weight for a palette is determined using the frequencies of the colors used in the palette. Finally, another user study is carried out to verify the results of the previous one, by generating palettes of different weights, and asking users to pick the best one. The results suggest that these weights indeed can be a good predictor of user preferences towards the color palettes.

We use color palettes from the second user study in their paper, which reveals the most preferred colors for each affect. Specifically, we rely on the palettes in Figure 7 from their paper [2] to determine the colors. Figure 10 in our article shows a congruent color palette sample for each of the eight affect categories. The color palettes for complex affects appear to follow a pattern based on their underlying core affects. As an example, the colors for *disturbing* seem to be a combination of colors for *negative* and *exciting*.

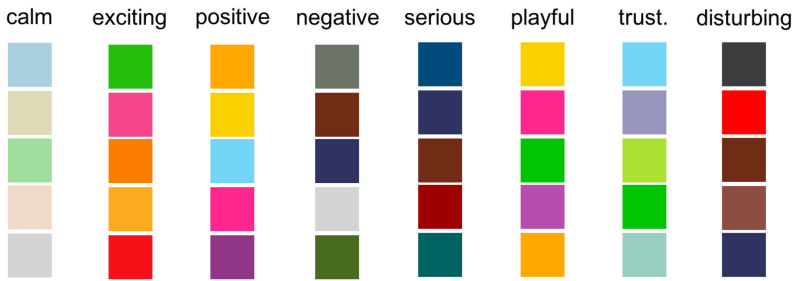


Fig. 10. Congruent color palette samples from Bartram et al. [2].

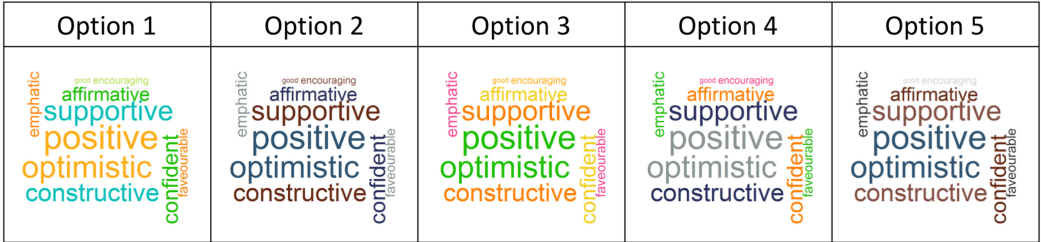


Fig. 11. A sample task from User Study III for the affect *positive*. The second and fifth images are generated using incongruent color palettes, the fourth one uses a mixed color palette, and the first and the third images use congruent color palettes. The users are asked to pick the image that best represents the words in the word cloud.

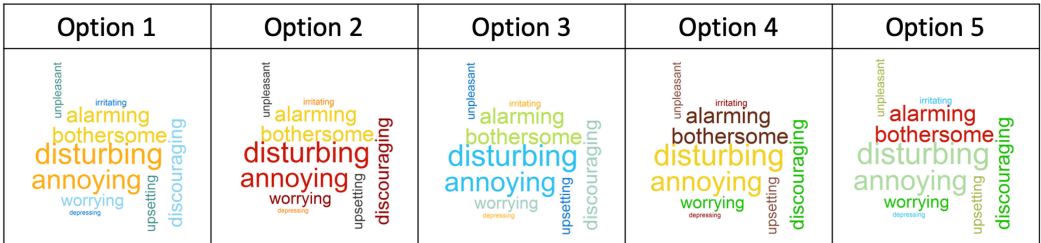


Fig. 12. A sample task from User Study III for the affect *disturbing*. The first and third images are generated using incongruent color palettes, the fifth one uses a mixed color palette, and the second and the fourth images use congruent color palettes. The users are asked to pick the image that best represents the words in the word cloud.

5.2 User Study III

Through a user study carried out on Mechanical Turk, we evaluate the impact of color palettes on word clouds with affective content.

Hypothesis. Word clouds using colors determined to be congruent are assessed as being more representative of pertinent affect-evoking words than word clouds using mixed or incongruent color palettes.

Participants and Method. The participant and method details are as in Study I and II (Sections 3.3 and 4.1), except that in this study, we use different color palettes in each option of the tasks, while keeping other signals, namely font and layout, the same across the presented choices. Sample tasks are shown in Figures 11 and 12. Two of the color palettes used in a task are *congruent*, two of them

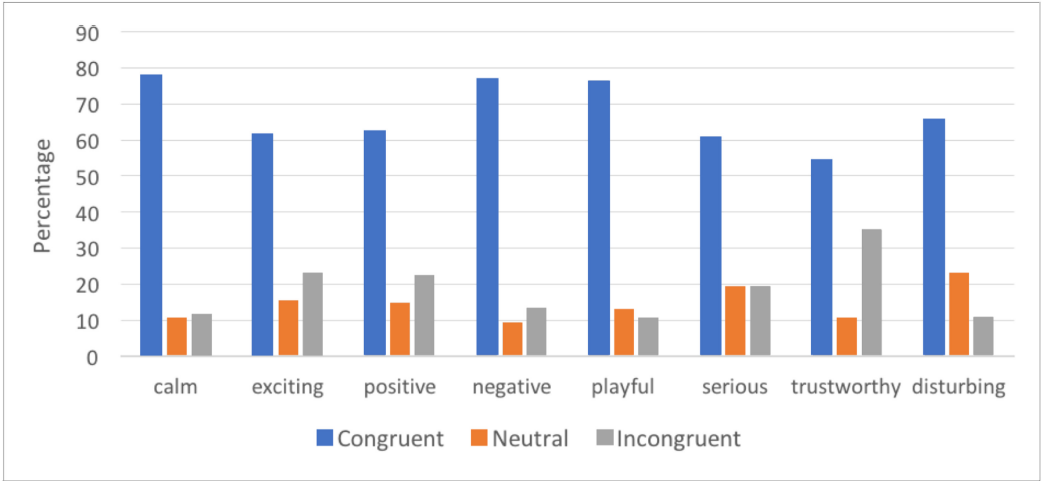


Fig. 13. Results of User Study III, providing percentages of user preferences for congruent, incongruent, and neutral color palette choices in the word clouds. For uniformly random selections, the respective expected values for the congruent, incongruent, and neutral options are 40%, 40%, and 20%.

incongruent, and one is *neutral*. The procedure to create the color palettes is as follows. We use Bartram et al.'s Figure 7 [2] to obtain a list of *congruent* colors, referred to as B_i , for each attribute i as defined in Table 1. To generate a *congruent* color palette for attribute i , we randomly pick five colors from B_i . Figure 1 provides samples of word clouds with congruent color palettes from this user study. To generate an *incongruent* color palette for attribute i , we randomly pick five colors from B_j , where j is the index of the opposite attribute. We use *exciting* vs. *calm*, *negative* vs. *positive*, *serious* vs. *playful*, and *disturbing* vs. *trustworthy* as opposite pairs. The color palette for *exciting*, e.g., provides incongruent colors for *calm*. To generate a *neutral* palette, we randomly pick two colors from B_i , two other colors from B_j , and one more color from $B_i \cap B_j$.

5.3 Results and Analysis

As summarized in Figure 13 and in Table 2, the results are consistent with our hypothesis: Similar to the results of our previous two user studies, results for each of the eight affective attributes are found to be statistically significant. This is particularly pronounced for *calm*, *negative*, and *playful* for User Study III, while the results appear comparably less strong for *exciting*, *serious*, *positive*, and *trustworthy*. These findings mirror those from Bartram et al. [2], in which *trustworthy* and *serious* were not strongly associated with specific colors, whereas the palettes for *calm* and *playful* had highly weighted colors, reflecting a strong preference for their respective affects.

6 AFFECTIVE FONTS AND COLOR PALETTES IN WORD CLOUDS

In Section 4 and Section 5, we carried out user studies to understand the effect of fonts and color palettes, respectively, on the affective impact of word clouds. Now we proceed to understand the relative affective power of fonts and color palettes on word clouds, and to what extent users prefer to combine these signals to create the intended affect in the word cloud. We use the same font and color datasets as described in the previous sections.

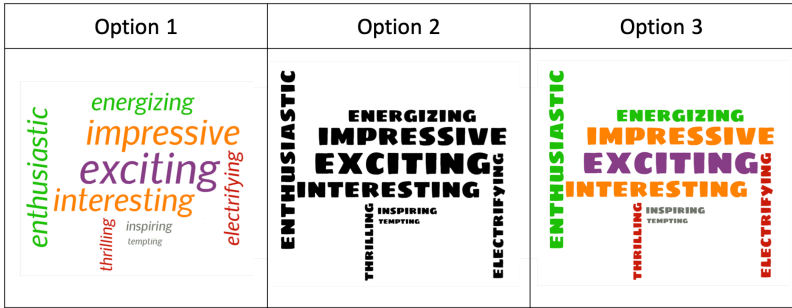


Fig. 14. A sample task from User Study IV for the affect *exciting*. The first option shows a word cloud with a congruent color palette and a neutral font. The second options shows a black-and-white word cloud with a congruent font. The third word cloud adopts the congruent color palette from the first option, and the congruent font from the second option.

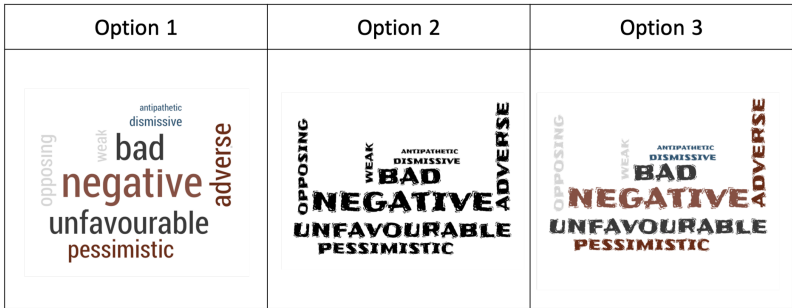


Fig. 15. A sample task from User Study IV for the affect *negative*. The first option shows a word cloud with a congruent color palette and a neutral font. The second options shows a black-and-white word cloud with a congruent font. The third word cloud adopts the congruent color palette from the first option, and the congruent font from the second option.

6.1 User Study IV

We carry out a user study on Mechanical Turk, with details similar to previous user studies. The main difference in this study is that this time we do not intend to test an hypothesis. Rather, the goal is to determine user preferences.

Participants and Method. We recruited 40 participants via Mechanical Turk, all from the United States, with at least 50 approved HITs and an overall approval rating of 95% or more. Participants were paid \$0.02 for each task. We rely on a within-subject design and perform counterbalancing by creating two identical assignments with a reversed task order and invoking the reversed task order for half of the participants. The study involves 27 tasks for each participant, consisting of 3 tasks for each of the 8 affective attributes (examples in Figures 14 and 15), and 3 additional validation tasks (see Figure 16). Each task includes 3 word clouds with the same content (10 affect-related words as described earlier) and layout. Each task includes the following three options to choose from:

- A word cloud with a congruent color palette and a neutral font (referred to as *the option with the congruent color*);

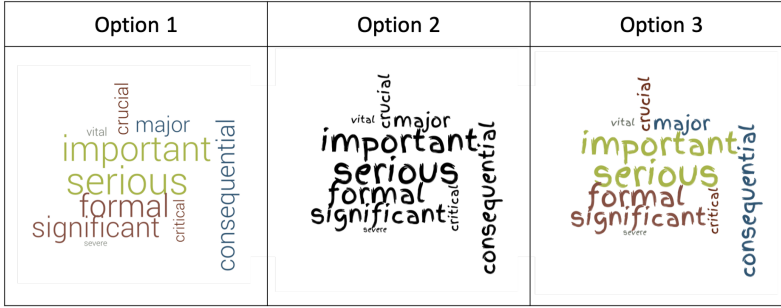


Fig. 16. A sample validation task from User Study IV. Different from the regular tasks in the experiment, this task uses an incongruent font in two of the options, leaving the option with just the congruent color palette as the only valid choice.

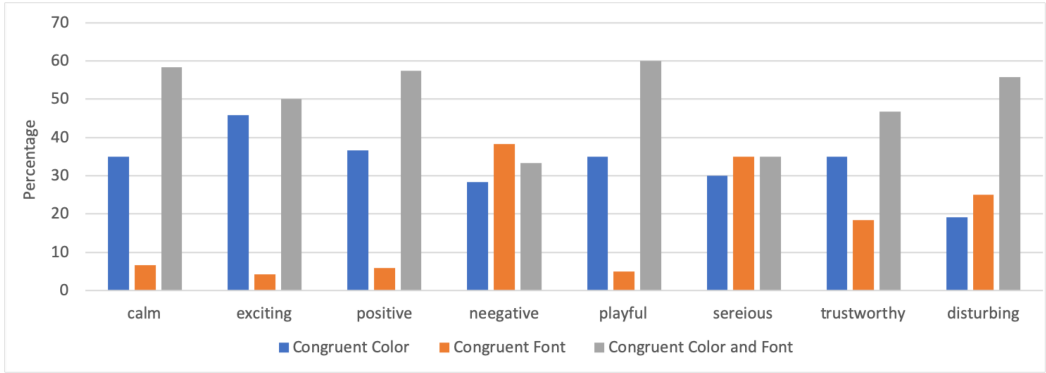


Fig. 17. Results of User Study IV, providing percentages of user preferences for congruent color with neutral font, black-and-white with congruent font, and congruent color palette with congruent font choices in the word clouds. For uniformly random selections, the expected value for each option is $\frac{1}{3}$.

- A black-and-white word cloud with a congruent font (referred to as *the option with the congruent font*);
- A word cloud with a congruent color palette and a congruent font (referred to as *the option with the congruent color and font*).

The order of these options are determined randomly for each task. For the details of font and color palette selection, please see Section 4.1 and Section 5.2, respectively. Participants were instructed to “Choose the image that best represents the words in the word cloud.”, with an additional detailed version given as “Select the image that you think best reflects the meaning of the words shown in the visualization.” From the original data, we excluded data from three participants who incorrectly answered all three validation tasks, as well as from two further participants who did not complete all tasks and incorrectly answered their validation questions. We recruited additional participants to extend the number of considered participants to 40 as planned.

6.2 Results and Analysis

The results are summarized in Figure 17 and in Table 2. Based on the results of the previous studies, one may reasonably expect the option with a congruent color palette and a congruent font (*Option 3*) to receive higher preference over the other two options, in which only one of these two



Fig. 18. A screenshot of our prototype Affect-Aware Word Cloud tool, visualizing an artificial placeholder text to show the affective power of fonts and colors.

signals are intentionally selected as congruent with the affect. The results for *calm*, *exciting*, *positive*, *playful*, *trustworthy* and *disturbing* seem to follow this pattern. Based on the χ^2 goodness of fit analysis summarized in Table 2, for *negative* and *serious*, the results do not meet the significance threshold 0.05, and are thus not considered statistically significant. This may stem from the association of the neutral font with these affects (leading users to select the option with the congruent color), as can be observed in Figure 15. Additionally, black-and-white colors might be a good fit for these affects (leading to users selecting the option with the congruent font). For all affects except *negative*, *serious*, and *disturbing*, the option with the congruent font was more frequently selected than the option with the congruent color. This may be interpreted as colors having a stronger bearing on the affect of a word cloud compared to fonts for the latter affects.

7 AFFECT-AWARE WORD CLOUD PROTOTYPE

In this section, we present our prototype tool that incorporates affects into the word cloud generation process by enabling the user to specify the intended affects such that congruent font and color palette options can be recommended. A screenshot of the tool is provided in Figure 18.

7.1 Implementation Details

7.1.1 User Input (Figure 18(a)). The tool takes as input a text, as well as the desired affective preferences. When no preference is provided, the input is considered to be *neutral*.

In the above user studies, we used a set of words that are synonymous with the affect word to prevent any confounding impact on the evaluation. However, on real-world data, multiple affects may be evoked within a single text. Hence, one challenge in font selection is to cope with such cases.

To overcome this, we introduce restrictions on possible combinations of the affects as follows: We group the affects in two groups: core affects (*calm*, *exciting*, *positive*, *negative*) and complex affects (*playful*, *serious*, *trustworthy*, *disturbing*). Using these groups, we only allow the following selections:

Positive	Calm & Positive	Exciting & Positive
positive	<i>calm+positive</i>	<i>exciting+positive</i>
POSITIVE	<i>calm+positive</i>	<i>exciting+positive</i>
positive	calm+positive	<i>exciting+positive</i>

Fig. 19. Samples of fonts with high scores for *positive* alone, and for its combinations with other core affects *calm* and *exciting*.

Negative	Calm & Negative	Exciting & Negative
negative	<i>calm+negative</i>	<i>EXCITING+NEGATIVE</i>
negative	calm+negative	<i>exciting+negative</i>
negative	calm+negative	<i>EXCITING+NEGATIVE</i>

Fig. 20. Samples of fonts with high scores for *negative* alone, and for its combinations with other core affects *calm* and *exciting*. The visual characteristics of the fonts are similar in a combination, whereas they are somewhat more diverse between different combinations.

- (1) A single complex affect;
- (2) A single core affect;
- (3) Two core affects, except for opposite pairs (*positive* and *negative*, *calm* and *exciting*).

7.1.2 Font Selection (Figure 18 (b)). In this part, we describe how we handle font selections for the different affect choices listed above. When a complex affect is chosen, we use the fonts that have the highest scores for that complex affect. When just a single core affect is selected, the system as well seeks out fonts that convey that affect. However, it makes the additional assumption that since the user did not select additional affects, the user presumably does not wish to convey those other affects. As an example, when only *positive* is selected, the intended affect is presumably one that does not encompass *calm* or *exciting*. As seen in Figure 3, a *positive* font could be *calm* or *exciting*, so simply using fonts with high *positive* values may not be an optimal solution. Thus, when only one core affect is selected, we filter fonts so as to only retain those that exhibit neutral behavior in the unspecified affect direction. Figure 19 shows some examples for the affect *positive*, which includes cases where neither *calm* nor *exciting* is selected along with it, or one of them is selected. Figure 20 provides similar examples for *negative*. Analyzing these figures in conjunction with Figure 2 reveals that the fonts that are congruent with multiple affects tend to combine characteristics of fonts for each of the participating affects. We also see that while having diverse visual characteristics in Figure 2, the top fonts for *negative* tend to be more similar when grouped by their scores for the unspecified affect dimension, as shown in Figure 20. The same is also true for *exciting*.

The final scenario is that of two core affects being selected. In such cases, for each font, we ensure that the values for each of the selected affects is above a certain threshold (e.g., 50%). If they are, then we use the average of these scores to determine the congruence of the font for this affect combination. Otherwise, we use the minimum of the affect values. This technique ensures that a high congruency score is indeed reflective of a font exhibiting characteristics of *both* selected

affects. If instead one were to generally consider the average, one could easily obtain fonts that reflect only one of the two affects very strongly, while having only a weak association with the other affect.

The actual font that is recommended by the tool is selected randomly among the top candidates using the above logic. The *Select* menu in the user interface, as marked by *Part b* in Figure 18, provides a list of recommended fonts to the user. Upon selecting a font, one may press the *Update* button to update the visualization to use the selected font. This facilitates exploring different designs for a specific affect.

7.1.3 Color Selection (Figure 18 (c)). Similar to the font selection methodology described above, in this part we explain how we handle color selections for the different affect choices. If only one affect is chosen, irrespective of whether it is a core affect or a complex one, we simply apply the corresponding color palette. If multiple core affects are selected, we use the corresponding complex affect, which is based on the relationships of emotions. Specifically, we use *trustworthy* for *calm* and *positive*, *serious* for *calm* and *negative*, *disturbing* for *exciting* and *negative*, and *playful* for *exciting* and *positive*. Reviewing Figure 7 from Bartram et al. [2] reveals that the color palettes for complex emotions are indeed very close to the combination of color palettes of their underlying core affects. Users can try different colors using the palettes provided in the user interface as marked by Figure 18(c).

7.1.4 Visualization (Figure 18 (d)). We preprocess the input to remove punctuation and so on as well as to obtain lemmatized versions of the words. To generate the word cloud, we rely on an external library⁵ that is developed using the D3 framework.⁶ The library allows for specifying preferred angle options for the words. It also allows for specifying a scale with which the sizes of the words are determined based on their distributions from the input. Based on the affect choices described earlier, we generate a word cloud with the automatically recommended font and color palette choices. Users have the option of trying further recommendations, or they can simply select any preferred fonts or colors.

7.1.5 Cloud Comparison (Figure 18 (e)). Our tool aims to encourage experimentation with different affective signals, as well as with different options for a specific affect. To facilitate this process, we provide the option to save a visualization in a panel on the right side of the user interface. This option obviates the need to try to remember previous designs, or save them to disk. Instead, one can rely on the comparison panel to observe the differences. In the screenshot provided in Figure 18, the comparison panel includes *playful*, *serious*, and *disturbing* versions of the input, respectively, from top to bottom. The rendering in the *Visualization* panel (Part d) shows the current visualization, which in this case uses *trustworthy* signals to visualize the same content.

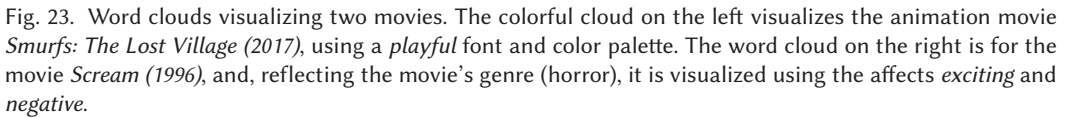
7.2 Output Samples

We showcase outputs of our prototype tool using data from several domains, while demonstrating the importance of affective visualizations for these domains.

7.2.1 Visualizing Restaurant Reviews. Word clouds are commonly used to visualize user-written reviews, such as of restaurants. A Google web search using keywords “restaurant review” and “word cloud” yields around eighteen thousand results, including academic papers [18]. Since a central goal of reviews is to convey the sentiment of a customer with respect to products or experiences, the affective perception of the word cloud is significant. Figure 21 portrays two word clouds generated

⁵<https://www.jasondavies.com/wordcloud/>.

⁶<https://d3js.org/>.



7.2.4 Visualizing Trustworthy Content. Another type of word cloud we wish to specifically analyze here are *trustworthy* ones. Word clouds are frequently used to describe the values of a corporation. Searches in websites providing professional design products, such as Shutterstock,⁸ return hundreds of word clouds specifically designed to reflect core values of corporations. A common goal of these corporations is to be perceived as *trustworthy*, which can be accounted for in their visualizations. Figure 24 provides a sample output of our tool using text⁹ related to the *United Nations*, visualized using the affective attribute *trustworthy*.

We discuss our results, their implications for users and use cases, and other potential research directions.

⁹The text is obtained from the Wikipedia page for the United Nations.

8.1 Affective Strengths of Fonts and Color Palettes

Across all experiments and attributes, congruent font and color palette choices were preferred by a plurality or majority, while incongruent choices were not preferred by a majority of responses. The findings in the first two user studies have shed light on the relationship between affective responses and fonts. For attributes such as *serious* and *trustworthy*, this relationship is found to be particularly strong. Interpreting these results together with the third user study, we observe that different signals exhibit different strengths in terms of their affective impact. Based on our experiments, color palettes prove particularly powerful for *calm*, *negative*, and *playful*. Thus, a *serious* or *trustworthy* perception appears easier to evoke with fonts, whereas a *calm* or *negative* appearance can arise from appropriate color palettes. For the remaining attributes, to achieve a more pronounced effect, a combination of both fonts and colors may be a compelling option.

8.2 Supporting Creativity

In a survey [48], 88% of users reported feeling creative when using Wordle, due to the use of font and color palette options that can be explored. Eighty-one per-cent of users reported they were trying it for fun. Hence, providing a medium in which users can feel creative ought to be an important aim of word cloud tools. We believe that allowing users to try applying different affects, or to try different options for the same affect, would substantially increase the creative potential to be explored by them, especially if they are given the chance to explore font and color palette options that are congruent with the affect(s) they specified. Nonetheless, a user study is needed to verify this hypothesis, since the users' sense of creativity is known to be hard to predict [50].

8.3 Word Clouds for Sentiment Analysis

Our output samples show that affective-aware choices of fonts are crucial for data from several domains. The same is more generally true of word clouds for sentiment analysis. There are many online resources, presentations, and academic papers [20, 33] that make use of word clouds for sentiment analysis, showcasing affective words that are used in the text. Most of them use a different word cloud for each emotion, or they use a comparison cloud to compare their intensities. In both cases, typographic signals could prove helpful to facilitate the perception of different emotions in sentiment visualizations.

8.4 Other Paralinguistic Signals and Visualizations

We have explored the affective usage of fonts and color palettes in word clouds. However, there are further affective signals to be explored in word clouds, as well as other text visualization methods to be investigated with respect to their incorporation of paralinguistic signals. For word clouds, one other such signal is the *layout*. We suspect that this could prove powerful, especially for attributes such as *disturbing* and *playful*. Again, this needs to be verified through a user study as well. In our experiments, in contrast, we keep the layout identical between different word cloud options to reduce the potential for confounding effects.

There are also several other visualizations that could benefit from an affect-aware approach. One line of such visualizations, as mentioned earlier, are visualization tools used for sentiment analysis. However, the list is not limited to this task. A tree map visualization, for instance, already conveys affect, since it relies on color palettes, and hence should perhaps be made affect-aware. In fact, for any visualization that makes use of fonts or color palettes, one may consider enabling such choices to be made more carefully and deliberately, in light of the affective nature of these signals.

8.5 Other Semantic Connections

Our study could be considered as a starting point towards exploring the connections between the semantics of the input and word clouds, or more generally, any visualization technique using the considered signals. More specific associations exist between fonts and semantic attributes [35], and these could be drawn upon to create word clouds with an even better thematic fit to the input. An example is using fonts found to be *technical* for *technical* content. Other more specific connections also exist between color and words, such as invoking the color *red* for the word *strawberry*, or *blue* for a word cloud relating to the *Smurfs*.

8.6 Regional and Cultural Differences

Previous studies reveal that regional and cultural differences affect color choices [15, 34]. A potential research direction is to enhance the word cloud tool to account for such differences based on user demographics. This might entail investigating whether such differences exist for font choices as well, especially with regard to the affective connections. Currently, in our tool, we provide users the opportunity to change the font and color choices without any restrictions, so that arbitrary personal preferences can be accommodated.

8.7 Input Sentiment Detection

Incorporating an automated sentiment detection method [11] or an emotion detection model [37, 43] into our tool could be a useful future direction. However, as discussed earlier, typographical signals are particularly useful for word clouds whose input includes additional modalities apart from text, such as music. Typographical signals help encode the affective nature of these additional input signals. In light of this, detecting the sentiment of a word cloud input using only its text may not be an optimal solution. This problem can perhaps be observed in the automated sentiment detection for the movie *Smurfs*. The genre of the movie is animation, and in general it has a *playful* nature (e.g., as easily revealed by its music). However, looking at its plot summary text in isolation may lead to inaccurate predictions, especially in the absence of further background information about the movie and animation series. Hence, for the case of affect-aware word clouds, obtaining affect preferences from users, or building a tool that accepts additional inputs based on the domain of interest (e.g., music for songs, the genre for movies, ratings for restaurant reviews) could be better solutions.

9 CONCLUSION

In this article, we study the affective connections of fonts and color palettes, specifically in the domain of word clouds. We have invoked a set of techniques to obtain font congruency values for several affective attributes based on a crowdsourced seed set. The results of our studies establish that such semi-automatically acquired font scores accord with human assessments of congruence, similar to previous studies that relied on human-chosen fonts.

Our findings reveal that both fonts and color palettes are potent signals in creating affective word clouds. Moreover, their respective strengths turn out to be complementary. Our last user study revealed that these signals should be used in conjunction to yield a stronger impact. Based on this work, we can conclude that fonts may be used as an additional dimension in visualizations to intentionally encode affect, and not only designers but also developers of computational tools need to account for the possibilities afforded by font and color associations with affective attributes.

ACKNOWLEDGMENTS

We thank Alex Borgida, Nil Tugce Kulahcioglu, and Min-Jeong Yang for helpful comments and discussions.

REFERENCES

- [1] Eric Alexander, Chih-Ching Chang, Mariana Shimabukuro, Steven Franconeri, Christopher Collins, and Michael Gleicher. 2018. Perceptual biases in font size as a data encoding. *IEEE Transactions on Visualization and Computer Graphics* 24, 8 (2018), 2397–2410.
- [2] Lyn Bartram, Abhisekh Patra, and Maureen Stone. 2017. Affective color in visualization. In *Proceedings of CHI 2017*. ACM, 1364–1374.
- [3] Sam Berlow, Stephen Coles, and Nick Sherman. 2017. Fonts In Use. <https://fontsinuse.com/>. Accessed: 2017-07-29.
- [4] Faber Birren. 2016. *Color Psychology and Color Therapy; A Factual Study of the Influence of Color on Human Life*. Pickle Partners Publishing.
- [5] Richard Brath and Ebad Banissi. 2016. Typographic sets: Labelled set elements with font attributes. In *Proceedings of the SetVR 2016 Workshop at Diagrams*. 29–43.
- [6] Michael Burch, Tobias Fluck, Julian Freund, Thomas Walzer, Uwe Kloos, and Daniel Weiskopf. 2016. Lyrics word clouds. In *Proceedings of the 20th International Conference on Information Visualisation (IV)*, 2016. IEEE, 51–56.
- [7] Michael Burch, Steffen Lohmann, Fabian Beck, Nils Rodriguez, Lorenzo Di Silvestro, and Daniel Weiskopf. 2014. RadCloud: Visualizing multiple texts with merged word clouds. In *Proceedings of the 18th International Conference on Information Visualisation (IV)*, 2014. IEEE, 108–113.
- [8] Terry L. Childers and Jeffrey Jass. 2002. All dressed up with something to say: Effects of typeface semantic associations on brand perceptions and consumer memory. *Journal of Consumer Psychology* 12, 2 (2002), 93–106.
- [9] Saemi Choi, Kiyoharu Aizawa, and Nicu Sebe. 2018. FontMatcher: Font image paring for harmonious digital graphic design. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces (IUI '18)*. ACM, New York, 37–41. DOI : <https://doi.org/10.1145/3172944.3173001>
- [10] Weiwei Cui, Yingcai Wu, Shixia Liu, Furu Wei, Michelle X. Zhou, and Huamin Qu. 2010. Context preserving dynamic word cloud visualization. In *Proceedings of the 2010 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 121–128.
- [11] Xin Dong and Gerard de Melo. 2018. A helping hand: Transfer learning for deep sentiment analysis. In *Proceedings of ACL 2018*. Association for Computational Linguistics, 2524–2534. DOI : <https://doi.org/10.18653/v1/P18-1235>
- [12] Cristian Felix, Steven Franconeri, and Enrico Bertini. 2018. Taking word clouds apart: An empirical investigation of the design space for keyword summaries. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 657–666.
- [13] Alysha Fligner. 2013. *The Effect of Packaging Typeface on Product Perception and Evaluation*. Ph.D. Dissertation. The Ohio State University.
- [14] Richard L Hazlett, Kevin Larson, A. Dawn Shaikh, and Barbara S. Chaparo. 2013. Two studies on how a typeface congruent with content can enhance onscreen communication. *Information Design Journal* 20, 3 (2013).
- [15] Johannes Itten. 1970. *The Elements of Color*. John Wiley & Sons.
- [16] Jaemin Jo, Bongshin Lee, and Jinwook Seo. 2015. WordlePlus: Expanding Wordle’s use through natural interaction and animation. *IEEE Computer Graphics and Applications* 35, 6 (2015), 20–28.
- [17] Samuel Juni and Julie S. Gross. 2008. Emotional and persuasive perception of fonts. *Perceptual and Motor Skills* 106, 1 (2008), 35–42.
- [18] David Kamerer. 2014. Understanding the Yelp review filter: An exploratory study. *First Monday* 19, 9 (2014).
- [19] Kyle Koh, Bongshin Lee, Bohyoung Kim, and Jinwook Seo. 2010. Maniwordle: Providing flexible control over Wordle. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1190–1197.
- [20] Kostiantyn Kucher and Andreas Kerren. 2014. Text visualization browser: A visual survey of text visualization techniques. *Poster Abstracts of IEEE VIS 2014* (2014).
- [21] Tugba Kuglacioglu and Gerard de Melo. 2020. Semantics-aware typographical choices via affective associations. *Language Resources and Evaluation* (2020).
- [22] Tugba Kulahcioglu and Gerard de Melo. 2018. FontLex: A typographical lexicon based on affective associations. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*. ELRA.
- [23] Tugba Kulahcioglu and Gerard de Melo. 2018. Predicting semantic signatures of fonts. In *Proceedings of the 2018 IEEE 12th International Conference on Semantic Computing (ICSC)*. IEEE, 115–122.
- [24] Tugba Kulahcioglu and Gerard de Melo. 2019. Paralinguistic recommendations for affective word clouds. In *Proceedings of ACM IUI 2019*. ACM, New York., 132–143. DOI : <https://doi.org/10.1145/3301275.3302327>
- [25] Lauren I. Labrecque and George R. Milne. 2012. Exciting red and competent blue: The importance of color in marketing. *Journal of the Academy of Marketing Science* 40, 5 (2012), 711–727.
- [26] Clive Lewis and Peter Walker. 1989. Typographic influences on reading. *British Journal of Psychology* 80, 2 (1989), 241–257.
- [27] Dongying Li and Xiaolu Zhou. 2017. “Leave your footprints in my words” –a georeferenced word-cloud approach. *Environment and Planning A* 49, 3 (2017), 489–492.

- [28] Steffen Lohmann, Florian Heimerl, Fabian Bopp, Michael Burch, and Thomas Ertl. 2015. Concentri cloud: Word cloud visualization for multiple text documents. In *Proceedings of the 19th International Conference on Information Visualization (IV), 2015*. IEEE, 114–120.
- [29] Colin M. MacLeod. 1991. Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin* 109, 2 (1991), 163.
- [30] Albert Mehrabian and James A. Russell. 1974. *An Approach to Environmental Psychology*. MIT Press.
- [31] Ravi Mehta and Rui Juliet Zhu. 2009. Blue or red? Exploring the effect of color on cognitive task performances. *Science* 323, 5918 (2009), 1226–1229.
- [32] Hayden Mills. [n.d.]. Font Pair. <http://fontpair.co/>. Accessed: 2017-07-29.
- [33] Saif M. Mohammad. 2012. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems* 53, 4 (2012), 730–741.
- [34] Jantawan Noiwan and Anthony F. Norcio. 2006. Cultural differences on attention and perceived usability: Investigating color combinations of animated graphics. *International Journal of Human-computer studies* 64, 2 (2006), 103–122.
- [35] Peter O'Donovan, Janis Libeks, Aseem Agarwala, and Aaron Hertzmann. 2014. Exploratory font selection using crowdsourced attributes. *ACM Trans. Graph.* 33, 4 (2014), 92:1–92:9.
- [36] Jack Qiao. [n.d.]. Fontjoy – Generate font combinations with deep learning. <http://fontjoy.com/>. Accessed: 2017-07-29.
- [37] Shahab Raji and Gerard de Melo. 2020. What sparks joy: The affectvec emotion database. In *Proceedings of the Web Conference 2020*. ACM, New York, 2991–2997. DOI : <https://doi.org/10.1145/3366423.3380068>
- [38] Johann Schrammel, Michael Leitner, and Manfred Tscheligi. 2009. Semantically structured tag clouds: An empirical evaluation of clustered presentation approaches. In *Proceedings of CHI 2009*. ACM, 2037–2040.
- [39] A. Dawn Shaikh. 2007. The effect of website typeface appropriateness on the perception of a company's ethos. *Usability News* 9, 2 (2007).
- [40] Audrey Dawn Shaikh. 2007. *Psychology of Onscreen Type: Investigations Regarding Typeface Personality, Appropriateness, and Impact on Document Perception*. Ph.D. Dissertation.
- [41] A. Dawn Shaikh, Barbara S. Chaparro, and Doug Fox. 2006. Perception of fonts: Perceived personality traits and uses. *Usability News* 8, 1 (2006), 1–6.
- [42] A. Dawn Shaikh, Doug Fox, and Barbara S. Chaparro. 2007. The effect of typeface on the perception of email. *Usability News* 9, 1 (2007), 1–7.
- [43] Abu Awal Md Shoeb and Gerard de Melo. 2020. Are emojis emotional? A study to understand the association between emojis and emotions. *ArXiv* 2005.00693 (2020). <https://arxiv.org/abs/2005.00693>.
- [44] Satyendra Singh. 2006. Impact of color on marketing. *Management Decision* 44, 6 (2006), 783–789.
- [45] Julie Steele and Noah Iliinsky. 2010. *Beautiful Visualization: Looking at Data Through the Eyes of Experts*. O'Reilly Media, Inc.
- [46] Carlos Velasco, Alejandro Salgado-Montejo, Fernando Marmolejo-Ramos, and Charles Spence. 2014. Predictive packaging design: Tasting shapes, typefaces, names, and sounds. *Food Quality and Preference* 34 (2014), 88–95.
- [47] Carlos Velasco, Andy T. Woods, Sarah Hyndman, and Charles Spence. 2015. The taste of typeface. *i-Perception* 6, 4 (2015).
- [48] Fernanda B. Viegas, Martin Wattenberg, and Jonathan Feinberg. 2009. Participatory visualization with Wordle. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009).
- [49] Ji Wang, Kyle D. Dent, and Chris L. North. 2013. Fisheye word cloud for temporal sentiment exploration. In *Proceedings of CHI 2013 Extended Abstracts*. ACM, 1767–1772.
- [50] Yunhai Wang, Xiaowei Chu, Chen Bao, Lifeng Zhu, Oliver Deussen, Baoquan Chen, and Michael Sedlmair. 2018. EdWordle: Consistency-preserving word cloud editing. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 647–656.
- [51] Alan J. Wecker, Joel Lanir, Osnat Mokryn, Einat Minkov, and Tsvi Kuflik. 2014. Semantize: Visualizing the sentiment of individual document. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. ACM, 385–386.
- [52] T. W. Whitfield and T. J. Whiltshire. 1990. Color psychology: A critical review. *Genetic, Social, and General Psychology Monographs* (1990).
- [53] Jin Xu, Yubo Tao, and Hai Lin. 2016. Semantic word cloud generation based on word embeddings. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis), 2016*. IEEE, 239–243.

Received September 2019; revised December 2019; accepted March 2020