SymbolFinder: Brainstorming diverse symbols using local semantic networks

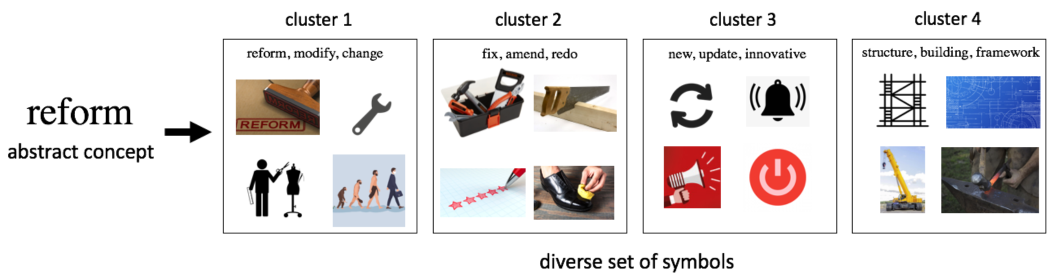

Fig. 1. SymbolFinder takes as input an abstract concept like *reform* and provides a set of related and diverse clusters like "*reform, modify, change*", "*fix, amend, redo*", "*new, update, innovative*", and "*structure, building, framework*" which capture different meanings and concepts associated with *reform*. Users can explore these clusters to find diverse representations of the concept.

Visual symbols are the building blocks for visual communication. They convey abstract concepts like *reform* and *participation* with concrete objects like *scaffolding* and *key*. Student designers struggle to brainstorm diverse symbols because they need to recall associations instead of recognizing them and they fixate on a few associations instead of exploring different related contexts. We present SymbolFinder, an interactive tool for finding visual symbols for abstract concepts. SymbolFinder molds symbol-finding into a recognition rather than recall task by introducing the user to diverse clusters associated with the concept. Users can dive into these clusters to find related, concrete objects that can symbolize the concept. We evaluate SymbolFinder with two studies: a comparative user study, demonstrating that SymbolFinder helps novices find more unique symbols for abstract concepts with significantly less effort than a popular image database and a case study demonstrating how SymbolFinder helped design students create symbol-blends for three cover illustrations of news articles.

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

Additional Key Words and Phrases: brainstorming, symbols, interactive tool, design

ACM Reference Format:
. 2018. SymbolFinder: Brainstorming diverse symbols using local semantic networks. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/10.1145/1122445.1122456

Fig. 2. Four visual symbols, from four domains: transportation hubs, human-computer interfaces, logos, public service announcements.

Visual symbols play a vital role in daily communication. They are placed on signs to help us locate services in transport hubs and public buildings, like the "lost and found" sign in Figure 2. They are used in human-computer interfaces to represent actions a user can take, such as the search icon symbol. They represent organizations and companies like the World Wildlife Fund's logo. Symbols are also combined in images to represent more complicated messages like "Earth is melting" in the public service announcement in Figure 2. Visual symbols are an important part of how we convey information.

While visual symbols come up in a wide range of different contexts, each symbol in Figure 2 share a key similarity: each represents an abstract idea with a concrete object. In the "lost and found" sign, the idea of *items commonly misplaced* is represented by an umbrella and a glove. In the look-up icon, *search* is represented by a magnifying glass. In the WWF logo, *endangered* is represented by a panda bear. Finally, in the public service announcement, *global warming* is represented by the Earth and an ice cream cone. These types of symbols are known as "representational icons," which consist of simple images of familiar objects that represent abstract ideas [36].

While there has been a great deal of work in the graphic arts and in icon design on how to create representational icons given a representative object [27] [16], the problem of how to find these representative objects in the first place has been relatively overlooked. Visual language is constantly evolving. New symbols are constantly being created to represent new experiences, organizations, and interactions on interfaces [24]. Novices with little to no experience in graphic design are also creating symbols and icons for logos, websites, slide decks, mobile apps and games. Novices have difficulty not only designing icons from concrete objects, but also finding concrete objects to represent the concepts they want to symbolize in the first place.

Even more difficult is finding symbols that can be combined or blended to create a new meaning. For example, a cover illustration for a news article on "police reform" could try to convey the message by combining symbols of *police* and symbols of *reform*. The two symbols must be combined in such a way that their shapes blend naturally and the combined design accurately reflects the emotional tone of the message [10] [21]. To accommodate such constraints and create many design alternatives, it is essential to find a diverse set of symbols for the abstract concepts being depicted. However, converting these abstract concepts into a diverse set of visual representations is hard for novice designers, preventing them from effectively combining them to convey a message.

In order to understand the challenges and workflow of novice designers, we conducted a formative study, where novice participants used Google Images, a popular image database, to find symbols for abstract concepts. We observed that novices relied almost exclusively on recalling their own associations about the concept to search for related images. They often had difficulty brainstorming many different related words, and ended up fixating on a narrow set of associations, which represented a limited aspect of the concept being symbolized. Novices needed help to explore diverse ideas, which is crucial to finding an effective and creative solution [40] [50]. Finally, novices struggled to convert abstract associations into concrete objects and actions that could visually represent the concept.

Inspired by these observations, we created SymbolFinder to help novices to find compelling visual symbols. SymbolFinder helps users brainstorm associations by providing related words from an expansive word association data set. By clustering the related words into groups, each of which represents a related but distinct aspect of the concept, SymbolFinder encourages users to explore a broad range of related contexts, rather than fixating on a narrow set of associations. To create these clusters, SymbolFinder constructs a semantic network of word associations and detects highly connected communities of words. Finally, SymbolFinder helps users find imageable objects and actions by organizing words related to each cluster by word-concreteness.

This paper presents the following contributions:

- SymbolFinder: an interactive interface for finding concrete images to represent abstract concepts.
- A technique for applying local semantic networks to help brainstorm distinct, concrete associations for abstract concepts.
- A preliminary study demonstrating that by organizing word-associations into clusters, users had to search less for diverse symbols and felt they were better able to brainstorm.
- A user study demonstrating that SymbolFinder enables users to find significantly more unique symbols with less perceived effort than with Google Images, a popular image database used to find symbols.
- A case study showing the use of SymbolFinder in practice, in which a team of 3 design students create cover illustrations, consisting of symbol blends, for 3 articles. SymbolFinder helped the team find a diverse set of symbols for the concepts of each article, which they successfully combined to form more than 10 different symbol-blend prototypes each.

2 RELATED WORK

2.1 Visual Symbols

Symbols are fundamental in visual communication and are used in a variety of contexts. They accompany headlines in news articles [24], represent actions in computer interfaces [46] [27], guide people in transportation hubs [41], represent corporations in logos [36], and form associations with products in advertisements [30]. There are many advantages in communicating ideas with symbols. Symbols often require less space to encapsulate an idea than using the word itself, saving space in interfaces, maps, and signs [23]. People can more quickly and easily recognize symbols than words because of our innate visual processes [29] [46]. Symbols are more universally understood than words across cultures, which is why they are used and designed for international transportation hubs [41] [35]. Finally, depicting ideas pictorially aids their memorability and recognition [4] [3]. For these reasons, we built SymbolFinder, to help convey more abstract ideas visually.

2.2 Brainstorming and Exploration Tools

Many tools have been created to help people brainstorm and explore related ideas. These systems are often designed to present a small set of related words or images to inspire new ideas. To present related textual ideas, *InspirationWall* [1] presents a few related topics from a knowledge graph, *V8Storming* [33] uses word embeddings to find similar words to suggest, and *CrowdBoard* [33] utilizes a real-time crowd to suggest more personalized ideas. Other tools like *Idea expander* [49] and *IdeaWall* [45] present a few related images based on the current spoken ideas of its users. Koch et. al. created a cooperative contextual bandit system which recommends a few images that match the user's current semantic and visual preferences [34]. While displaying closely related words and images is very helpful for finding symbols, it is also necessary that these recommendations encapsulate different aspects of the concept. SymbolFinder organizes a network of words into clusters capturing distinct associations, enabling users to explore diverse contexts and images for an abstract concept.

Clustering is a popular method used to help users understand and explore large datasets. *Scatter/Gather* enables users to interactively choose clusters to find and explore specific documents in a large collection [12]. *Exploratory Lab Assistant* presents clusters of documents to users as a preliminary step to help them label groups of documents themselves [18]. *Recipescape* clusters recipes for a dish based on the structure of its preparation, enabling users to find recipes with similar or different steps [9]. SymbolFinder clusters word associations to present users with diverse ideas related to the concept being symbolized. Word association data sets are often analyzed as networks, where words are nodes and edges represent associations between them [15] [14]. In this format, they are referred to as semantic networks. We construct a "local" semantic network, consisting of words near the concept being symbolized, and cluster it using a popular network clustering algorithm [2].

2.3 Query Expansion and Exploratory Image Search

The queries that users enter when searching for images are often ambiguous and can refer to many different real-world entities. Many researchers have recognized this problem and created tools to help users either clarify their search or find what they're looking for by providing a diverse set of image results. Textual query suggestions are a common technique for helping a user clarify their search. Keywords can be extracted from the most relevant documents associated with the query [51] [7] or taken from commonly occurring pairs of queries from search logs [28] [20]. Zha et. al. improved upon this technique by showing clusters of visually similar images for each keyword to help users preview and compare the images for each keyword [52]. *IGroup* employs a similar technique, by extracting common phrases (n-grams) from the most relevant documents associated with the query and presenting clusters of images for each of these phrases [31]. While keyword suggestions help users disambiguate their queries, they do not let them explore the broader associated meanings and contexts of their search. By using a word associations dataset, SymbolFinder presents broader contexts that expand the users idea of the query to help them brainstorm.

Instead of using the documents associated with the images, other tools expand queries with knowledge bases to capture diverse intentions. *PARAgrab* takes synonyms, hyponyms, and hypernyms from WordNet [39] and presents these as related searches to users [32]. Hoque et. al. use both the incoming and outgoing of links of the query's Wikipedia page to provide a list of related queries [26]. A separate knowledge base is used to cluster these associations into categories like person, place, and location. *CIDER* adds to this work by spatially arranging the images from these different queries based on their visual attributes [25]. These tools serve to quickly disambiguate a user's search, like separating Denzel Washington the actor from Washington D.C., the place. However, the organization of these related concepts does not capture different meanings and greater contexts associated with the query. For example, for a query like *reform*, instead of returning a list of specific types of *reform*, SymbolFinder presents a set of diverse clusters, each encapsulating a different sense of *reform* like "fix, amend, redo" and "new, update, innovative", to help the user brainstorm.

There exist also a multitude of exploratory image search tools that help users explore diverse results by clustering images. Cai et. al. use text, link, and visual features to cluster a query's image results [6]. Leuken et. al. create a similar system, involving a dynamic weighting function for the visual features, creating clusters that better align with a human's idea of image diversity [48]. Fan et. al. create a visual "summary" of image results on Flickr by creating a topic network from user-generated tags. This enabled users to view an overview of the various images connected to their query and explore highly connected clusters [17]. By providing clusters of word-associations associated with the query, SymbolFinder also presents an overview of diverse contexts related to the query. However, a crucial component of SymbolFinder is the ability to dive deeper into each cluster and explore concrete words. By incorporating concreteness and in-depth exploration of each cluster, SymbolFinder helps users find objects to symbolize their abstract query.

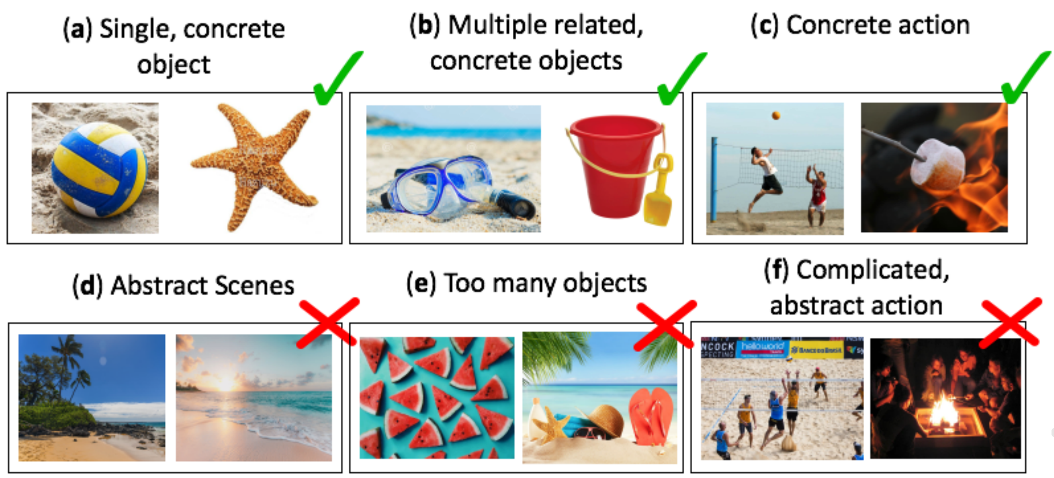

Fig. 4. The rules for what makes a good symbol, derived from theory on icons and symbols, and explained with the concept: *summer*. These rules were shown to both raters and participants.

3 BACKGROUND: WHAT MAKES A GOOD SYMBOL?

According to the theory of symbols, there are three basic types of symbols: abstract, directly representational, and indirectly representational [37] (Figure 3). A symbol is abstract when an abstract pattern represents the idea, like the radioactive symbol. A symbol is directly representational when its content is an exact representation of its idea, like the telephone symbol in Figure 3. A symbol is indirectly representational when the image content is associated with but not an exact representation of the idea, like the coat hanger, which represents a *coat check* (Figure 3). SymbolFinder was built to help people find indirectly representational symbols for abstract terms that have a variety of meanings and contexts associated with them. These types of symbols do not require a new design like the *radioactive* symbol and are difficult to find with current image databases, unlike directly representational symbols, as these databases do not enable an exploration of various ideas related to the concept.

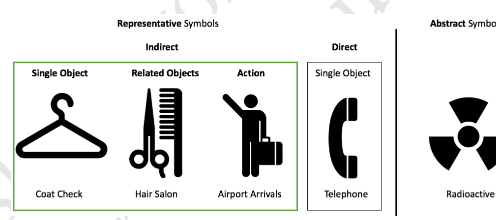

Fig. 3. The types of symbols include: representative (indirectly and directly), as well as abstract (radioactive symbol).

A representational symbol can contain three things: a single object, a few related objects, or an action [27]. For example, the coat hanger is the most essential object related to a *coat check*, and thus makes a good single object symbol. Sometimes an extra object makes a symbol more specifically related to the idea it represents. For example, a scissor and a comb together represent a *hair salon* better than either one alone. The two of them together effectively represent the tools a hair stylist uses. Finally, a symbol can also contain an action, like the *airport arrivals* symbol, in which there is a man hailing a taxi. These three categories make up the vast majority of the content displayed in representational symbols.

A good representational symbol is simple and concrete in the content it depicts [16] [27]. The most essential quality of a symbol is that it is recognizable. Its content should contain no more than what is necessary to depict the idea. Visual complexity and extra entities only make them slower to interpret and recognize. From this symbol theory, we establish a set of rules to help users of our system find good symbols (Figure 4). A good, indirectly representational symbol can be:

- **A single concrete object**. The object must be able to represent the concept on its own (Figure 4a).
- **Multiple related objects**. The objects should be related to the concept and to each other, like the combination of the scissor and comb in Figure 3. However, they should not be the same object, like the watermelon slices (Figure 4e), since one is enough to convey the idea. Symbols with more than two representative objects, like the collection of unrelated beach objects in (Figure 4e) are too complicated and can be separated into separate symbols.
- **A concrete action**. The action should be concrete and shown clearly, like the volleyball spike in (Figure 4c), as opposed to the more complex volleyball scene in Figure 4f.
- **No abstract scenes**. Symbols should depict a concrete object or action, instead of abstract landscapes (Figure 4d).

4 FORMATIVE STUDY

To better understand the pitfalls novices face when searching for symbols and how to help them, we conducted a formative study in which we observed participants search for symbols with a popular image database, Google Images. Google Images is often used by novice and professional icon designers to look up visualizations of concepts [53]. Its interface also has a number of features to help users search for particular images, including a list of suggested searches that appear above the image results, each consisting of a related term and a representative image. It also includes the ability to filter images by color and type (clip Art and line drawing), under the "tools" section. Because of its popularity, usage by professional icon designers, and its features, we study how novices use Google Images when searching for symbols.

We recruited 5 participants (3 male, 2 female, average age 24.8). They found symbols for three abstract concepts: *old*, *exciting*, and *innovation*. These concepts were selected randomly from a visual messaging dataset, containing the most common concepts symbolized in online messages [30]. Prior to finding symbols, users were shown a slide-deck detailing the task: to find 20 unique symbols, and shown the good symbol rules (Figure 4). They had 10 minutes to find symbols for each concept. While searching they were asked to think aloud to convey their thought process. After each concept, users were asked to explain the benefits and drawbacks of Google Images, what search terms helped their brainstorming, and their general strategy.

All five participants were frustrated by the lack of conceptual diversity in the images presented when searching the "concept" as is on Google Images. P1 and P2 both mentioned that the results for *old* predominantly contained images of old people. Similarly, upon seeing the image results for *excited*, P1 states, "These are all images of the word 'excited'. Or just people looking excited." While there was generally a couple representations of the concept in the first set of images produced by Google, users found that they needed to brainstorm on their own to find different symbols.

The most common strategy to find different images was to search terms related to the concept and scan the image results for new visualizations. For example, P1 searched *ancient*, which he recalled on his own, and met many images of the Parthenon, the Colosseum, and pyramids. This turned out to be a fruitful context, from which he was able to collect an additional three symbols for *old*. Similarly, when seeing only images of excited people in the results for *exciting*, P2 subsequently searched *fun* and *adventure*. In doing so, he found other contexts related to *exciting* like extreme sports. Users had to recall these associations on their own. Therefore, our first design goal for SymbolFinder was to **help users brainstorm related words**, in order to enable recognition over recall.

Users however also struggled to find related words that presented different images and concrete contexts related to the concept. For example, when searching for symbols of *exciting*, P2 searched for images of *adventure* and *explore* and was met with similar images of hiking and camping. While he was able to collect a number of symbols from these searches, it was difficult for him to think of another related word that encapsulated a different flavor of *exciting*. Eventually, he searched the word *suspenseful* and found images of horror movies and theatre which inspired more symbols. From this issue we formed our second design goal: when helping users brainstorm associations, we should ensure that we present diverse ideas in order to help them collect **diverse symbols**.

Once users found a fruitful context, their strategy shifted to searching concrete objects and actions that they would select as their symbols. For example, while searching for symbols of *innovation* P2 started searching for advanced technology like virtual reality goggles and hovercrafts. Similarly for *old*, P1 searched for objects old people use like canes and wheelchairs. While more abstract searches like *elder* and *technology* served as inspiration, these highly concrete searches contained the images that would end up being their symbols. When exploring related contexts, users should be able to explore concrete words within these contexts to find representative objects and actions. Thus, our third design goal was to help users **concretize abstract concepts**.

Design Goals. In summary, from the formative study we formed three design goals for the SymbolFinder:

D1: Help brainstorm related words to encourage recognition over recall. Users often recalled related terms to see new visualizations of the concept. By relying on their own memory, they miss obvious symbols and contexts associated with the concept.

D2: Symbol diversity. When helping users brainstorm related terms we should present them a variety of diverse associations so that they can collect diverse symbols from these associations.

D3: Concretize abstract concepts. As well as enabling users to explore diverse associations, they should also be able to explore related concrete terms for each association. This way, users can better find objects and actions to represent the concept.

5 SYMBOLFINDER

To address these design goals, we present SymbolFinder, an interactive tool that enables novices to find symbols for abstract concepts, by facilitating an exploration of diverse contexts associated with a concept. SymbolFinder's interface consists of two phases. Phase 1 is a breadth-first exploration of clusters of associations related to the concept; users select clusters they would like to explore further (Figure 5). In phase 2, users select symbols while exploring words and images associated with the clusters they selected in phase 1 (Figure 6).

5.1 Phase 1: Breadth-first concept exploration

Phase 1 addresses D1 (help brainstorm related words) and D2 (symbol diversity). To help users brainstorm associations, we enable users to explore different words related to the concept. To encourage symbol diversity, similar related words are clustered together into associations that encapsulate a unique aspect of the concept. Each cluster is represented by its three most important words. For example, two clusters for *control* are "rule, government, governance" and "handle, lever, knob" (Figure 5). Users scroll through the clusters and select clusters to explore further. For each cluster, the user is posed the following question: "Could symbols of [word 1], [word 2], [word 3] represent [concept]?". The user is instructed to answer this question based on the words. If they think there are potential symbols for these words that could represent the concept, then they press "yes". Below this question are also 5 images related to these words (Figure 5b). Users have the option to select an image if they think it is a good symbol. These images come from three Google Image searches, one for each of the words, where each query is formulated as follows: "[concept] [word]". This is done to keep the results relevant to the concept. The queries for top cluster in figure 5 were: "control rule", "control government", and "control governance". By having users explore each cluster briefly, we quickly expose them to a number of different associations, preventing fixation on any single association.

5.2 Phase 2: Image selection within clusters

Phase 2 further supports D1 (help brainstorm related words) as well as D3 (concretize abstract concepts). In phase 2, users further explore the clusters they selected from phase 1 (Figure 6a) and select symbols (Figure 6c). The key part of this interface is the sidebar on the left which is where users explore the clusters (Figure 6a) and recursively explore concrete words related to them (Figure 6b). To support D3, when users select the top level cluster words, they are shown related words sorted by concreteness (Figure 6b). In Figure 6, the user selected the "rule, government, governance" cluster. They then expanded *regulation*, one of the cluster words, and selected *referee*, a related concrete term. As well as exploring the clusters, users can also type associations they think of themselves in the "write your own" text boxes and view images related to their entry. In this way, the sidebar enables users to recognize good symbols as well as use their own thought processes.

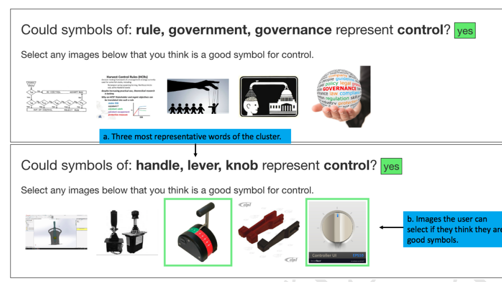

Fig. 5. Phase 1 for the concept: *control*. Users select relevant clusters they would like to explore further in phase 2. Each cluster conveys a different association related to *control*, like the government (top) or physical tools we use to control machines. (bottom).

The second key part of this interface is the set of Google Image search results that populate the screen when a word is selected (Figure 6c). Four queries are made per word, and they include the word on its own [referee], the word and its parent [regulation referee], the concept and the word [control referee], and finally the word and "icon" appended to the search [referee icon]. We include the parent and concept queries as they help keep images on topic. We include the icon query as they often provide simple images of the action or item we are looking for. Together, the sidebar of clusters and the multiple image searches effectively help users to find concrete symbols.

6 IMPLEMENTATION

6.1 D1: Helping users brainstorm related words

In the formative study, participants searched words related to the concept in order to find new images of symbols. This was a good strategy, but novices had difficulty recalling related words on their own, and missed many useful words. Therefore our first design goal was to help users brainstorm related words by making it a recognition rather than a recall task. First, in phase 1, SymbolFinder presents clusters of word associations for users to browse and choose from. Then, in phase 2, users can explore more related words associated with each cluster. We explored two different options for creating word associations: (1) Glove word embeddings, trained on Common Crawl [43] and (2) Small World of Words (SWOW), a crowd-sourced word association database

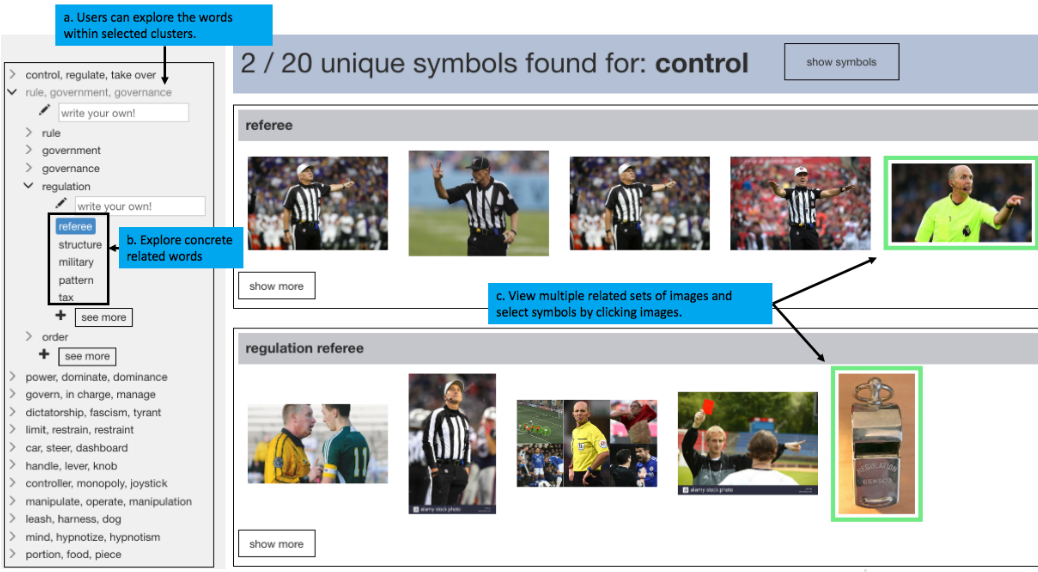

Fig. 6. Phase 2: Users explore the clusters they chose from the first phase. On the left sidebar are the clusters, where users can explore related words. On the right are a few Google Image searches for the selected word, which here is “referee”. The user selected two images, indicated by the green boxes.

[13]. Word embeddings are commonly used for comparing the similarity between words [38] and have been used in a number of brainstorming tools to compare the similarity of ideas [45] [8]. SWOW is a large English word association dataset. The dataset was created by having thousands of participants complete a free word association task, in which each participant records the first three words they think of when seeing a cue word.

In initial testing, we found that SWOW produced words that were more relevant and concrete than those by Glove. For example, for the abstract word *help*, the most related words that Glove produces include words like: *helping* and *need*, which are related, but are not specific or concrete. Meanwhile SWOW produces terms like: *donation*, *red cross*, and *tutor*, giving specific actions, organizations, and people that help. These words are actually helpful for visualizing *help*. At the same time, word embeddings tend to also return antonyms when querying similar words, which was less of a problem with SWOW [42]. For these reasons, we chose SWOW to be our dataset for providing related words.

6.2 D2: Diversity using Local Semantic Networks

In the formative study, users struggled to think of related words that represented new associations with the concept they were symbolizing. They explored redundant terms and fixated on only a few related contexts. To support D2 and help users find diverse symbols, we cluster a large set of word associations and present a diverse set of contexts from which users can find symbols.

Word association datasets are often analyzed as networks, where each word in the dataset is a node and each association is an edge. In this format, they are referred to as semantic networks. Researchers analyze these networks to identify important words in languages via network centrality measures as well as highly connected clusters of words [15] [14]. We leverage this technique to find clusters of words close to the concept we are symbolizing. We create a “local” semantic network, consisting of words in the concept's neighborhood. We then run a popular network clustering algorithm on this network to identify sets of highly connected words that represent distinct associations of the concept.

6.2.1 Constructing a local semantic network. Our high-level goal when constructing the local semantic network was to create a network consisting of a variety of words to generate a diverse set of clusters, without making the clusters too large so that they became indecipherable. To ensure a variety of words, we construct a network with two levels of nodes. This network consists of the root concept, its list of related words (level 1), and the related words of nodes in level 1 (level 2). This is shown in Figure 7, where the root is *control*. The words of level 2 are still relevant to the root concept but also representative of a more specific association to the root. For example, in Figure 7, from *control* we get *manage* which yields *boss*. While *boss* is more directly associated with *manage*, it is still very related to *control*. Each word in SWOW has a set of associated words, each with a count indicating the number of unique participants who made that association. For every pair of nodes in the network, we create an undirected edge between them if they are associated, where the weight of that edge is the associated count. We do not construct a third level of words from the second as we found this includes many irrelevant nodes in the network, making the clusters less interpretable. By incorporating two levels of words, we ensure the words in the network are related to the concept and diverse.

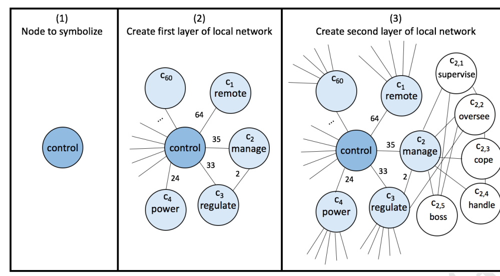

Fig. 7. The construction of the local semantic network.

To ensure that the clusters do not get too large or indecipherable we set two constraints. The first constraint is that we take up to the first c related words for the root concept to include in level 1. We empirically determined 60 to be a suitable value for c; this leads to clusters containing on average 21 words. The size of the clusters will not vary much for small changes in c, but generally when set to 90 or higher, the clusters tend to get large, containing on average 35 words, becoming less interpretable. That being said, the average number of related words per concept is 20.32, with 6.1% of the dataset containing more than 60, so we generally take all of the root concept's related words. The second constraint pertains to the nodes we add to level 2. For each of the c nodes, we take their first g most related words not present in the network. We set g to be 5. We keep g relatively low compared to c because some level 1 nodes, like *remote* in figure 7, introduce words unrelated to the root concept, *control*. Also by setting c to 60 and g to 5, we ensure a maximum number of nodes that does not exceed 360, which keeps the cluster sizes in the next step manageable. While simple, these settings produce a network that is reasonably sized and a set of clusters that are coherent but also diverse.

6.2.2 Clustering the network. Our goal is to present users a list of diverse and distinct ideas related to the concept. The local semantic network we create can contain up to a few hundred words. While containing many different associations, this is too large a space for users to explore without guidance. To condense this information into a manageable set of diverse associations, we cluster the network. Our goal is to create a set of clusters that contain highly related words that capture a distinct association of the concept. To cluster the network we considered two algorithms: the Clauset-Newman-Moore [11] and Louvain [2] network clustering algorithms. Both are very popular hierarchical clustering algorithms. From initial experimentation we determined that the Louvain

Exciting		Future	
Related	Concrete	Related	Concrete
Fun	Motorcycle	Past	Crystal Ball
Thrill	Race car	Tomorrow	Hovercraft
Happy	Roller coaster	Crystal Ball	Robot
Interesting	Package	Someday	Car
New	Firework	Prediction	Spacecraft

Table 1. Associations sorted by strength of association (related) and by concreteness (concrete) for two abstract concepts: *Exciting* and *Future*. Concreteness surfaces imageable objects that can serve as symbols.

algorithm produced more interpretable clusters, as the Clauset-Newman-Moore algorithm tended to produce fewer clusters with a greater number of words, often combining clusters that the Louvain algorithm separated.

The Louvain algorithm optimizes modularity, a measure which compares the edge density of the nodes in a cluster to the edge density of the same nodes in a randomly generated network. In our case, the algorithm identifies communities of highly related words by grouping words connected with high edges weights. The algorithm returns a hierarchy of clusters. We use the highest level of clusters (i.e. largest cluster size). For our dataset, the final pass of the algorithm generates about 12 to 20 clusters, which is small enough for a user to explore and large enough to contain a variety of unique ideas related to the concept.

6.3 D3: Concretize abstract concepts

In the formative study, even when users found relevant associations, they still had trouble identifying concrete objects and actions that could visually represent the concept. Therefore, our third design goal was to enable users to explore concrete words related to contexts associated with the concept. To support D3, users can explore related concrete words for each of the words within the clusters. For example, in Figure 6c, the user expanded *regulation*, one of the cluster words, to view its related concrete words. The most concrete word related to *regulation* is *referee*, which led to two good symbols: a soccer referee and a whistle. Thus, by exploring these related concrete terms, users can find objects and actions related to abstract clusters like "rule, government, governance".

Concreteness is a measure of how physically perceptible something is. For example, words like *seashell* are more concrete than words like *fun*. To incorporate concreteness, we use a crowd-sourced dataset that includes concreteness ratings for 40,000 English words and phrases [5]. Crowd-workers rated words on a scale from 1 (abstract) to 5 (concrete). When resorting the words, we incorporate both concreteness and strength of association, as sorting by concreteness alone leads to random words at the top. To include concreteness, we normalize the word's association strength and multiply this value by the word's concreteness score. We sort by this product. Consider the examples shown in Table 1. Instead of abstract related terms like *fun*, the concrete lists provide imageable words like *motorcycle* that can symbolize the abstract concept: *exciting*.

7 PRELIMINARY CLUSTER STUDY

To address D2 (symbol diversity), SymbolFinder presents clusters of unique associations in order to enable users to easily explore a variety of symbols. In this study we evaluate whether the clusters represent coherent ideas related to the concept and whether they help users find different symbols more efficiently than a simple list of related words. We conducted a comparative user study where participants used two different versions of the SymbolFinder: (1) cluster interface, the tool as described in the SymbolFinder section, and (2) list interface. The list interface is exactly like the cluster interface except for two key differences: (1) there is no phase 1 exploration of clusters, (2) the sidebar consists of the concept's related words, sorted by strength of association. In the list-interface, users can still explore related, concrete terms for each of the concept's related words. After using both versions of the interface, participants completed a semi-structured interview which gauged their preference of the interfaces and their opinions of the clusters' usefulness.

7.0.1 Methodology. We recruited 6 participants (2 male, 4 female, average age of 24.7). Each participant searched for symbols of two different concepts of similar concreteness, randomly chosen from the same visual message data set used in the formative study: *Halloween* (concreteness = 0.64) and *energy* (0.62). Half of the participants found symbols for *energy* with the cluster interface and *Halloween* with the list interface, while half did the opposite. Order of interface usage and concept assignments were randomized. They were given ten minutes to find 20 different symbols per concept. Prior to finding symbols for the two test concepts, they practiced finding symbols for *summer* with each system for up to 10 minutes. Afterwards, they completed a semi-structured interview.

7.0.2 Results. **Participants found that most of the clusters were coherent and related to the concept.** When using the cluster interface, participants kept on average 16.3 ($\sigma = 1.2$) of the 19 *energy* clusters and 12.3 ($\sigma = 0.9$) of the 16 *Halloween* clusters, indicating that the majority of clusters were deemed relevant. Generally, for both concepts, there were a couple clusters that users felt were unrelated, like the "harness, horse, dog" cluster for *energy*. But in the interview afterwards, all confirmed it was generally easy for them to rule out these clusters in phase 1.

Participants had to search less to find symbols with the cluster interface than with the list interface. Users expanded on average 42.5($\sigma = 5.7$) words with the list-interface and 25.47($\sigma = 4.9$) words with the cluster-interface. This difference is predominantly due to phase 1 (Figure 5) of the cluster-interface, where participants collected on average 9.5($\sigma = 3.4$) of the 20 symbols they were asked to find. Subsequently, in phase 2 (Figure 6), users were able to narrow their search and explore fewer words. All users preferred the cluster-interface over the list-interface because they appreciated searching less with the cluster-interface. P1 noted that exploring so many disparate words with the list-interface was "more frustrating. Everything was disorganized. I had to jump from one idea to another". Users had to think more when using the list-interface.

Clusters helped users brainstorm more associations. In the interviews, participants noted that clusters provided greater context to the related words, enabling them to think of more associations. For instance, for P2, the clusters were "a jumping off point. I thought of different types of fuel and vehicles when I saw the 'fuel gas car' cluster." Meanwhile, words in the list-interface suffered from a lack of context. For example, when seeing *green*, a word associated with *energy*, P5 stated that she immediately thought of the color, instead of *renewable energy*, its actual association. Clusters formed more coherent ideas with clearer connections to the concept than single words.

To summarize: participants found the clusters to be coherent and related to the concept, searched less for symbols with the cluster-interface than with the list-interface, and more easily formed their own connections with the clusters than with the list-interface.

8 EVALUATION

In this study we evaluate whether users can find more unique symbols with SymbolFinder than with Google Images for both very abstract and less abstract concepts. Google Images is a good baseline as both professional icon designers and novice designers alike use it to browse visualizations of ideas [53]. We also compare the perceived difficulty of finding symbols with SymbolFinder to finding symbols with Google Images.

Fig. 8. Users collected more unique symbols for each concept with SymbolFinder. Above are results for *control*, where the SymbolFinder user collected 15 unique symbols and the Google Image user collected 8.

8.0.1 Methodology. We recruited 10 graduate students via e-mail through a local university: 2 female and 8 male, with an average age of 26.6. Each participant found symbols for six concepts and after each concept, completed a NASA-TLX questionnaire to measure perceived effort [22]. Prior to finding symbols, each participant was shown a slide-deck explaining the task: find 20 different symbols for each term. They are also shown the symbol rules in Figure 4; we emphasize that the symbols should be unique, displaying different concrete objects and activities.

We compare SymbolFinder with Google for concepts of varying levels of concreteness. The six concepts were randomly selected from the same visual messaging dataset [30] from three levels of concreteness: most concrete (*fast*, concreteness=0.66, *art*, 0.83), less concrete (*dangerous*, 0.46, *rugged*, 0.55), least concrete (*control*, 0.38, *simple*, 0.32). Users found symbols for the concepts in the following order: *fast*, *dangerous*, *control*, *art*, *rugged*, *simple*. To counterbalance the experiment, we had half of the participants use SymbolFinder for the first three concepts and half use it for the last three. For each tool, users find symbols for concepts from three different levels of concreteness. We hypothesize that **users will find more unique symbols across all concepts with SymbolFinder**. We also hypothesize that **SymbolFinder will help more with the most abstract concepts**, as concepts like *simple* and *control* apply to multiple diverse contexts, and SymbolFinder presents a diverse set of unique associations (D2).

Prior to finding symbols, users practiced with each interface for ten minutes to collect symbols for the concept *summer*. During this practice, users were guided through the features of each tool. In the Google condition, participants used the standard Google Image web-page. They were shown the related search terms that appear above the images as well as the filter tools. While practicing, users could ask any questions about the goodness of their symbols. When finding symbols for the six concepts, they had a "cheat sheet", containing the rules of figure 4, but were no longer helped by the experimenter. Participants had 10 minutes to find 20 or more symbols for each concept.

After finding symbols for each concept, participants completed a NASA Task Load Index (NASA TLX) survey, to understand their perceived effort for each word-tool combination [22]. The NASA TLX is a standard measure for subjective workload when using a tool, measuring: mental demand, physical demand, temporal demand, performance, effort, and frustration level. We hypothesize that **users will find SymbolFinder easier to use** as it helps users brainstorm associations (D1) and find concrete objects (D3), requiring less mental demand and effort.

We recruited two graduate students in design to annotate the collected images for unique, good symbols. They annotated good symbols based on the rules of Figure 4, which are rooted in symbol theory. Along with these rules, annotators were instructed to reject images that seemed unrelated

	SymbolFinder	Google	P-value
Total Symbols	26.1 (9.6)	16.2 (4.96)	**<0.001**
Good Symbols	22.7 (8.26)	14 (4.49)	**<0.001**
Unique Symbols	14.8 (5.5)	9.92 (2.89)	**<0.001**

Table 2. The total number of symbols, good symbols, and unique good symbols found by participants in the second user study. Bold P-values indicate statistical significance. In parentheses is the standard deviation. Participants found significantly more unique symbols with SymbolFinder than with Google.

	SymbolFinder	Google	P-value
Mental Demand	5.13 (1.41)	6.8 (1.97)	**<0.001**
Physical Demand	1.97 (1.43)	3.97 (2.58)	**<0.001**
Temporal Demand	5.13 (2.55)	6.17 (2.44)	0.10
Performance	6.77 (2.03)	5.9 (1.8)	0.061
Effort	4.87 (1.67)	7.43 (1.52)	**<0.001**
Frustration	3.0 (1.93)	4.33 (1.92)	**0.034**

Table 3. NASA-TLX Questionnaire results. Bold P-values indicate statistical significance. In parentheses is the standard deviation. SymbolFinder required significantly less mental demand, effort, physical demand and frustration.

Concept	Art		Fast		Rugged		Dangerous		Control		Simple	
Concreteness	0.83		0.66		0.55		0.43		0.39		0.32	
Condition	SF	G	SF	G	SF	G	SF	G	SF	G	SF	G
Unique Symbols	13.4	10.1	**18.4**	10.7	11.1	9	18.4	12	13.4	10.2	**14.3**	7.3

Table 4. With SymbolFinder (SF), participants found more unique symbols for all concepts, across all levels of concreteness than with Google (G). For *fast* (a less abstract term) and *simple* (a very abstract term) this difference is statistically significant.

to the concept. As well as symbol goodness, the annotators also labeled duplicate symbols. They were instructed to label two images as duplicates if they conveyed the same object or activity, regardless of style, background or color. Because of the natural subjectivity of this task, we had the annotators label two practice sets of images for good and unique symbols together. They then annotated the images collected during the evaluation separately.

8.0.2 Results. We report the percent agreement between the two raters for determining good and unique symbols (and the Cohen's Kappa correlation coefficients) to be 94%(0.74) and 96%(0.75) respectively. To calculate the results in Table 2, we averaged the number of good and unique symbols determined by the raters for both conditions.

Participants found more unique symbols with SymbolFinder. We conducted paired t-tests and found with SymbolFinder, participants collected significantly more unique, good symbols than with Google Images, finding an average of 14.8 unique symbols across the six concepts with SymbolFinder and 9.92 unique symbols with Google ($t = 4.39$, $p < 0.001$) (Table 2). Figure 8 shows results for *control*; the SymbolFinder user almost doubles the amount of unique symbols found by the Google Image user. SymbolFinder users collected significantly more symbols, many of them good. Participants on average collected 26.1 symbols per concept with SymbolFinder compared to 16.2 with Google ($t = 4.5$, $p < 0.001$) , with an average of 22.7 good symbols per concept, compared to 14 with Google ($t = 4.75$, $p < 0.001$) . In both conditions, users had a tendency to select multiple symbols of the same object or activity, leading to a drop off from good symbols to unique symbols.

SymbolFinder helped users find more symbols for both very abstract and less abstract concepts. Participants were able to find more unique symbols for each of the six concepts with SymbolFinder than with Google, regardless of the concept's concreteness (Table 4). We conducted unpaired t-tests and found the difference in unique symbols was statistically significant for two concepts: *fast* a less abstract term and *simple*, a more abstract term. Participants collected an average of 18.4 unique symbols for *fast* with SymbolFinder and 10.7 with Google ($t = 6.4$, $p < 0.001$). For *simple*, participants collected 14.3 unique symbols with SymbolFinder and 7.3 with Google ($t = 2.4$, $p = 0.04$). For both these terms, despite their difference in abstractness, SymbolFinder presented a number of associations Google users did not consider. For example, when finding symbols for *simple* with Google, participants fixated on a single meaning of *simple*, "easily understood" and collected symbols of shapes and simple mathematics. Meanwhile, SymbolFinder users found symbols for this context and many others, including "primitive" and "pure", collecting symbols like the caveman wheel and a water droplet. A similar effect occurred with *dangerous*, where SymbolFinder users found 53% more symbols than Google users. Meanwhile, SymbolFinder was less useful for *art*, the most concrete concept. While the clusters provided many different associations, like sculpture, painting, film, and more, it was relatively easy for users to brainstorm many concrete items on their own with Google Images.

SymbolFinder was easier to use than Google. A summary of the NASA-TLX results are shown in Table 3. Participants found using SymbolFinder to be significantly less mentally demanding than Google Images, reporting an average mental demand of 5.13 with SymbolFinder and 6.8 with Google Images ($t = -3.98$, $p < 0.001$). Similarly, frustration was also significantly lower with SymbolFinder (3) than with Google (4.33), ($t = -2.23$, $p = 0.034$). Users often hit dead-ends of redundant symbols with Google, increasing frustration and mental demand as they brainstormed unassisted for new ideas. Physical demand was significantly lower for SymbolFinder, with an average of 1.97, compared to 3.97 with Google ($t = -4.85$, $p < 0.001$). This is likely due to users having to copy and paste images from Google into a slide-deck. For the same reasons as mental and physical demand, effort was also significantly lower for SymbolFinder (4.87) than Google (7.43), ($t = -6.41$, $p < 0.001$). Temporal demand was lower with SymbolFinder (5.13) than Google (6.17), but not significantly so ($t = 1.7$, $p = 0.1$). When users ran out of ideas with Google, they felt a greater time pressure to find 20 symbols. Some users felt more time pressure with SymbolFinder as they wanted to explore all the related words and images. Finally, participants felt they had performed better with SymbolFinder (6.77) than with Google (5.9), but this difference was not significant ($t = 1.95$, $p = 0.06$).

9 CASE STUDY

To evaluate how SymbolFinder helps people in practice, we found a group of three students who make cover illustrations for a school science publication. We asked them to use SymbolFinder as a group in their process over a three month period to help them make symbol blends that could serve as cover illustrations. We remotely observed them in three 90-minute sessions where they used SymbolFinder together as a part of their process for making blends for three different articles, one from their school science publication and two chosen from The New York Times for content diversity:

(1) "Public Health Messaging in Minority Communities and COVID-19's Neurological Effects"
Concept pair: *Diversity* (concreteness = 0.45) + *Neurology* (0.5)

(2) "The N.Y.P.D. Has Rejected Reform for Decades. It Can't Anymore." (NYT)
Concept pair: *Police* (0.96) + *Reform* (0.4)
(3) "When the World Shut Down, They Saw It Open - The pandemic has made work and social life more accessible for many. People with disabilities are wondering whether virtual accommodations will last." (NYT)
Concept pair: *Disability* (0.69) + *Participation* (0.52)

From our observations we wanted to answer the following four questions:

(1) **End-to-end process**. Where does SymbolFinder fit in the process of creating cover illustrations that blend symbols?
(2) **Picking concepts for articles**. What kind of concepts do they enter into SymbolFinder and how do they choose them?
(3) **Picking symbols**. What is their process of finding symbols and how does SymbolFinder help?
(4) **Combining symbols**. How do they use the symbols to make the blends?

9.1 Results

9.1.1 End-to-end process. We observed their end-to-end process of making symbols blends for cover illustrations, which generally consisted of five steps: (1) finding an article that needs a cover illustration, (2) picking two words that represents the article's key concepts, (3) use SymbolFinder to find images to represent the two concepts, (4) match these symbols to ideate blends, (5) create prototypes of these blends with either PowerPoint for lo-fi blends or Photoshop for hi-fi blends.

To answer the rest of the questions, we will discuss one of the case studies from beginning to end, highlighting how the team used SymbolFinder and the constraints that emerged when making blends. We study how they made blends for the NYT article: "The N.Y.P.D. Has Rejected Reform for Decades. It Can't Anymore."

9.1.2 Picking concepts for articles. **The team often picked one or two very abstract concepts to symbolize in the cover illustration.** A key part of their process was to scan both the article title and text to extract potential concepts to symbolize and combine. Their goal was to identify concepts that best capture the meaning of the article, and often these words are abstract. For example, while working on the second article: "The N.Y.P.D. Has Rejected Reform for Decades. It Can't Anymore", the team quickly picked *police*, a relatively concrete concept, as one of the keywords, as it was the clear subject of the article. For the second concept, they had a few candidates, including *reform* (concreteness = 0.4), *law* (0.51), and *scrutiny* (0.45), which are all very abstract words. In this case, they chose both *reform* and *law*. Similarly, for the other articles, they picked abstract concepts, like *diversity* (0.45), *participation* (0.52), and *neurology* (0.5). Given the content the articles, the team often found themselves symbolizing abstract concepts.

9.1.3 Picking symbols. We observed that when they collected symbols, **their focus was to find different representations of the concept**, as opposed to finding the perfect image for one particular representation. SymbolFinder helped them find different representations in two ways: (1) by exposing them to multiple different ideas through the clusters and (2) by presenting them with concrete objects within clusters to find symbols from these different ideas.

In the first phase of SymbolFinder for *police*, the team found multiple distinct contexts associated with *police*, many of which led to symbols in phase 2. They selected 16 of 20 clusters shown to them in the first phase. Among these was the first (and most obvious) cluster "police, cop, officer" as well as others that captured different aspects of police such as "questioning, interrogation, skeptical" and "brutality, riots, violence". Ultimately the first obvious cluster provided the most symbols, about

37%, but the team found many useful symbols from the less obvious clusters, where 64% of their symbols were spread across 11 other clusters. The second most fruitful cluster was "brutality, riots, violence", containing 10% of their total number of symbols. On average, users used 6.83 clusters presented in SymbolFinder, demonstrating that **the clusters were useful for finding multiple visual representations of the concepts**.

In the second phase of SymbolFinder for *police*, while diving into clusters, **the team took advantage of the concrete sub-words to collect many different objects associated with that cluster**. While searching for symbols of police, the team first explored the most obvious cluster "police, cop, officer" in phase 2. The image searches generated for each of these terms consisted of images of police officers in different poses and settings, but not specific objects related to police officers. However, when drilling into this cluster and viewing related subwords of its terms, they were presented with many concrete, associated objects like *siren*, *handgun*, *badge*, and *baton*. They then selected each of these objects and quickly collected 3 or 4 images of each one, including realistic images, icons, and objects of different shapes. Remarking on this process, P3 explained that when collecting images of an object, it is useful to gather images with different styles because "it sparks more ideas later on when we're matching symbols. We're not really sure what kind of images we'll need, so having symbols with different colors and shapes helps in imagining blends." Overall, SymbolFinder helped the team collect objects with different stylistic properties thanks to concreteness and SymbolFinder's many image search results; these two features helped them collect a greater diversity of symbols.

In the first phase of SymbolFinder for *reform*, the team found that **while the clusters introduced new ideas, some were not a great match with the content of the article**. They selected only 7 of 18 clusters shown to them. The clusters they did not pick were related to reform, but brought in connotations that did not fit with the overarching message: "police reform". For example, they rejected "political party, progressive, republican", which while related to reforming politics, is not so relevant to police reform. Instead, the team opted for clusters that contained different perspectives on what reform can be, like "reform, modify, change", "fix, amend, redo", and "new, update, innovative". Therefore, **the content of the article constrains the symbol search space, and phase 1 of SymbolFinder helped the team navigate this constraint by quickly eliminating irrelevant clusters**.

In the second phase for *reform*, a second constraint on the symbol space became apparent: the connotation and tone of the symbol. From the first cluster "reform, modify, change" they found a wrench image through modify's concrete sub-words and an image of the evolutionary progression of man from transform, another word within the cluster. They continued on to the "fix, amend, redo" cluster, where they found more tool-related concepts and collected images of a *toolbox*, *saw*, and *screwdriver* - all tools that can be used to reform something physically. From the "new, update, innovative" cluster they found symbols like the "update bell icon" used in social media interfaces, as well as the "cycle refresh button". Although both of these symbols did not have the tone or gravity they wanted for a blend conveying "police reform", they collected them anyway, thinking they could be useful for future blends with *reform*. However in a later cluster "structure, building, framework" they found the sub-word *scaffolding* and collected multiple images of buildings being constructed or renovated with scaffolding. They were excited by this symbol as its tougher tone and "New Yorkness" fit with the article tone well. **Thus, while the team is predominantly seeking a variety of representations for each concept, they do keep in mind the tone of the article when looking for appropriate symbols**.

9.1.4 Combining symbols. To blend two symbols, the team employs a **matching strategy**. After finding symbols for both *police* and *reform*, the team placed the images side-by-side to ideate

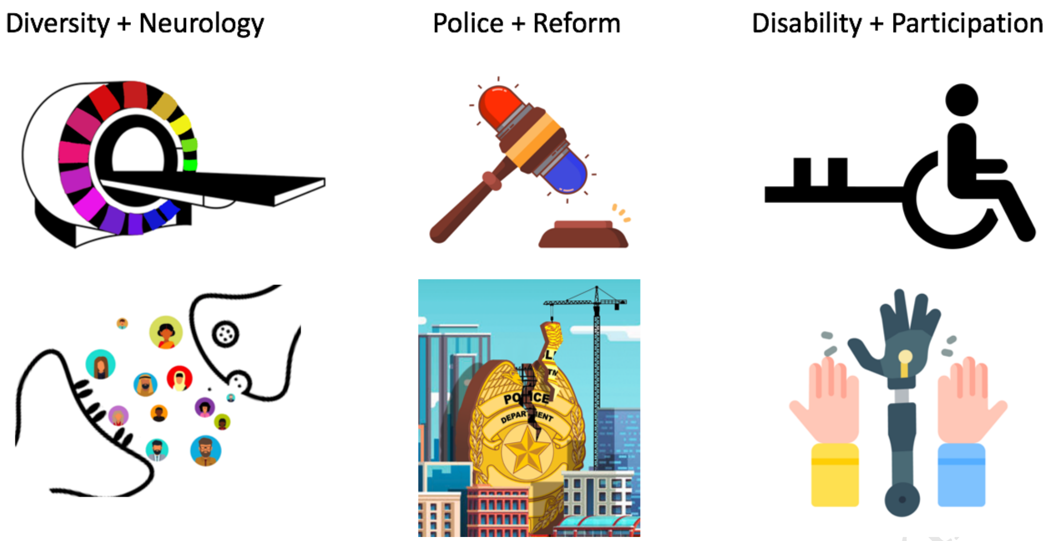

Fig. 9. Cover illustrations consisting of two blended symbols for the three articles. The team of student designers found each symbol idea from SymbolFinder and used PowerPoint or Photoshop to create these prototypes.

blends between them. Commenting on their overall process, P2 explained "we start by choosing a symbol we like from one concept. Then we match that symbol with one from the other concept, usually based on shape or function." They ultimately made 18 initial prototypes, 2 of which are shown in Figure 9. 10 of these prototypes combined symbols of *police* with symbols of reform, with the other 8 combining *police* with symbols of *law*. Across these 18 prototypes, they used 8 unique symbols of *police*, which came from 4 different clusters. They combined these *police* symbols with 8 unique symbols of *reform*, which came from 5 different clusters. They also used 6 unique symbols of *law*, which came from 3 clusters. **By having multiple diverse symbols, the team is (1) able to successfully find a match between two concepts with a higher probability and (2) create a diverse set of prototypes to show their client, using 6-8 unique symbols from each concept.**

Diverse symbols helped the team match symbols by giving them the assets to deal with emerging constraints. For example, the first blend for police + reform, consisted of scaffolding, symbolizing *reform*, applied to a police badge. The team was excited after finding scaffolding as a symbol for reform, and proceeded to try and match it with a *police* symbol. They needed something that looked building-like, and the police badge was one of the few symbols that was relatively rectangular and tall (Figure 9). By having many diverse symbols for *police*, they were able to meet this shape-constraint.

A similar situation occurred when the team blended a symbol of *police* with a symbol of *law*. They wanted to find a police symbol to combine with a gavel, a classic symbol of law. They first tried applying stick-shaped symbols like the baton to replace the handle of the gavel. However this blend was not so appealing as the police symbol seemed less prominent in the image. After scanning the police images once more, they found a symbol containing a red and blue siren, which they thought could replace the ends of the gavel. They liked this idea more as they thought the police symbol was more salient, thanks to the red and blue color of the siren, while maintaining the gavel's silhouette (Figure 9). By allowing the team to find diverse symbols for each concept, SymbolFinder gave them the ability to connect them even when other design constraints emerged. **In both cases, when they chose a symbol they liked from one concept, they were able to ideate a blend that made sense shape-wise, tonally, and stylistically**.

10 LIMITATIONS AND FUTURE WORK

In the following section we discuss limitations and future work to improve the system. We also discuss applying SymbolFinder to other visual media.

10.1 Emerging vocabulary

Currently, SymbolFinder is limited to the associations present in the SWOW dataset. And while the dataset is quite large, this can be a problem when searching for new or esoteric concepts. For example, in the past, the team worked on an article where one of the concepts was COVID-19, which did not exist in SWOW. In order to find symbols, the team brainstormed on their own, used Google Image Search, and also tried inputting related terms like *virus* into SymbolFinder. A solution to this problem could involve extracting related keywords for the new concept from web search results or from a frequently updated knowledge base like Wikipedia. These keywords could then be linked to current entries in the SWOW dataset. By doing so, the new concept would have a set of SWOW associations from which a local semantic network could be created and clustered.

10.2 Finding the perfect image

While SymbolFinder is effective for finding many diverse representations of an abstract concept, it is less useful for finding a specific image, once a particular representation is chosen. In the case study, after the team came up with an idea for a blend using symbols they found with SymbolFinder, they would sometimes perform a secondary search with Google Images to find a particular version of the symbol. For example, while making the police badge and scaffolding blend (Figure 9), the designer did not use the images of scaffolding they found with SymbolFinder. After imagining the symbol blend, she had a specific idea for how she wanted the scaffolding to look. She wanted a "consistent background color so it would be easy to remove it". As well as a removable background, she wanted a 2d image that was neither "super busy", containing "overlapping scaffolding", nor too simple and "unnatural" looking. She ended up scanning many images to find the image she used. As well as finding the perfect image that contains the right visual detail, the team also mentioned another factor in the secondary search was finding images that are free to use. Fundamentally, SymbolFinder is a brainstorming tool, but in the future, we can incorporate tools to help users find particular versions of an image that fits their purpose.

10.3 Applying SymbolFinder to other visual media and databases

As well as images, SymbolFinder can be slightly altered to search Google for GIFS and videos instead of images. SymbolFinder's core feature is the clustered local semantic network, which enables users to explore a number of diverse and related contexts associated with a concept. Similarly, though built with Google Images as its image database, SymbolFinder can be applied to any other image database, depending on the user's needs, such as the Noun Project [44], Flickr [19], or Shutterstock [47]. Although these database are already included Google Image Search, users may want to restrict their search to focus on these databases, given their goals.

11 CONCLUSION

This paper presents SymbolFinder, an interactive tool that enables users to find diverse symbols for abstract concepts. In our user study we show that users can find significantly more unique symbols for abstract concepts with significantly less effort with SymbolFinder than with Google Images. We also conduct a case study showing how SymbolFinder is useful for creating cover illustrations of news articles. In the future, SymbolFinder can be applied to other media types, like GIFs, and other image databases. Also SymbolFinder could include tools to help users find a perfect image after a representation is chosen and expand its word association dataset automatically with new concepts.

REFERENCES

[1] Salvatore Andolina, Khalil Klouche, Diogo Cabral, Tuukka Ruotsalo, and Giulio Jacucci. 2015. InspirationWall: Supporting Idea Generation Through Automatic Information Exploration. 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, 103–106. https://doi.org/10.1145/2757226.2757252

[2] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), P10008.

[3] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. 2016. Beyond Memorability: Visualization Recognition and Recall. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 519–528.

[4] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. 2013. What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2306–2315.

[5] Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods* 46, 3 (2014), 904–911.

[6] Deng Cai, Xiaofei He, Zhiwei Li, Wei-Ying Ma, and Ji-Rong Wen. 2004. Hierarchical Clustering of WWW Image Search Results Using Visual, Textual and Link Information. In *Proceedings of the 12th Annual ACM International Conference on Multimedia* (New York, NY, USA) *(MULTIMEDIA '04)*. Association for Computing Machinery, New York, NY, USA, 952–959. https://doi.org/10.1145/1027527.1027747

[7] Claudio Carpineto, Renato de Mori, Giovanni Romano, and Brigitte Bigi. 2001. An Information-Theoretic Approach to Automatic Query Expansion. *ACM Trans. Inf. Syst.* 19, 1 (Jan. 2001), 1–27. https://doi.org/10.1145/366836.366860

[8] 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

[9] Minsuk Chang, Leonore V. Guillain, Hyeungshik Jung, Vivian M. Hare, Juho Kim, and Maneesh Agrawala. 2018. RecipeScape: An Interactive Tool for Analyzing Cooking Instructions at Scale. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article 451, 12 pages. https://doi.org/10.1145/3173574.3174025

[10] 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

[11] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. 2004. Finding community structure in very large networks. *Physical review E* 70, 6 (2004), 066111.

[12] Douglass R. Cutting, David R. Karger, Jan O. Pedersen, and John W. Tukey. 1992. Scatter/Gather: A Cluster-Based Approach to Browsing Large Document Collections. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Copenhagen, Denmark) *(SIGIR '92)*. Association for Computing Machinery, New York, NY, USA, 318–329. https://doi.org/10.1145/133160.133214

[13] 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.

[14] Simon De Deyne, Daniel J Navarro, and Gert Storms. 2013. Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior research methods* 45, 2 (2013), 480–498.

[15] Simon De Deyne and Gert Storms. 2008. Word associations: Network and semantic properties. *Behavior research methods* 40, 1 (2008), 213–231.

[16] Ron S Easterby. 1970. The perception of symbols for machine displays. *Ergonomics* 13, 1 (1970), 149–158.

[17] Jianping Fan, Yuli Gao, Hangzai Luo, Daniel A. Keim, and Zongmin Li. 2008. A Novel Approach to Enable Semantic and Visual Image Summarization for Exploratory Image Search. In *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval* (Vancouver, British Columbia, Canada) *(MIR '08)*. Association for Computing Machinery, New York, NY, USA, 358–365. https://doi.org/10.1145/1460096.1460155

[18] Cristian Felix, Aritra Dasgupta, and Enrico Bertini. 2018. The Exploratory Labeling Assistant: Mixed-Initiative Label Curation with Large Document Collections. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. Association for Computing Machinery, New York, NY, USA, 153–164. https://doi.org/10.1145/3242587.3242596
[19] Flickr. 2020. *Flickr.* https://www.flickr.com/
[20] Bruno M. Fonseca, Paulo Golgher, Bruno Pôssas, Berthier Ribeiro-Neto, and Nivio Ziviani. 2005. Concept-Based Interactive Query Expansion. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management* (Bremen, Germany) *(CIKM '05)*. Association for Computing Machinery, New York, NY, USA, 696–703. https://doi.org/10.1145/1099554.1099726
[21] Charles Forceville. 1994. Pictorial metaphor in advertisements. *Metaphor and symbol* 9, 1 (1994), 1–29.
[22] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
[23] Kathleen Hemenway. 1982. Psychological Issues in the Use of Icons in Command Menus. In *Proceedings of the 1982 Conference on Human Factors in Computing Systems* (Gaithersburg, Maryland, USA) *(CHI '82)*. Association for Computing Machinery, New York, NY, USA, 20–23. https://doi.org/10.1145/800049.801748
[24] Nigel Holmes and Rose De Neve. 1985. *Designing pictorial symbols.* Rac Books.
[25] Enamul Hoque, Orland Hoeber, and Minglun Gong. 2013. CIDER: Concept-based image diversification, exploration, and retrieval. *Information Processing Management* 49, 5 (2013), 1122 – 1138. https://doi.org/10.1016/j.ipm.2012.12.001
[26] Enamul Hoque, Grant Strong, Orland Hoeber, and Minglun Gong. 2011. Conceptual query expansion and visual search results exploration for Web image retrieval. In *Advances in Intelligent Web Mastering–3*. Springer, 73–82.
[27] William K Horton. 1994. *The icon book: Visual symbols for computer systems and documentation.* John Wiley & Sons, Inc.
[28] Chien-Kang Huang, Lee-Feng Chien, and Yen-Jen Oyang. 2003. Relevant term suggestion in interactive web search based on contextual information in query session logs. *Journal of the American Society for Information Science and Technology* 54, 7 (2003), 638–649.
[29] WH Huggins and Doris R Entwisle. 1974. *Iconic Communication: An Annoted Bibliography.* Johns Hopkins University Press.
[30] Zaeem Hussain, Mingda Zhang, Xiaozhong Zhang, Keren Ye, Christopher Thomas, Zuha Agha, Nathan Ong, and Adriana Kovashka. 2017. Automatic understanding of image and video advertisements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1705–1715.
[31] Feng Jing, Changhu Wang, Yuhuan Yao, Kefeng Deng, Lei Zhang, and Wei-Ying Ma. 2006. IGroup: Web Image Search Results Clustering. In *Proceedings of the 14th ACM International Conference on Multimedia* (Santa Barbara, CA, USA) *(MM '06)*. Association for Computing Machinery, New York, NY, USA, 377–384. https://doi.org/10.1145/1180639.1180720
[32] Dhiraj Joshi, Ritendra Datta, Ziming Zhuang, WP Weiss, Marc Friedenberg, Jia Li, and James Ze Wang. 2006. PARAgrab: a comprehensive architecture for web image management and multimodal querying. In *VLDB*, Vol. 6. 1163–1166.
[33] Yui Kita and Jun Rekimoto. 2018. V8 Storming: How Far Should Two Ideas Be?. In *Proceedings of the 9th Augmented Human International Conference* (Seoul, Republic of Korea) *(AH '18)*. Association for Computing Machinery, New York, NY, USA, Article 14, 8 pages. https://doi.org/10.1145/3174910.3174937
[34] 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, Article 633, 12 pages. https://doi.org/10.1145/3290605.3300863
[35] Paul A Kolers. 1969. Some formal characteristics of pictograms. *American scientist* 57, 3 (1969), 348–363.
[36] Aaron Marcus. 1984. Corporate identity for iconic interface design: The graphic design perspective. *Interfaces in Computing* 2, 4 (1984), 365–378.
[37] Siné JP Mcdougall, Martin B Curry, and Oscar de Bruijn. 1999. Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers* 31, 3 (1999), 487–519.
[38] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
[39] George A. Miller. 1995. WordNet: A Lexical Database for English. *Commun. ACM* 38, 11 (Nov. 1995), 39–41. https://doi.org/10.1145/219717.219748
[40] Murli Nagasundaram and Alan R Dennis. 1993. When a group is not a group: The cognitive foundation of group idea generation. *Small Group Research* 24, 4 (1993), 463–489.
[41] American Institute of Graphic Arts. 1974. *Symbol Signs.* National Technical Information Service.
[42] Masataka Ono, Makoto Miwa, and Yutaka Sasaki. 2015. Word embedding-based antonym detection using thesauri and distributional information. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for*

Computational Linguistics: Human Language Technologies. 984–989.

[43] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[44] Noun Project. 2020. *Noun Project*. https://thenounproject.com/

[45] 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

[46] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. 2016. *Designing the user interface: strategies for effective human-computer interaction.* Pearson.

[47] Shutterstock. 2020. *Shutterstock*. https://www.shutterstock.com/

[48] Reinier H. van Leuken, Lluis Garcia, Ximena Olivares, and Roelof van Zwol. 2009. Visual Diversification of Image Search Results. In *Proceedings of the 18th International Conference on World Wide Web* (Madrid, Spain) *(WWW '09)*. Association for Computing Machinery, New York, NY, USA, 341–350. https://doi.org/10.1145/1526709.1526756

[49] Hao-Chuan Wang, Dan Cosley, and Susan R. Fussell. 2010. Idea Expander: Supporting Group Brainstorming with Conversationally Triggered Visual Thinking Stimuli. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, Georgia, USA) *(CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 103–106. https://doi.org/10.1145/1718918.1718938

[50] Hao-Chuan Wang, Susan R. Fussell, and Dan Cosley. 2011. From Diversity to Creativity: Stimulating Group Brainstorming with Cultural Differences and Conversationally-Retrieved Pictures. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (Hangzhou, China) *(CSCW '11)*. Association for Computing Machinery, New York, NY, USA, 265–274. https://doi.org/10.1145/1958824.1958864

[51] Jinxi Xu and W. Bruce Croft. 2017. Quary Expansion Using Local and Global Document Analysis. *SIGIR Forum* 51, 2 (Aug. 2017), 168–175. https://doi.org/10.1145/3130348.3130364

[52] Zheng-Jun Zha, Linjun Yang, Tao Mei, Meng Wang, Zengfu Wang, Tat-Seng Chua, and Xian-Sheng Hua. 2010. Visual Query Suggestion: Towards Capturing User Intent in Internet Image Search. *ACM Trans. Multimedia Comput. Commun. Appl.* 6, 3, Article 13 (Aug. 2010), 19 pages. https://doi.org/10.1145/1823746.1823747

[53] 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