MetaMap: Supporting Visual Metaphor Ideation through Multi-dimensional Example-based Exploration

Youwen Kang*
The Hong Kong University of Science and Technology
ykangae@connect.ust.hk

Zhida Sun*
The Hong Kong University of Science and Technology
zhida.sun@connect.ust.hk

Sitong Wang†
Columbia University
sw3504@columbia.edu

Zeyu Huang
The Hong Kong University of Science and Technology
zhuangbi@connect.ust.hk

Ziming Wu
The Hong Kong University of Science and Technology
zwual@connect.ust.hk

Xiaojuan Ma‡
The Hong Kong University of Science and Technology
mxj@cse.ust.hk

ABSTRACT

Visual metaphors, which are widely used in graphic design, can deliver messages in creative ways by fusing different objects. The keys to creating visual metaphors are diverse exploration and creative combinations, which is challenging with conventional methods like image searching. To streamline this ideation process, we propose to use a mind-map-like structure to recommend and assist users to explore materials. We present MetaMap, a supporting tool which inspires visual metaphor ideation through multi-dimensional example-based exploration. To facilitate the divergence and convergence of the ideation process, MetaMap provides 1) sample images based on keyword association and color filtering; 2) example-based exploration in semantics, color, and shape dimensions; and 3) thinking path tracking and idea recording. We conduct a within-subject study with 24 design enthusiasts by taking a Pinterest-like interface as the baseline. Our evaluation results suggest that MetaMap provides an engaging ideation process and helps participants create diverse and creative ideas.

CCS CONCEPTS

• **Human-centered computing → Interactive systems and tools**.

KEYWORDS

Design; Visual Metaphor; Creativity Support Tool

ACM Reference Format:
Youwen Kang, Zhida Sun, Sitong Wang, Zeyu Huang, Ziming Wu, and Xiaojuan Ma. 2021. MetaMap: Supporting Visual Metaphor Ideation through Multi-dimensional Example-based Exploration. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3411764.3445325

*These authors contributed equally to this work.
†The work was done during Sitong's internship at the Hong Kong University of Science and Technology.
‡The corresponding author.

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445325

1 INTRODUCTION

Visual metaphor is a powerful [13] and intriguing [38] means to communicate ideas or concepts, which is widely used in graphic design (*e.g.,* advertising and editorial design). Unlike linguistic metaphors, visual metaphors deliver messages through symbolism by visually fusing different objects [8, 38]. A good visual metaphor can explain its underlying meaning in a resourceful way, and further, stimulate reflection among the audience through a salient relationship between the visual representations and the conveyed message [16]. Previous research has come up with various definitions of visual metaphor by dissecting its effects in design [29] with aspects ranging from aesthetic education [1] to art [16]. In this paper, we take a general view of visual metaphor and define it as visual representations which convey particular associations by fusing different elements that share similar points in various dimensions (*e.g.,* semantics, colors, shapes) [7, 40]. For example, in Figure 1, (a) is a Nestle coffee advertisement, a typical visual metaphor. It replaces the body of an alarm clock with a cup of coffee to deliver the message of "Nestle coffee wakes you up". Audiences can effectively understand the visual metaphor because "coffee" and "alarm clock" share similarities in various dimensions. At the semantic level, coffee's primary function is to wake people up in the morning, so does an alarm clock. Moreover, the alarm clock and the cup are circular with Nestle's red color to strengthen the connection. Thus, the visual metaphor with elements sharing multi-dimensional features can quickly deliver the core message.

Creating good visual metaphors is not a trivial task for designers, especially for amateur designers who have design experience but do not receive professional training on visual design and design thinking methodology. An act of creation often requires designers to bridge the gap between the new and the familiar, and generate an "a-ha" effect by presenting "off-the-wall" ideas. In this process, taking inspiration from existing examples is commonplace [18].

[1]NESCAFÉ, print ads design.
[2]Lori Miller, *Global Warming Awareness*, poster campaign.
[3]Pei-Ling Ou, *Freedom of Speech*, red-dot.
[4]John Holcroft, a satirical illustration.

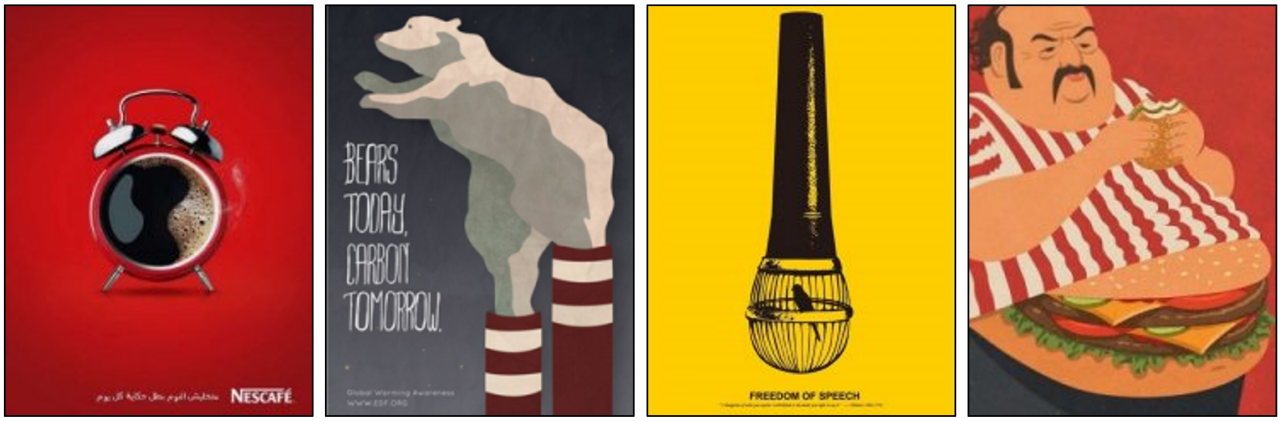

Figure 1: Thumbnail examples of visual metaphors: (a) Commercial advertisement from Nescafe Print Ad[1]; (b) Public service announcement for environment protection[2]; (c) Creative poster for freedom of speech[3]; and (d) Satirical illustration to criticize unhealthy eating habits[4].

Prior research suggests that designers can better understand the possible design space and the potential means to realize design ideas by exploring examples [15]. Designers also find it more efficient to repurpose related materials from previous creation than constructing everything from scratch [4, 48]. It is thus an established routine for many designers to collect exemplars during the ideation stage of a design task from online resources including but not limited to image search engines (*e.g.,* Google Image Search) and design sharing communities (*e.g.,* Pinterest). This, however, can be very time-consuming [49]. Human-computer interaction researchers have designed various tools to facilitate example searching in domains such as website design [27], app design [22], and logo design [55]. These tools propose to recommend items based on style [43] or similarity in layout to users' descriptions presented in the form of text or sketch [22, 43], or to some previous designs liked by users [54].

Nevertheless, existing tools built for other types of designs may not sufficiently support example exploration during visual metaphor ideation. For one thing, designers, especially amateurs, usually do not have a concrete idea in the early design stage, and it is difficult to articulate their thoughts in words [4]. Ideation is even more challenging for visual metaphor because it is often employed to illustrate concepts that are poorly or vaguely represented by a verbal description [8]. For another, the core of visual metaphor is an innovative fusion of associated elements [8], which cannot be simply characterized as a style or layout. Existing works that support visual metaphor design focuses primarily on the prototyping stage (*e.g.,* [11]), and they require a lot of mental efforts to conduct image search and information collection [26]. There is still a pressing demand for effective tools to assist in visual metaphor ideation – the basis for the remaining design process [35].

This paper presents MetaMap, an ideation support tool for inspiring visual metaphor ideas and streamlining their development through multi-dimensional example exploration. We first conduct a formative study by interviewing nine designers from various backgrounds who have created visual metaphors in their previous design activities. We can then understand their practices, barriers, and needs when designing visual metaphors. According to our findings from a formative study as well as related exploration on a design element framework [32, 55], we then propose to support amateur designers' ideation process by recommending exemplars from three dimensions (*i.e.,* semantics, color, shape). These are imperative for establishing the association, similarity, or analogy between design elements [8], and to organize these examples in a mind-map like manner. This is the core design of MetaMap. To facilitate divergence and convergence activities in the ideation process, MetaMap provides 1) assembly of image search results based on automatic keyword association and color theme filtering; 2) exploration of examples related to a seed image from the semantics, color, and shape dimensions; and 3) support for thinking paths tracking and ideas recording. Using MetaMap, designers can continuously expand and iteratively refine their ideas with inspirations from prior works. Next, we evaluate MetaMap regarding the outcome satisfaction, engagement level, and usefulness through a within-subjects study with 24 design enthusiasts by taking a Pinterest-like searching interface as the baseline. Each participant joined two separate and counterbalanced ideation sessions to brainstorm for visual metaphors that can be applied in an activity poster design; one session's theme was health and the other was about music. We collected 185 ideas in total and obtained quantitative and qualitative feedback about the participants' experience. Our evaluation results suggest that our design enthusiasts perceive MetaMap to be significantly more useful for ideation than the baseline due to its support for engaging interaction, diverse exploration, and a trackable thinking path. Analysis of the output ideas also shows that our tool indeed helps amateur designers generate more diverse and creative visual metaphors than the baseline. Furthermore, we identify different ideation patterns among the participants from interview

responses and discuss how MetaMap can be further improved to support varying users' needs.

Overall, the main contributions of this work are three-fold:

- We propose a three-dimensional recommendation framework (*i.e.,* semantics, color, shape) utilizing both image and keyword information for visual metaphor example exploration;
- We present the MetaMap, an interactive system for users to inquire into visual metaphor design materials and organize them in a mind-map-like layout to track users' thinking paths;
- We evaluate the efficacy of MetaMap through a user study with amateur designers and provide an in-depth understanding of whether and how designers gain inspiration from multi-dimensional example-based exploration.

2 RELATED WORK

2.1 Design based on Exemplars

Previous research suggests that, by exploring a large amount of exemplars, designers can gain a better understanding of possible design space and potential means to implement the design ideas [15]. Example-based exploration is being found as an efficient way to create new ideas than constructing everything from scratch [48]. By receiving timely stimuli from exemplars, designers can generate more ideas than brainstorming by themselves [49]. However, when people have a strong tendency to rely on existing knowledge elements and exemplars before coming up with their new ideas [24], the problem of design fixation appears. In addition, the problem can not be simply solved by giving irrelevant inspirations, which is also proved to reduce the ideation efficiency [10]. Thus, it is essential to keep the example-based recommendation inner-related but also diverse enough to support designers' creativity [37].

2.2 Creativity with Association

Psychologists have long been interested in investigating the relationship between human creativity and association [20, 34, 45]. Though creativity is a complex activity, it can be described as consisting of evocation and a recombination of previous knowledge to generate new properties based on previous ones [4, 53]. However, a strong ability of association is required to generate creative ideas rather than imitating previous work [6]. According to the prior experiment [4], professional designers have acquired a stronger fluency than novices in the use of analogical reasoning to build up connections between diverse concepts. Thus, amateur designers need more robust support in their evocation process to create a new combination of concepts, which is essential for visual metaphor creation.

Evocation is a physiological concept indicating the intuitive association between concepts based on human perception. It measures how one concept evokes another in the human mind. The connection between concepts has different strengths with specific directions. For example, "beer" strongly evokes "glass", but the "glass" to "beer" connection has a lower evocation strength for most people. Through collecting responses after a cue is given, free word association experiments provide clues to reflect the evocation result of how people store and retrieve concepts in memory [33]. Researchers have collected and published free association datasets [14] to help generalize the evocation process for general crowds. The small world project[5] is currently the largest free word association resource in English with over 12,000 cue words, which provide us with a solid database to support people's conceptual level evocation. In addition, facilitating the evocation process at the concept level is essential for creative design [4, 24].

2.3 Design with Visual Metaphors

Visual metaphors are widely used in graphic design (*e.g.,* advertising and editorial design), to deliver messages in intriguing ways by implying it through the symbolism and juxtaposition of the symbols [38]. In Phillips and McQuarrie's pioneering work, the embedded associations are analyzed from dimensions of visual structure and meaning operation [39]. Analogous to verbal metaphors, which link two different words or concepts together, visual metaphors incorporate multiple objects which are disparate in real life, in one image space. However, visual metaphors do not essentially involve the interaction of words; they mobilize the interaction of concepts which includes all sorts of information [8]. In our paper, we define the visual metaphor as being a general analogy representation by fusing two objects to deliver an implicit message.

The analogy between the two objects can involve interactions of various categories and concepts, which are not limited to their semantic relevance. In addition to semantic analogy, color [25] and shape [51] are important visual information to establish a metaphorical relationship. In a visual image, the color distribution brings a specific atmosphere to audiences [32]. Previous work investigates the semantic meaning of colors [23, 25, 54], and try to build up the connections between colors and semantics. Observations from designers' behaviors [25] also suggest that color features are useful to represent objects and deliver underlying semantic messages. Besides, shape features have a critical role in visual metaphor. For designers, they seek elements with similar shapes for potential visual blending [11, 32]. For audiences, more commonalities in shapes also facilitate their understanding of the semantic messages behind. The overlapping visual features will evoke the semantic knowledge behind two objects [51]. For example, (c) in Figure 1, the similar shapes between a microphone and a birdcage are used to express the freedom of speech; even when there is no obvious semantic connection between a birdcage and a microphone.

2.4 Creativity Support Tool

As the initial step for design, ideation has received much scholarly attention in the creativity support field [17]. Numerous computational pipelines have been constructed to support the idea generation progress [2, 11, 26, 32, 46, 55], which incorporate various design elements that are significant to designers. For example, by considering the factors of semantics, style, and space, Zhao *et al.* [55] built an automatic system of generating icon suggestions to support compound icon ideation; Secondary features of the image - color, shape, and composition - are also identified as crucial aspects in the early stages of the design process by [32]. Furthermore, different forms of ideation support tools are adopted as well. By comparing three modes of visual stimulus, Shi *et al.* [46] find dynamic cells - a

[5]Link to the small world project: https://smallworldofwords.org/en/project/home.

clustering structure based on associations - the preferred one in collaborative brainstorming. Recent work in aiding creative ideation [2] adopts mind maps to mimic the process of making associations. As a visual collage of useful materials in the ideation stage, mood board creation is also enhanced with computational ability in [26].

Pioneering work supporting visual blending [11] provides a design workflow for creating visual blending, which focuses more on the implementation step. The system can generate a blending result given two symbolic images. However, the brainstorming and material searching process still follows conventional methods of image searching. Nevertheless, existing tools may not be able to sufficiently support example exploration during visual metaphor ideation.

3 FORMATIVE INTERVIEW

In order to understand the process of visual metaphor creation and the difficulties users face, we conducted semi-structured interviews with nine participants from various design backgrounds. We include both amateur and professional designers in the formative study, and distinguish them by checking whether they have acquired professional training or degrees in the visual design area. All of them have at least some experience in using visual metaphors in their design work. The semi-structured interviews are designed to have a direct conversation with designers and investigate their needs in ideation practices under real-world scenarios. Participants are invited to showcase their work and demonstrate the process of creation. In spite of lacking a strict control over the design process, the interviews on real-world showcases help us learn the specific scenario and stage to focus on before deriving a specific design task.

3.1 Interview Setup and Process

We recruited nine participants (7 female, 2 male, and aged 20-30) with at least some prior experiences in visual metaphor design. Three of them have received professional visual design education as their major (*i.e.,* visual communication, advertising design). Others are amateur designers having used visual metaphors in their posters. Participants provided 2-3 examples from their prior work containing visual metaphors and shared the creative process. Some of the examples are shown in Fig 2. Interview questions cover how they came up with their initial ideas, where they got their inspiration, and how they prototyped and implemented them. They were also asked to identify any difficulties they encountered during the process and raise their needs to support visual metaphor creation. Finally, we asked them an open question to get a better sense of what would pop into their minds in the beginning stages, given a specific topic. The questions are designed to let participants articulate their needs during the process of design creation.

3.2 Findings

During the interview, all participants would iteratively go through three general steps: ideation, elements collection/making, and implementation. Ideation was identified as the most challenging (8/9) and time-consuming (5/9) step for them. Five of them mentioned that *"Ideation is an exhausting process requiring diverse inspiration and iterative refinement."*. In addition, two amateur designers emphasized it is difficult for those lacking in professional design training to brainstorm systematically. We further decompose the ideation process and address their difficulties and needs in 3 sub-steps: concept association, exemplar searching, and prototyping. For instance, (a) Figure 2 fuses the reflection of a flamingo with an eighth note (Flamingo is the theme of a singing competition). In the concept association stage, the designer (an amateur designer in college) thought of many related concepts and elements related to music (*e.g.,* notes, guitar, stave). The designer then searched for different exemplars with the above concepts and found the shape similarity between the flamingo and the eighth note. When the flamingo reflects music note, it becomes the companion with music. Finally, the message "You are not alone with music" becomes embedded. As we can see, after the concept-level association, this designer discovered an unexpected metaphorical relationship between two objects according to the analogy in shapes and semantics during the process of exemplar searching. Of course, designers would also iteratively go through the three stages to polish ideas and generate more during the ideation process.

3.2.1 Concept Association. This is the step of thinking of concepts related to the central topic. Most of them reported that it is easy to get stuck in the concept level and not know what to search for later. One amateur designer also added *"Brainstorming always begins from a very abstract concept, it is challenging to think of concrete representations"*. Five participants suggested that it would be helpful to provide them with association concepts for reference.

3.2.2 Exemplar Search. This stage was identified as a necessary step by all the participants to be creative and learn from others. Some popular websites for exemplar searching frequently mentioned by participants are Pinterest (7/9), Google (4/9), and Behance (3/9). Pinterest is reflected as the most popular tool to find exemplars. They all look for diverse stimuli from others' work, including content, color usage, style, composition, and the like. New ideas pop into their minds when being exposed to rich information. However, when talking about visual metaphor technique, five amateur designers reported little knowledge about it, though they have used them unintentionally in prior work. Six participants suggested a diverse image recommendation would be helpful. Also, two participants mentioned that they would easily get lost in image searching. Due to the lack of exploration tracking, they sometimes forget where their ideas come from and feel lost when they want to check previous results.

3.2.3 Prototyping. This stage needs designers to form and record their ideas, which relies more on the designers' side. It requires them to trace back to the thinking path and outline details of their ideas to make them more concrete. Participants would choose hand-drawing (5/9) or Photoshop (3/9).

4 SYSTEM DESIGN AND IMPLEMENTATION

4.1 Design Requirements

To support the process of visual metaphor ideation, we present our system - MetaMap. According to our literature review and formative study, amateur designers who lack knowledge and experience

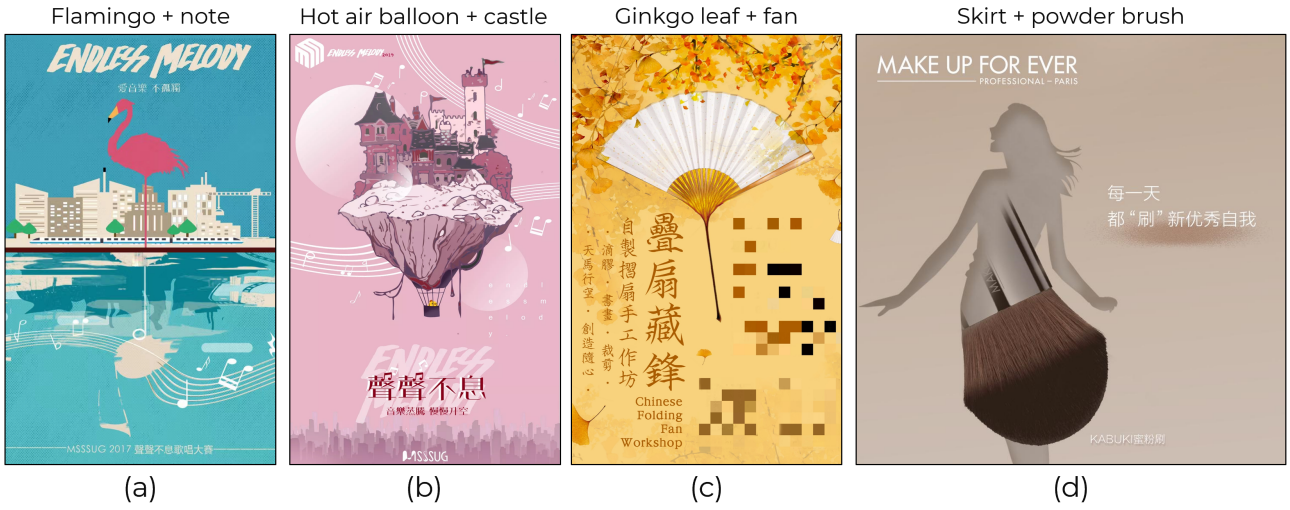

Figure 2: Some design work collected from interviewees: (a) and (b) are posters for college singing competitions (MSSSUG, HKUST). (a) merges the reflection of a flamingo with a note to express "you are not alone with music". (b) uses a rising castle to bring people the sense of rising with the power of music. (c) is a poster for a fan making workshop (CFAS, HKUSTSU), utilizing the shape similarity between the fan and the leaf. (d) is an ad (Make Up For Ever on Woman's Day) to promote feminine charm and their products.

in visual metaphor creation have a stronger need for exemplar-based ideation support [4]. Thus, we derive the following design requirements for our system to especially solve their problems:

R1 Search exemplars with concept association. Creativity relies heavily on association [20, 34]. At the conceptual level, automatic word association is helpful to evoke designers' imagination [6, 45]. Exemplars should be retrieved based on various concepts.

R2 Explore example-based recommendations in diverse directions. Designers create new ideas through association and recombination from previous exemplars [4, 53]. In addition, the creation of visual metaphors involve analogy between all kinds of information [8], including semantics [8, 38, 39], color [25, 54], and shape [32, 51]. By recommending exemplars based on multi-dimensional features (*i.e.,* semantics, color, shape), designers can explore diverse materials with any additional knowledge behind the recommendation.

R3 Record ideas and iteratively brainstorm. Ideation is an iterative process [17]. Designers need to polish their original thoughts and continuously generate new ideas [21]. New exemplars can trigger new possibilities to recombine previous ideas. Thus, it is necessary to enable an iterative search and idea recording at any time for users.

R4 Keep track of historical thinking path. According to our formative interview, designers can easily get lost in the image searching process during ideation without tracking their exploration path. Nevertheless, rich history keeping is an essential design principle to support creativity generation [47]. The tracked thinking path can remind users of the previous exploration and build connections between new thoughts and previous ones.

4.2 System Pipeline

To facilitate divergence and convergence activities in the ideation process, MetaMap provides 1) assembly of image search results based on automatic keyword association and color theme filtering; 2) exploration of examples related to a seed image from semantics, color, and shape dimensions; and 3) support for thinking paths tracking and ideas recording. As shown in the pipeline (see Figure 3), the conventional image searching process is supported by word association to inspire users with more concepts (**R1**). After identifying an image of interest, users can further explore the recommendation exemplars based on 3-dimensional features: semantics, color, and shape (**R2**). Users can save useful images and type comments to record their thoughts before a new brainstorming iteration (**R3**). All the saved mind maps and images record the thinking path generated by users for a quick recollection (**R4**).

4.3 Implementation

4.3.1 *Database Preparation.*

Keywords Association Dataset Construction. Previous work has proved that, by collecting responses after giving a cue, free word association experiments provide clues to reflect the evocation results on how people store and retrieve concepts in memory [33]. Our keywords database is built upon the small world dataset [14], which is currently the largest free word association resource in English with over 12,000 cue words. In this dataset, each cue is responded to by 100 participants, and each participant provides three words that immediately pop into their minds based on the cue. In this way, we compute the association strength for each cue-response link; that is, we use the response frequency divided by the total number of responses [9]. For example, there are 300 responses

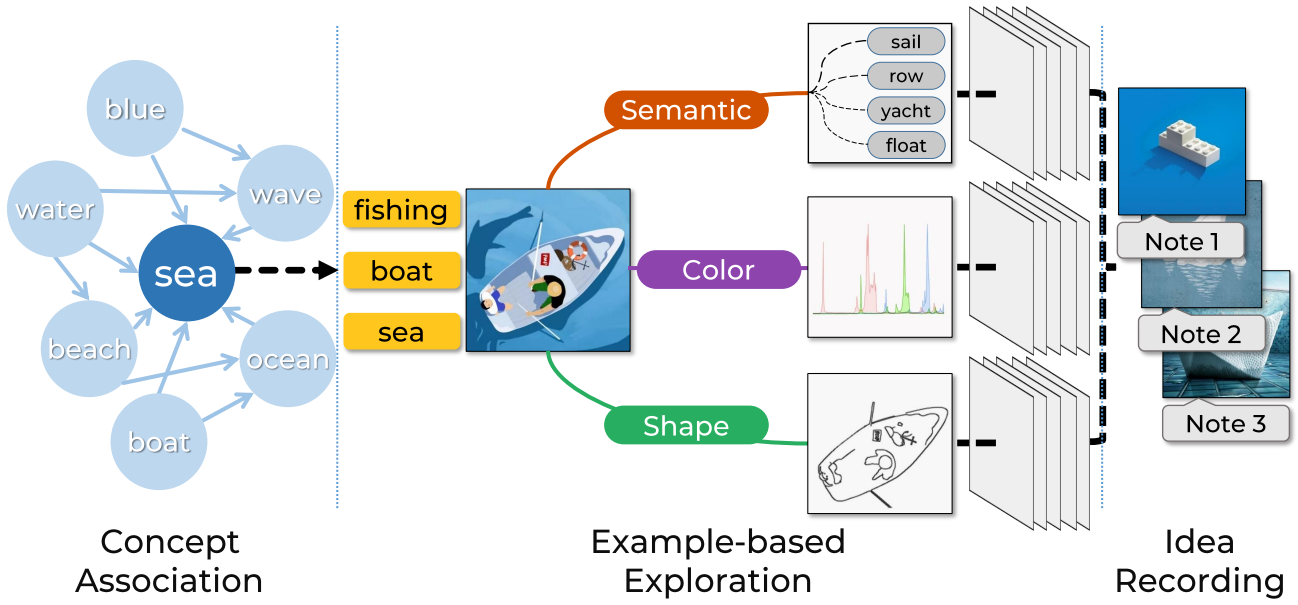

Figure 3: The system pipeline of MetaMap.

to "bike", and if "health" occurs three times, the association strength for "bike"-"health" is 0.01.

Next, we construct our keywords association dataset with the forward association (*i.e.,* from responses to cues) using a directed graph, where we take the association strengths as weights. The intuition behind this is that designers need concept associations to inspire them with more unexpected and understandable concepts. Some associations are intuitive in one direction but are not when interpreting them in reverse [33]. For instance, given the cue "vitamin", people will make a strong connection with "health", which indicates a straightforward association for the audience. However, "vitamin" does not immediately come to mind when given the cue of "health", which lowers the chances of designers developing it. Thus, the forward association is adopted in our dataset to help evoke related concepts to the users. In our current system, we leverage the small world dataset and select all words within a radius of two from the central topic words for evaluation (*i.e.,* association distance from "health" and "music" less than or equal to two). We finally construct our keywords association dataset comprising 7,407 words.

Image Collection. According to our formative study, Pinterest is the most popular website for designers to find exemplars. To test our proposed method for ideation, we crawl image data from Pinterest which is only presented on the first-page view for each query, which indicates a higher chance that designers notice. In order to give amateur designers a clear mind of what kind of design they are aiming for, we focus on collecting creative advertising posters. To ensure the relatedness of images and visual metaphors, we use "advertising creative" and "advertising metaphor" as the suffixes to make the query. The suffixes are decided by different experiments under a small amount of data with human judgment on the overall image quality. We collect 98,931 for our initial image dataset.

Data Cleaning. We build the keyword-image relationship utilizing the small world dataset and the searching function on Pinterest, which contains low-quality images and nonsense words. First, we filter out keywords with less than ten images retrieved from Pinterest, or five associated keywords in our word association dataset, to ensure the necessary amount of results provided under each search. In addition, words without concrete meaning like preposition and single characters are filtered out. If an image has many textual contents in it, it is very likely to be a capture of a newspaper or a magazine [28]. The text information rather than visual objects will dominate the image. Tesseract OCR [6] [50] is adopted as the state-of-the-art method to extract words on images. Images containing over 60 characters are filtered out from the database to ensure the visual quality based on empirical experiments. After the above data cleaning, we obtain 4,861 keywords and 76,686 images in our final database for evaluation.

4.3.2 Recommendation Algorithm Design. We propose to recommend exemplars from three dimensions (*i.e.,* semantics, color, shape) to provide related and diverse examples. Thus, we extract the corresponding features from the seed image and retrieve the top-ranked exemplars from the above three dimensions (see Figure 3).

Semantics. We recommend semantically related images based on its forward association strengths. Moreover, according to our formative study, designers look forward to related and concrete concepts to inspire them. Concreteness and imageability are two psycholinguistic norms [42]. Concreteness measures how strongly a word is related to some perceptible concept, while imageability

[6] Tesseract OCR project: https://github.com/tesseract-ocr/tesseract.

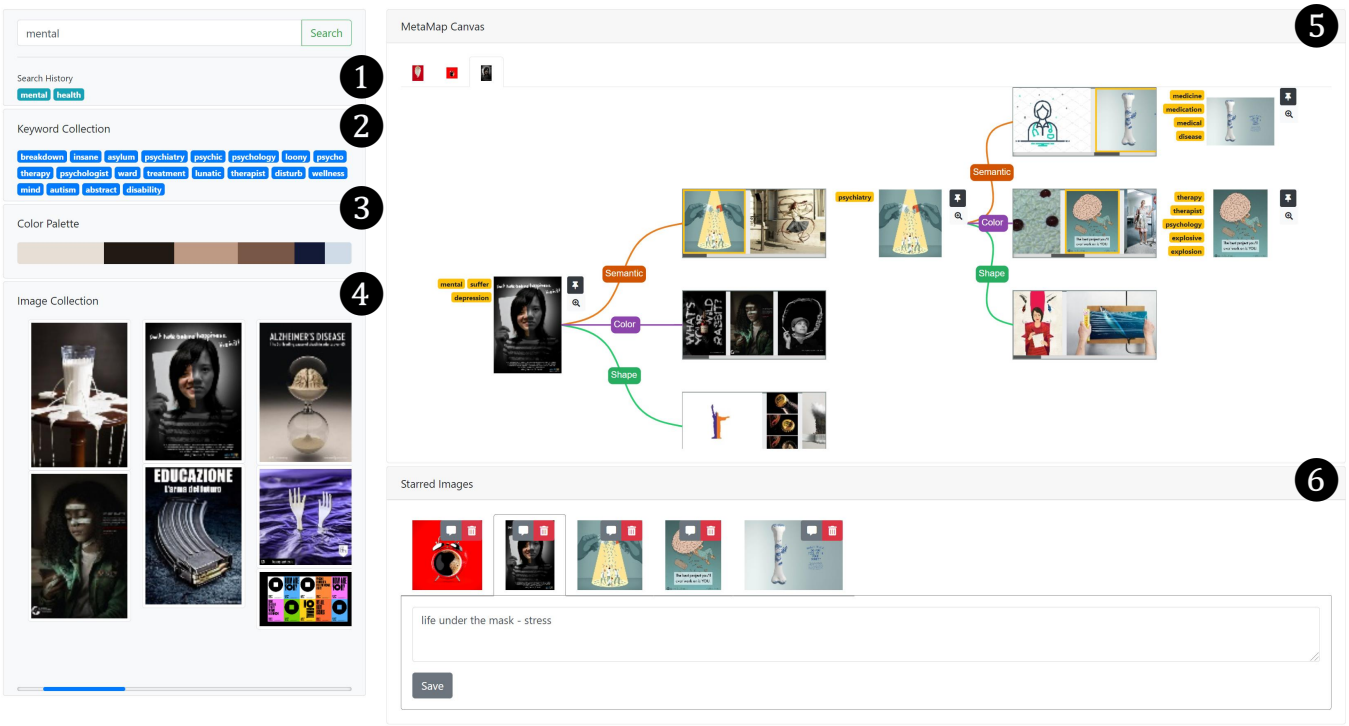

Figure 4: The interface of MetaMap includes three parts: 1) image search area (❶❷❸❹); 2) image exploration area (❺) and 3) idea tracking area (❻).

measures the extent to which the item evokes a mental image. Previous work has used concreteness to retrieve high-quality images which make more sense to humans [28]. We adopt concreteness and imageability scores (scale from 0 to 1) from [30], which is collected through human annotation and synonym expansion. For semantic exemplar recommendation, we only select concepts with the sum of concreteness and imageability larger or equal than 1.5. Assigning higher priority concepts with a stronger association, we randomly retrieve the top 10 images from the qualified candidates. The random selection is to increase the diversity of returned exemplars when the semantic correlation is on the same level.

Color. We recommend images with similar colors based on color histogram correlation between images. However, to facilitate the search speed and make sure that the image has some semantic relationship with the original, we only search among neighboring concepts. All images under neighboring concepts are compared with their color histogram extracted using OpenCV. We select the top 10 images for the recommendation.

Shape. Similar to color recommendation, we search images with similar shapes among neighboring concepts. To extract the shape of the main object, we first use canny edge detection to obtain the images' outline. Noises are eliminated with iterations of erosion and dilation. At last, the contour with the largest area is found and compared with OpenCV. The top 10 images with the most similar contours are recommended in the direction of the shape. A similar approach has been adopted in [44] to identify the shape of objects in images.

4.3.3 User Interface and Interaction. Based on the collected requirements (**R1-4**), the user interface of MetaMap includes three parts: 1) image searching area; 2) image exploration area and 3) idea tracking area (Figure 4). By typing the keywords in the search box (Figure 4 ❶), users can obtain a list of sample images with the corresponding keywords suggestions (Figure 4 ❷). In the color collection area (Figure 4 ❸), users can see the color summary[7] of retrieved images. By clicking on a specific color, users can re-rank the images based on the color similarity. Meanwhile, the search history records users' searched keywords and present the history under the search box for reference (Figure 4 ❶). After that, users can browse the returned image list and click one for further exploration. The selected image will appear in the image exploration area (Figure 4 ❺). Users can explore recommended exemplars based on the selected image in three dimensions (*i.e.,* semantics, color, shape) and keep expanding along any branch by clicking on the corresponding keyword buttons. During the process, users can pin any image that they find useful or inspiring to the idea tracking area (Figure 4 ❻) and leave comments for post-inquiries.

[7]Color palette extraction: https://github.com/obskyr/colorgram.py.

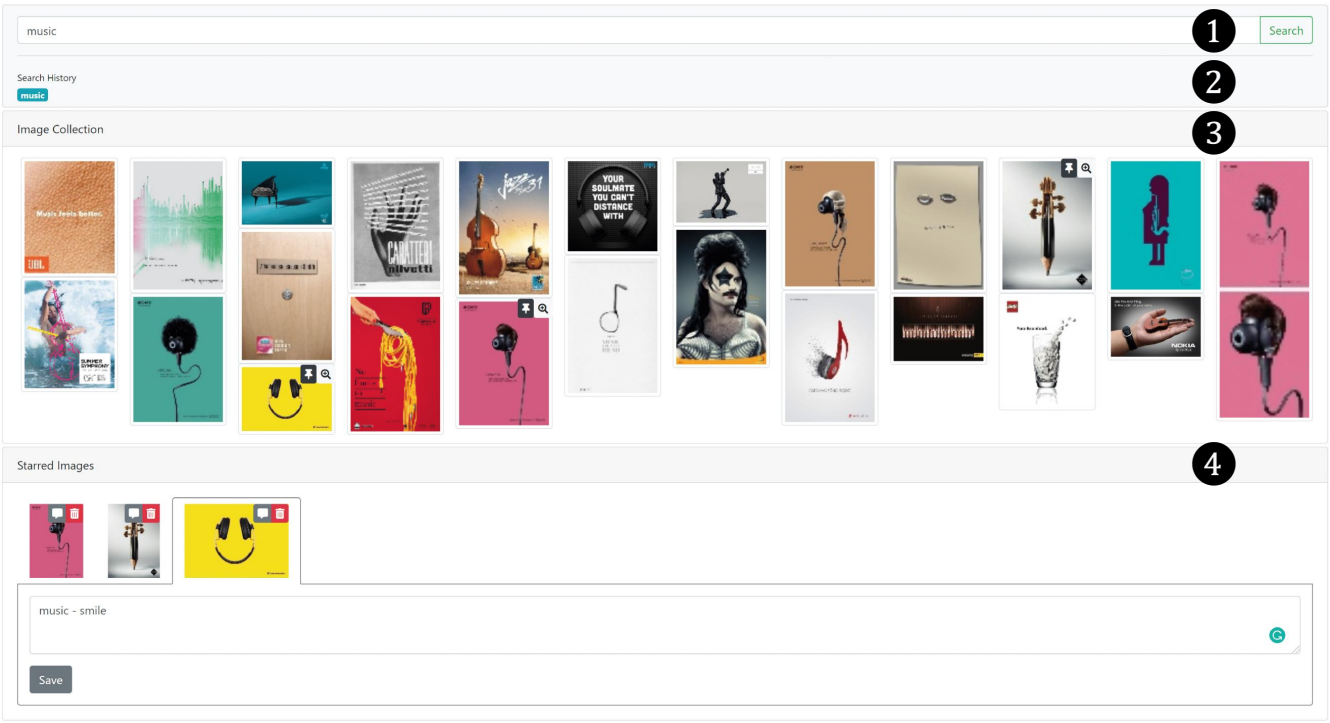

Figure 5: The interface of Pinterest-like baseline includes two parts: 1) image search area (❶❷❸) and 2) idea tracking area (❹).

5 HYPOTHESES

Previous research suggests that a diverse but related exemplar-exploration can improve the ideation outcome [15, 24]. Also, engagement and creativity are mutually supportive of each other. Reid and Solomonides [41] suggest that designers can become more creative when they get involved and engaged in the design process. Such an effect would further improve their overall experience. Furthermore, a competitive tool should be useful and usable [21]. However, there does exist a trade-off between the functionality and the usability [19]. Therefore, we make the following hypotheses:

H1 Compared to a Pinterest-like baseline system, MetaMap significantly increases amateur designers' satisfaction level (***H1***) with their overall design outcome (*H1a*), the number of generated ideas (*H1b*), the diversity of ideas (*H1c*), and the creativity of ideas (*H1d*).

H2 Compared to a Pinterest-like baseline system, MetaMap provides a more engaging experience to inspire creativity (***H2***), especially for improving users' concentration(*H2a*), sense of ecstasy (*H2b*), clarity of the task (*H2c*), confidence of the ability (*H2d*), sense of serenity (*H2e*), feeling of timelessness (*H2f*), and intrinsic motivation (*H2g*).

H3 Compared to a Pinterest-like baseline system, MetaMap reduces its usability with the additional functionalities (***H3***).

H4 Compared to a Pinterest-like baseline system, MetaMap is significantly more useful for amateur designers (***H4***).

6 EVALUATION

To test the hypotheses above, we conduct a within-subjects controlled experiment which compares MetaMap and a Pinterest-like baseline with 24 amateur visual designers. Since ideation is a complicated process, we design a within-subjects study to reduce the possible effect of individual differences in the ideation process. Based on the formative study findings, Pinterest is the most commonly used website to find inspiration, which could be a baseline system in ideation. We implement a baseline rather than using Pinterest directly to minimize the possible influences caused by different UI styles and data sources. The baseline is a Pinterest-like interface with a basic search history, image searching (with the same database), and image saving functions (see Figure 5). Participants completed two tasks with each system: brainstorming, giving feedback, and implementing one idea draft. Before finishing the drafts, we ask all the participants to evaluate the two systems with questionnaires on ideas generated, outcome satisfaction, user engagement, system usability, and tool usefulness. We also interview them based on their experiences. Additionally, we conduct semi-structured interviews with three visual design experts to give comments to MetaMap and participants' representative ideas.

6.1 Participants

We recruit 24 amateur designers (see Table 1; 17 female, 6 male, 1 prefer not to say; age range 20-30, $M = 22.75, SD = 2.52$) through online advertising and word-of-mouth. The criteria is that participants have experience in some design activities but do not receive

ID	Gender	Age	Use visual metaphors?	Exp.	Previous Design Activities
1	Female	22	Somewhat Agree	4	Activity Poster, Product Advertisement Promotion
2	Male	23	Somewhat Agree	3	Activity Poster, Mobile App UI
3	Female	22	Somewhat Agree	3	Activity Poster, Class Uniform
4	Male	22	Agree	3	Activity Poster, Game
5	Female	22	Somewhat Agree	4	Activity Poster
6	Female	19	Agree	2	Activity Poster, Product Advertisement Promotion
7	Female	22	Somewhat Disagree	1	Activity Poster
8	Male	19	Somewhat Agree	5	Phone Case Design
9	Female	20	Neutral	3	Activity Poster, Illustration Design
10	Female	27	Neutral	2	Activity Poster, Mobile App UI
11	Female	23	Not Applicable	5	Activity Poster, Architectual Design
12	Female	23	Somewhat Disagree	1	Activity Poster, Product Advertisement Promotion
13	Male	23	Somewhat Agree	4	Activity Poster, Mobile App UI
14	Male	30	Somewhat Agree	10	Mobile App UI, Industrial Design
15	Female	23	Somewhat Agree	8	Mobile App UI
16	Prefer not to say	22	Neutral	2	Activity Poster, Mobile App UI
17	Female	27	Somewhat Agree	6	Activity Poster, Mobile App UI
18	Female	24	Agree	5	Mobile App UI
19	Female	24	Somewhat Agree	5	Activity Poster, Mobile App UI
20	Female	23	Not Applicable	1	Activity Poster
21	Female	21	Neutral	2	Activity Poster
22	Female	20	Disagree	2	Activity Poster, Mobile App UI
23	Male	24	Neutral	1	Activity Poster
24	Female	21	Neutral	2	Activity Poster

Table 1: Demographics of all the participants. This table includes participants' ID, gender, age, their self-assessment on designing with visual metaphors and design experiences.

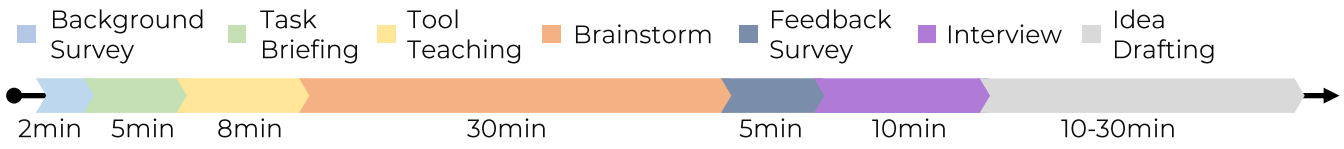

Figure 6: The flow chart of one task. Notice that the background survey only appears in the first task. In the second task interview session, they will be asked additional questions to compare experiences in two tasks together. In the idea drafting session, participants are told to express the selected idea in any format (*e.g.*, hand-drawing, Photoshop), and they can leave once they finish the draft.

professional visual design education. All participants have no color vision deficiency, and they can communicate in English fluently. All participants are familiar with Pinterest for image search.

6.2 Tasks and Procedure

During the experiment, each participant is asked to complete two tasks. In each task, they are asked to brainstorm under given topics: "health" or "music", which is an open topic for them to deliver any creativity visual design related to the topic. Before the brainstorming session, they are told with the basic concept of visual metaphor and encouraged to generate such ideas during the brainstorming session. During the 30 minutes of brainstorming, they are asked to think of as many ideas as possible. They complete the two tasks by using MetaMap and the baseline system separately. Feedback questionnaires and comments regarding the ideation experience are collected right after. In the end, they are asked to pick one favorite idea to make a draft in the implementation session (10-30 minutes). The task ends once they submit their draft. After being counterbalanced with Latin Square, there are four combinations: (a) health (MetaMap) - music (Baseline), (b) music (MetaMap) - health (Baseline), (c) health (Baseline) - music (MetaMap) and (d) music (Baseline) - health (MetaMap). The whole process of one task is shown in Figure 6 without exceeding 1.5 hours.

Category	Factor	MetaMap		Baseline		Statistics			Hypotheses
		Mean	SD	Mean	SD	Z	p	Sig.	
Satisfaction	Overall outcome	5.92	.65	4.83	1.34	-3.21	.001	**	*H1a* accepted
	Amount	5.87	.90	4.96	1.37	-2.87	.004	**	*H1b* accepted
	Diversity	5.63	1.25	4.5	1.69	-2.98	.003	**	*H1c* accepted
	Creativity	5.54	.88	4.75	1.45	-2.50	.013	*	*H1d* accepted
Engagement	Concentration	6.13	.85	4.83	1.58	-3.04	.002	**	*H2a* accepted
	Sense of ecstasy	5.75	.94	4.21	1.72	-3.44	.001	**	*H2b* accepted
	Clarity	5.75	.99	4.92	1.41	-2.56	.011	*	*H2c* accepted
	Doability	5.71	1.08	4.71	1.20	-3.18	.001	**	*H2d* accepted
	Sense of serenity	5.25	1.19	4.50	1.10	-2.87	.004	**	*H2e* accepted
	Timelessness	5.71	1.08	4.88	1.36	-2.17	.030	*	*H2f* accepted
	Intrinsic motivation	5.67	1.05	5.17	1.37	-1.63	.104	-	*H2g* rejected
Usability		77.29	12.29	77.4	12.32	-.37	.708	-	*H3* rejected
Usefulness		6.08	.72	5.12	1.19	-3.03	.002	**	*H4* accepted

Table 2: The statistical user feedback with MetaMap and the baseline system, where the p-values (−: $p > .100$, +: $.050 < p < .100$, ∗: $p < .050$, ∗∗: $p < .010$, ∗ ∗ ∗: $p < .001$) is reported.

6.3 Results Analysis

In the experiment, we collect participants' ratings on the ideas generated, outcome satisfaction, engaging level, usability, and usefulness of the two systems (*i.e.,* MetaMap and Pinterest-like baseline). After obtaining the ratings, we interview the participants to understand the reasons behind the scores. We report the collected feedback in this section.

6.3.1 Ideas Generated with MetaMap and the Baseline. After each session, we count the starred images and the ideas generated from the participants. During the ideation process, participants starred 206 useful images ($M = 8.58, SD = 5.68$) for ideation in the MetaMap system. They saved 191 useful images ($M = 7.96, SD = 4.91$) when using the baseline system. The total number of ideas generated by using MetaMap is 101 ($M = 4.21, SD = 1.98$), while this number drops to 84 when using the Pinterest-like system ($M = 3.50, SD = 2.06$). Overall, participants were exposed to more useful images and generated more ideas when using the MetaMap system. Here we present four representative ideas among 48 drafts generated with MetaMap and the baseline system in Figure 7. These four ideas are generated by two participants under different conditions. They demonstrate different traits of the ideation process with the two systems. Ideas generated with MetaMap have more diverse exposure with clear logic flow. We further analyze these ideas in detail with design experts' comments in section 6.3.6.

6.3.2 Satisfaction of Ideation Outcome. We collect participants' questionnaire data (7-points Likert scale) on their ideation experience with MetaMap and the Pinterest-like baseline system. As shown in Table 2, Wilcoxon signed-rank tests are used to compare the differences in the satisfaction level of using these two systems with detailed factors. Overall, participants were significantly more satisfied with ideas generated with MetaMap ($\eta^2 = -.46$); *H1a* is accepted. There are significant differences between MetaMap and the baseline regarding the satisfaction of amount ($\eta^2 = -.41$), diversity ($\eta^2 = -.43$), and creativity ($\eta^2 = -.36$); *H1b*, *H1c*, and *H1d* are accepted. Thus, ***H1*** is fully accepted. Many of them reflected, *"MetaMap provides diverse exemplars which can expand my imagination."* (P1-2, P5-6, P11, P13-16, P23-24). In addition, when they were exploring exemplars, recombination and expansion happened to create new ideas [4, 53]. Some stated that *"unexpected images"* opened their minds so that they could generate more creative ideas (P5, P14, P24).

6.3.3 Engagement in Ideation. We also investigate the engagement level of participants in the ideation tasks. We utilize the flow theory for a positive experience [12], a commonly used theory evaluating user engagement [36] in 7 dimensions: concentration (how concentrated), sense of ecstasy (how special to you), clarity (how clear while doing), doability (how confident with your ability), sense of serenity (forget yourself), timelessness (time passes quickly), and intrinsic motivation (how self-rewarding). Wilcoxon signed-rank tests are used to determine whether there elicits a statistically significant difference in user engagement by using MetaMap and the baseline system. As shown in Table 2, there are significant differences in: concentration ($\eta^2 = -.44$), sense of ecstasy ($\eta^2 = -.50$), clarity ($\eta^2 = -.37$), doability ($\eta^2 = -.46$), sense of serenity ($\eta^2 = -.41$), and timelessness ($\eta^2 = -.31$); *H2a*, *H2b*, *H2c*, *H2d*, *H2e*, and *H2f* are accepted; only *H2g* is rejected. Overall, ***H2*** is partially accepted. It suggests that MetaMap improves their concentration, clarity, and ability to finish the task and provides them a more exceptional, selfless, and timeless experience. Three participants highlighted that the exploration process on MetaMap is intriguing and engaging itself to make them more focusing on brainstorming (P5, P14, P24). One even reported *"Using MetaMap is a bit like playing a game that triggers a new level, it tells you something new, and then you cannot help but keep exploring."* (P5).

6.3.4 System Usability. To test two systems' usability, we adopt the standard System Usability Scale (SUS) questionnaire [5]. The SUS

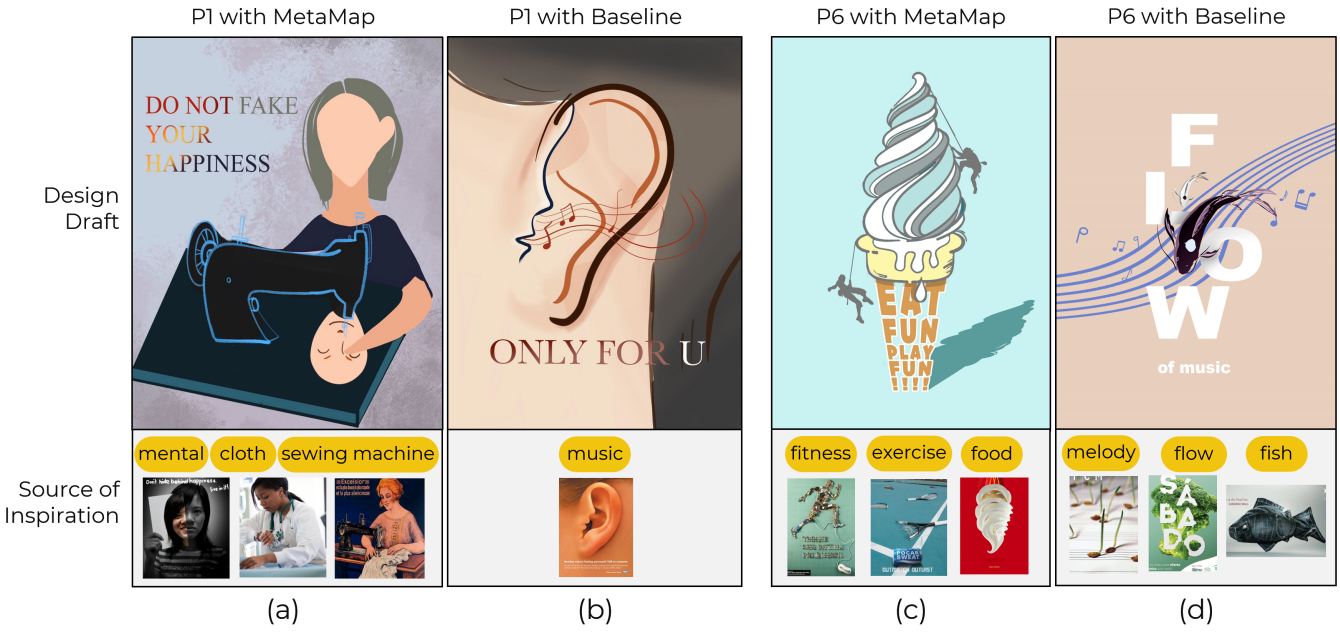

Figure 7: Examples of 4 drafts generated by 2 participants (P1, P6) with their sources of inspiration: (a) utilizes the concept of "sewing your mask" to express the mental pressure under work. (b) uses the tragus as the shape of a side face to express "music only for you". (c) treats the ice-cream as a hill to deliver the message of "eat fun, play fun", (d) shows a whale swimming in the "flow" of music.

rating of the MetaMap system is 77.29, while the baseline system rates for 77.39. These scores indicate that both systems achieve an acceptable level (B+) [3]. A Shapiro-Wilk test shows a significant departure from normality for MetaMap system ($W(24) = .90, p = .018$). Therefore, we perform a Wilcoxon signed-rank test to compare the differences in telling the significant difference of usability scores. There is no statistical significance between MetaMap ($M = 77.29, SD = 12.29$) and the Pinterest-like baseline ($M = 77.40, SD = 12.32$) in the usability score (Table 2). Thus, the two systems are close regarding the usability scores. The Pinterest-like system did receive positive feedback for the *"clear layout"* (P7, P21-22) and *"simplicity to use"* (P12, P16). However, MetaMap still achieves around the same level of usability with the Pinterest-like system. Overall, we manage to add more functionalities in MetaMap without reducing the usability level; thus ***H3*** is rejected.

6.3.5 Usefulness of Functions. At last, we compare the usefulness of functions between the two systems. This dependent variable is measured at the ordinal level, and our independent variable only consists of two categories, then we need to know whether the distribution of the differences between the two related groups is symmetrical in shape or not. A Wilcoxon signed-rank test was used to compare the differences in telling the usefulness of the two systems (see Table 2). We can tell that MetaMap is significantly more useful than the baseline system; ***H4*** is accepted. As reported by participants, MetaMap is useful for them because of its *"diverse recommended examples"* (P1-3, P6, P11, P13, P19, P24) and *"track of thinking paths"* (P4, P6, P15, P18-19).

We also evaluate the inner functions within MetaMap to see which functions are comparatively more useful: Keyword Association ($M = 5.96, SD = .69$), Color Collection ($M = 4.17, SD = 1.52$), Image Collection ($M = 5.79, SD = .98$), Canvas Exploration ($M = 5.83, SD = .96$), Semantic Recommendation ($M = 5.38, SD = 1.50$), Color Recommendation ($M = 4.25, SD = 1.54$), Shape Recommendation ($M = 4.75, SD = 1.29$), Starred Images ($M = 5.83, SD = .82$). As we can see, the keyword association, exemplar exploration, and starred images are recognized as the most useful functions in MetaMap. However, we can observe functions with sizeable standard deviation: color collection and semantics/color/shape exploration. Combining with our observations, this kind of disagreement in usefulness evaluation reflects different user preferences with various ideation patterns. We will discuss different ideation patterns and user preferences in section 7.1.

6.3.6 Additional Comments. We conducted semi-structured interviews after each task to collect feedback from users. This section summarizes how MetaMap supports their ideation process. We consolidate highly similar experssion into same quotations. The debating ideas with suggestions to further improve MetaMap is addressed in section 7.

Tools Preference. Overall, 21 participants out of 24 stated that they would like to use MetaMap as their ideation tool over the baseline system. Two other participants preferred to combine the two systems to view image searching results in a waterfall flow view, but explore exemplar recommendations on another interface. Though the baseline Pinterest-like system provides a more intuitive

view of images, most participants still prefer MetaMap for its diverse exposure to exemplars with tracking their historical thinking paths.

Inspiring Image Searching. Participants recognize the overall quality of the image searching results. Six participants stated that the images returned are very creative and of high quality compared to their previous experience using a search engine (P3, P5-6, P10, P14, P21). Their feedback suggests that adding certain suffixes to search queries and filtering out images with too much text effectively fits the creative design scenario. In addition, the concept association function is highlighted by 11 participants. Five participants stated this function is *"very useful to open my mind"* (P1, P6, P15, P21, P24). Furthermore, four users reflected that they *"would like to use this function when having no clues in mind"* (P2, P7, P12, P17).

Idea Expansion with Logical Organization. All participants who preferred to have MetaMap as their ideation tool found that the exemplar recommendation is helpful. The diversity was highly praised when exploring the exemplars: *e.g., "the diverse exploration is very important to generate creative ideas"* (P3, the same meaning expressed by P6, P11, and P19), *"it gives more information during the process, which can reduce the mental effort to brainstorm"* (P2), *"I get some unexpected images which may trigger my innovative thoughts in the process"* (P4). Moreover, the logical layout of 3 recommendation directions was also recognized by 2 participants (P2, P13). *"It is a very nice framework of design elements to help me think systematically."* said P14. Another one reported *"It provides me information about why this image is recommended so I can know what to focus on when viewing them"* (P2).

Thinking Path Tracking and Idea Iteration. This is a unique experience reported by 6 participants (P4, P6, P15, P18-19). They found it very helpful to keep the mind map exploration structure to trace back to their original thoughts. By keeping the exploration history in mind maps, they can *"review them and recombine different ideas"* (P4, P6, P18, P24). One participant added *"It records my thoughts and helps me regulate my thinking path. More importantly, when I know that I have those images recorded, I feel safe and free to keep exploring without worrying about getting lost."* (P19). Another participant also shared how the mind map structure helps him to iterate his own ideas: *"I also have a "mindmap" in my mind being triggered at the same time when exploring on MetaMap. When I encounter new images, my "mindmap" will combine and fuse together with the displayed mind map to trigger new ideas."* (P4).

6.3.7 Expert Interview. To get more insights into the system design and usage from a professional designer's view, we conduct semi-structured interviews with three design experts to analyze the drafts collected from our participants. We also invite them to share their views regarding MetaMap afterwards. The background of the three design experts are 1) E1: male, a visual designer in the industry, focusing on graphic design, holding a master's degree in visual design, with eight years of practice; 2) E2: female, a visual designer in the industry, focusing on advertising design, received a master's degree in design theory, with seven years of design practice; and 3) E3: female, an associate professor, focusing on art and design, with 22 years of practice.

System Design of MetaMap. The 3-dimensional recommendation framework is recognized as reasonable by all experts (E1-3). E2 raised that *"This structured brainstorm should be very useful for novices to familiarize themselves with the basic design elements in the beginning stage"*. E1 reported that even this tool for him is *"intriguing to interact with"*, and he will give it a try. Moreover, E1 gave high praise for the idea of tracking the thinking path with mind maps during exploration. The professor suggested that all materials need to be connected and recombined to generate new ideas, *"mind map is one way to achieve that"* (E3).

Showcase Analysis. We also presented design drafts from 10 participants with a high degree of completion to make comments. Figure 7 shows 4 representative ideas from two participants (P1, P6) as examples. Example (a) is generated with MetaMap, P1 explored concepts with a "mental" first and saw plenty of work using "mask" as a metaphor of mental illness. The "sewing machine" was automatically suggested by the system in Figure 4 ❷ and inspired P1. This ideation process has a *"clear logic flow behind"* under the interaction between the user and the system (E2). *"The sewing machine distinguishes this idea from some others' work using the "mask" concept and makes it more creative."* (E2). Another idea (b) generated by P1 using the baseline system has heavy imitation marks by referring to a very similar design. Also, *"this idea is hard to understand at a glance"* (E2-3). The idea (c) from P6 generated using MetaMap, is thought to be intriguing by experts (E1-2). P6 got the inspiration to connect "ice-cream" and "hill" based on the similar shapes and concepts of "exercise" explored previously. *"Treating ice-cream as a hill is a quite creative idea for me. However, the message behind is not very direct to understand."* (E1). Another idea (d), though with a relatively high aesthetic degree, is criticized by all experts. The metaphor of *"music is like water"* is so common in design, and this idea *"did not express it in a innovative way"* (E1-3). Overall, we see MetaMap build up connections to link multiple elements together, which helps generate creative ideas through the recombination from existing work.

7 DISCUSSION

Overall, our evaluation results suggest that MetaMap is useful in visual metaphor ideation with a more engaging process and satisfying outcomes, without reducing the usability level comparing to the Pinterest-like system. In our experiment, 87.5% (21/24) participants preferred MetaMap to the baseline system, suggesting that amateur designers may accept our tool as a promising visual metaphor ideation support. In this section, we discuss different user behavioral patterns and the possibility of generalizing MetaMap. We also identify the limitations and future work of our research.

7.1 Usage Patterns and User Preferences

During the evaluation of MetaMap, we identified different user preferences on specific functions, which are reflected not only in our participants' comments but also in their behavioral patterns during the brainstorm session. Such preferences may indicate different user approaches to ideation. By coordinating the participants' self-reported activities with their interaction trajectories logged by our system, we derive two main ideation patterns: concept-focused

ideation (*e.g.,* P7, P10, P16, P19, P21-22, P24) and exemplar-focused ideation (*e.g.,* P1, P5, P8, P17-18, P20, P23).

Participants who tended to prioritize concept level association would use MetaMap's search function very frequently at more than 10 times ($M = 13.86, SD = 1.95$). They found the keyword association function particularly useful for evoking their imagination of related concepts. When exploring related images on MetaMap Canvas (Figure 4 ❺), they also paid more attention to the semantic features of exemplars. They usually have distinctive views of the three dimensions: *"Semantic dimension is handy"* (P7, the same meaning expressed by P10, P19, and P24), *"Color is not an element to consider at the ideation stage"* (P7, P10, the same meaning expressed by P16 and P24), and *"Shape information is useless and confusing"* (P10, the same meaning expressed by P19 and P21-22). To better support this type of ideation pattern, one design expert (E1) suggested that the keyword association function could be displayed in a network format (currently as a word list) to show more information at a conceptual level, *e.g.,* strength of connection, association direction, word frequency and concreteness, and so on.

In contrast, exemplar-focused participants preferred to explore on MetaMap Canvas (Figure 4 ❺) without continuously searching many concepts. All of them used the search function of MetaMap no more than three times. On MetaMap Canvas, they would either do breadth-first exploration by going through every recommendation in all three directions before taking the next step, or depth-first exploration by immediately diving deeper into something they found intriguing. This group of participants highly welcomed the diverse visual exposure in the exploration. To better support this type of user, experts suggested functions to provide more visual information (*e.g.,* summarize a color palette from images user saved and give further recommendations (E1), allowing users to interactively select shapes to explore (E2)).

Despite individual differences in dimension preference, several participants shared the same concern about *"not being able to distinguish or relate the three dimensions together"* (P8, the same meaning expressed by P3, P5-6, P20, and P24). Surely, all kinds of features of design materials are inter-related when constructing a metaphorical relationship [8]. For example, as proposed in the expert interview (E1, E3), by incorporating color-semantics [23, 54] and shape-semantics [51, 52] relationships to enrich the connections among the three dimensions, users might be able to obtain more conceptual-level inspirations from color and shape or get perceptual stimulation from lexical meanings.

7.2 Generalization of MetaMap

Our evaluation results suggest that amateur designers may accept our tool as a good visual metaphor ideation support. However, both the amateur designers and the experts further pointed out in their interviews the possibilities of applying the mind map-like ideation service to other design scenarios.

7.2.1 Supporting ideation for more general visual design. Our proposed 3-dimensional exploration framework has initially been designed to create a visual metaphor. We represent it in a mind-map like structure to meet the needs of establishing connections between different design materials of interests as discovered in our formative interview. In this way, the mind map is also a record of exploration history, to remind users of their thinking paths to generate metaphorical associations. However, for designers, this need – recombining previous knowledge and recording a historical thinking path – is not unique to visual metaphor design; instead, it is commonly shared across all kinds of visual design [15, 17]. As the design professor (E3) suggests, *"designers should be exposed to all kinds of information (e.g., texts, photos, illustration, typography design) to open their minds"*. Our proposed ideation tool could potentially be adapted to inspire the production of a broader kind of creative work (*e.g.,* photography, painting, slogan, *etc.*) by 1) adding more types of exemplars into the database, and 2) taking more dimensions (*e.g.,* style, composition, *etc.*) into consideration for exemplar recommendation. We can provide users the flexibility to customize the design materials and exploration dimensions based on the application domain and personal interests.

7.2.2 Supporting ideation for professional designers. In our interviews with professional designers, we noticed that they have very different needs and concerns about ideation in their practices from our participants. Amateur designers, without a solid foundation of design theories and enough exposure of exemplars [4], need organized input to establish better associations, which could be supported by MetaMap. Expert designers acknowledged this point and added that MetaMap could even serve as a *"reference for design education"*, assisting novices in systematically dissecting critical visual features in good design examples as in textbooks (E2-3). However, it is less important to give structured information to professional designers, because they *"have already had established methodologies to look for ideas on their own"* (E1-2). They feel that they might enjoy receiving fuzzy, less organized recommendations of serendipitous examples from our systems more even if the algorithm behind is somewhat a "black box" (E1). All design experts we interviewed expressed the importance of "accumulating design materials as a long-term practice in everyday life" (E1-3). They thus demand a tool that can help them manage the large number of examples collected and ideas generated along the way. They postulated that they might benefit more from our tool if a system allowed them to input new data and automatically construct an editable mind map of all or a specified subset of their curated design materials.

7.3 Limitations and Future Work

This work has several limitations which we plan to address in the future.

7.3.1 Performance of recommendation algorithms. This work demonstrates a proof-of-concept framework of recommending visual metaphor exemplars based on three different features (*i.e.,* semantics, color, and shape). However, the current visual algorithms we employed to compute example similarity in each dimension are mostly rule-based by setting specific thresholds with feature extraction, which may not be very accurate given more complex visual design works. For instance, some small characters are not recognized by the OCR algorithm due to the low resolution of the image and deformation of artistic fonts. Also, the rule-based algorithms failed to connect different features to make joint recommendations. Currently, an example that ranks high in one dimension might not

be as relevant in other dimensions, which sometimes confuses users. In the future, we would like to expand our dataset and explore deep learning methods [26] to improve the accuracy and integrity of recommendation algorithms.

7.3.2 Constraints on system. The current MetaMap mind map has a rather fixed structure. Users could only rescale or move the entire tree around, as well as show or hide a branch. The system currently does not support the rearrangement of individual nodes in the mind map. Also, while users are able to explore and record their ideas on MetaMap, the idea recording function is still stiff. Users can only make comments on saved images without further interactions. To enable a more flexible idea recording, we could adopt a self-defined mind map editor for the personalized organization of one's own saved images and comments in the future. Moreover, users can only take notes of their generated ideas in words. These ideas could be better communicated through quick, low-fidelity prototypes generated based on the associated example images saved in the system [11]. In the future, we will design models to learn the composition patterns of given exemplars and automatically assemble user input materials to generate low-fidelity prototypes.

7.3.3 Alternative evaluations. This work implements a baseline to minimize the possible influences caused by different UI styles and data sources. In the comparative study, the mind-map like exploration is suggested to be useful and engaging compared to the Pinterest-like baseline system. After further generalizing and improving the system, future work may evaluate the ideation experience with other different existing methods (*e.g.*, moodboarding [26], metaphor cards [31]) to further investigate the design process. Moreover, since we focus on the design idea rather than the aesthetics of results, we do not constraint the level of finest on ideation drafts as long as the idea can be conveyed. So that the results are evaluated by experts who would not be influenced by the aesthetic feeling of the design. However, previous work [38] has identified the common errors in humans interpretations of visual metaphors. Though visual metaphor could be creative and intriguing, it could also be challenging for the general public to understand if the associations between visual stimuli and the underlying message are not intuitive. A good visual metaphor should be easily understood by people as well as being creative [38]. Future work may look into the question of how to create visual metaphors that are more understandable by lay audiences.

8 CONCLUSION

In this paper, we introduce MetaMap, a design tool to support amateur designers in the visual metaphor ideation process. Specifically, we incorporate in MetaMap a 3-dimensional recommendation framework (*i.e.,* semantics, color, shape) in a mind-map structure to aid ideation with diverse exemplar exploration and historical thinking path tracking. To evaluate the tool, we conduct a user study with 24 amateur designers. The results suggest that compared to a Pinterest-like interface, MetaMap can inspire participants to generate more diverse and creative ideas in a more engaging interaction. We also provide insights into supporting more general design ideation in the future.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers and participants for their helpful suggestions and useful feedback. We especially thank Prof. Zhong WANG from Changsha University of Science and Technology for her valuable inputs. This research was supported in part by the HKUST-WeBank Joint Laboratory Project Grant No.: WEB19EG01-d.

REFERENCES

[1] Virgil C. Aldrich. 1968. Visual Metaphor. Journal of Aesthetic Education 2, 1 (1968), 73–86. http://www.jstor.org/stable/3331241
[2] Suyun Sandra Bae, Oh-Hyun Kwon, Senthil Chandrasegaran, and Kwan-Liu Ma. 2020. Spinneret: Aiding Creative Ideation through Non-Obvious Concept Associations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376746
[3] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of usability studies 4, 3 (2009), 114–123.
[4] Nathalie Bonnardel and Evelyne Marmèche. 2005. Towards supporting evocation processes in creative design: A cognitive approach. International Journal of Human-Computer Studies 63, 4 (2005), 422 – 435. https://doi.org/10.1016/j.ijhcs.2005.04.006 Computer support for creativity.
[5] John Brooke. 1996. SUS: a "quick and dirty" usability scale. CRC press, Florida, USA, 189.
[6] David C Brown. 2008. Guiding computational design creativity research.
[7] Noel Carroll. 1994. Visual Metaphor. Springer Netherlands, Dordrecht, 189–218. https://doi.org/10.1007/978-94-015-8315-2_6
[8] Noël Carroll. 2001. Visual Metaphor. Cambridge University Press, Cambridge, UK, 347–368. https://doi.org/10.1017/CBO9780511605970.023
[9] Andrew Cattle and Xiaojuan Ma. 2017. Predicting Word Association Strengths. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Copenhagen, Denmark, 1283–1288. https://doi.org/10.18653/v1/D17-1132
[10] Joel Chan, Pao Siangliulue, Denisa Qori McDonald, Ruixue Liu, Reza Moradinezhad, Safa Aman, Erin T. Solovey, Krzysztof Z. Gajos, and Steven P. Dow. 2017. Semantically Far Inspirations Considered Harmful? Accounting for Cognitive States in Collaborative Ideation. In Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition (Singapore, Singapore) (C&C '17). Association for Computing Machinery, New York, NY, USA, 93–105. https://doi.org/10.1145/3059454.3059455
[11] Lydia B. Chilton, Savvas Petridis, and Maneesh Agrawala. 2019. VisiBlends: A Flexible Workflow for Visual Blends. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300402
[12] Mihaly Csikszentmihalyi. 1990. Flow: The Psychology of Optimal Experience. Vol. 1990. Harper & Row, New York, USA.
[13] Marcel Danesi. 1990. Thinking is seeing: Visual metaphors and the nature of abstract thought. Semiotica 80, 3-4 (1990), 221–238. https://doi.org/10.1515/semi.1990.80.3-4.221
[14] Simon De Deyne, Danielle J Navarro, Amy Perfors, Marc Brysbaert, and Gert Storms. 2019. The "Small World of Words" English word association norms for over 12,000 cue words. Behavior research methods 51, 3 (2019), 987–1006.
[15] Claudia Eckert and Martin Stacey. 2000. Sources of Inspiration: A Language of Design. Design Studies 21 (09 2000). https://doi.org/10.1016/S0142-694X(00)00022-3
[16] Hermine Feinstein. 1985. Art as Visual Metaphor. Art Education 38, 4 (1985), 26–29. http://www.jstor.org/stable/3192822
[17] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3290605.3300619
[18] Milene Gonçalves, Carlos Cardoso, and Petra Badke-Schaub. 2014. What inspires designers? Preferences on inspirational approaches during idea generation. Design Studies 35, 1 (2014), 29 – 53. https://doi.org/10.1016/j.destud.2013.09.001
[19] Nancy C Goodwin. 1987. Functionality and usability. Commun. ACM 30, 3 (1987), 229–233.
[20] H. Gough. 1976. Studying creativity by means of word association test. The Journal of applied psychology 61 3 (1976), 348–53.
[21] Tom Hewett, Mary Czerwinski, Michael Terry, Jay Nunamaker, Linda Candy, Bill Kules, and Elisabeth Sylvan. 2005. Creativity support tool evaluation methods

and metrics. , 10–24 pages.
[22] Forrest Huang, John F. Canny, and Jeffrey Nichols. 2019. Swire: Sketch-Based User Interface Retrieval. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3290605.3300334
[23] Ali Jahanian, Shaiyan Keshvari, S. V. N. Vishwanathan, and Jan P. Allebach. 2017. Colors – Messengers of Concepts: Visual Design Mining for Learning Color Semantics. ACM Trans. Comput.-Hum. Interact. 24, 1, Article 2 (Jan. 2017), 39 pages. https://doi.org/10.1145/3009924
[24] David G. Jansson and Steven M. Smith. 1991. Design fixation. Design Studies 12, 1 (1991), 3 – 11. https://doi.org/10.1016/0142-694X(91)90003-F
[25] EunJin Kim and Hyeon-Jeong Suk. 2017. Thoughts and Tools for Crafting Colors: Implications from Designers' Behavior. In Proceedings of the 2017 Conference on Designing Interactive Systems (Edinburgh, United Kingdom) (DIS '17). Association for Computing Machinery, New York, NY, USA, 321–331. https://doi.org/10.1145/3064663.3064769
[26] Janin Koch, Andrés Lucero, Lena Hegemann, and Antti Oulasvirta. 2019. May AI? Design Ideation with Cooperative Contextual Bandits. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300863
[27] Ranjitha Kumar, Jerry O. Talton, Salman Ahmad, and Scott R. Klemmer. 2011. Bricolage: Example-Based Retargeting for Web Design. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Vancouver, BC, Canada) (CHI '11). Association for Computing Machinery, New York, NY, USA, 2197–2206. https://doi.org/10.1145/1978942.1979262
[28] Mackenzie Leake, Hijung Valentina Shin, Joy O. Kim, and Maneesh Agrawala. 2020. Generating Audio-Visual Slideshows from Text Articles Using Word Concreteness. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376519
[29] Jiunde Lee. 2007. The Effects of Visual Metaphor and Cognitive Style for Mental Modeling in a Hypermedia-Based Environment. Interact. Comput. 19, 5–6 (Dec. 2007), 614–629. https://doi.org/10.1016/j.intcom.2007.05.005
[30] Ting Liu, Kit Cho, G. Aaron Broadwell, Samira Shaikh, Tomek Strzalkowski, John Lien, Sarah Taylor, Laurie Feldman, Boris Yamrom, Nick Webb, Umit Boz, Ignacio Cases, and Ching-sheng Lin. 2014. Automatic Expansion of the MRC Psycholinguistic Database Imageability Ratings. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). European Language Resources Association (ELRA), Reykjavik, Iceland, 2800–2805. http://www.lrec-conf.org/proceedings/lrec2014/pdf/178_Paper.pdf
[31] Nick Logler, Daisy Yoo, and Batya Friedman. 2018. Metaphor Cards: A How-to-Guide for Making and Using a Generative Metaphorical Design Toolkit. In Proceedings of the 2018 Designing Interactive Systems Conference (Hong Kong, China) (DIS '18). Association for Computing Machinery, New York, NY, USA, 1373–1386. https://doi.org/10.1145/3196709.3196811
[32] Lucero, AA (Andrés). 2009. Co-designing interactive spaces for and with designers:supporting mood-board making. https://doi.org/10.6100/IR641288
[33] Xiaojuan Ma. 2013. Evocation: analyzing and propagating a semantic link based on free word association. Language Resources and Evaluation 47, 3 (01 Sep 2013), 819–837. https://doi.org/10.1007/s10579-013-9219-2
[34] Thomas Merten and Ines Fischer. 1999. Creativity, personality and word association responses: associative behaviour in forty supposedly creative persons. Personality and Individual Differences 27, 5 (1999), 933 – 942. https://doi.org/10.1016/S0191-8869(99)00042-2
[35] Donald A. Norman. 2002. The Design of Everyday Things. Basic Books, Inc., USA.
[36] Heather O'Brien. 2016. Theoretical Perspectives on User Engagement. Springer International Publishing, Cham, 1–26. https://doi.org/10.1007/978-3-319-27446-1_1
[37] Matti Perttula and Pekka Sipilä. 2007. The idea exposure paradigm in design idea generation. Journal of Engineering Design 18, 1 (2007), 93–102. https://doi.org/10.1080/09544820600679679 arXiv:https://doi.org/10.1080/09544820600679679
[38] Savvas Petridis and Lydia B. Chilton. 2019. Human Errors in Interpreting Visual Metaphor. In Proceedings of the 2019 on Creativity and Cognition (San Diego, CA, USA) (C&C '19). Association for Computing Machinery, New York, NY, USA, 187–197. https://doi.org/10.1145/3325480.3325503
[39] Barbara J Phillips and Edward F McQuarrie. 2004. Beyond visual metaphor: A new typology of visual rhetoric in advertising. Marketing theory 4, 1-2 (2004), 113–136.
[40] Elisabeth El Refaie. 2003. Understanding visual metaphor: the example of newspaper cartoons. Visual Communication 2, 1 (2003), 75–95. https://doi.org/10.1177/1470357203002001755 arXiv:https://doi.org/10.1177/1470357203002001755
[41] A. Reid and Ian Solomonides. 2007. Design students' experience of engagement and creativity. Art, Design and Communication in Higher Education 6 (2007), 25–39.
[42] John T. E. Richardson. 1975. Concreteness and Imageability. Quarterly Journal of Experimental Psychology 27, 2 (1975), 235–249. https://doi.org/10.1080/14640747508400483 arXiv:https://doi.org/10.1080/14640747508400483
[43] Daniel Ritchie, Ankita Arvind Kejriwal, and Scott R. Klemmer. 2011. D.Tour: Style-Based Exploration of Design Example Galleries. In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (Santa Barbara, California, USA) (UIST '11). Association for Computing Machinery, New York, NY, USA, 165–174. https://doi.org/10.1145/2047196.2047216
[44] Sam H. Ross, Ecenaz Jen Ozmen, Maria V. Kogan, and Lydia B. Chilton. 2020. WordBlender: Principles and Tools for Generating Word Blends. In Proceedings of the 25th International Conference on Intelligent User Interfaces (Cagliari, Italy) (IUI '20). Association for Computing Machinery, New York, NY, USA, 38–42. https://doi.org/10.1145/3377325.3377527
[45] Albert Rothenberg. 1973. Word Association and Creativity. Psychological Reports 33, 1 (1973), 3–12. https://doi.org/10.2466/pr0.1973.33.1.3 arXiv:https://doi.org/10.2466/pr0.1973.33.1.3 PMID: 4728480.
[46] Yang Shi, Yang Wang, Ye Qi, John Chen, Xiaoyao Xu, and Kwan-Liu Ma. 2017. IdeaWall: Improving Creative Collaboration through Combinatorial Visual Stimuli. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (Portland, Oregon, USA) (CSCW '17). Association for Computing Machinery, New York, NY, USA, 594–603. https://doi.org/10.1145/2998181.2998208
[47] Ben Shneiderman. 2007. Creativity Support Tools: Accelerating Discovery and Innovation. Commun. ACM 50, 12 (Dec. 2007), 20–32. https://doi.org/10.1145/1323688.1323689
[48] Pao Siangliulue, Kenneth C. Arnold, Krzysztof Z. Gajos, and Steven P. Dow. 2015. Toward Collaborative Ideation at Scale: Leveraging Ideas from Others to Generate More Creative and Diverse Ideas. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (Vancouver, BC, Canada) (CSCW '15). Association for Computing Machinery, New York, NY, USA, 937–945. https://doi.org/10.1145/2675133.2675239
[49] Pao Siangliulue, Joel Chan, Krzysztof Z. Gajos, and Steven P. Dow. 2015. Providing Timely Examples Improves the Quantity and Quality of Generated Ideas. In Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition (Glasgow, United Kingdom) (C&C '15). Association for Computing Machinery, New York, NY, USA, 83–92. https://doi.org/10.1145/2757226.2757230
[50] R. Smith. 2007. An Overview of the Tesseract OCR Engine. In Proceedings of the Ninth International Conference on Document Analysis and Recognition - Volume 02 (ICDAR '07). IEEE Computer Society, USA, 629–633.
[51] Lisanne Weelden, Alfons Maes, Joost Schilperoord, and Marc Swerts. 2012. How Object Shape Affects Visual Metaphor Processing. Experimental psychology 59 (07 2012), 1–8. https://doi.org/10.1027/1618-3169/a000165
[52] T Wiegers, Y Song, JSM Vergeest, et al. 2010. Shape Ideation Through Shape Metaphors. In DS 60: Proceedings of DESIGN 2010, the 11th International Design Conference. The Design Society, Dubrovnik, Croatia, 481–490.
[53] Merryl J. Wilkenfeld and Thomas B. Ward. 2001. Similarity and Emergence in Conceptual Combination. Journal of Memory and Language 45, 1 (2001), 21 – 38. https://doi.org/10.1006/jmla.2000.2772
[54] Ziming Wu, Zhida Sun, Taewook Kim, Manuele Reani, Caroline Jay, and Xiaojuan Ma. 2018. Mediating Color Filter Exploration with Color Theme Semantics Derived from Social Curation Data. Proc. ACM Hum.-Comput. Interact. 2, CSCW, Article 187 (Nov. 2018), 24 pages. https://doi.org/10.1145/3274456
[55] Nanxuan Zhao, Nam Wook Kim, Laura Mariah Herman, Hanspeter Pfister, Rynson W.H. Lau, Jose Echevarria, and Zoya Bylinskii. 2020. ICONATE: Automatic Compound Icon Generation and Ideation. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376618