Sketch-Based Creativity Assistant

by

Ezgi Dede

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Computer Science and Engineering

KOÇ ÜNİVERSİTESİ

September 19, 2023

Sketch-Based Creativity Assistant

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Ezgi Dede

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

Prof. Dr. T. Metin Sezgin (Advisor)

Prof. Dr. Ergun Akleman

Assoc. Dr. Burcu Ayşen Ürgen

Date:

I will be forever grateful to Birkan Çelik for supporting me from beginning to end. He was an excellent assistant and an even better friend. He carried this project with me and contributed to every step of it. Therefore, I dedicate this thesis to Birkan Çelik. In a way, it's his thesis as well.

ABSTRACT

Sketch-Based Creativity Assistant
Ezgi Dede
Master of Science in Computer Science and Engineering
September 19, 2023

Creativity is the ability to come up with new ideas. Yet, even creative people may find themselves stuck with overused ideas, a phenomenon known as 'design fixation.' To overcome the fixation and discover fresh ideas, experts in visual design suggest sketching out the initial idea. Sketches, being abstract representations, introduce an element of ambiguity. Exploring and reinterpreting these ambiguous shapes helps break away from cliches and foster creative thinking. In our project, we introduce a computerized intelligent system that leverages sketching to assist non-designers in visual design tasks. This system assists users in reinterpreting their sketches creatively. To accomplish this, we've developed a unique semantic network called the 'Sketch Net,' which connects ambiguous sketches to multiple crowdsourced interpretations. Our contributions include 1) Sketch Net, 2) a method for generating ambiguous sketches, and 3) an intelligent system that provides diverse interpretations of the user's sketch to enhance the creative process by associating it with Sketch Net elements. The effectiveness of this assistant in boosting creativity is evaluated through a user study. The results show that the Sketch-Based Creativity Assistant significantly improves the quality of the final product.

ÖZETÇE

Çizim Temelli Yaratıcılık Asistanı
Ezgi Dede
Bilgisayar Bilimleri ve Mühendisliği, Yüksek Lisans
19 Eylül 2023

Yaratıcılık, yeni fikirler üretebilme yeteneğidir. Ancak yaratıcı kişiler dahi sıradan fikirlere takılıp kalabilir. Bu takılıp kalma halinden kurtulmak ve yeni fikirler bulmak için görsel tasarım uzmanları akla gelen ilk fikri çizmeyi önerirler. Zira çizimler soyutlama gerektirir ve soyutlanmış temsiller çok anlamlı yorumlanmaya açıktır. Çok anlamlı şekilleri yorumlamak ve yeniden yorumlamak ise klişelerden kurtulmaya yardımcı olur. Üstelik çok anlamlı görsellerin başka anlamlarının keşfedilmesi yaratıcı düşünmeyi tetikler. Bu çalışma tasarım deneyimi olmayan kullanıcılara görsel tasarım konusunda yardımcı olmak ve onları daha yaratıcı çözümler bulma konusunda desteklemek amacıyla hazırlanmıştır. Çizim Temelli Yaratıcılık Asistanı, kullanıcıların çizimlerini farklı şekillerde yorumlayabilen ve takılıp kalma durumunu yok eden akıllı bir bilgisayar sistemidir. Bu sistem, çok anlamlı çizimleri ve onlara atfedilmiş anlamları birbirine bağlayan, kitle kaynak çalışmaları ile oluşturulmuş Sketch Net adlı bir anlamsal ağ üzerine kurulmuştur. Çalışmamızın katkıları şu şekildedir: 1) Çok anlamlı çizimleri uygun anlamlarla ilişkilendiren Sketch Net, 2) çok anlamlı çizim oluşturma yöntemi ve 3) kullanıcının çizimini Sketch Net'teki çok anlamlı görsellerle ilişkilendirerek yeniden yorumlayabilen, bu sayede de yaratıcı süreçleri destekleyen akıllı bir bilgisayar sistemi. Çizim Temelli Yaratıcılık Asistanı'nın yaratıcılığı etkin şekilde arttırıp arttıramadığı bir kullanıcı çalışmasıyla test edilmiştir. Sonuçlar asistanın son üründeki fikirsel değerini kayda değer miktarda geliştirdiğini göstermiştir.

ACKNOWLEDGMENTS

First, I wholeheartedly thank Prof. T. Metin Sezgin for entrusting me with his brilliant project. He believed in me, even though I didn't have an engineering background. Over time, he devoted countless hours to instructing me in the ways of an engineer and a researcher. Birkan Çelik embarked on this project as a first-year undergraduate student and contributed to it in every conceivable manner. Kamile Aslı Ağılönü, upon graduating, chose to remain at Koç University solely to lend support to this project. We formed the Sketch Net together, and she taught me how to be a better computer engineer. Birkan and Aslı extended an incredible amount of assistance for which I am profoundly thankful. Alara Zindancıoğlu collaborated with me on a review article stemming from our sketch-related studies and provided guidance on academic writing. Aleyna Kütük partnered with me on a related, more extensive project. Caner Şetlek, Demir Ekin Arıkan, İrem Karaca, Kaan Türkmen, Furkan Bulgur, Serkan Berk Bilgiç, and Gülay Canbaloğlu got excited by this research idea and volunteered to be a part of it. Melis Çağla Kıroğlu, İrem Güngör, Aslıhan Cemre Çetin, and Dilay Tetik volunteered to rate the Sketch Net interpretations.

Alpay Sabuncuğlu, Gökçe İymen, Gürkan Soykan, Egecan Esen, and all members of the Intelligent User Interfaces Lab touched this project in one way or another. Some by giving valuable feedback, some by participating in our early experiments to provide us with insight, and some by giving mental support when I needed it. I want to acknowledge their effort and sincerely thank each one of them.

This research project is supported by grants from The Scientific and Technological Research Council of Turkey (TÜBİTAK; Project No. 120E489).

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

AI	Artificial Intelligence
ANOVA	Analysis of Variance
CGA	Control Group Assistant
CNN	Convolutional Neural Network
CSI	Creativity Support Index
CSP	Creative Sketching Partner
EEG	Electroencephalogram
HLDA	Hierarchical Latent Dirichlet Allocation
LDA	Latent Dirichlet Allocation
LSTM	Long Short-Term Memory
QD	Google's Quick Draw
RNN	Recurrent Neural Network
SBCA	Sketch-Based Creativity Assistant
SMHYS	Show Me How You See
Tok-Dict	Dictionary Learned Tokenization
YAKE!	Yet Another Keyword Extractor

Chapter 1

INTRODUCTION

"Creativity is the art of misunderstanding in a meaningful way"
[Takala, 1993]

The standard definition of creativity is the ability to find original and effective ideas [Runco and Jaeger, 2012, Eckert and Stacey, 2000, Rhodes, 1961, Gero, 1990, Diedrich et al., 2015]. However, the scope, the assessment, and even the definition of creativity changes with the field of the study [Plucker and Makel, 2010, Benedek et al., 2020, Niu and Sternberg, 2001, Sternberg, 2006].

In the field of visual design, the quantity of interpretations generated for a single ambiguous visual is an acceptable indicator of creative proficiency. [Stones and Cassidy, 2010, McKim, 1980]. This is due to the fact that each interpretation initiates a process of reinterpretation. Over time, this cyclic process yields novel and imaginative concepts. Therefore, ambiguous shapes trigger creativity. Goldschmidt has used the term *"seeing as and seeing that cycle"* to describe this iterative relationship between perception and ambiguous sketches. [Goldschmidt, 1991, Buxton, 2010]. According to Goldschmidt, seeing an ambiguous shape as something else and seeing another meaning in that newly perceived shape is a way to trigger creative thoughts in design. Designers are trained to leverage this continuous cycle of interpretation and reinterpretation in visual design as a means of discovering innovative ideas [Ferguson, 1994, Yang, 2009].

A commonly employed technique in visual design is sketching the idea in mind frequently to reinterpret it and boost creative thinking that way [McKim, 1980, Yang, 2009, Farthing, 2011, Finke and Slayton, 1988, Van Der Lugt, 2002]. Sketches, serving as abstract depictions of these thoughts, inherently possess a degree of ambiguity [Goel, 1995, Ferguson, 1994]. This inherent ambiguity fosters a fertile ground for new ideas [Stones and Cassidy, 2010, Suwa and Tversky, 1997, Kavakli and Gero, 2001, Schon and Wiggins, 1992]. Nevertheless, the human mind occasionally succumbs to design fixation, an unavoidable preoccupation with conventional or before-thought ideas. Overcoming design fixation is a challenging task, particularly when attempting to swiftly reevaluate an ambiguous shape after perceiving it in a certain form [Crilly, 2015]. One of the fastest ways to overcome fixation is to collaborate with another person [Sawyer and DeZutter, 2009, Nijstad and Stroebe, 2006, Paulus and Brown, 2007, Tversky, 2015]. This collaborative approach proves beneficial as exposure to fresh interpretations aids in overcoming fixation, thereby empowering collaborators to unearth alternative ideas. However, the feasibility of finding a suitable collaborative partner is not always guaranteed. Hence, the integration of a digital assistant capable of reinterpreting ambiguous sketches would enhance the overall visual design experience.

Numerous initiatives have been undertaken to create design tools utilizing computer-based technologies [Lee et al., 2011, Ge et al., 2020, Adobe Inc., 2019]. Nevertheless, a majority of these tools are primarily designed for two distinct purposes: Refining preconceived ideas or swiftly generating products through predefined templates. Only a limited number of research initiatives have sought to facilitate users in the initial stages of the design process [Davis et al., 2016, Karimi et al., 2019]. Furthermore, an even smaller fraction of these initiatives concentrate on harnessing the potential of sketching to trigger the interpretation and reinterpretation cycle [Karimi et al., 2020, Rezwana et al., 2021].

Professional digital sketching tools are predominantly employed by individuals with a background in design. Consequently, previous research endeavors have mostly catered to the needs of expert users. However, it is imperative to acknowledge that individuals lacking a specialized background in design also require support for their design-related tasks. In fact, nonspecialists may find themselves in greater need of assistance during the ideation phase, primarily due to their limited exposure to the concept of the reinterpretation cycle. It is challenging for nonspecialists to perceive the potential of ambiguity and generate possible interpretations of an ambiguous shape. Notably, individuals with formal training in visual arts often excel in this capacity due to their heightened perception of visual details [Kozbelt, 2001, Kozbelt and Seeley, 2007]. Therefore, the capacity of the Sketch-Based Creativity Assistant to provide multiple interpretations of a single sketch holds significant potential in facilitating the creative process for non-specialists.

Our Sketch-Based Creativity Assistant aims to facilitate the interpretation and reinterpretation cycle to generate new ideas on early design sketches (i.e., ideation sketches). Since collaboration is helpful in reducing fixation, we propose simulating a collaborative relationship with a digital assistant. In this relationship, the assistant acts as a group of collaborators, rather than a single collaborator. In other words, designing with the help of Sketch-Based Creativity Assistant gives users a chance to ask what their sketches look like to more than one hundred people in an instant.

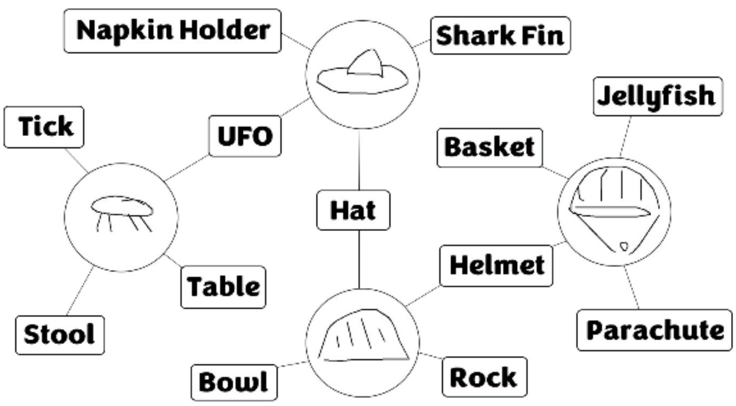


Figure 1.1: A brief sample of Sketch Net

The development of the Sketch-Based Creativity Assistant required the establishment of a semantic network that connects ambiguous shapes with plausible interpretations, referred to as the "Sketch Net." The methodology chapter elucidates the technique of generating ambiguous sketches of Sketch Net and crowdsourcing steps for interpretations of Sketch Net sketches. The rest of the chapter expounds upon how the Sketch-Based Creativity Assistant leverages this semantic network to provide inspiring visual and textual stimuli to its users. The Experiment chapter details the user experiment that is conducted to measure how effective Sketch-Based

Creativity Assistant is in providing inspiring stimuli and collaborative assistance. Results indicate that Sketch-Based Creativity Assistant significantly improves the quality of the design process and the creative value of the end product.

Chapter 2

RELATED WORK

It is possible to come across many software developed to assist people in design in the literature. However, these have either aimed at a different goal than ours or had to impose limitations on their methods due to a lack of resources. Our study has found a way to overcome the shortage of resources regarding the dataset and thus managed to avoid being constrained in previous methods. Additionally, the field that it aims to contribute to is an area in which computer science studies are comparatively scarce. Our project is targeted to solve the problems related to the ideation phase of visual design, in which the initial design ideas are generated. Those issues are solved by utilizing the interpretation and reinterpretation cycle. Therefore, our project has been able to make a unique contribution in a distinct area that is separate from the domain of well-known and successful design software, such as Adobe Illustrator, Autodesk, and AutoCAD.

To our knowledge, Creative Sketching Partner (CSP) appears to be the most similar to our project [Karimi et al., 2020]. First of all, CSP also focuses on the ideation step of design. However, CSP works with a single pre-defined target category in its design task (e.g., bridge) and assists its users with a singular sketch, which is chosen from Google's Quick Draw dataset [Jongejan et al., 2016]. The selection is made based on two specific parameters. Those two parameters range from 1 to 10, and their values are set by the users manually. The first parameter is visual similarity. Its value is utilized on a CNN-LSTM model to find the most similar or distant sketch to the user's sketch. The conceptual similarity parameter, on the other hand, is calculated as the distance between word2vec vectors [Mikolov et al., 2013]. Those vectors are sketch categories of Google's Quick Draw and the target category of the design task. In the user study, users were instructed to design the target object and rely on the suggested sketch for inspiration. The session ended when the user got inspiration from the visual thrice. It should be noted that the user study was carried out exclusively with design students.

Another similar project is the Creative PenPal [Rezwana et al., 2021]. In 2021, researchers proposed it as a conversational creativity assistant. However, the results section of their study has yet to be completed and published. Their take on the conversational model was to focus on emotional responses. To achieve that goal, they planned to include an avatar of the assistant on their interface. The avatar was a pen with expressive facial features, and it would inquire whether its presented sketch stimulus was useful whenever it suggested a new sketch. Based on the answer, the avatar would look sad or happy. The mode of the suggested sketch would be selected from three options: a sketch that closely resembles the user's sketch, a sketch that is conceptually related to the target category, or an unrelated sketch that belongs to the target category itself. These options would be presented on the interface as three buttons. The users would be required to choose one of them to get assistance. The proposal doesn't give information on the dataset of suggested sketches. However, they do not explicitly mention intentionally designing the assistant with ambiguity in mind. Therefore, there is no indicator that PenPal is a similar study to ours.

The sketches are inherently ambiguous; therefore, they have the potential to trigger the interpretation & reinterpretation cycle. However, in previous studies, the researchers could not handle this cycle to better their models [Gross and Do, 1996]. That's why they removed hard-to-recognize and hard-to-process sketches from the datasets. In other words, issues related to ambiguity have been mostly ignored in computerized sketch systems [Eitz et al., 2012]. However, the models developed to assist design should not ignore ambiguity. In fact, they must embrace ambiguity. There are clever methods to cope with the limitations of the current computer technologies and the lack of an unprocessed dataset. These methods achieve processing ambiguity in a computerized system but they are extremely rare and their focus is not on creative design [Zhang et al., 2018, Sarvadevabhatla et al., 2018]. Our solution to the lack of a proper dataset is to expose humans to ambiguous sketches and obtain many possible interpretations for each. In this project, we use the dataset to improve our model's ability to handle ambiguity. Yet, ours is not the only study that records interpretations with ambiguous sketches to form a dataset. There is a similar dataset collected in a Pictionary game [Sarvadevabhatla et al., 2018]. In the Pictionary Game study, the researchers re-constructed Quick Draw sketches stroke by stroke, and whenever they added a new stroke, they asked their participants what that object could be. Although similar to our approach, their method is not aimed at finding the most efficient ambiguous visuals and their interpretations. Therefore, their approach is different than ours in creating an ambiguous sketch dataset.

There are studies similar to ours in terms of focusing on the needs of non-experts, too. For example, Shadow Draw is developed to assist individuals without expertise in drawing by providing them with hints regarding the correct proportion of object parts due to their lack of precise perceptive abilities [Lee et al., 2011]. However, Shadow Draw assists in refining the existing ideas. It is not designed to give new ideas to its users. Shadow Draw does not interfere with users' sketches directly unlike similar software. Most of the refinement tools modify the users' sketches automatically and directly. Therefore, they are criticized for being intrusive to design processes [Fan et al., 2019, Davis et al., 2016, Motzenbecker and Phillips, 2017]. Our model differs from those, too because it does not interfere with the sketching process of the user. Sketch-Based Creativity Assistant offers its help next to the canvas of the user for the user to look at and get inspiration from.

Co-creative design assistants are recently getting more attention since conversational AI models such as Chat-GPT are publicly recognized and proved successful [Brown et al., 2020, OpenAI, 2023, Rezwana and Maher, 2022]. Previous works didn't include complex language models in their automatic co-creative design systems [Karimi et al., 2019, Davis et al., 2016, Rezwana et al., 2021]. Even though our work emphasizes the interaction between the model and the user, we didn't utilize GPT models, too. In the future, our architecture would greatly benefit from integration with Chat-GPT. Yet, in its present form, our interface creates the sensation of conversing without GPT assistance. Our design allowed users to engage in open-ended conversations with the model and express their ideas. By doing so, it managed to distinguish itself from previous studies.

As confirmed in this section, there is significant work that paves the way for us.

However, our Sketch-Based Creativity Assistant has the potential to make valuable improvements in each of the aforementioned aspects of the previous work. As a result, it makes three unique contributions to the literature: 1) Sketch Net, a unique ambiguous sketches dataset with multiple interpretations, 2) a method of generating ambiguous sketches and 3) a sketch-based creativity assistant that is specifically designed to trigger interpretation and reinterpretation cycles to assist ideation step of visual design.

Chapter 3

METHODOLOGY

Sketch-Based Creativity Assistant (SBCA) is designed to trigger the interpretation & reinterpretation cycle as frequently as possible. There are two channels through which this cycle can be triggered. SBCA can either present new sketches to be interpreted or generate a new interpretation for the sketch at hand. For the assistant to do both, it needs an experience in this matter, that is, a dataset it can be trained on. Even though there is data on sketch interpretation, as we mentioned before, there is no available dataset that is suitable and aligned with our objective.

Our objective was to discover the frequently used ambiguous shapes because we previously provided evidence that most sketches intersect in a few basic shapes in the early stages, even though they represent different objects when completed [Tirkaz et al., 2012]. Therefore, working on those intersection sketches would provide us with ambiguous and mainly used sketches. Also, this practice ensures that the set of ambiguous shapes we base our work on will be efficient enough to cover a wide range of sketches with minimal resources. When ambiguous shapes are generated, the next step is to collect as many interpretations as possible for each. This is the process we followed to create the Sketch Net dataset.

SBCA utilizes the Sketch Net dataset to trigger the interpretation & reinterpretation cycle. To determine the best stimuli for each user, SBCA asks various questions. Those questions are also designed to provoke new ideas. According to the users' answers, sketches of the users, and the theme of the task, the three best-fitting options are calculated and selected among those ambiguous shapes. Those shapes are presented to the user as visual stimuli. For each ambiguous shape, a list of interpretations that belong to it is also provided. Those lists of words are ordered based on the information that SBCA received from the user, and they act as textual stimuli.

3.1 Sketch Net: Designing Ambiguous Sketches

Multiple techniques exist in the literature for creating ambiguous visuals. In the Psychology field, researchers commonly utilize ambiguous visual stimuli that can be interpreted in numerous ways, such as Mooney images and Rorschach tests.[Mooney, 1957, Verhallen and Mollon, 2016, Exner Jr, 1993]. That's why it is also possible to come across some articles on generating ambiguous visuals with different methods such as extracting edges, outlines, and shadows [Wagemans et al., 2008]. However, our aim was to pinpoint the most commonly drawn sketches. Therefore, we had to invent our own methods to extract those ambiguous shapes.

SBCA heavily relies on the Sketch Net dataset because to assist users in their tasks and simulate a human-like collaborative conversation, we needed to understand what they drew and suggest sketches of similar ideas. Additionally, the suggested sketches should have a high potential of triggering the interpretation & reinterpretation cycle for users. That's why, they should be open to numerous interpretations, as it is explained before. Therefore, any visual that doesn't meet these criteria wouldn't help SBCA's performance.

In order to pinpoint the most commonly drawn ambiguous shapes during sketching, we employed Google's Quick Draw (QD) dataset. Since QD is the biggest sketch dataset with over 50 million sketches in 345 categories, it is an extensive source to go for [Jongejan et al., 2016, Wilkinson, 1998, Stiny, 2006, Gross and Do, 1996]. Our previous studies showed that people tend to draw intersecting ambiguous shapes in early sketching and then evolve those shapes into distinct objects. So, we had to extract incomplete sketches to generate ambiguous sketches [Tirkaz et al., 2012]. We took a sample of 150 sketches from QD per category. Thus, our sample set had 51,750 sketches which are correctly classified by Google's classifier. Then, we reconstructed all of those sketches, stroke by stroke without shuffling their order. In other words, we reconstructed seven partial sketches from a sketch with seven strokes, progressively (See Figure 3.1).

For the purpose of extracting the visual feature vectors of sketches, we applied Sketchformer, which is a state-of-the-art sketch processing model [Ribeiro et al.,

Figure 3.1: The partial sketches extracted from a complete alarm clock sketch of QD (sketch ID: 4646577031348224)

2020]. Sketchformer's encoding capabilities are proved to be satisfactory for our purposes, in Tok-Dict version [Kütük and Sezgin, prep]. Therefore, all of the partial sketches are given to the model separately to get their Tok-Dict feature vectors. After that, all feature vectors are clustered via the k-clustering algorithm (k= 1245) [MacQueen, 1965]. This way, all sketches, and their partial versions were grouped under 1245 clusters based on their feature vectors.

Figure 3.2: Sample sketches from a cluster that contains only two QD classes: Cellphone and Calculator

When the contents of these clusters are examined, it is observed that some of them contain sketches coming from only one or two QD categories. As an example, the vectors from the sketches of a phone and a calculator were merged together in one of the clusters (See Figure 3.2). Therefore, these clusters are assumed to represent a shape that can only be interpreted as one or two objects. Due to their limited ambiguity, those clusters are removed from the analysis. On the other hand, some clusters were found to be containing sketches from more than 50 categories of QD. Those clusters represent basic shapes without distinctive features such as a line, a rectangle, or a circle. Our aim is to capture more complex and reasonably ambiguous shapes. Therefore, these clusters were also ignored in the analysis.

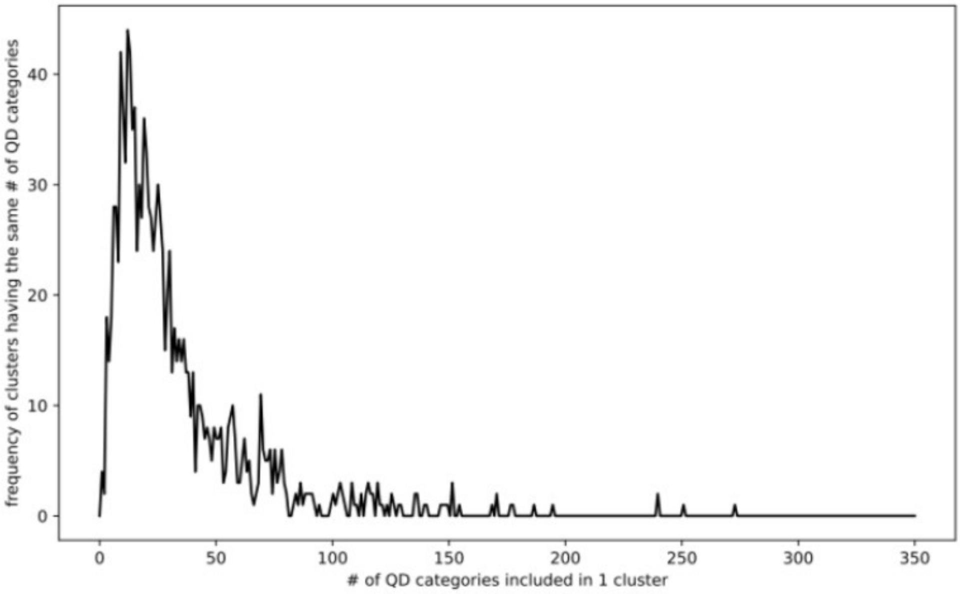


Figure 3.3: The graph of categories per cluster distribution

The remaining clusters are assumed to represent ambiguous shapes since they contain feature vectors of an adequate number of categories. The centroid vectors of each cluster are calculated to measure distances between clusters. Those centroid vectors are treated as the vectors of the ambiguous shapes those clusters represent. According to their Euclidian distance score, 220 clusters that are far from each other are selected to be candidate ambiguous shapes.

An expected follow-up step would be reconstructing those 220 centroid vectors to see the representation of the cluster's ambiguous shape. However, the marginal performance of the Sketchformer in the reconstruction phase and the concern that this method will produce artificial sketches prompted us to use a different method. To see what ambiguous shape the cluster represents, the closest vector to the centroid vector is selected as a representer. Therefore, every representer (i.e. ambiguous shape) is guaranteed to be drawn by a human and therefore to be a genuine sketch.

Selected 220 representers are reviewed by 11 volunteers. Volunteers were asked to interpret those sketches as many times as possible. Those sketches that were not able to elicit at least 4 distinct interpretations are removed from the set. From the rest, we selected 122 representers to be in our main ambiguous shape set. The restriction on the number of shapes was due to budget limitations. Two shapes are utilized in the attention check test. The remaining 120 are the shapes that form the main ambiguous sketches of the Sketch Net dataset.

3.2 Sketch Net: Deciding on Context Words

After generating ambiguous sketches for the dataset, we had to assign a large number of diverse labels to each. However, research on memory and fixation shows that when a mind interprets an ambiguous shape once, it will be hard for it to perceive another interpretation for a while [Malmberg et al., 2012, Jansson and Smith, 1991]. To overcome this fixation, the perception can be triggered with a new stimulus. We call this stimulus our context words.

Context words serve to distract a person's mind from the interpretation that is most likely to come to mind (i.e. cliche) [Murnane and Phelps, 1994, Kornmeier and Bach, 2012]. These context words could be chosen as subject categories such as fashion, sports, and food. However, since memory and attention are encoded or altered depending on the place where memories are formed, we anticipated that selecting context words based on locations would be more beneficial for our purposes [Bar and Aminoff, 2003, Torralba, 2003, Biederman, 2017, Tversky, 2009, Barenholtz, 2014, Rodrigues et al., 2019, Robin and Olsen, 2019]. In other words, objects will be easier to recall when a person is given a cue about where they can encounter those objects [Stokes et al., 2012, Summerfield et al., 2011, Luck and Hollingworth, 2008]. That's why, forming a context word list from location names would help people to perceive ambiguous sketches in different ways.

To form the list, we extracted words from the dictionary that include terms such as "location, place, room, district, area, region" in their definitions to create a comprehensive list of places. That step resulted in a list of 3533 words with examples like "Zimbabwe" and "National Park." We filtered out the proper nouns, and the rest of the words are ordered based on their frequency value. After that, we started from the top of the list and grouped words to form context words until our context word list reached 100 words. This grouping technique was based on the work of Tversky & Hemenway [Tversky and Hemenway, 1983]. In this way, we identified 100 location words that are most likely to contain different objects. Then, we added 20 proper nouns that are most likely to trigger memories of distinct cultures and objects such as "Paris" and "Mexico." In total, the context word list included 120 locations, which can be seen in detail in Appendix A.

3.3 Sketch Net: Collecting Interpretations

An interface has been designed to enable the recording of as many interpretations as possible for each ambiguous shape. For this purpose, each ambiguous sketch is rotated 0, 90, 180, or 270 degrees randomly before being shown to the participant. Also, a context word is randomly selected from the context list to be presented alongside the sketch. Participants are asked what the object might be in the given context. If they could think of more than one object, they had the chance to add another interpretation to their answer for the same sketch-context pair (See Figure 3.4). Participants weren't instructed to try being creative in their interpretations because fMRI studies show that being exposed to ordinary ideas during collaboration benefits people more than being exposed to creative or brilliant ideas[Fink et al., 2012]. That's why, we didn't want our assistant to provide creative interpretations to the user and hinder their creative process. Therefore, we didn't want our dataset to be biased toward creative interpretations.

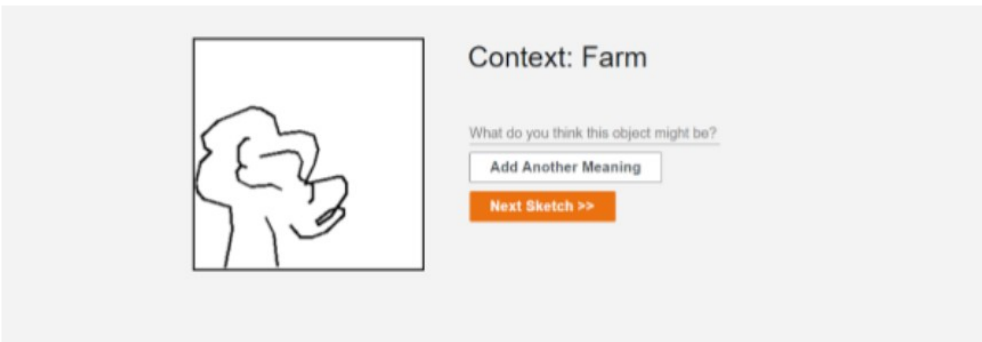


Figure 3.4: The interface of Sketchnet interpretation collection

It was important to limit the participants' answers to two words when collecting interpretations. Because perception studies show that the human mind is inclined to perceive an unfamiliar shape as a combination of familiar shapes [Biederman, 1987]. Hence, if there weren't any limitations on participants' answers, there was a possibility that they may describe their interpretation in great detail instead of associating the ambiguous sketch with an existing object. That would make the post-processing step open to subjectivity. That's why we allowed our participants to get help from Google Search and Google Images when they could think of an object but couldn't remember its actual name [Feng et al., 2009].

Initially, we tried crowd-sourcing via Amazon Mechanical Turk to collect interpretations because it's known that people coming from diverse cultural and personal backgrounds would perceive ambiguous shapes differently [Zhao et al., 2021]. Therefore, their contributions would be highly valuable. However, the preliminary research we conducted showed that the data we collected from Amazon Mechanical Turk was of very low quality. This is because more than 90% of the users who provided this data were Workers who participated in the study by changing their location with a VPN, but in reality, they did not even know English. Similar reports exist in the literature, warning against those Workers [Heer and Bostock, 2010, Kittur et al., 2008]. Even though we used only 2 sketches of the 122 sketches we put in our main set in an attention check test, the resources spent on failed attention checks were too big for our budget. Thus, we had to continue our research with participants from Koç University, Computer Engineering Department instead of Amazon Mechanical Turk Workers.

129 juniors and seniors participated in our study. We used the same 2 sketches as the attention test check again. All of the participants succeeded in that test and were qualified to participate in the actual study. In the main task, we asked them to interpret every single sketch we had in the Sketch Net dataset. All of the 120 sketches were presented in a random rotation, random order with a random context word. We followed the previously discussed instructions on creativity, web search, and word limitations. All participants recorded at least one interpretation per sketch. In the end, they have written 11.000 different interpretations in total

for the set of 120 sketches.

Of course, some of the interpretations were synonyms and some of them were irrelevant to the sketch. All low-quality data that did not meet the standard were removed from the dataset while all synonyms were grouped under one word to describe that interpretation (See Section 3.4). After these adjustments, the Sketch Net dataset has 3654 interpretations in total for 120 ambiguous sketches.

3.4 Sketch Net: Quality Adjustments

The process of selecting, organizing, and evaluating over 11,000 interpretations collected for the Sketch Net dataset was completed in several steps.

3.4.1 Sketch - Interpretation Ratings

Independent raters were recruited from the Psychology department of Koç University to review all 11,000 interpretations. The raters were seniors, therefore they were trained in being precise on data examinations, making evaluations in an unbiased manner, and being blind judges [Diedrich et al., 2015]. Since we especially want to work with raters from the Psychology department, 8 students volunteered to be a rater. However, only 5 of them were able to complete evaluating all interpretations.

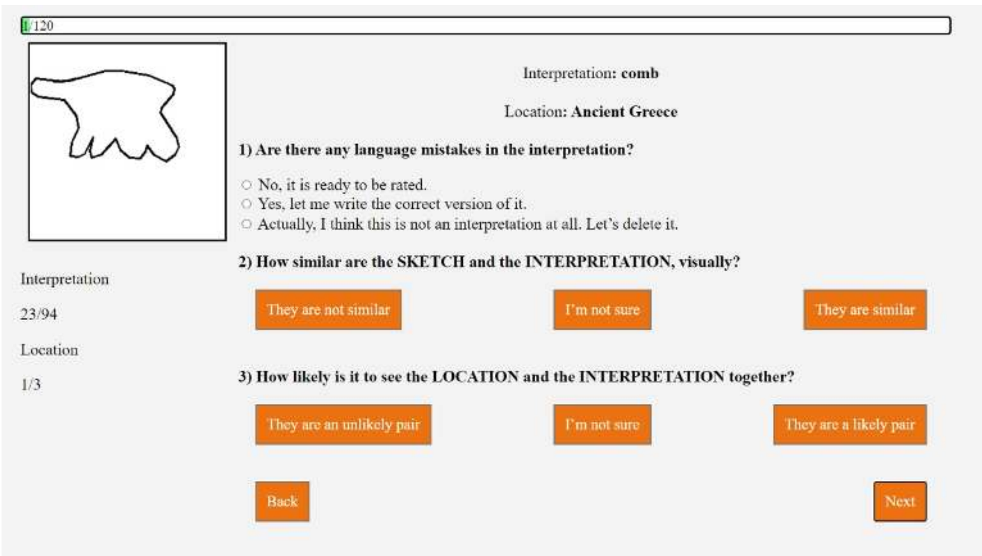


Figure 3.5: The interface of rating Sketch Net data

Raters evaluated the data by answering three questions for each sketch-interpretation-context group (See Figure 3.5):

1) Are there any language mistakes in the interpretation?: This question has three options for answer. "A) No, it's ready for rating.", "B) Yes, let me write the correct version of it.", and "C) I don't think it's an acceptable interpretation. We should delete it." If they choose option B, a box appears for raters to write the correct version of the word. If they choose option C, they skip the following questions for this sketch-interpretation-context group. We asked raters to click on option C only if the interpretation is highly irrelevant such as "who knows" or "don't know". Due to the open-ended nature of the task, some participants typed irrelevant two-word answers. Option C is designed to eliminate them.

2) How similar are the SKETCH and the INTERPRETATION, visually?: For this question, the rater should initially choose between three buttons: "They don't look similar", "I'm not sure", and "They look similar." When they press a button, the corresponding part of the 11-point Likert scale becomes visible. For example, when they click on "They look similar", they see a 5-point Likert scale from "6: Slightly similar" to "10: Exactly the same." Therefore they should again make a decision on a 5-point Likert scale on how much they look similar or dissimilar. On the other hand, clicking on the "I'm not sure" button marks point 5 on the 11-point Likert scale automatically, setting it as a neutral element for the scale.

3) How likely is it to see LOCATION and INTERPRETATION together?: Similar to the previous one, in this question, the rater should initially choose between three buttons: "Not likely", "I'm not sure", and "Likely" When they press a button, the corresponding part of the 11-point Likert scale becomes visible. For example, when they click on "Not likely", they see a 5-point Likert scale from "0: Impossible" to "4: Slightly unlikely." Therefore they should again make a decision on a 5-point Likert scale on how likely or unlikely they are to be seen together. Again, clicking on the "I'm not sure" button among the first three buttons automatically marks 5 on the 11-point Likert scale of likelihood measure.

3.4.2 Manual Inspection

After the evaluations of psychology students were completed, all the interpretations and evaluations were manually reviewed together by the researchers.

Interpretations that are marked by the C option of the first questions are reviewed and deleted by the researcher in this step. Typos were corrected by the researcher if at least 3 raters agreed on the correct version of the word.

If an interpretation has a lower inter-rater reliability than 80% on the initial decision question of the second question (i.e. Similar - Not Dissimilar decision), that interpretation is left out of the final dataset. In other words, if 4 raters out of 5 don't make the same decision between the "They look similar" and "They don't look similar" options in the second question, that interpretation is discarded. Rarely, an interpretation has less than 5 ratings for the inter-rater reliability check due to the "I'm not sure" button. In that case, the main researcher included her rating to compensate. The elements that the researcher intervened in their ratings are clearly marked in the dataset.

The interpretations that got an average score lower than six on the similarity measure (i.e. second question) are also omitted from the dataset. That step ensured that all interpretations were indeed resembling the shown ambiguous shape.

The remaining interpretations are reviewed to be grouped if they are synonyms, near-synonyms, or closely related words. For example, "rock" and "rocks" are grouped to be "rock(s)." Similarly, "toy ball" and "plastic ball" are grouped under the interpretation "ball." The record of the information about each element of the created groups exists in the dataset. However, if the group has been formed under an already existing interpretation, the ratings assigned to the other interpretations have not been used during the evaluation of the main interpretation. For instance, if there is the interpretation "ball" with an average similarity rating of 7, and we combine "toy ball" with it, which has an average similarity score of 8, then the newly formed "ball" interpretation group has a similarity score of 7. Because that interpretation is already rated by the raters. That's why we don't change it. However, if there were only "rock" and "rocks" grouped under the newly formed "rock(s)" interpretation, then we take the average of all ratings assigned to both interpretations. Because in that case there aren't readily assigned ratings for this particular interpretation, "rock(s)".

3.4.3 Show Me How You See (SMHYS)

During the evaluations, an issue related to Sketch Net surfaced: It was not easy to evaluate and rate the accuracy of some interpretations because the minds of the raters and researchers were also prone to fixation. Therefore, sometimes it was hard for the evaluators to see an interpreted meaning in the ambiguous sketch even if the sketch actually resembles it. It is anticipated that the SBCA's performance will be negatively affected by these unreliable ratings, so a new dataset, Show Me How You See (SMHYS) has been created in an attempt to reduce the risk.

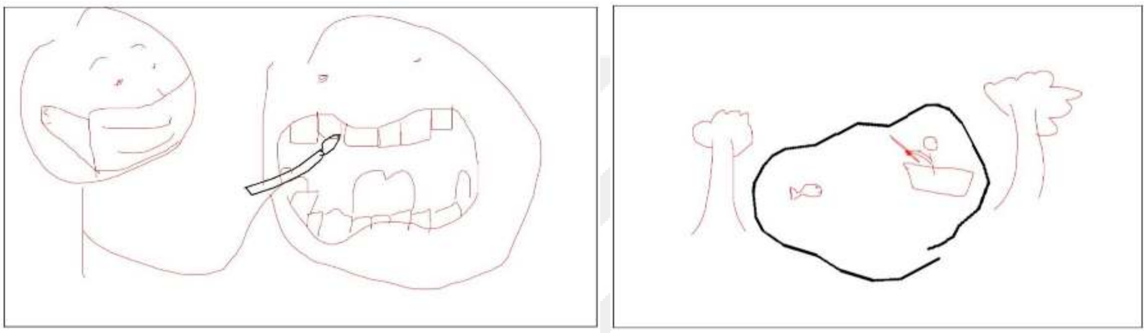
(a) (b)

Figure 3.6: (a) An ambiguous sketch from Sketchnet. (b) The SMHYS version of the ambiguous shape with "tie" interpretation.

59 volunteers from the Computer Engineering Department of Koç University have contributed to the SMHYS dataset. At the beginning of the study, the participants saw one of the ambiguous sketches in the center of an otherwise blank canvas. The interpretation for that sketch is written below. The participants could rotate, drag, and change the size of the sketch on the canvas. We asked them to use the sketch in a scene where anyone could immediately perceive the assigned interpretation. The given sketch was drawn in black ink, and the participants drew their own

strokes using red ink (See Figure 3.7).

SMHYS dataset is utilized only by the researcher during the manual inspection step (See Section 3.4.2) for the final decision. In other words, actual raters weren't able to see those to adjust their ratings accordingly. SMHYS data is only available for 3654 interpretations that are grouped and mostly post-processed because it took a significant amount of time to collect this data. Therefore, we only focused on the list of interpretations that we thought were essential for SBCA'a performance.



(a) (b)

Figure 3.7: (a) The SMHYS version of the ambiguous shape with "dentist mirror" interpretation. (b) The SMHYS version of the ambiguous shape with "lake" interpretation.

During the collection of the SMHYS dataset, our observations hinted to us that when participants are asked to create a scene, if the given ambiguous shape is rotatable and moveable on the canvas, it is easier for them to see serendipitous interpretations in the ambiguous shape. We don't have scientific data to support this claim. Also, since we didn't want creative interpretations in the Sketch Net dataset, we haven't explored this method of collecting interpretations thoroughly. However, based on our observations, we noted SMHYS interface could have been a great tool to trigger new interpretations in a future study.

3.5 Assistant: Suggesting Sketches

There are two ways to trigger the interpretation & reinterpretation cycle. You either draw a new sketch to interpret or you find a way to re-interpret the existing sketch.

SBCA should be able to help the cycle by assisting both ends. In other words, the assistant should help the user re-interpret their sketch or it should generate some fresh sketches for the user.

In order to be in an interpretation & reinterpretation cycle, an initial sketch to interpret is essential. Therefore, users are expected to sketch at least one idea before calling the assistant for help. We considered giving a random initial stroke to trigger the first idea as well, but we were hesitant about intervening too much in the user's creative process [Sun et al., 2014, Ge et al., 2020]. In the end, we decided not to intervene until participants willingly asked for help from the assistant.

Users were expected to sketch at least one idea, but they were also allowed to sketch more. In that case, they need to select a single sketch with the selection tool to get help from the assistant. When called for help, SBCA initially presents three ambiguous sketches to the user (See Figure 3.8). Those sketches are presented in their original rotation and they function as clickable buttons. When clicked, they reveal corresponding Sketch Net interpretations.

Figure 3.8: SBCA suggests three sketches according to the user's sketch.

Sketches that are shown by SBCA are selected from the Sketch Net dataset. Therefore, there are only 120 ambiguous sketches to choose from. Those sketches are selected based on their visual similarity with the user's sketch. In other words, SBCA offers the most similar sketches of Sketch Net to the user's sketch. By doing so, it creates a perception that SBCA contributed to the user's idea by sketching something similar to the user's work. If suggested sketches are very similar to the user's sketch, the re-interpretation side of the cycle is triggered. Otherwise, new sketches trigger the interpretation side of the cycle (For further information, see Section 3.6).

The visual similarity between sketches is calculated as the Euclidian distance between feature vectors, encoded by the Sketchformer model as described in Section 3.1 [Ribeiro et al., 2020].

Users are expected to treat the suggested sketches as visual stimuli and get inspiration from them. Since studies in the literature suggest that visuals are better inspiration sources than other types of stimuli, SBCA initially offers visual ideas [Gonçalves et al., 2014, Sun et al., 2013, Sarkar and Chakrabarti, 2008]. If users like the sketch as it is, they can't copy and paste it on their canvas. To use the sketch, they need to sketch it themselves. While drawing, they will probably change the original sketch a bit and this will create the possibility of triggering a new interpretation and reinterpretation cycle.

3.6 Assistant: Suggesting Interpretations

Previous studies similar to ours have preferred to use a single type of stimulus [Karimi et al., 2020, Ge et al., 2020, Rezwana et al., 2021]. However, while scientific research has focused on finding the superior type among them, it has also shown that all of them are beneficial in creative processes. Therefore, we preferred using multiple types of stimuli in our work to utilize their full potential. Suggested ambiguous sketches initially served to act as visual stimuli, and the following interpretation lists act as textual stimuli.

All of the processed and confirmed interpretations of the suggested sketches are retrieved from the Sketch Net dataset. They are sorted and ordered by their average similarity rating. By doing so, the most similar interpretations to the sketch are placed at the top of the list, while the relatively controversial ones are located at the end. All of them are initially written in the default black color.

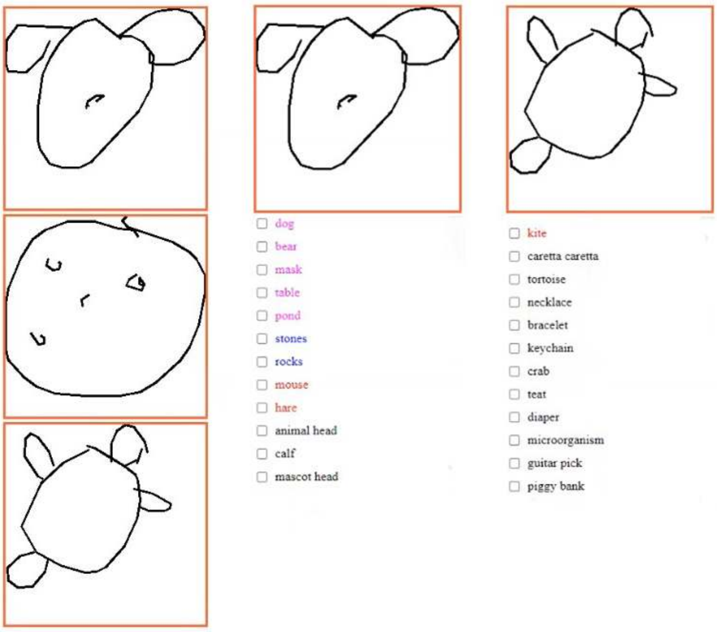


Figure 3.9: SBCA lists interpretations when clicked on the sketch.

Users may freely explore the interpretation lists of all three sketches. They may also ignore them altogether and just only take inspiration from the sketches. There is no restriction in that regard.

Interpretations are recorded with the information at which rotation the interpretation was perceived in that sketch. Therefore, for each interpretation, rotating the sketch with that amount makes perceiving the interpretation easier. Instead of using SMHYS data to help users perceive, we preferred using rotations in this step. When the user clicks on an interpretation from the list, the sketch rotates accordingly to help the user perceive it. We didn't utilize SMHYS data for this purpose because that would introduce a confounding variable in our experiments.

3.7 Assistant: Listening to the User

Once the user calls the assistant for help, the assistant asks three questions. Questions are both displayed written on the interface and spoken via speakers by Google's text-to-speech module. After the model stops reading the question, a ”Rec” button appears on the screen and records the answer of the user. The user answers the question aloud. Their answers are recorded and converted to text with Google’s text-to-speech module for further analysis. When the user answers the question, they need to press the ”Next Question” button to stop recording and hearing the next question. These questions are designed to be aligned with the nature of the design task and explained in detail in Section 4.2.2.

After the user answers all three questions, the answers are converted to text and translated into English. There are two reasons for the translation: 1) People feel less stressed while talking about their ideas if they are allowed to speak in their native language, and 2) NLP models work best in languages with the biggest available corpus, which is English. That’s why we allowed users to interact with the SBCA in Turkish but we processed all of the information in English. The translation is also done with the help of Google’s translator module. This is the only step in the entire study in which we used a translator. For example, interpretations are collected, processed, and presented in English at all times (See Figure 3.9. Since all students of Koç University are advanced in English, there wasn’t a need for translations.

Recorded and translated answers are pre-processed with regular NLP steps such as removing stopwords, lemmatization, and tokenization before being given to the language model. Often, users either repeated themselves or said things that didn’t convey much meaning. Only a small portion of what they said was valuable in terms of containing useful information. Our goal was to identify those valuable words. That’s why we used the off-the-shelf keyword extractor YAKE! to identify them [Campos et al., 2020].

Extracted keywords are encoded in word vectors as well as the interpretations of the suggested sketches [Mikolov et al., 2013]. Pairwise distances for each keyword and interpretation set are calculated. The interpretations with the minimum distance to the keywords are painted red and moved to the top of the list. That means red interpretations are related to the user’s answers and therefore ideas.

3.8 Assistant: Associating with Task Domain

SBCA's performance also benefits from giving attention to the task domain when determining stimuli. In the previous work, the assistant is generally designed to work in a single, limited domain such as bridges [Karimi et al., 2020]. We believe, having an assistant that can offer help in any domain without requiring new data would be more beneficial in terms of cultivating new ideas with inter-domain references. However, for the sake of the experiment, we intentionally restricted our model to work on only two domains related to the task we've given. Because that domain limitation gives us time efficiency we needed that in user experiments. We chose those two domains to be rather broad topics: Science & Technology and Travel & Tourism.

Having knowledge about the domain and being familiar with it enhances the performance of the SBCA in predicting the users' thoughts and goals accurately. To cover the knowledge and familiarity with the domain, we used topic modeling algorithms.

Topic modeling algorithms are designed to group documents based on their topics and predict the topic of a given document [Jelodar et al., 2019]. HLDA version is designed to build a hierarchical relationship between topics and applies it to the basic LDA algorithm, which is the foundation of topic modeling algorithms [Griffiths et al., 2003]. HLDA models have the ability to inherently learn topics and inter-topic relations while sorting documents. They use the bag of words technique, so they also learn the relationship between words and phrases as well. We used this information to calculate the distance between Sketch Net interpretations and the task domain. In the end, we had a domain-relatedness score for each interpretation. Since we had only two domains, every interpretation had two scores showing how much they were related to that particular domain.

In order to get those scores, the HLDA algorithm is trained on a specially crafted document set. Since we were only interested in the relationship between domains and interpretations, we trained the model with the documents only related to those. In other words, irrelevant documents such as the ones on celebrities, mathematical foundations of theoretical physics, and philosophy were not in the training set. Every document is hand-selected from Wikipedia based on the Sketch Net dataset.

Every processed interpretation in the Sketch Net dataset is manually searched on Wikipedia to find the related article. The article is included in the dataset if it is indeed relevant to the interpretation. However, some interpreted objects didn't have a sufficient Wikipedia page. For example, "MickeyMouse ears" was an interpretation in the dataset and it is a significant figure in modern culture and fashion. Yet, it doesn't have its own Wikipedia article. In those cases, the researchers tried and find written, explanatory information about the object from other sources. In some cases, information about one interpretation can be found scattered across different Wikipedia articles. That issue is solved by including all related articles as a document in the training set. For example, the interpretation "seat" has two articles: Theatre seats and airplane seats, because both of them cover the seat interpretation in Sketch Net. In total, only six interpretations have been outsourced for documents. The remaining 3648 interpretations are assigned at least one Wikipedia article.

To increase the performance of the model in the chosen domains, 15 articles per domain are also hand-selected and added to the training set. It's possible to choose and add 15 articles per domain in the set at this step to allow SBCA to work efficiently with other domains, as well. Another interesting idea would be to train the model with the whole of Wikipedia and other corpora as well. Budget and similar resource limitations, as well as the time-efficiency issue forced us to train the smallest model with the smallest dataset possible. That's why we limited our training set to articles on 3654 interpretations of the Sketch Net dataset and 30 articles on task domains.

After the documents are pre-processed to be fed to the model, HLDA is trained until it has balanced log-likelihood and perplexity values [Kütük and Sezgin, prep]. Once the training is complete, probability scores of the interpretations in the topic that gives the highest possibility to the domain words are taken as the relatedness score.

The domain that we will refer to as Technology includes topics such as technological devices, scientific equipment, and science fiction. Other domain that we will refer to as Travel includes topics such as travel, tourism, and entertainment. The interpretations with the highest probability score on the corresponding domain topic are painted blue and moved to the top of the interpretation list. That means blue interpretations that are listed by SBCA are the words related to the task domain.

3.9 Assistant: Coloring the Interpretation List

The ordering of the interpretation list and the colors of its items have been arranged based on a hypothesis. The hypothesis is that users will find the suggested interpretations most helpful when they are related to their own ideas and the domain they are working on. This is because the user will think that SBCA understands their ideas and tries to support them. Once the user is assured that the assistant is there to assist, SBCA has to offer new ideas. Those are the ones that are related to the user's task and ideas but haven't been investigated by the user yet. After all those, new suggestions can be offered that are somewhat related to the user's idea but give the opportunity to roam the creative space.

As stated before, all interpretations are sorted based on their similarity ratings with the suggested sketch. At that phase, all words were colored black. Secondly, interpretations related to the user's thoughts are identified based on their answers and word vectors. Those elements were painted red and placed at the top of the list. Thirdly, interpretations related to the task domain are identified based on the topic modeling algorithm. Those interpretations are colored blue and placed at the top of the interpretations list. As a last step, those elements that are chosen to be colored both blue and red are painted purple instead and placed at the top of the interpretation list. As a result, the user sees the list that starts with purple elements, followed by the blue ones, which are followed by red ones, and eventually the black interpretations.

This way, when the users click on a sketch button, they see interpretations of that sketch that are related to their answers and their task domain at the same time. Those purple interpretations make the user feel understood by SBCA. After

Figure 3.10: SBCA suggests colorful words related to suggested sketches.

that, the user reads blue interpretations. Those are the ones related to the domain. However, since they are not purple, they are not related to the user's answers and ideas. Therefore, blue ones have the potential to give new ideas to the user. So, the user will feel assisted while reading these. The following interpretations are painted red because they are related to the user's answers but not related to the task domain. Those interpretations are expected to be the least likely to assist the user since the user was already thinking about them. Lastly, the user peruses black interpretations. Those have the biggest potential to trigger the interpretation & reinterpretation cycle. Because they are neither related to the domain of the task, nor the user's ideas. Yet, they have an acceptable level of similarity rating with the suggested sketch, and that sketch is visually similar to the user's sketch. Therefore, they can create a chain reaction between sketches, perceived meanings, and ideas.

The order of the interpretation list is carefully designed. Its aim is to initially show support to the user so that the user won't reject the assistant's help right away. After the smooth introduction, it offers some helpful and reasonable new ideas to show its willingness and ability to assist. Lastly, it finds a chance to offer some wild ideas to trigger the interpretation & reinterpretation cycle. Those ideas are not irrelevant, since they are visually similar to the suggested sketch and have the potential to diverge the user's mind from their current thoughts and ideas.

Chapter 4

EXPERIMENT

The benefits of the Sketch Net dataset and the performance of SBCA are tested with a user experiment.

4.1 Participants

41 students from the Computer Engineering Department of Koç University volunteered to be in this study. These students didn't participate in any of the other studies that we mentioned in the previous sections. Therefore, they have never been exposed to the Sketch Net dataset. We did not include the data of 5 students in the analysis due to reasons such as carelessness and failure to follow the instructions. Consequently, the study is conducted with the remaining 36 participants.

In the pre-test stage, participants answered a few questions about their personal background. According to their answers, 7 of them designed a visual before. But only 2 of them have done something related to sketching. 5 of the participants reported that they felt anxious about being about to draw something. And 24 of them stated that they think they are creative people.

We could use an 11-item Short Scale of the Creative Self to measure self-reported creativity [Karwowski, 2012]. However, since we only used the background check as descriptive information, we didn't want to spend too much of the experiment time on the background questionnaires.

4.2 Experiment Design

Participants were recruited with an invitation and granted 2.5 points to one of their Computer Engineering classes upon completing the task. They are informed that the experiment is about designing logos, but drawing skills and aesthetics of the design are not crucial for participation. They had to come to a lab reserved for this experiment.

In the lab, a research observer was waiting for them. Firstly, they approved their voices and screens being recorded. After signing the Informed Consent, the setup is introduced. Their voice was recorded via the microphone of the laptop. Therefore, they didn't have to use a headset. The main researcher was also connected to the experiment online via Zoom call and observed the experiment on a one-way shared screen setup. The sound of the interface and the voice of the main researcher were audible via the speakers.

The experiments were done on a WACOM Cintiq 22HD tablet. Participants were not allowed to use the mouse at all, they were only allowed to use the pen and their voices to interact with the assistant.

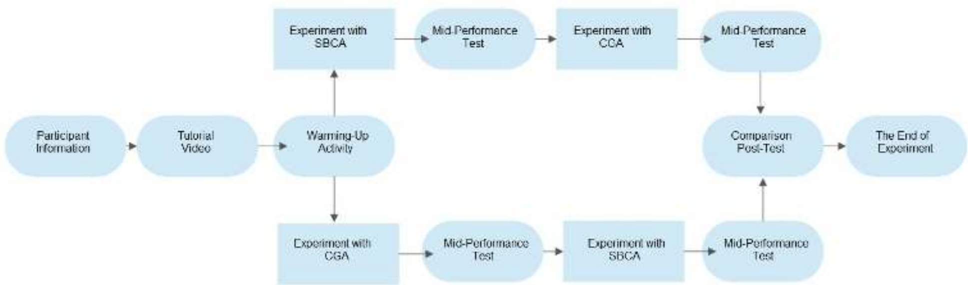


Figure 4.1: Structure of the experiment design.

The experiment started with a pre-test questionnaire on their current feelings and personal background. Once they finished answering those, they were instructed to watch the tutorial on using the interface. After that, they tackled the warming-up activity to try what they had seen and learned in the tutorial. In the warming-up task, participants were asked to design a logo for a food company. Drawing a plate, fork, or knife was forbidden. We asked them to represent the menu or the ambiance of the food company with their logo. While doing that, they interacted with selection, drawing, and eraser tools. When they asked for assistance from SBCA, they were offered three buttons with basic shapes on them: A star, a square, and a hexagon. The interpretation word list was a list of placeholders. All the items on the list were "a word" to not offer any stimulus to the user for the upcoming experiment.

This warming-up task was recommended by Van der Lugt for non-designer participants. Van der Lugt also recommended emphasizing that the interpretation and exploration of ideas are more important than the result [Van Der Lugt, 2002]. Following his advice, we initially calmed the anxious participants by telling them it wouldn't be a problem if they came up with a cliché. We also underlined that the experiment is about seeing how creative thinking is influenced by the stimulus we offer. Therefore, we wanted them to think aloud about their creative process for us to be able to record them. This method has also proved helpful and frequently used by the community [Fonteyn et al., 1993].

When they completed the warming-up activity, participants were divided into two groups for counterbalancing [Brooks, 2012]. In the end, every participant designed two logos. One was for a science & technology firm, and the other for a tourism & travel agency. One of those logo designs was assisted by SBCA, and the other was assisted by Control Group Assistant (CGA). The logo type (i.e. tourism or travel) order was randomized. However, the assistant type (i.e. SBCA or CGA) was controlled with a counter-balanced design. 18 participants started the experiment with SBCA, while the other 18 initially got help from CGA. For the second logo design, they got help from the other assistant in both cases. 20 participants initially designed a science & technology company logo, and 16 started designing with tourism & travel agency task. Therefore 19 participants designed a science & technology logo with SBCA, while 17 participants designed a tourism & travel logo with it.

After each logo design, the participants filled out a Mid-Performance Test Questionnaire. In this questionnaire, participants were reminded about the visual and verbal suggestions of the assistant they had just interacted with. They have rated the performance of the assistant accordingly.

After completing both logo design tasks and filling out the Mid-Performance Test Questionnaire for both, participants were asked to fill out one last questionnaire to compare the performances of both assistants. For the details of the questionnaires, see Section 4.3.1.

After filling out the Comparison Post-Test Questionnaire, the session ended.

4.2.1 Task

Participants were instructed to design a logo for a Tourism & Travel Agency or a Science & Technology Firm. They were required to think aloud about their designs and ideas. We emphasized three times that the visual quality of the logos is not important, and the main objective is to find creative ideas. They knew they only had 10 - 15 minutes to design a single logo, and in that period, they had to interact with the assistant at least once.

The detailed instructions for the logo were:

- **Science & Technology**: "This is a science and technology company. The company does not want to focus on products such as computers, tablets, or mobile phones in its logo. It prefers to highlight the aspect of science that improves human life. It wants a simple design rather than a design with many parts or details."

- **Tourism & Travel**: "This is a travel-tourism and entertainment company. The company doesn't want to focus on transportation vehicles such as airplanes and buses in its logo. It prefers to highlight the discovery and pleasure aspects of travel. It wants a simple design rather than a design with many parts or details."

We forbid some objects to start with because it is known that, rather counter-intuitively, a restriction may boost creative thinking [Runco and Pritzker, 2020]. In our early studies, we have seen that participants tend to initially think of a plane or a bus and then stuck with that idea. Restricting those ideas that we've mostly encountered in the early experiments helped them to discover different ideas and suggestions of the assistants.

Similar design studies gave 20 minutes to 1 hour to their participants for a design task [Stones and Cassidy, 2010]. However, their participants were mostly designers. Also, their participants had to perfect their ideas because they rated the aesthetic quality of the last design as well. We had to limit the design task to 15 minutes per logo. We observed that it was enough for most of our participants since most of them reported that they were done designing and pressed the "Finish Designing" button themselves.

4.2.2 Questions of the Assistant

All of the assistants, including the one on the warming-up activity asked participants 3 questions when they clicked on the "Call Assistant" button on the interface for the first time.

Q1: "What kind of thoughts or associations come to your mind when you look at the logo you created?"

Q2: "What makes your logo stand out from a cliché?"

Q3: "Could you think of a feature that could make the company more attractive or different from others?"

The first question aims to record the user's thoughts for the assistant to process. Even though we recorded all the processes during the experiment with think aloud method, the SBCA processes only the information it gets with these questions.

The second question serves two purposes. If the user is initially stuck with a cliché, the question implies that it's not acceptable and desirable in design. But if the idea is not cliché, then the assistant has the opportunity to focus on aspects of the idea, which makes the design far from being cliché.

The third question follows a similar path to the second one. If the user hasn't thought about the company that will use the logo, the question implies that it's an excellent aspect to think about to find new ideas and get away from the cliché. On the other hand, if the user has already thought about the company, the question is aimed at recording the critical aspects of the company to find better suggestions for the user.

If the user interacts with the assistant more than once, the assistant asks only one question: "What makes this sketch different from the previous one we were discussing?" Doing so reduces the time required to interact with the assistant for the same design idea. Therefore, it makes calling the assistant more appealing to the users. It is important to note that participants were asked to interact with the assistant at least once. One participant called the assistant five times for the same design task and that was the record of the experiment.

As stated, all assistants (SBCA, CGA, and Warming-up activity assistant) asked those three or four questions accordingly and recorded the user's answers. However, only SBCA processed the responses to adjust its suggestions. The questions are designed for SBCA to stimulate new ideas, as mentioned above. However, asking them with one assistant and not the other would create a bias. For similar reasons, asking different questions with different assistants would introduce a confounding variable. Therefore, we used the same questions with all assistants. As a result, it may have boosted the performance scores of the CGA, too.

4.2.3 Control Group Assistant

Control Group Assistant (CGA) acts as a control condition of the experiment. Since Google's Quick Draw (QD) dataset is considered helpful for creative work and the Sketch Net sketches are extracted from that dataset, we compare SBCA's suggestions with QD items as creative stimuli [Ha and Eck, 2017].

CGA offers three sketches as buttons, similar to SBCA. To suggest those sketches, we first selected 120 categories out of 345 classes of QD. Then, for each selected category, we randomly chose one sketch. All chosen sketches were labeled to have been correctly recognized by Google's classifier. To sum up, we formed a set of 120 QD sketches for CGA to sample from. CGA randomly selects three sketches from this pool every time the user calls for assistance.

For each button, we set a random number between 10-20 to be our word list length. After setting the length randomly, we form these lists by randomly selecting that number of 345 QD categories. Selected category names are used as textual stimuli. All of the category names are randomly colored to be red and blue. If a word is colored both red and blue, it is painted purple. The order of the word color presentation is the same as SBCA. The word list starts with purples, followed by blues, reds, and blacks.

As stated in the previous section, asking questions is not designed for the CGA, and the CGA doesn't process the users' answers. However, in order to avoid introducing bias in the experiment design, we also included assistant questions in CGA. In other words, the users also answer the same questions in control condition but their answer is not processed.

We expected CGA to trigger creative ideas to an extent because sketches are ambiguous, and ambiguity triggers creativity. Also, CGA asks the same questions that we designed to trigger deeper thought. That's why we hypothesized that CGA would help the design process as well. However, SBCA would trigger significantly more distinct ideas than CGA while creating a sense of collaboration. Since SBCA is designed to create a feeling of understanding the user, we also hypothesized that it would be preferred over CGA by the users.

4.3 Metrics

We used three measures to test the effectiveness and benefits of the SBCA.

4.3.1 Perceived Benefit

First, the perceived benefit of the SBCA. During the experiment, we asked our participants to fill out four questionnaires. The first one was the demographics questionnaire, so it is not included in the perceived benefit score. The rest are two Mid-Performance Test Questionnaires and one Comparison Post-Test Questionnaire. Those were subjective measures of effectiveness. Yet, it allowed us to measure the perceived benefit.

In the Mid-Performance Test Questionnaire, participants answered five questions:

1) Did sketches of the assistant give you new ideas?

2) Did the word list of the assistant give you new ideas?

3) How related were the assistant's suggestions to your design or idea?

4) How helpful was the assistant to your creation process, in general?

5) Is there any comment you wish to add?

The last question was open-ended. All the others were 11-point Likert scales. 0 represents None or Never, and 10 represents Too Much or Always for all questions. These questionnaire items were inspired by the Creativity Support Index (CSI) [Cherry and Latulipe, 2014]. We didn't use CSI to measure the benefits of the assistant due to three reasons: 1) We specifically wanted to learn about the contributions of visuals and texts independently, 2) CSI is reported to take 5 minutes that we didn't spare, and 3) Since we conduct our study in Turkish, the metric should've been translated to Turkish and that would require reliability analysis, too.

Comparison Post-Test Questionnaire asks six questions:

1) Which assistant's suggestions were more beneficial to you?

2) During which logo design did the assistant provide you with more new ideas?

3) During which logo design did you feel the assistant understood you better?

4) Compare both assistants' performance in giving you new ideas.

5) Compare both assistants according to their level of creativity in giving suggestions.

6) Is there any comment you wish to add?

The last question was open-ended. The first three questions required selecting between "Science & Technology" and "Travel & Tourism" because participants didn't know if they were interacting with SBCA or CGA. They were also not informed about there being two different assistants until they saw the Comparison Post-Test Questionnaire. Question 4 offered two different 11-point Likert Scales. One scale was for "Science & Technology," and the other was for "Travel & Tourism." So that their performances can be compared based on a numeric rating. Lastly, the fifth question placed "Travel & Tourism" at the left end and "Science & Technology" to the right end on an 11-point Likert Scale. The participants had to choose one side over the other.

4.3.2 The Quality of the Creative Process

To measure the quality of the creative process, we collaborated with three expert designers. They are doctors and professors of the Arts & Design Department with at least eight years of work experience.

Raters reviewed the creative process of each eligible participant and rated the performance of both the users and the assistants. They worked on a transcription document rather than the recording of the whole experiment. The voice and screen recordings are transcribed to present the entirety of the creative processes. When necessary, the screenshots are attached to the transcription to make all processes evident to the raters. Still, they could also ask for the recording of a particular part if there is confusion about it.

Expert ratings are one of the literature's most common creativity measurement techniques [Zhao et al., 2021, Saggar et al., 2015, Barbot et al., 2011, Finke and Slayton, 1988]. Therefore, we employed this measurement technique to compare our study with the existing work. However, we also included some scale items unique to us in order to identify the contributions of each individual component to the overall creative processes.

For this purpose, experts individually rated the creation processes for 72 logos in 8 different items.

- **User Creativity - Divergent Thinking and Flexibility Score**: The user's ability to uncover innovative ideas is assessed by taking into account various factors, including the number of distinct ideas generated, the extent of their divergence and abstraction, and their overall productivity.

- **User Creativity - Convergent Thinking and Elaboration Score**: The user's proficiency in refining and improving existing ideas is evaluated by analyzing their assessment of the assistant's stimulus to enhance their initial concepts, as well as their skill in converging on a solution without getting fixated.

- **Assistant Creativity Score**: The assistant's ability to inspire new ideas through textual and visual stimuli is assessed by examining the extent to which the presented stimuli diverge within the creative design space and how abstract they are, thus creating opportunities for new ideas. This evaluation is conducted solely based on the assistant's performance, without considering the user's contributions.

- **User-Assistant Interaction Score**: The quality of the interaction between the user and the assistant in facilitating creative processes is evaluated by analyzing the extent to which they worked together to diverge from less productive ideas and converge on more valuable ones, as well as their collective ability to achieve abstraction.

- **Most Original Idea Score**: The degree of creativity in the most innovative idea generated through collaboration with the assistant.

- **Last Logo Design Creativity Score**: The extent of creativity exhibited in the final logo design.

- **Close Fit with the Theme**: The level of alignment with the design domain (i.e. technology or travel) throughout the creative process. It is evaluated based on the ability to stay within the domain, generate novel ideas related to the field, and the relevance of the stimuli provided by the assistant to the chosen domain.

- **Clarity of Expression**: The degree of visual expression accomplished by the user during the logo design process. This assessment considers the visual quality and appeal of the final logo design, as well as how effectively the intended idea is conveyed with clarity.

Every item was an 11-point Likert scale ranging from 0 to 10. For each, 0 represented the inability and poor performance while 10 represented the perfect score with fulfilling performance.

4.3.3 Number of Creative Ideas

Since we claim that SBCA offers a simulation for group collaboration, counting the distinct ideas it triggered is a good measure of its performance. That is also an acceptable creativity measurement method in the literature.

To reduce bias, we wanted experts to count the number of ideas generated during the interaction. There wasn't an established method of discerning different ideas in a chain of thought. Therefore, experts counted the ideas if they were thoroughly explained or visually depicted during the process. It's important to emphasize that merely mentioning a keyword without providing additional context or details was not regarded as an idea creation. When it comes to the chain of ideas, the thought process was segmented at specific points where the idea evolved into something different and distinct from the previous step. Each time the idea transformed into a new form, the counter was incremented by one.

The experts worked collectively to monitor the count of ideas and engaged in ongoing discussions during the logo creation process. Together, they evaluated whether each idea represented an evolution or a departure from the preceding concept.

Chapter 5

RESULTS

5.1 Perceived Creativity Analysis

The results show that for every questionnaire item, users preferred SBCA or rated its performance higher with a statistically significant difference.

Repeated Measures ANOVA test showed that in the Mid-Performance Test Questionnaire, participants reported that sketches of the SBCA give them new ideas more than CGA did ($F_{1,35} = 13.7$, p <.001, $\eta^2 = .28$). This difference returned a large effect size. Similarly, word suggestions of the SBCA are perceived to trigger more ideas than CGA ($F_{1,35} = 11.7$, p <.005, $\eta^2 = .25$). This difference also returned a large effect size.

When users were questioned about the relevance of the assistant's suggested words to their ideas and thoughts, they consistently rated SBCA higher in terms of word relatedness ($F_{1,35} = 16.0$, p <.001, $\eta^2 = .31$). This time, the difference returned an even larger effect size.

In the Mid-Performance Test, users rated the helpfulness of the assistant higher for SBCA ($F_{1,35} = 6.7$, p <.05, $\eta^2 = .16$) with a large effect size.

In the Comparison Post-Test Questionnaire, when users compared the performance of the assistants in providing them with new ideas, they rated SBCA higher ($F_{1,35} = 13.4$, p <.001, $\eta^2 = .28$).

Mauchley's Test of Sphericity returned non-significant results for all those tests reported above. Therefore, the statistical reliability of the results is also guaranteed.

Since 36 participants are insufficient to conduct a Chi-Square test, we report the statistics of the other Comparison Post-Test Questionnaire results as a descriptive measure.

Out of the 36 participants, the majority (26 participants) indicated that they found SBCA's suggestions more helpful. Similarly, 26 participants stated that SBCA provided them with more ideas during the design process. A substantial number of participants (30) reported that they felt that SBCA understood them better. Moreover, when participants were asked to compare the creativity level of the two assistants, the results consistently favored SBCA (See Figure 5.1).

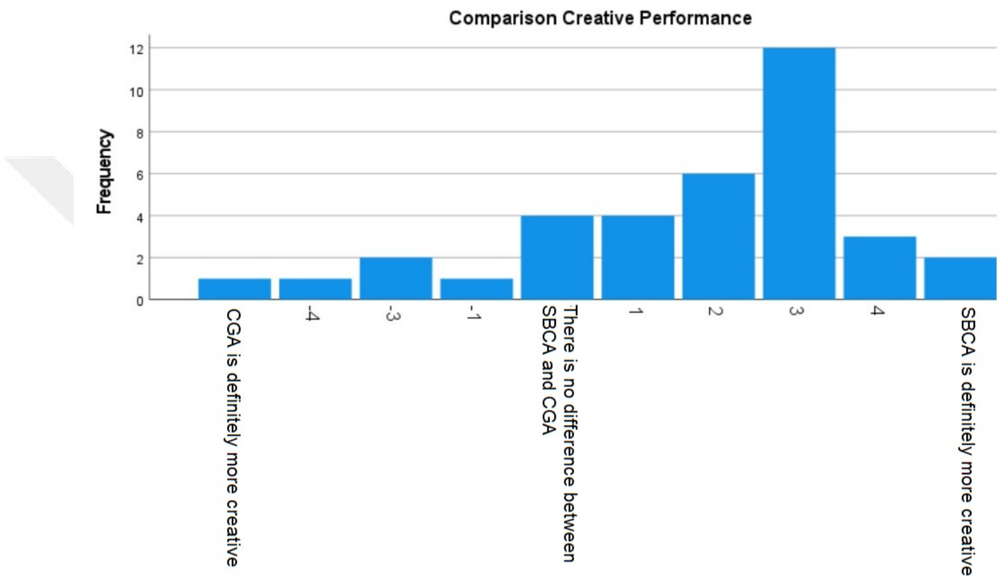


Figure 5.1: The frequency distribution of answers given to the Comparison Pots-Test Questionnaire Question 5: "Compare both assistants according to their level of creativity in giving suggestions."

We also conducted tests to examine whether the type of logo had an impact on the perceived benefit of the assistants. The Independent Samples T-test results indicated that for SBCA, there was no statistically significant difference. In simpler terms, whether participants were designing a logo for the travel & tourism agency or the science & technology industry did not significantly alter the perceived benefit of SBCA.

Similarly, participants' level of experience in visual design or sketching did not affect the perceived benefit of SBCA. Additionally, factors such as self-perceived creativity or anxiety related to drawing did not have a statistically significant effect on the perceived benefit of SBCA. In other words, whether individuals considered themselves creative or not, and whether they felt anxious or at ease with drawing, did not lead to significant changes in the perceived benefit of SBCA.

In the SBCA condition, we conducted a comparison between the frequency of colorful and black word suggestions that were marked as inspiring. As depicted in the Figure 5.2, the analysis of means indicates a tendency for users to favor colorful words. Upon closer examination of individual colors, it was observed that users showed a preference for purple words, followed by blue words, and lastly, red words. However, it's essential to note that none of these differences reached statistical significance. This lack of significance can be attributed to the limited sample size and unstable variance differences.

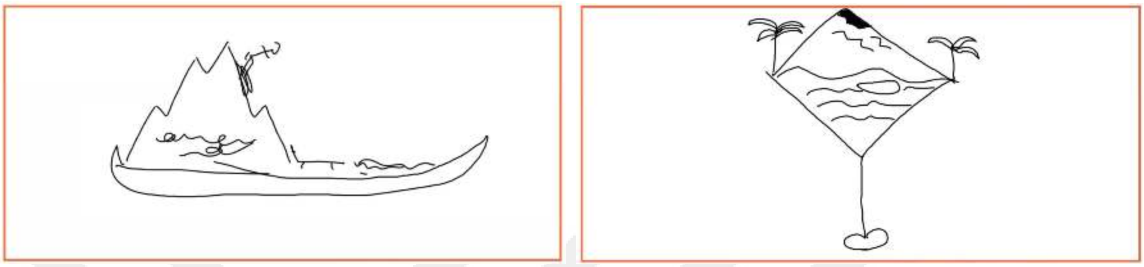
(a) (b)

Figure 5.2: (a) Boxplot of preferred words in SBCA condition. The orange represents colorful words: purple, blue and red. (b) Boxplot of preferred colorful words in SBCA condition.

5.2 *Expert Designer Ratings*

We conducted a Repeated Measures ANOVA test to determine which assistant was more beneficial for the creative processes, as assessed by experts. To achieve this, we calculated an average rating based on evaluations from all three raters for each item.

In essence, each logo underwent individual evaluations by raters based on eight distinct criteria. For example, Logo A was assessed in terms of convergence, divergence, assistant contribution, interaction, the most original idea, final logo creativity, theme, and expression (See Section 4.3.2). However, it's important to note that these criteria were rated independently by three different experts. Consequently, Logo A, in reality, received a total of 24 scores, with each criterion rated by a different rater.



(a) (b)

Figure 5.3: (a) A travel agency logo. Ship with a mountain at the top. (b) Another travel agency logo. Cocktail glass with a mountain inside.

Initially, our objective was to evaluate the inter-rater reliability to ensure that the ratings were consistent and aligned measures of creativity. Once we confirmed the reliability, we calculated the average score for each criterion by aggregating the individual ratings. It's noteworthy that all criteria exhibited acceptable inter-rater reliability, with moderate-to-high agreement levels (ranging from .7 to .9) between raters, as indicated by the correlation analysis. Consequently, we obtained an average score for each of the eight criteria for each logo, serving as the foundation for subsequent analyses.

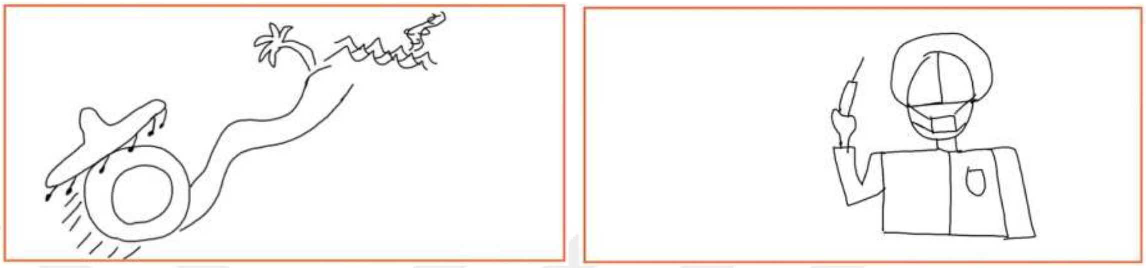
Combined average scores are used to conduct Repeated Measures ANOVA. The results are as follows:

There isn't a statistically significant difference between SBCA and CGA in terms of User Creativity - Divergence and Flexibility Score ($F_{1,35} = 1.21$, ns) as expected (See Chapter 6). Similarly, User Creativity - Convergent Thinking and Elaboration Score analysis also didn't result in a statistically significant difference between SBCA and CGA ($F_{1,35} = 2.07$, ns).

The performance of SBCA (M=5.12; SD=1.09) is significantly better than the performance of CGA (M=4.67; SD=1.04) in terms of The Assistant Creativity Score

($F_{1,35} = 4.86$, $p < .05$, $\eta^2 = .12$)

The User-Assistant Interaction Score returned a marginally significant ($F_{1,35} = 3.80$, $p = .059$) result to indicate that SBCA (M=5.37, SD=1.57) interacted better with the user than CGA (M=4.79, SD=1.23).



(a) (b)

Figure 5.4: (a) A travel agency logo. A single tire with a sombrero, going for an adventure. (b) A science firm logo. A robotic nurse.

Even though the SBCA (M=5.01, SD=1.57) has a higher mean score for the Most Original Idea Score than the CGA (M=4.61, SD=1.35), the difference between them is not statistically significant ($F_{1,35} = 1.82$, ns).

However, the Last Logo Design Creativity Score analysis returned a statistically significant result ($F_{1,35} = 4.15$, $p < .05$, $\eta^2 = .11$). According to that, the final logos that are created after interacting with SBCA (M=5.09, SD=1.58) is more creative than the ones created with the help of CGA (M=4.52, SD=1.35).

SBCA (M=5.59, SD=1.47) is rated higher than CGA (M=4.66, SD=1.50) for the Close Fit With the Theme Score and this difference is statistically significant ($F_{1,35} = 7.45$, $p < .01$, $\eta^2 = .18$).

SBCA (M=5.59, SD=1.62) also helped its users to clearly express their ideas in their design with better performance than CGA (M=4.85, SD=1.40). Clarity of Expression scores also showed a statistically significant difference between the two assistants ($F_{1,35} = 6.49$, $p < .05$, $\eta^2 = .16$).

It is also important to note that Mauchly's Test of Sphericity returned a perfect, non-significant score for each item; showing that the results were reliable. Also, adding the type of logo variable as a confounding variable to the ANOVA model didn't return a statistically significant interaction score.

5.3 Number of Generated Ideas

The experts were asked to collaboratively count the ideas generated during the creative process. A Repeated Measures ANOVA is conducted to see if the assistants were able to trigger significantly more ideas from each other.

The results indicated that CGA (M=3.31, SD=1.23) was able to generate a significantly greater quantity of ideas compared to SBCA (M=2.83, SD=1.02) ($F_{1,35}$ = 6.07, p <.05, η^2 = .15).

Chapter 6

CONCLUSION & DISCUSSION

Results of the user study showed that Sketch-Based Creativity Assistant is an effective tool for supporting the early visual design process.

The perceived benefit of SBCA appears to be significant, as indicated by the effect size scores. In fact, all the significant findings showed a substantial effect size. Participants reported finding both visual stimuli and word lists beneficial. Furthermore, many participants noted that attempting to respond to the assistant's questions also contributed to improving their ideas.

According to experts, SBCA indeed played a more substantial role in enhancing the creative processes. It aided users in generating more original ideas, sketching more creative logos, and maintaining alignment with the task theme. There was no difference in user creativity scores, which encompassed both divergence and convergence, between SBCA and CGA. Those measures solely examined user contributions. This outcome was expected, given that it was a Within-Group experiment. In other words, users exhibited similar levels of creativity when utilizing both CGA and SBCA. However, when using SBCA, they achieved the creation of higher-quality logos in terms of creativity. Consequently, the non-significant results of user creativity scores further substantiate our claim that SBCA enhances creativity and stimulates valuable ideas in design tasks. It's important to note that this enhancement in creativity is not a permanent attribute of the user. Additionally, SBCA assists users in articulating their ideas with precision in a detailed and visually captivating design. This was an unexpected but fortuitous outcome of SBCA.

SBCA didn't elicit as many ideas as CGA did. However, it appears that quantity does not necessarily equate to quality in terms of ideas and creativity. Prior studies in the literature have often quantified creativity by counting the number of ideas. Our findings, however, suggest that this might not be a reliable measure for assessing creativity. There are two plausible interpretations of these results. One, SBCA could be prompting fewer but more valuable ideas. Alternatively, users might have been sufficiently satisfied with SBCA's initial suggestions and didn't feel compelled to generate new ideas. Both interpretations hold merit, but further evidence is needed to definitively support either claim, as our current data does not offer conclusive proof either way.

To sum up, SBCA proves to be useful in finding creative ideas when it comes to visual design. It can assist non-experienced designers and its benefits can be measured even in a task that takes 10-15 minutes. SBCA's contributions to the creative process are both clearly perceived by the users and acknowledged by the experts. These are encouraging results to improve our model further.

6.1 Limitations & Further Research

There were many limitations in our research which can be studied with further steps.

For example, our participants were not promised a reward for finding creative ideas. They were only rewarded with a bonus point upon completing the experiment tasks. If we transformed the experiment into a competition that rewards the most original idea, that would act as an incentive to draw more distinct logos.

Another limitation could be the time limitation. Even though most of our participants finished their session design in 10-15 minutes, some were told to finalize their designs when they hit the time limit. Giving more time to design a single logo could increase the interaction with the assistant. Also, having more time would also allow us to present the CSI questionnaire to the participants to get more comparable results.

We tested our interface with two students from the Arts & Design department to get insight and expert opinion. Naturally, their results are not included in our experiments. However, their input was valuable. They suggested providing users with the option to utilize both CGA and SBCA. In this proposed setup, CGA would function as a random stimulus generator, serving as a creative spark when a designer is unable to sketch anything. Since SBCA requires users to sketch an initial idea before being activated, CGA could play a role in the very first step of the design process. However, it's crucial to acknowledge that the effectiveness of randomized stimuli depends on their relevance to the design task and the user's background [Boden et al., 2004]. This is why we currently encourage users to create their initial sketches themselves. SBCA is specifically designed to assist in the creative process when users seek collaboration on an idea that is already initiated by the user. As a potential avenue for future research, we may consider conducting a similar experiment involving both designer and non-designer participants, offering them the option to choose between assistants for the same design task each time they call for an assistant.

Yet, we found Facebook's approach brilliant to trigger the initial sketch [Ge et al., 2020]. They have given users a single randomly generated stroke to kickstart the creative process. Drawing inspiration from this approach, we could consider offering our users the option to receive a similarly generated initial stroke upon request. This would open up an interesting avenue for comparison between CGA as an optional assistant and the provision of a random initial stroke, enabling us to discern which feature is more beneficial for our users. Such a comparative study could yield valuable insights into enhancing the user experience and effectiveness of our system.

We collected the SMHYS dataset to handle a potential risk. Eventually, we didn't utilize it in our final design. Nevertheless, we saw that rotating a sketch on the canvas, adding new strokes on top of it, and dragging the sketch to change its position in a scene sometimes triggered serendipitous findings. That's probably due to psychological concepts such as constructive memory and situatedness [Gero, 1999]. It's possible to combine the SMHYS interface with the SBCA to record even better results for creative thinking. Also, suggesting scene sketches of the SMHYS dataset could also be utilized as another type of visual stimuli.

In another interesting study, we could potentially incorporate the designer's interpretations of ambiguous sketches into Sketch Net. In the current state of Sketch Net, all interpretations were crowdsourced from non-designers. Recent research has indicated that when a non-designer collaborates with a designer on a design task, they tend to generate more creative ideas compared to scenarios involving exclusively non-designer teams or exclusively designer teams [Ou et al., 2023]. Since SBCA is composed to simulate a collaborative assistant in design, improving its data set with the designers' interpretations could be useful.

It is known that receiving feedback and criticism is an important step in design. However, SBCA only offers assistance on the interpretation & reinterpretation cycle. It may be possible to integrate the Chat-GPT into SBCA to give feedback to the users, as well. It would be an intriguing addition. However, we observed via the think-aloud method that the users tend to utter incomplete sentences and raw ideas with poor grammatical structure. Their thought processes were sometimes hard to understand even for a human who is knowledgeable about the task, the context, and the informal cues of the language. Therefore, we anticipate that Chat-GPT, in its current form, would fail to understand users if they thought aloud. Also, since we transcribed the creative processes for the experts, we tried asking Chat-GPT to translate them. The model wasn't able to catch the semantics of most of the utterances.

Another performance improvement of SBCA is possible by removing the translator component. We realized that sometimes the translator missed or misunderstood some of the words while translating from Turkish to English. That would only change the color of the words in the interpretation list. Yet, it is a limitation that we should take note of.

Sketch Net contains only 120 sketches due to budget limitations. However, we have written the algorithm for extracting ambiguous sketches and seen that it's working. Therefore, it is easy and possible for us to increase the number of ambiguous sketches in the dataset and assist users with a more extensive range of visual stimuli. It's a simple limitation that we can easily overcome.

It is known that people tend to change the style of their sketches while talking and collaborating [Dede et al., 2023, Çelik et al., 2022]. For example, they draw deictic strokes to represent some gestures on paper while talking. Our research topic and experiment setting were able to inhibit these variations. Yet, handling those could be another improvement to make SBCA a fully collaborative assistant.

Numerous avenues lay ahead, offering opportunities for the advancement of SBCA and the utilization of Sketch Net to enhance sketch processing technologies. We are eager to embark on further research endeavors within these domains.

BIBLIOGRAPHY

[Adobe Inc., 2019] Adobe Inc. (2019). Adobe illustrator.

[Bar and Aminoff, 2003] Bar, M. and Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, 38(2):347–358.

[Barbot et al., 2011] Barbot, B., Besançon, M., and I Lubart, T. (2011). Assessing creativity in the classroom. *The Open Education Journal*, 4(1).

[Barenholtz, 2014] Barenholtz, E. (2014). Quantifying the role of context in visual object recognition. *Visual Cognition*, 22(1):30–56.

[Benedek et al., 2020] Benedek, M., Bruckdorfer, R., and Jauk, E. (2020). Motives for creativity: Exploring the what and why of everyday creativity. *The Journal of Creative Behavior*, 54(3):610–625.

[Biederman, 1987] Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2):115.

[Biederman, 2017] Biederman, I. (2017). On the semantics of a glance at a scene. In *Perceptual organization*, pages 213–253. Routledge.

[Boden et al., 2004] Boden, M. A. et al. (2004). *The creative mind: Myths and mechanisms*. Psychology Press.

[Brooks, 2012] Brooks, J. L. (2012). Counterbalancing for serial order carryover effects in experimental condition orders. *Psychological methods*, 17(4):600.

[Brown et al., 2020] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler,

D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners. *CoRR*, abs/2005.14165.

[Buxton, 2010] Buxton, B. (2010). *Sketching user experiences: getting the design right and the right design*. Morgan kaufmann.

[Campos et al., 2020] Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., and Jatowt, A. (2020). Yake! keyword extraction from single documents using multiple local features. *Information Sciences*, 509:257–289.

[Çelik et al., 2022] Çelik, B., Dede, E., and Sezgin, T. M. (2022). A criticism on popular sketch datasets. In *2022 30th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4. IEEE.

[Cherry and Latulipe, 2014] Cherry, E. and Latulipe, C. (2014). Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(4):1–25.

[Crilly, 2015] Crilly, N. (2015). Fixation and creativity in concept development: The attitudes and practices of expert designers. *Design studies*, 38:54–91.

[Davis et al., 2016] Davis, N., Hsiao, C.-P., Singh, K. Y., and Magerko, B. (2016). Co-creative drawing agent with object recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 12, pages 9–15.

[Dede et al., 2023] Dede, E., Zindancıoğlu, A., and Sezgin, T. M. (2023). A critical review of sketch collection methods: Remembering how humans really sketch. In Preparation - prep.

[Diedrich et al., 2015] Diedrich, J., Benedek, M., Jauk, E., and Neubauer, A. C.

(2015). Are creative ideas novel and useful? *Psychology of Aesthetics, Creativity, and the Arts*, 9(1):35.

[Eckert and Stacey, 2000] Eckert, C. and Stacey, M. (2000). Sources of inspiration: a language of design. *Design studies*, 21(5):523–538.

[Eitz et al., 2012] Eitz, M., Hays, J., and Alexa, M. (2012). How do humans sketch objects? *ACM Transactions on graphics (TOG)*, 31(4):1–10.

[Exner Jr, 1993] Exner Jr, J. E. (1993). *The Rorschach: A Comprehensive System: Basic Foundations, Vol. 1.* John Wiley & Sons.

[Fan et al., 2019] Fan, J. E., Dinculescu, M., and Ha, D. (2019). Collabdraw: an environment for collaborative sketching with an artificial agent. In *Proceedings of the 2019 on Creativity and Cognition*, pages 556–561.

[Farthing, 2011] Farthing, S. (2011). The bigger picture of drawing. In *Thinking through drawing: practice into knowledge. Proceedings of an interdisciplinary symposium on drawing, cognition and education. New York: Columbia Uni-versity*, pages 21–25.

[Feng et al., 2009] Feng, D., Besana, S., and Zajac, R. (2009). Acquiring high quality non-expert knowledge from on-demand workforce. In *Proceedings of the 2009 Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources (People's Web)*, pages 51–56.

[Ferguson, 1994] Ferguson, E. S. (1994). *Engineering and the Mind's Eye.* MIT press.

[Fink et al., 2012] Fink, A., Koschutnig, K., Benedek, M., Reishofer, G., Ischebeck, A., Weiss, E. M., and Ebner, F. (2012). Stimulating creativity via the exposure to other people's ideas. *Human brain mapping*, 33(11):2603–2610.

[Finke and Slayton, 1988] Finke, R. A. and Slayton, K. (1988). Explorations of creative visual synthesis in mental imagery. *Memory & cognition*, 16(3):252–257.

[Fonteyn et al., 1993] Fonteyn, M. E., Kuipers, B., and Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative health research*, 3(4):430–441.

[Ge et al., 2020] Ge, S., Goswami, V., Zitnick, C. L., and Parikh, D. (2020). Creative sketch generation. *arXiv preprint arXiv:2011.10039*.

[Gero, 1990] Gero, J. S. (1990). Design prototypes: a knowledge representation schema for design. *AI magazine*, 11(4):26–26.

[Gero, 1999] Gero, J. S. (1999). Constructive memory in design thinking. In *Design thinking research symposium: Design representation*, volume 4, pages 29–35. MIT, Cambridge Cambridge.

[Goel, 1995] Goel, V. (1995). *Sketches of thought*. MIT press.

[Goldschmidt, 1991] Goldschmidt, G. (1991). The dialectics of sketching. *Creativity research journal*, 4(2):123–143.

[Gonçalves et al., 2014] Gonçalves, M., Cardoso, C., and Badke-Schaub, P. (2014). What inspires designers? preferences on inspirational approaches during idea generation. *Design studies*, 35(1):29–53.

[Griffiths et al., 2003] Griffiths, T., Jordan, M., Tenenbaum, J., and Blei, D. (2003). Hierarchical topic models and the nested chinese restaurant process. *Advances in neural information processing systems*, 16.

[Gross and Do, 1996] Gross, M. D. and Do, E. Y.-L. (1996). Ambiguous intentions: a paper-like interface for creative design. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, pages 183–192.

[Ha and Eck, 2017] Ha, D. and Eck, D. (2017). A neural representation of sketch drawings. *arXiv preprint arXiv:1704.03477*.

[Heer and Bostock, 2010] Heer, J. and Bostock, M. (2010). Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 203–212.

[Jansson and Smith, 1991] Jansson, D. G. and Smith, S. M. (1991). Design fixation. *Design studies*, 12(1):3–11.

[Jelodar et al., 2019] Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., and Zhao, L. (2019). Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78:15169–15211.

[Jongejan et al., 2016] Jongejan, J., Rowley, H., Kawashima, T., Kim, J., and Fox-Gieg, N. (2016). The quick, draw!-ai experiment. *Mount View, CA, accessed Feb*, 17(2018):4.

[Karimi et al., 2019] Karimi, P., Maher, M. L., Davis, N., and Grace, K. (2019). Deep learning in a computational model for conceptual shifts in a co-creative design system. *arXiv preprint arXiv:1906.10188*.

[Karimi et al., 2020] Karimi, P., Rezwana, J., Siddiqui, S., Maher, M. L., and Dehbozorgi, N. (2020). Creative sketching partner: an analysis of human-ai co-creativity. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pages 221–230.

[Karwowski, 2012] Karwowski, M. (2012). Did curiosity kill the cat? relationship between trait curiosity, creative self-efficacy and creative personal identity. *Europe's Journal of Psychology*, 8(4):547–558.

[Kavakli and Gero, 2001] Kavakli, M. and Gero, J. S. (2001). Sketching as mental imagery processing. *Design studies*, 22(4):347–364.

[Kittur et al., 2008] Kittur, A., Chi, E. H., and Suh, B. (2008). Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 453–456.

[Kornmeier and Bach, 2012] Kornmeier, J. and Bach, M. (2012). Ambiguous figures–what happens in the brain when perception changes but not the stimulus. *Frontiers in human neuroscience*, 6:51.

[Kozbelt, 2001] Kozbelt, A. (2001). Artists as experts in visual cognition. *Visual cognition*, 8(6):705–723.

[Kozbelt and Seeley, 2007] Kozbelt, A. and Seeley, W. P. (2007). Integrating art historical, psychological, and neuroscientific explanations of artists' advantages in drawing and perception. *Psychology of Aesthetics, Creativity, and the Arts*, 1(2):80.

[Kütük and Sezgin, prep] Kütük, A. and Sezgin, T. M. (In Preparation - prep). Human perception based sketch processing. Master's thesis, Koç University.

[Lee et al., 2011] Lee, Y. J., Zitnick, C. L., and Cohen, M. F. (2011). Shadowdraw: real-time user guidance for freehand drawing. *ACM Transactions on Graphics (ToG)*, 30(4):1–10.

[Luck and Hollingworth, 2008] Luck, S. J. and Hollingworth, A. (2008). *Visual memory*. OUP USA.

[MacQueen, 1965] MacQueen, J. (1965). Some methods for classification and analysis of multivariate observations. In *Proc. 5th Berkeley Symposium on Math., Stat., and Prob*, page 281.

[Malmberg et al., 2012] Malmberg, K. J., Criss, A. H., Gangwani, T. H., and Shiffrin, R. M. (2012). Overcoming the negative consequences of interference from recognition memory testing. *Psychological Science*, 23(2):115–119.

[McKim, 1980] McKim, R. H. (1980). Experiences in visual thinking. brooks. *Cole, Monterey, CA*.

[Mikolov et al., 2013] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[Mooney, 1957] Mooney, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 11(4):219.

[Motzenbecker and Phillips, 2017] Motzenbecker, D. and Phillips, K. (2017). Google creative lab - autodraw.

[Murnane and Phelps, 1994] Murnane, K. and Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? a global activation model. *Memory & Cognition*, 22:584–590.

[Nijstad and Stroebe, 2006] Nijstad, B. A. and Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and social psychology review*, 10(3):186–213.

[Niu and Sternberg, 2001] Niu, W. and Sternberg, R. J. (2001). Cultural influences on artistic creativity and its evaluation. *International journal of psychology*, 36(4):225–241.

[OpenAI, 2023] OpenAI (2023). Gpt-4 technical report.

[Ou et al., 2023] Ou, X., Goldschmidt, G., and Erez, M. (2023). The effect of disciplinary diversity on design idea generation in dyadic teams. *Design Studies*, 86:101184.

[Paulus and Brown, 2007] Paulus, P. B. and Brown, V. R. (2007). Toward more creative and innovative group idea generation: A cognitive-social-motivational perspective of brainstorming. *Social and Personality Psychology Compass*, 1(1):248–265.

[Plucker and Makel, 2010] Plucker, J. A. and Makel, M. C. (2010). Assessment of creativity.

[Rezwana and Maher, 2022] Rezwana, J. and Maher, M. L. (2022). Designing creative ai partners with cofi: A framework for modeling interaction in human-ai co-creative systems. *ACM Transactions on Computer-Human Interaction*.

[Rezwana et al., 2021] Rezwana, J., Maher, M. L., and Davis, N. (2021). Creative penpal: A virtual embodied conversational ai agent to improve user engagement and collaborative experience in human-ai co-creative design ideation. In *IUI Workshops*.

[Rhodes, 1961] Rhodes, M. (1961). An analysis of creativity. *The Phi delta kappan*, 42(7):305–310.

[Ribeiro et al., 2020] Ribeiro, L. S. F., Bui, T., Collomosse, J., and Ponti, M. (2020). Sketchformer: Transformer-based representation for sketched structure. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14153–14162.

[Robin and Olsen, 2019] Robin, J. and Olsen, R. K. (2019). Scenes facilitate associative memory and integration. *Learning & Memory*, 26(7):252–261.

[Rodrigues et al., 2019] Rodrigues, S., Figueiras, A., and Alexandre, I. (2019). Once upon a time in a land far away: guidelines for spatio-temporal narrative visualization. In *2019 23rd International Conference Information Visualisation (IV)*, pages 44–49. IEEE.

[Runco and Jaeger, 2012] Runco, M. A. and Jaeger, G. J. (2012). The standard definition of creativity. *Creativity research journal*, 24(1):92–96.

[Runco and Pritzker, 2020] Runco, M. A. and Pritzker, S. R. (2020). *Encyclopedia of creativity*. Academic press. In 'Novelty'.

[Saggar et al., 2015] Saggar, M., Quintin, E.-M., Kienitz, E., Bott, N. T., Sun, Z., Hong, W.-C., Chien, Y.-h., Liu, N., Dougherty, R. F., Royalty, A., et al. (2015). Pictionary-based fmri paradigm to study the neural correlates of spontaneous improvisation and figural creativity. *Scientific reports*, 5(1):1–11.

[Sarkar and Chakrabarti, 2008] Sarkar, P. and Chakrabarti, A. (2008). The effect of representation of triggers on design outcomes. *Ai Edam*, 22(2):101–116.

[Sarvadevabhatla et al., 2018] Sarvadevabhatla, R. K., Surya, S., Mittal, T., and Babu, R. V. (2018). Pictionary-style word guessing on hand-drawn object sketches: Dataset, analysis and deep network models. *IEEE transactions on pattern analysis and machine intelligence*, 42(1):221–231.

[Sawyer and DeZutter, 2009] Sawyer, R. K. and DeZutter, S. (2009). Distributed creativity: How collective creations emerge from collaboration. *Psychology of aesthetics, creativity, and the arts*, 3(2):81.

[Schon and Wiggins, 1992] Schon, D. A. and Wiggins, G. (1992). Kinds of seeing and their functions in designing. *Design studies*, 13(2):135–156.

[Sternberg, 2006] Sternberg, R. J. (2006). The nature of creativity. *Creativity research journal*, 18(1):87.

[Stiny, 2006] Stiny, G. (2006). *Shape: talking about seeing and doing*. MIt Press.

[Stokes et al., 2012] Stokes, M. G., Atherton, K., Patai, E. Z., and Nobre, A. C. (2012). Long-term memory prepares neural activity for perception. *Proceedings of the National Academy of Sciences*, 109(6):E360–E367.

[Stones and Cassidy, 2010] Stones, C. and Cassidy, T. (2010). Seeing and discovering: how do student designers reinterpret sketches and digital marks during graphic design ideation? *Design studies*, 31(5):439–460.

[Summerfield et al., 2011] Summerfield, J. J., Rao, A., Garside, N., and Nobre, A. C. (2011). Biasing perception by spatial long-term memory. *Journal of Neuroscience*, 31(42):14952–14960.

[Sun et al., 2014] Sun, L., Xiang, W., Chai, C., Wang, C., and Huang, Q. (2014). Creative segment: A descriptive theory applied to computer-aided sketching. *Design Studies*, 35(1):54–79.

[Sun et al., 2013] Sun, L., Xiang, W., Chai, C., Wang, C., and Liu, Z. (2013). Impact of text on idea generation: an electroencephalography study. *International Journal of Technology and Design Education*, 23:1047–1062.

[Suwa and Tversky, 1997] Suwa, M. and Tversky, B. (1997). What do architects and students perceive in their design sketches? a protocol analysis. *Design studies*, 18(4):385–403.

[Takala, 1993] Takala, T. (1993). A neuropsychologically based approach to creativity. *Modeling Creativity and Knowledge-Based Creative Design*, 91:108.

[Tirkaz et al., 2012] Tirkaz, C., Yanikoglu, B., and Sezgin, T. M. (2012). Sketched symbol recognition with auto-completion. *Pattern Recognition*, 45(11):3926–3937.

[Torralba, 2003] Torralba, A. (2003). Contextual priming for object detection. *International journal of computer vision*, 53:169–191.

[Tversky, 2009] Tversky, B. (2009). Spatial cognition: Embodied and situated.

[Tversky, 2015] Tversky, B. (2015). Affording design, affording redesign. *Principia Designae-Pre-Design, Design, and Post-Design: Social Motive for the Highly Advanced Technological Society*, pages 91–102.

[Tversky and Hemenway, 1983] Tversky, B. and Hemenway, K. (1983). Categories of environmental scenes. *Cognitive psychology*, 15(1):121–149.

[Van Der Lugt, 2002] Van Der Lugt, R. (2002). Brainsketching and how it differs from brainstorming. *Creativity and innovation management*, 11(1):43–54.

[Verhallen and Mollon, 2016] Verhallen, R. and Mollon, J. (2016). A new mooney test. *Behavior research methods*, 48(4):1546–1559.

[Wagemans et al., 2008] Wagemans, J., De Winter, J., de Beeck, H. O., Ploeger, A., Beckers, T., and Vanroose, P. (2008). Identification of everyday objects on the basis of silhouette and outline versions. *Perception*, 37(2):207–244.

[Wilkinson, 1998] Wilkinson, S. (1998). Focus group methodology: a review. *International journal of social research methodology*, 1(3):181–203.

[Yang, 2009] Yang, M. C. (2009). Observations on concept generation and sketching in engineering design. *Research in Engineering Design*, 20:1–11.

[Zhang et al., 2018] Zhang, J., Chen, Y., Li, L., Fu, H., and Tai, C.-L. (2018). Context-based sketch classification. In *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch-Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering*, pages 1–10.

[Zhao et al., 2021] Zhao, T., Yang, J., Zhang, H., and Siu, K. W. M. (2021). Creative idea generation method based on deep learning technology. *International Journal of Technology and Design Education*, 31:421–440.

Appendix A

CONTEXT WORD LIST

Ancient Greece, Adolescent room, Airport, Herbalist, Alley, Amusement park, Apartment, Aquapark, Arabic Lands, Art studio, Atlantis, Australia, Baby room, Bakery, Ball, Bank, Bar, Basement, Bathroom, Battlefield, Blacksmith, Booth, Business center, Cafe, Camping site, Car park (parking lot), Carnival, Carpenter, Casino, Castle, Cave, Concert hall, Construction site, Council, Court, Craft workshop, Dentist, Desert, Deserted building, Dorm, Dungeon, Egypt, Electronics store, Excavation site, Factory, Far East, Farm, Festival, Garage, Garbage dump, Gas station, Golf course, Graveyard, Hairdresser, Harbor, Hawaii, Hell / Heaven, Highway, Hogwarts, Hospital, Hotel, India, Inn, Island, Istanbul (Constantinople), Italy, Jungle, Kitchen, Laboratory, Living room, London, Mexico, Middle Earth, Military base, Mine, Mountain, Movie house, Movie set, Museum, Music store, Nursing home, Ocean, Office supply store, Palace, Paris, Park, Pawnshop, Pet shop, Pirate ship, Police station, Post office, Printing house, Restaurant, Ritual place, River, Russia, School, Sea, Shopping center, Skating rink, Ski resort, Souvenir shop, Space, Spain, Sports center, Stadium, Suburbs, Supermarket, Swamp, Tailor, Temple, Texas, Theatre, Tower, Train station, Tribal camp, VIP Lounge, Waiting room / Lobby, Prison, Zoo.