Keywords: Co-creativity, AI-based Co-creative System, Ideation, Design.

Abstract: AI-based co-creative design systems enable users to collaborate with an AI agent on open-ended creative tasks during the design process. This paper describes a co-creative system that supports design creativity by providing inspiring design solutions in the initial idea generation process, based on the visual and conceptual similarity to sketches drawn by a designer. The interactive experience allows the user to seek inspiration collaborating with the AI agent as needed. In this paper, we study how the visual and conceptual similarity of the inspiring design from the AI partner influences design ideation by examining the effect on design ideation during a design task. Our findings show that the AI-based stimuli produce ideation outcomes with more variety and novelty when compared to random stimuli.

1 INTRODUCTION

Computational co-creative systems are a growing research area in computational creativity. While some research on computational creativity has a focus on generative creativity (Colton et al., 2012; Gatys et al., 2015; Veale, 2014), co-creative systems focus on computer programs collaborating with humans on a creative task (N. M. Davis, 2013; Hoffman & Weinberg, 2010; Jacob et al., 2013). Co-creative systems have enormous potential since they can be applied to a variety of domains associated with creativity and encourage designers’ creative thinking. Understanding the effect of co-creative systems in the ideation process can aid in the design of co-creative systems and evaluation of the effectiveness of co-creative systems. However, most research on co-creative systems focuses on evaluating the usability and the interactive experience (Karimi et al., 2018) rather than how the co-creative systems influence creativity in the creative process. In this paper we focus on ideation rather than the user experience in order to understand the cognitive effect of AI inspiration.

Ideation, an idea generation process for conceptualizing a design solution, is a key step that can lead a designer to an innovative design solution in the design process. Idea generation is a process that allows designers to explore many different areas of the design solution space (Shah et al., 2003). Ideation has been studied in human design tasks and collaborative tasks in which all participants are human. Collaborative ideation can help people generate more creative ideas by exposing them to ideas different from their own (Chan et al., 2017). Recently, the field of computational creativity began exploring how AI agents can collaborate with humans in a creative process. We posit that a co-creative system can augment the creative process through human-AI collaborative ideation.

We present a co-creative sketching AI partner, the Collaborative Ideation Partner (CIP), that provides inspirational sketches based on the visual and conceptual similarity to sketches drawn by a designer. To generate an inspiring sketch, the AI model of CIP computes the visual similarity based on the vector representations of visual features of the sketches and the conceptual similarity based on the category names of the sketches using two pre-trained word2vec models. The turn-taking interaction between the user and the AI partner is designed to facilitate communication for design ideation. The CIP was developed to support an exploratory study that evaluates the effect of an AI model for visual and

https://orcid.org/0000-0003-3597-6153
conceptual similarity on design ideation in a co-creative design tool.

In this paper, we emphasize the effect of the AI-based inspirations based on the visual and conceptual similarity. The main contributions of this paper to the HCI and co-creativity community are (1) a methodology for evaluating the impact of AI inspiration on ideation and (2) the impact of AI-based visually and conceptually similar designs on ideation.

2 COMPUTATIONAL CO-CREATIVE SYSTEMS

Computational co-creative systems are one of the growing fields in computational creativity that involves human users collaborating with an AI agent to make creative artifacts. The distinction of co-creativity from computational creativity is that co-creativity is a collaboration in which multiple parties contribute to the creative process in a blended manner (Mamykina et al., 2002). Co-creative systems have been applied in different creative domains such as art, music, dance, drawing, and game design. Some co-creative systems directly perform actions on a shared artifact or contribute to a performance whereas others provide suggestions to inspire users for generating novel ideas. This distinguishes how a co-creative AI agent contributes to the creative process. One co-creative interaction paradigm is an AI agent performing actions with a user simultaneously. Shimon (Hoffman & Weinberg, 2010) is a robotic marimba player that listens and responds to a musician in real time. This improvisational robotic musician performs accompaniment with the users’ musical performance simultaneously. Another co-creative interaction paradigm is a turn-taking action between a user and an AI agent in a shared artifact. Drawing Apprentice (N. Davis et al., 2015) is a co-creative drawing system in which the computational partner analyzes the user’s sketch and responds to the user’s sketch. Viewpoints AI (VAI) is a co-creative dance partner that analyzes the user’s dance gestures and provides complimentary dance in real-time by a virtual character projected on a large display screen (Jacob et al., 2013). These co-creative interaction paradigms are examples of an AI agent participating in a creative activity by performing the same type of action as a user. Another co-creative interaction paradigm is providing suggestions to the user. Sentient Sketchbook (Yannakakis et al., 2014) and 3Buddy (Lucas & Martinho, 2017) are co-creative systems for game level design. In both systems, the AI agent provides feedback and additional ideas to develop the game design rather than creating game level directly.

3 THE COLLABORATIVE IDEATION PARTNER (CIP)

The Collaborative Ideation Partner (CIP) as a co-creative design system builds on previous projects (Karimi et al., 2019, 2020) that interpret sketches drawn by a user and provides inspirational sketches based on visual similarity and conceptual similarity. We developed the CIP to explore the effect of an AI model for visual and conceptual similarity on design ideation in a co-creative design tool.

The user interface of CIP is shown in Figure 1. There are two main spaces in the CIP interface: the drawing space (pink area) and the inspiring sketch space (purple area). The drawing space consists of a design task statement, undo button, clear button, and user’s canvas. The design task statement in the drawing space includes the object to be designed as well as a context to further specify the objects’ use and environment. The user can draw a sketch in the drawing space and edit the sketch using the undo and clear button. The inspiring sketch space includes an “inspire me” button, the name of the inspiring object, and a space for presenting the AI partner’s sketch. When the user clicks the “inspire me” button after sketching their design concept, the AI partner provides an inspiring sketch based on visual and conceptual similarity. An ideation process using CIP involves turn-taking communications between the user and the AI partner. Another part of the CIP interface in addition to the two main spaces is the top area (grey area) including a hamburger menu and an introductory statement. The hamburger menu on the top-left corner of the interface includes four design tasks (i.e. sink, bed, table, chair) and allows the experiment facilitator to select one of the design tasks. Each design task provides different categories of ideation stimuli.

Figure 1 shows an example of an inspiring sketch and how participants communicate with an inspiring sketch to develop their design. The design task shown in Figure 1 is to design a chair for a gaming computer desk. The participant drew a basic chair with back, seat, legs, and small wheels before requesting inspiration from the AI partner. The sketch suggested from the AI models is a bulldozer: visually similar and conceptually different to the participant’s sketch. After getting the inspiring sketch, the participant
made the wheels much bigger for better mobility and added a leg rest for comfort. During the retrospective protocol, the participant described that “I decided to go with bigger wheels here, just thinking of bulldozer, little more heavy duty. I mean, I also noticed the little lift gate or whatever that is. And that kind of made me think that I needed to add like some kind of leg support and that kind of made sense.”

3.1 Dataset

For the source of inspiring sketches, CIP uses a public benchmark dataset called QuickDraw! (Jongejan et al., 2016), which was created during an online game where players were asked to draw a particular object within 20 seconds. The dataset includes 345 categories with more than 50 million labelled sketches, where sketches are the array of the x and y coordinates of the strokes. The system uses the simplified drawing json files that use Ramer–Douglas–Peucker algorithm (Douglas & Peucker, 1973; Ramer, 1972) to simplify the strokes, and position and scale the sketches into a 256 X 256 region. The stroke data associated with these sketches are used to calculate the visual similarity and the corresponding category names are used to measure the conceptual similarity.

3.2 AI Models for Visual and Conceptual Similarity

The CIP has 2 distinct components for measuring similarity between the user’s sketch and the sketches in the dataset: one component for calculating visual similarity and another component for calculating conceptual similarity. The visual similarity component selects sketches from the sketch dataset based on a representation of the stroke data in the image file. The conceptual similarity component computes the degree of similarity between the category names of the objects in design tasks and the category names in the objects in the sketch dataset.

For the visual similarity component, we used a pre-trained CNN-LSTM model from the precedent with 3 convolutional layers, 2 LSTM layers, and a softmax output layer on the QuickDraw dataset (Karimi et al., 2019, 2020). For the conceptual similarity component, we considered sketch category names in the QuickDraw dataset as the concepts of the sketches that contain 345 unique categories. We used two pre-trained word2vec models, Google News (Mikolov et al., 2013) and Wikipedia (Rehurek & Sojka, 2010), and calculated cosine similarities for measuring the conceptual similarities between the object categories of the design tasks and the categories of inspiring sketches from the dataset. For each category of the design tasks, we generated two sorted lists of conceptually similar category names, one for each word2vec model, and then used human judgement to compare the sorted lists and select the top 15 common conceptually similar category names that appear in both lists. This final step of using human judgement improved the alignment between the conceptual similarities of AI models and human perception. The conceptual similarity component of CIP uses the common list of category names for sorting the sketches based on the conceptual similarities.

We use these two AI-based components of the CIP to generate sequences of sketches with combinations of visual and conceptual similarity to...
the user’s current sketch and design task to inspire the user during their design process and measure the effect of visual and conceptual similarities on ideation.

3.3 AI-based Inspiration in CIP

To support an exploratory study that measures ideation when co-creating with CIP, the interaction with CIP has four distinct modes of inspiration that vary the visual and conceptual similarity. Each of the four modes appears as a design task (i.e. sink, bed, table, chair) in the CIP interface. One of the modes (i.e. sink) uses a random sketch selection while three other modes use AI models to select an inspiring sketch as inspiration in CIP.

- **Random**: Inspire with a random sketch (sink): The CIP selects a sketch randomly from the sketch dataset to be displayed on the AI partner’s canvas.

- **Similar**: Inspire with a visually and conceptually similar sketch (bed): The CIP selects a sketch from a set of sketches where each one is similar visually and conceptually to the user’s sketch (e.g. user sketch - a bed, AI sketch - a similar shape of bed to the user’s sketch).

- **Conceptually Similar**: Inspire with a conceptually similar and visually different sketch (table): The CIP selects a sketch from a set of sketches where each one is conceptually similar but visually different to the user’s sketch (e.g. user sketch - a square table, AI sketch - a round table).

- **Visually Similar**: Inspire with a visually similar and conceptually different sketch (chair): The CIP selects a sketch from a set of sketches where each one is visually similar but conceptually different to the user’s sketch (e.g. user sketch - a circular chair back, AI sketch - a face).

4 EXPLORATORY STUDY

The goal of the exploratory study is to explore the effect of AI inspiration on ideation through an analysis of the correlation between conceptual and visual similarity with characteristics of ideation. Specifically, we are interested in the relationship between the users’ ideation and sources of AI inspiration.

4.1 Study Design

The type of study is a mixed design of between-subject and within-subject design. There are 3 groups of within-subject design (i.e. A&B, A&C, A&D) in this study and each group has a control condition (i.e. condition A) and one of 3 treatment conditions (i.e. condition B, C, D). The control condition (condition A) for each group is the same but the treatment condition for each group is different (condition B or C or D). The control condition and 3 treatment conditions are the different types of inspirations presented in Section 3.3:

- **Condition A (control condition)**: randomly (sink)
- **Condition B (treatment condition)**: visually and conceptually similar (bed)
- **Condition C (treatment condition)**: conceptually similar and visually different (table)
- **Condition D (treatment condition)**: visually similar and conceptually different (chair)

The protocol including the informed consent document has been reviewed and approved by our IRB and we obtained informed consent from all participants to conduct the experiment. We recruited 12 students from human-centered design courses for the participants: each participant engaged in 2 conditions: a control condition and one of the treatment conditions, with 4 participants for each of the 3 groups of within-subject design (i.e. A&B, A&C, A&D). The experiment is a mixed design with N=4 and a total of 12 participants.

The task is an open-end design task in which participants were asked to design an object in a given context through sketching. Different objects for the design task were used for each condition: a sink for an accessible bathroom (condition A), a bed for a senior living facility (condition B), a table for a tinkering studio, a collaborative space for designing, making, building, etc. (condition C), a chair for a gaming computer desk (condition D).

The procedure consists of a training session, two design task sessions, and two retrospective protocol sessions. In the training session, the participants are given an introduction to the features of the CIP interface and how they work to enable the AI partner to provide inspiration during their design task. After the training session, the participants perform two design tasks in a control condition and a treatment condition. The study used a counterbalanced order for the two design tasks. The participants were given as
much time as needed to perform the design task until they were satisfied with their design. The participants are free to click the “inspire me” button as many times as they would like to get inspiration from the system. However, the participants were told to have at least 3 inspirational sketches (i.e. clicking the “inspire me” button at least 3 times during a design session), a minimum number of inspirations, from the system. Once the participants finish the two design task sessions, the participants are asked to explain what they were thinking while watching their design session recording as time goes on, and how the AI’s sketches inspired their design in the retrospective protocol session.

4.2 Data Collected

Two types of data were collected for analyzing the study results: a set of sketches that participants produced during the design tasks and the verbalization of the ideation process during the retrospective protocol. We recorded the entire design task sessions and retrospective sessions for each participant. The sketch data collected from the recordings of design task sessions shows the progress of design and the final design visually for each design task session. The verbal data collected from the recordings of retrospective sessions records how the participants came up with ideas collaborating with the AI partner and applied the ideas to their design.

4.3 Data Segmentation and Coding

To analyze the verbal data collected from the retrospective sessions, we adapted the FBS coding scheme for characterizing cognitive issues during a design process (Gero, 1990; Gero & Kannengiesser, 2004). An idea can be variously defined as a contribution that contains task-related information, a solution in the form of a verb–object combination, and a specific benefit or difficulty related to the task (Reinig et al., 2007). The FBS coding scheme provides a segmentation into individual ideas associated with specific cognitive issues in design. First, the verbal data of all retrospective protocol sessions was transcribed. The transcripts were segmented based on the inspiring sketches the participant clicked. A segment starts with an inspiring sketch and ends when the inspiration is clicked for the next sketch. To identify each idea in an inspiring sketch segment, we segmented the inspiring segments again based on FBS ontology (Gero, 1990; Gero & Kannengiesser, 2004) as an idea segment, since an inspiring sketch segment includes multiple ideas. The idea segments were coded based on FBS ontology (Gero, 1990; Gero & Kannengiesser, 2004) as requirement (R), function (F), expected behavior (Be), behavior from structure (Bs), and structure (S). A segment coded R is an utterance that talks about the given requirement in the statement of design task (e.g. accessible bathroom); a segment coded F is an utterance that talks about a purpose or a function of the design object (e.g. more accessible); a segment coded Be is an utterance that talks about an expected behaviors from the structure (e.g. water could automatically come out); a segment coded Bs is an utterance that talks about a behavior derived from the structure (e.g. pressing on); a segment coded S is an utterance that talks about a component of the design object (e.g. button). The result of this coding scheme is a segmentation of the verbal protocol into individual ideas, each associated with one code: R, F, Be, Bs, S.

Two coders coded the idea segments individually based on the coding scheme above then came to consensus for the different coding results. The coding instruction was given to the coders included how to segment inspiring sketch segments and idea segments, how to code each idea segment with the coding scheme, and how to code new and repeated ideas. The two coders coded a design session together to make an initial agreement for segmentation and coding before coding individually then coded all design sessions individually. Once each coder completed coding all data individually, the two coders discussed each of the different coding results and came to consensus.

4.4 Analysis of Exploratory Study: Measuring Ideation

To evaluate the effect of AI inspiration on ideation, we adapted the metrics from Shah et al. (2003) for measuring ideas in a design process. We applied four types of metrics for measuring ideation effectiveness, used for evaluating idea generation in design: novelty, variety, quality, and quantity of design ideas. We developed the four metrics based on (Shah et al., 2003) to analyze the coded data of the retrospective protocol session.

Novelty. Novelty is a measure of how unusual or unexpected an idea is as compared to other ideas (Shah et al., 2003). In this study, a novel idea is defined as a unique idea across all design sessions in a condition. For measuring novelty, we counted how many novel ideas in the entire collection of ideas in a
design session (personal level of novelty) and a condition (condition level of novelty). We removed the same ideas across all design sessions in a condition then counted the number of ideas.

The results showed that all treatment conditions (B, C, D) have more novel ideas than the control condition (A) in the total number of novel ideas. Specifically, 10 participants out of 12 participants produced more novel ideas in a treatment condition than the control condition. When comparing the novelty of 3 groups, the group A&C showed the largest difference between the control condition and the treatment condition where condition C selected inspiring sketches that are conceptually similar and visually different. As shown in Figure 2, all participants in the group of A&C produced more novel ideas in the condition C than the condition A while one of the participants (i.e., P4) in the group A&B and one of participants (i.e., P9) in the group A&D produced fewer novel ideas in the treatment condition than the control condition. This result can indicate that the conceptual similarity of inspiring sketches may be associated with the novelty of ideas in the ideation with CIP.

Variety. Variety is a measure of the explored solution space during the idea generation process (Shah et al., 2003). Each idea segment was coded whether it is a new idea or a repeated idea in a design session. For measuring variety in this study, only the number of new ideas coded as R/F/B/S is counted in a design session while the metric of quantity includes both new ideas and repeated ideas.

The results showed that the variety of ideas in condition C is higher than in condition A. Figure 3 shows the results of codes comparing the control condition (A) and one of the treatment conditions (C). The results of the group A&C show some distinct patterns in function. All participants produced more functions in condition C than in condition A. The number of function ideas showed a large difference for all participants between condition A and C. This result indicates that the conceptual similarity inspired the participants to produce more various functions associated with the context of the design.

Quality. Quality is a subjective measure of the design (Shah et al., 2003). In this study, quality is measured using the Consensual Assessment Technique (CAT) (Amabile, 1982), a method in which a panel of expert judges is asked to rate the creativity of projects. Two judges, researchers involved in this study, individually evaluated the final design in each condition as low/medium/high quality, in two evaluation rounds. In the first-round of evaluation, each judge evaluated the final designs identifying some criteria for evaluating the quality of ideas. Once the judges finished the first-round of evaluation, they shared the criteria they identified/used, not sharing the results of the evaluation, then made a consensus for the criteria that will be used for the second-round evaluation. The criteria that the judges agreed for evaluating the quality of ideas in this study are the number of features, how responsive the features are to the specific task, how creative the design is. In the second-round evaluation, each judge evaluated the final design again using the agreed criteria.

Table 1: Quality evaluation results of each judge in the group of A&D.

<table>
<thead>
<tr>
<th></th>
<th>Condition A</th>
<th>Condition D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>Judge 2</td>
<td>Judge 1</td>
</tr>
<tr>
<td>P3</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>P5</td>
<td>low</td>
<td>medium</td>
</tr>
<tr>
<td>P9</td>
<td>medium</td>
<td>medium</td>
</tr>
<tr>
<td>P12</td>
<td>low</td>
<td>low</td>
</tr>
</tbody>
</table>

The results showed that the quality of ideas in condition D is higher than in condition A, where condition D selects sketches that are visually similar
and conceptually different for inspiration. Table 1 shows the result of the quality evaluation that each judge made for each design in condition A and condition D. Three out of four participants produced higher quality in condition D than condition A. P3 produced much higher quality in condition D than condition A (i.e. low to high). P5 and P9 produced higher quality in condition D than in condition A (i.e. P5: low to medium, P9: medium to high). This result indicates that the visual similarity of inspiring sketches may be associated with the quality of ideas in the ideation with CIP.

**Quantity.** Quantity is the total number of ideas generated (Shah et al., 2003). For measuring quantity in this study, the number of ideas both new ideas and repeated ideas coded as R/F/B/S is counted in a design.

![Figure 4: Quantity of ideas in the group of A&C.](image)

Figure 4 shows the results of the quantity of ideas in the group of A&C. The results show a similar pattern to the result of variety with some distinct patterns. First, for the total number of ideas, 3 out of 4 participants (i.e. P2, P8, P6) generated more ideas in condition C than in condition A. Second, 3 out of 4 participants (i.e. P2, P8, P6) generated more ideas of F (function) and S (structure) in condition C than in condition A. This result indicates that the conceptual similarity of inspiring sketches facilitates producing new functions and the emerging functions were transferred to structures of the design.

Our exploratory study does not have a sufficient number of participants to allow us to generalize the results for all cases of ideation from AI-based visual and conceptual similarity. However, we did a significance test on the results to see if there are significant trends to look for in a more robust study. A paired t-test was conducted to determine the significance of our results between the control condition and the treatment conditions in novelty, variety, and quantity. The results showed a significant difference in variety and quantity. For variety, participants in condition C (M=24.25, SD=10.21) produced more functions than in condition A (M=15.50, SD=9.81), t(3)=−5.14, two tail p=0.014253. For quantity, participants in condition B (M=28.75, SD=12.76) produced more functions than in condition A (M=19.00, SD=13.24), t(3)=−3.30, two tail p=0.045732. This exploratory study does not have enough participants to measure or check for statistical significance, but the trends of the results show the potential for further analysis of the effect of an AI model for visual and conceptual similarity on design ideation with the metrics we identified for measuring ideation.

### 5 DISCUSSION

In this paper, we presented a co-creative design system called CIP and an exploratory study that explores the effect of an AI model for visual and conceptual similarity on design ideation in a co-creative design tool. To evaluate the effect of AI inspiration on ideation, we applied four metrics (i.e. novelty, variety, quality, quantity) to measure the ideation in an exploratory study. Overall our findings show that the AI-based stimuli produce different ideation outcomes when compared to random stimuli. More specifically, we found that different types of AI-based stimuli show potential for different types of ideation. Novel ideation is associated with AI-based conceptually similar stimuli. Idea variety and quantity is associated with both AI-based visual and conceptual similarity of the inspiration. Idea quality is associated with visual similarity.

In addition to measuring ideation, we observed the video stream data to see how participants develop their design ideas communicating with the inspirations. The participants' responses to inspirations showed different patterns of users on the use of CIP in an ideation process. In an evolution of the participant’s sketch, participants in each condition start with a basic shape of the target design then develop the design with inspiration from the AI partner. Participants explored many inspiring sketches in condition A but did not have many design changes; while participants in conditions B, C, and D developed their design in response to fewer inspiring sketches. This observation suggests further analysis of ideation to understand the cognitive process of ideation when co-creating with the CIP.
6 CONCLUSION

This paper presents a co-creative design tool called Collaborative Ideation Partner (CIP) that supports idea generation for new designs with stimuli that vary in similarity to the user’s design in two dimensions: conceptual and visual similarity. The AI models for measuring similarity in the CIP use deep learning models as a latent space representation and similarity metrics for comparison to the user’s sketch or design concept. The interactive experience allows the user to seek inspiration when desired. To study the impact of varying levels of visual and conceptual similar stimuli, we performed an exploratory study with four conditions for the AI inspiration: random, high visual and conceptual similarity, high conceptual similarity with low visual similarity, and high visual similarity with low conceptual similarity. To evaluate the effect of AI inspiration, we evaluated the ideation with CIP using the metrics of novelty, variety, quality and quantity of ideas. We found that conceptually similar inspiration that does not have strong visual similarity leads to more novelty, variety, and quality during ideation. We found that visually similar inspiration that does not have strong conceptual similarity leads to more quality ideas during ideation. Future AI-based co-creativity can be more intentional by contributing inspiration to improve novelty and quality, the basic characteristics of creativity.

REFERENCES


