IDETC2023-115108

SIMILARITIES AND DIFFERENCES IN HUMAN VS. COMPUTATIONAL REPRESENTATIONS OF NON-SEMANTIC INSPIRATIONAL DESIGN STIMULI

Elisa Kwon
Department of Mechanical Engineering
University of California, Berkeley
Berkeley, CA
Email: elisa.kwon@berkeley.edu

Kosa Goucher-Lambert
Dept. of Mechanical Engineering
University of California, Berkeley
Berkeley, CA
Email: kosa@berkeley.edu

ABSTRACT

As inspirational stimuli can assist designers with achieving enhanced design outcomes, supporting the retrieval of impactful sources of inspiration is important. Existing methods facilitating this retrieval have relied mostly on semantic relationships, e.g., analogical distances. Increasingly, data-driven methods can be leveraged to represent diverse stimuli in terms of multi-modal information, enabling designers to access stimuli in terms of less explored, non-text-based relationships. Toward improved retrieval of multi-modal representations of inspirational stimuli, this work compares human-evaluated and computationally derived similarities between stimuli in terms of non-text-based visual and functional features. A human subjects study (n=36) was conducted where similarity assessments between triplets of 3D-model parts were collected and used to construct psychological embedding spaces. Distances between unique part embeddings were used to represent similarities in terms of visual and functional features. Obtained distances were compared with computed distances between embeddings of the same stimuli generated using artificial intelligence (AI)-based deep-learning approaches. When used to assess similarity in appearance and function, these representations were found to be largely consistent, with highest agreement found when assessing pairs of stimuli with low similarity. Alignment between models was otherwise lower when identifying the same pairs of stimuli with higher levels of similarity. Importantly, qualitative data also revealed insights regarding how humans made similarity assessments, including more abstract information not captured using AI-based approaches. Toward providing inspiration to designers that considers design problems, ideas, and solutions in terms of non-text-based relationships, further exploration of how these relationships are represented and evaluated is encouraged.

Keywords: Inspirational stimuli, Data-driven design, Design representation

1. INTRODUCTION

To support early-stage design processes, external sources of inspiration can assist designers toward achieving desirable outcomes. The impact of inspirational stimuli, most notably analogies, on design processes has been well studied due to their potential to retrieve relevant concepts from long-term memory and aid conceptual design [1]. By studying these processes, features of inspirational stimuli that can lead to beneficial outcomes such as increased novelty, feasibility, or innovativeness of ideas (e.g., [2–4]) can be determined. As methods to retrieve inspirational stimuli for designers rely increasingly on large datasets and data-driven techniques [5], improved understanding of how to discover and select relevant stimuli is needed. In particular, Jiang et al. in a recent review on data-driven design-by-analogy (DbA) propose that, beyond textual data, other modalities such as visual information (2D-image or 3D-model datasets) should be utilized to support visual or multi-modal DbA [5].

One implementation of data-driven methods for the retrieval of multi-modal inspirational stimuli was explored in prior work from our team [6]. In this work, deep-neural networks modeling visual and functional relationships between 3D-model parts were used in a multi-modal search platform for inspiration discovery. In two subsequent user studies, designers using this system searched for stimuli in terms of appearance and function-based similarities to a specified input and were frequently returned results they did not expect [7]. These findings motivate the aim of the present study to improve representation of non-text-based measures of similarity, which have not been widely studied in interactive settings. Currently, there is limited knowledge regarding how relationships defined in terms of non-textual properties of inspirational stimuli align with human representations. Increased availability, interest, and use of 2D-image and 3D-model datasets encourage the development of tools enabling discovery of design stimuli related to an input by *non-text-based features* rather

than semantic distances. Our approach is to compare human and AI-based representations of non-text-based definitions of similarity to increase understanding of these less explored measures of similarity. Representations of similarity are compared through one central research aim: How *consistent* and *aligned* are human and AI-based representations of non-text-based similarities of inspirational stimuli?

This research question is studied considering 3D-model parts as a source of design relevant inspiration, which contain both visual and functional information. A human subjects study was conducted (n=36) where human assessments of similarity of 3D-model parts by visual and functional features were collected in a triplet rating task. The alignment and consistency across human and computational representations of visual and functional similarities are evaluated. Additional insight into a comparison of these representations is revealed through qualitative analysis. Understanding how computational methods may agree with and differ from human representations of non-text-based relationships between inspirational stimuli can support the effective retrieval of relevant and impactful sources of inspiration for designers based on non-textual information.

2. RELATED WORKS

Motivating the study of multi-modal representations of similarity in this paper, prior work on multi-modal inspirational design stimuli and methods and tools enabling their retrieval is reviewed.

2.1 Impact of Multi-Modal Inspirational Stimuli

The modality in which a stimulus is represented to designers can differently influence design outcomes. Several examples of past work have investigated designer's interactions with multi-modal stimuli. Borgianni et al. studied the impact of stimulus form on idea generation by presenting textual, pictorial, or combined stimuli [8, 9]. Findings encourage the presentation of multiple forms of stimuli to designers due to the diversity and limited overlap of ideas generated by participants exposed to different forms of stimuli. Designers tend to prefer visual information [10, 11], which Linsey et al. found can lead to increased idea novelty [12]. Han et al. suggest that images combined with unrelated semantic elements can promote creative idea generation [13], while using pictorial stimuli was found by Malaga et al. to outperform the use of words alone for enhanced creativity of ideas [14]. In general, interacting with visual stimuli can importantly trigger formation of new mental images, which can support generation of new design ideas [15]. These studies demonstrate the value of providing multi-modal, e.g., pictorial, stimuli to designers. In the current study, 3D-model parts are proposed as another form of inspirational stimuli containing both visual and functional attributes. Human and computational representations of relationships between these parts are investigated to support designers' interactions with these stimuli. In the next section, methods enabling their representation and retrieval are explored.

2.2 Enabling Retrieval of Inspirational Stimuli

To provide designers with relevant sources of inspiration, similarity relationships between designer inputs and potential stimuli need to be defined. Defining similarities to support data-driven DbA has been most widely studied in the context of deriving analogical distances between source and target domains [5]. Computational methods are often used to retrieve design stimuli with varying analogical distances to a given design problem or designer specified input. Similarity relationships specifically relying on textual information can be derived. For instance, text-based processing has been used to define function-based similarity between design problems and solutions from patents [16], to define contextual similarity between patents [3, 17], or to assign function-based topics to patents based on different semantic themes [18, 19]. Semantic networks used during engineering design activities can facilitate exploration and retrieval of analogies consisting of common words, such as in WordNet or ConceptNet [20], or technology-based knowledge from patent texts in the Technology Semantic Network (TechNet) [21].

However, beyond processing of textual information, there is increasing interest in using AI to represent and retrieve stimuli from 2D-image and 3D-model datasets [5]. These stimuli can support *multi-modal analogy* for design inspiration. Sketch-based retrieval of visually similar examples can importantly support visual analogy [22, 23]. Zhang and Jin used an unsupervised deep-learning model to construct a latent space for a dataset of sketches [23]. Image-based search using visual similarity can also extract relevant examples from sources such as patent documents [24, 25]. Jiang et al. constructed a convolutional neural network-based model to derive a vector space where feature vectors embed visual and technology-related information from patent images [25]. Other sketch-based user interfaces include DreamSketch, which provides designers with 3D-modeled design solutions based on early stage 2D-sketch-based designs [26] or SketchSoup, which inputs rough sketches and generates new sets of sketches to inspire further concept generation [27]. Design ideas represented in 3D can be recognized by tools such as the InspireMe interface, which provides suggestions for new components to add to a designer's initial 3D model [28]. Kim et al. developed a co-creative sketching artificial-intelligence (AI) partner that provides inspirational sketches related by visual and conceptual similarity to designer-drawn sketches [29]. The effects of providing sketches with varying levels of visual and conceptual similarity to the designer's sketch were investigated [30]. In our prior work, deep learning was applied to develop deep-neural networks modeling visual and functional relationships between 3D-model parts in a large dataset [6]. These neural networks were used to construct a multi-modal search platform, through which designers' search for inspiration was examined.

Using AI to represent multi-modal stimuli in terms of non-text-based features can increase their utilization as sources of design inspiration. As well, designers' interactions with multi-modal inputs (e.g., sketch or 3D-model based) can be better enabled. Ensuring that computational methods used to define non-text-based relationships appropriately represent how humans perceive these similarities is the primary aim motivating this work. This aim is achieved by conducting a human subjects study, as described in the following section, to model human representations of visual and functional similarities between inspirational stimuli and by performing a comparison with AI-based representations.

3. METHODS

To investigate the representation of inspirational stimuli in terms of multi-modal information, visual and functional similarities between 3D-model parts are explored in this work. Similarity is described by distances between stimuli in embedding spaces derived using two approaches. The first approach uses deep learning to construct neural networks modeling these relationships, resulting in *computational embedding spaces* for a large dataset of 3D-model parts (developed in prior work [6]). Presented in the current work, the second approach uses human-evaluated similarities of a selection of 3D-model parts to build *psychological embedding spaces* of parts. A human subjects study (n=36) was conducted to collect similarity assessments used to develop the psychological embedding spaces for visual and functional relationships between 3D-model parts. Methods used to conduct the study, construct the psychological embedding spaces, and analyze post-task qualitative data are described in this section.

3.1 Experimental Design

This study consisted of two main tasks: a triplet rating task and a categorization task, each completed twice (for each similarity type explored). For one similarity type, participants completed 25 trials of the triplet rating task followed by the categorization task. The same two tasks were then repeated for the other stimulus set. The order of similarity type (visual or functional) presented was counterbalanced across participants. Results of the categorization task are not reported in the present work. After completing each set of 25 triplet ratings, participants were additionally asked to provide open-ended responses describing the specific criteria used to assess visual or functional similarity. Experimental details of the triplet-rating task are fully described in Sec. 3.1.3. To determine which stimuli to present in these tasks, two distinct sets of 16 3D-model parts were selected from

the computational embedding spaces with varying pairwise distances in either visual and functional similarity.

3.1.1 Participants. For this study, 36 participants (13 female, 22 male, 1 non-binary) were recruited including 14 graduate students, 16 undergraduate students, and 6 industry professionals (with <1 to 9 years of experience). In prior work from the authors, any impact of expertise when engaging with inspirational stimuli was in their utilization in a structured design task (not relevant to the current study) [31]. For the tasks completed, no particular level of engineering design knowledge or experience was required and no analysis of differences in expertise was conducted. Participants were recruited via email from among current students in Mechanical Engineering as well as participants who previously completed research studies related to engineering design. Participants were compensated with $10 for their completion of the 30 minute study. This human subjects research study has been approved by the Institutional Review Board at the University of California, Berkeley.

3.1.2 Selection of Task Stimuli. The stimulus sets provided to participants in the study (see Fig. 1) were selected by considering distances between 3D-model parts in the deep-learning-based computational embedding spaces. These neural networks were trained on 573,585 part instances belonging to 26,671 3D-model object assemblies across 24 object categories. To encode visual similarity of 3D-model parts, the deep-learning model takes 2D snapshots from various angles of each part to understand its geometric and physical form. The functional network is developed by considering neighboring parts within a part's respective object assembly such that two parts are similar if they share similar neighbors (e.g., a chair leg and back are functionally similar because a chair seat is a common neighbor). These similarity definitions exclude the use of semantic information of
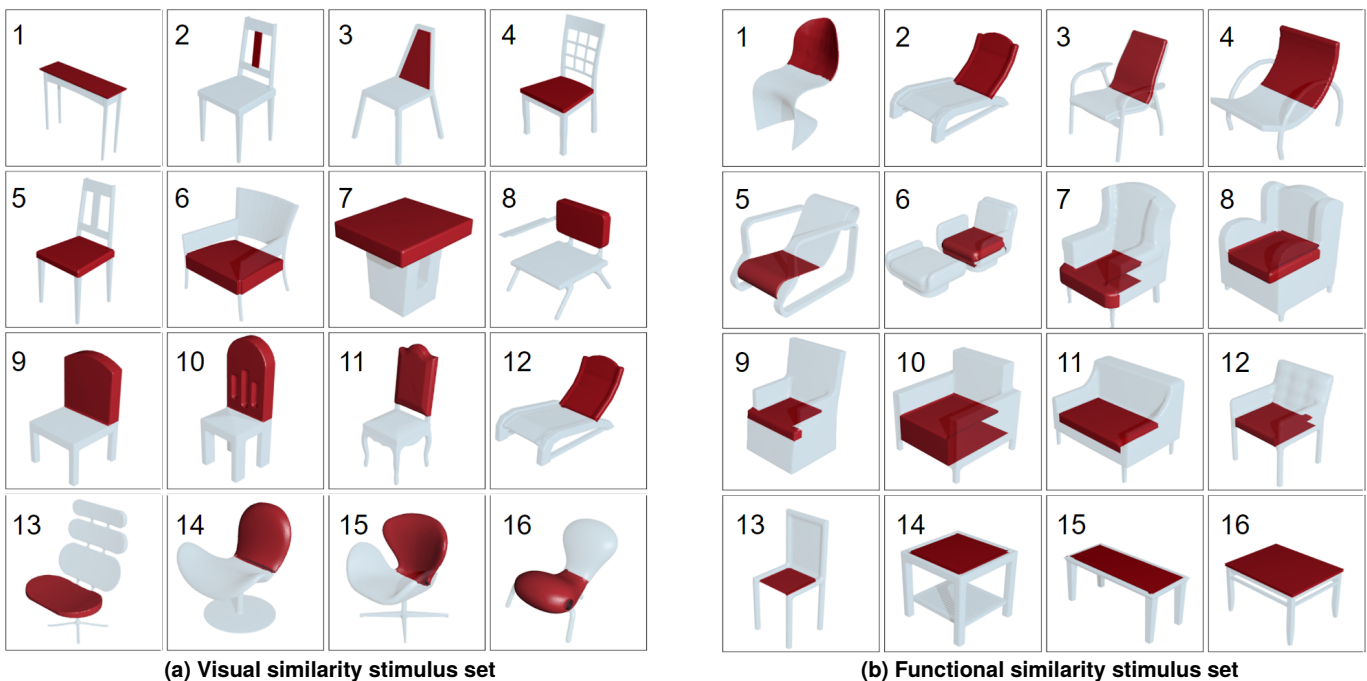

(a) Visual similarity stimulus set   (b) Functional similarity stimulus set

FIGURE 1: STIMULI PRESENTED DURING TRIPLET RATING TASK

individual parts. The development of these neural networks is fully described in our past work [6].

Given the size and diversity of the full dataset, candidate stimuli were restricted to "chair" and "table" object categories, specifically considering chair seats, chair backs, and tabletops (as labelled within the PartNet dataset [32]), resulting in 2043 possible parts. This was done to reduce the potential difficulty of rating similarity between and categorizing very diverse objects (e.g., bottles and tables). Although task complexity is reduced as a result, ultimately, the aim of this selection of stimuli was to encourage the assessment of similarity in terms of visual and functional features only. This aim could be better achieved without the influence of semantic information, including product category. Potential limitations of the present findings related to stimuli selection are discussed in Sec. 5.2.

The full 16-part stimulus sets selected to present in the triplet rating and categorization tasks are shown in Fig.1, chosen based on visual similarity in Fig.1a and functional similarity in Fig.1b. Euclidean distances between parts in the computational embedding spaces are used to represent how similar (low distance) or dissimilar (high distance) parts are. While distances between neighbors are not constant, neighboring parts (e.g., 1 and 2) are always nearer in terms of pairwise distance than non-neighboring parts (e.g., 1 and 5). By maintaining consistency in pairwise distances, we ensure that the stimulus sets used contain a diversity of distances where all parts belong to both low and high distance pairs.

3.1.3 Triplet Rating Task. Developing a psychological embedding space that models human representations of a given stimulus set requires the collection of many trials of human judgments. A common task used to elicit these judgments is a triplet rating task where one of two options is selected as being more similar to a given reference. Prior work by Nandy & Goucher-Lambert and Ahmed et al. have also used triplet similarity ratings to generate embedding spaces for human representations of design stimuli [33, 34]. Preceding each triplet rating task of 25 trials, participants were told that "In the [first/second] section of this study, you will consider the [function/appearance] of parts when assessing similarity". At the beginning of each trial, participants were asked to "Select the option with the most similar part in [function/appearance] to the reference part" with two options presented, such as in the example shown in Fig. 2. Participants were instructed to make this selection based on the red-highlighted 3D-model part in the object assembly. When considering functional similarity, participants were told to consider the object the red part belongs to, other neighboring parts in the object, and that parts with high functional similarity may be used in the same object and/or neighbor similar parts. For visual similarity trials, no further detail was provided.

For the number of parts in each stimulus set (16), a total of 1680 unique triplet trials are possible. Ahmed et al. recommend that a minimum of 30% of the full stimulus set is needed to construct a robust embedding space of human representations [34]. In our study, 36 participants completed 25 triplet ratings for each stimulus set. Due to data collection errors and exclusion of data from one participant who failed the attention check for the visual similarity triplet rating task, a total of 801 trials for visual
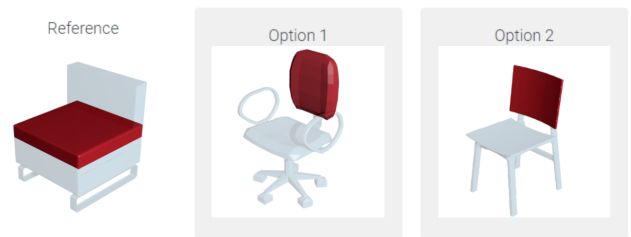

FIGURE 2: EXAMPLE TRIPLET OF 3D-MODEL PARTS SHOWN TO PARTICIPANTS

similarity and 826 trials for functional similarity were included, constituting 48% and 49% of all potential trials.

3.2 Construction of Psychological Embedding Spaces

Using outcomes from the triplet rating tasks, psychological embedding spaces were constructed. These models include two layers: an embedding layer representing multidimensional features, and a similarity kernel. The Python library PsiZ was used to generate these models, which specifically handles behavioral data such as triplet ratings to infer psychological embeddings (https://github.com/psiz-org/psiz). The similarity kernel consists of a distance function (weighted Minkowski distance) and a similarity function (exponential decay in similarity with increased distance). The use of this two-component kernel is motivated by psychological theory and has been used to successfully represent psychological embeddings [35]. Data was split into a training set (80% of trials), test set (10%) and validation set (10%). The number of dimensions for each model was determined by training models with dimensions varying from two to ten. The highest value at which validation set losses stopped improving for increasing values of dimensionality was selected. The final psychological embedding spaces for both visual and functional similarity are two dimensional with consistent training/validation/test set losses of 0.45/0.51/0.45 and 0.39/0.43/0.49, respectively. Constructing these embedding spaces importantly enables the measurement of distances between stimuli in terms of human representations for comparison against computational representations of visual and functional attributes.

3.3 Analysis of Qualitative Data

Following the completion of 25 trials of triplet ratings of stimuli based on visual and functional similarity, participants provided written open-ended responses to describe the criteria they used to assess similarity. While evaluation criteria used when employing deep learning can be speculated, exact definitions for each dimension of these models are unknown. However, deeper insight can be gained regarding how humans represent visual and functional information through qualitative post-task data obtained. This analysis may help to inform future deployment of computational methods to represent relationships based on multimodal information by understanding the features of inspirational stimuli emphasized using each method. Criteria used across participants to evaluate visual and functional similarity were coded from the open-ended responses provided following an inductive category formation approach [36]. Using this method of qualitative content analysis, criteria were defined to code responses

where new criteria were formed if responses could not be subsumed under previously defined criteria. Multiple criteria from a participant's response could be assigned to the same codes. This process continued for all responses collected from all 36 participants and repeated for both similarity types evaluated.

Two coders, mechanical engineering graduate students each with at least one publication in an engineering design journal, coded the full dataset. Coder 1 manually coded all responses using the described inductive category formation process. Coder 2 then validated the assignment of responses to the criteria identified by Coder 1 independently. Across both coding processes, 0.98 Cohen's Kappa interrator reliability was achieved for visual similarity and 0.82 for functional similarity, suggesting high consistency between coders [37]. Differences in assigned codes were discussed and resolved between coders. Details of both coding outcomes are described in Sec. 4.3 to identify human evaluation criteria used for visual and functional similarity.

4. RESULTS

The main research aim of this work is to compare the representation of stimuli within embedding spaces constructed using human and AI-based assessments of visual and functional similarity. First, the newly constructed psychological embedding spaces modeling human similarity assessments are presented in Sec. 4.1. Two sets of analyses are then conducted to compare these psychological embeddings with computational embeddings of stimuli. The first analysis considers the consistency in defining *pairwise distances* from embedding spaces of both models (Sec. 4.2.1). In the second analysis, stimulus pairs are ordered in terms of pairwise distances, and assigned to *levels of similarity* (1-5). Alignment between methods in assigning pairs to the same levels is then investigated (Sec. 4.2.2). Supporting these findings, qualitative findings are presented to uncover features of of non-text-based similarities that may be specific to human or computational representations (Sec. 4.3).
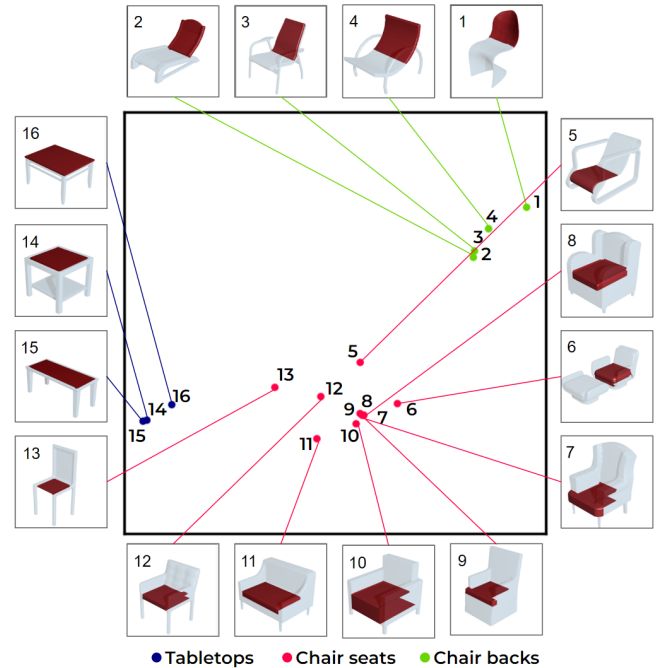
4.1 Examining Psychological Embedding Spaces

To enable the comparison of human and AI-based representations of inspirational stimuli by non-text-based relationships, psychological embedding spaces are constructed, as described in Sec. 3.2. These are visualized in Fig. 3 where plotted points are numbered corresponding to parts in stimulus sets in Fig. 1. Numbering of parts reveals how relationships between parts are represented in computational spaces (as described in Sec. 3.1.2) such that parts 1 and 2 are closer in distance than 1 and 5.

Inspecting psychological embedding spaces, several visually related stimuli separated by low Euclidean distances in the computational embedding space are more distant in Fig. 3a. Part 12, for instance, is closer to parts 9 and 10 in the psychological embedding space than 13, which is a nearest neighbor to 12 in the computational embedding space. This may suggest low agreement between visual similarity relationships represented by both models. Parts are colored in terms of edge curvature, which is one potential criteria used to evaluate visual similarity. Criteria used by human participants to make both visual and functional similarity assessments are presented in Sec. 4.3. Based
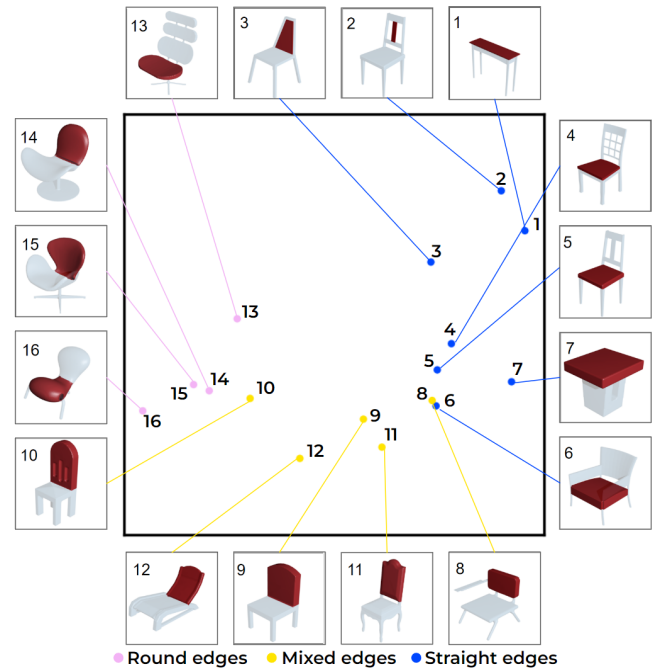
(a) Visual similarity of stimuli in Fig.1a by edge curvature

(b) Functional similarity of stimuli in Fig.1b by object part

FIGURE 3: VISUALIZATIONS OF PSYCHOLOGICAL EMBEDDING SPACES REPRESENTING HUMAN-EVALUATED SIMILARITIES

on embedding space distances, computationally derived relationships appear more preserved in terms of functional similarity, as demonstrated by clusters of closely numbered parts in Fig. 3b. It is evident from the separation of tabletops, chair backs, and chair seats in Fig. 3b that humans relied on the object part when making functional similarity judgments. Overall, by representing these distances using human evaluations of similarity obtained experimentally, the alignment with deep-learning methods can be determined.

In the following section, further examination of agreement between these representations is conducted to gain insight into defining relationships in terms of multi-modal information.

4.2 Agreement between Human and Computational Representations of Similarity

Responding to the main research question posed in this work, this study focuses on determining the agreement between human and AI-based representations of visual and functional similarity. Two sets of stimuli consisting of 16 3D-model parts are considered, as shown in Fig. 1. Across these stimuli, there are 120 unique pairs in each stimulus set, where visual attributes are evaluated for one stimulus set (Fig. 1a) and functional attributes for the other (Fig. 1b). In total, four models are considered: one computational and one psychological embedding space for each of the visual and functional similarity stimulus sets. The agreement between these representations of similarity is evaluated for each similarity type using measures of consistency and alignment. *Consistency* is assessed in terms of computing pairwise distances and *alignment* in assigning pairs to low to high ranging levels of similarity.

4.2.1 Consistency between Pairwise Similarities. The first method used to describe the agreement between human and computational representations of visual and functional similarities is by comparing pairwise distances. Two metrics are used to compare range-normalized pairwise distances derived from psychological and computational embedding spaces: Pearson correlation, r, and Cronbach's alpha coefficient, α. Both metrics, in the context of interrater reliability, measure the consistency between raters in measuring a common dimension [37]. To compare embedding space distances, these metrics reveal whether the two models are consistent in what is being assessed, but not necessarily that the computed distances exactly agree. The relationship between range-normalized pairwise distances in psychological and computational embedding spaces are visualized in Fig. 4, for visual similarity (Fig. 4a) and functional similarity (Fig. 4b). Assignment of pairs by distance to five similarity levels, through a process described in Sec. 4.2.2, is also shown. There are significant positive correlations between pairwise distances modeling both visual similarity, $r(118)=0.74$, $p<0.001$, and functional similarity, $r(118)=0.79$, $p<0.001$. These relationships are confirmed visually by the positive linear correlations of pairwise embedding space distances. High Cronbach's alpha values are also observed, for distances in visual similarity, $\alpha=0.82$, and functional similarity, $\alpha=0.85$. In general, these results demonstrate that, considering all pairs, embedding spaces are consistent in their representations of visual and functional similarity.

Range-normalized pairwise distances between stimuli labelled 1-16 are differently visualized in heatmaps in Fig. 5. Larger distances between stimuli are darker and represent lower similarity between parts. In the first column, distances derived from computational models are represented (labelled CV, CF), and from psychological models in the second column (PV, PF). Heatmaps in the third column (DV, DF) represent differences (computational - psychological) between these distances to directly compare which pairs of stimuli are represented by a larger distance in one embedding space than the other. The first row of
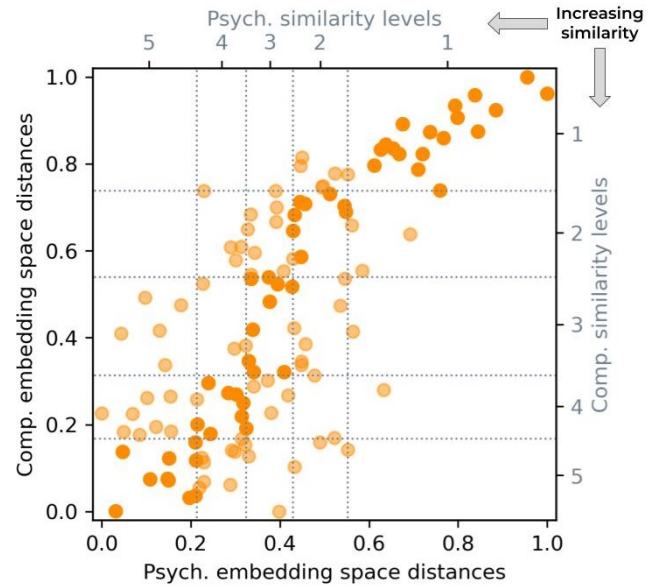
(a) Visual similarity embedding space distances
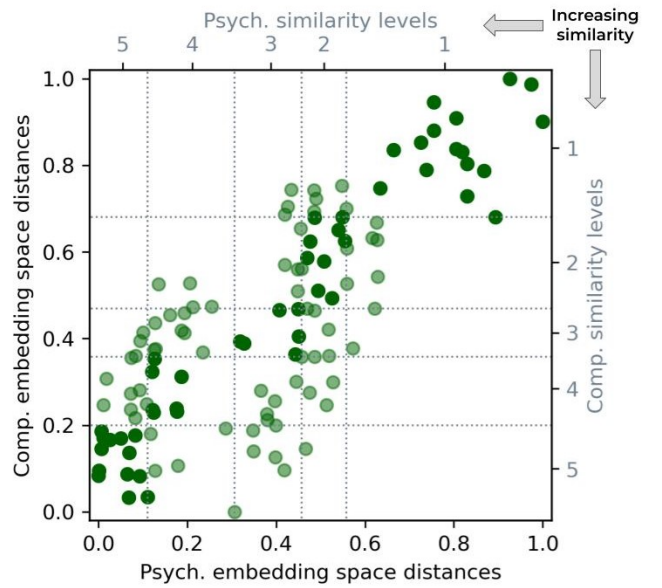
(b) Functional similarity embedding space distances

FIGURE 4: RANGE-NORMALIZED PAIRWISE PSYCHOLOGICAL AND COMPUTATIONAL EMBEDDING SPACE DISTANCES WITH ASSOCIATED SIMILARITY LEVELS SHOWN. DARKER POINTS INDICATE OVERLAP OF PAIRS ASSIGNED TO SIMILARITY LEVELS

heatmaps represent distances in terms of visual features (CV, PV, DV) while the second row represents functional features (CF, PF, DF). As noted in Sec. 3.1.2, stimuli were selected based on computational embedding space distances, which explains the visual consistency in the blue heatmaps (CV, CF) showing increasing distances between farther separated pairs (e.g., pair 5 and 15, compared to pair 5 and 6).

To further investigate where there is more and less agreement between models, the third column of heatmaps (DV, DF) in Fig. 5 is examined. Blue-colored squares indicate pairs separated by a higher distance in the computational than psychological
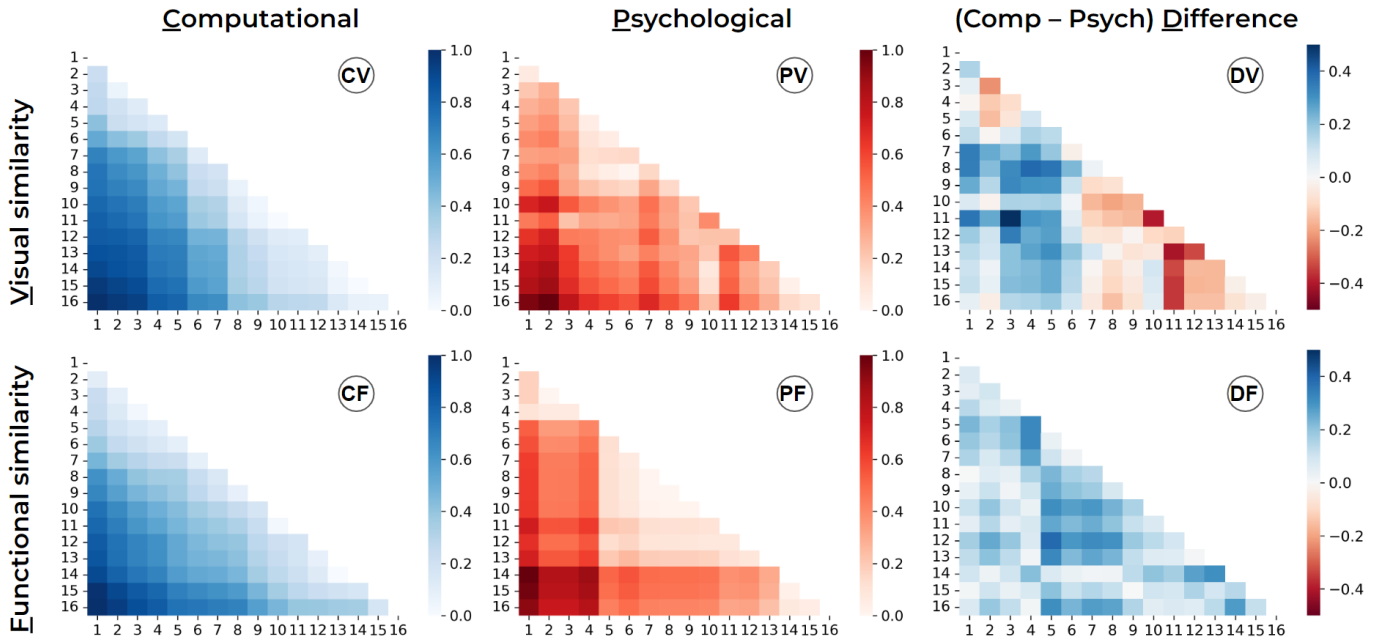
FIGURE 5: RANGE-NORMALIZED PAIRWISE DISTANCES BETWEEN STIMULI 1-16 IN COMPUTATIONAL (C) AND PSYCHOLOGICAL (P) EMBEDDING SPACES REPRESENTING VISUAL (V) AND FUNCTIONAL SIMILARITY (F). DIFFERENCES IN COMPUTATIONAL AND PSYCHOLOGICAL DISTANCES SHOWN IN (D)

embedding spaces, and vice versa for red-colored squares. As an example, the dark blue square in heatmap-DV shows that parts 3 (triangular chair back) and 11 (irregularly curved chair back) in the visual similarity stimulus set (Fig. 1a) are more distant in the computational embedding space. By contrast, the red squares at the intersections of parts 10 (round-edged chair back) and 11 or 11 and 13 (rounded chair seat) are evaluated as more distant, and less similar, by humans than by the computational model. Deeper insight into how humans made visual similarity judgments is explored in Sec. 4.3.1.

Interestingly, comparing heatmap DF to DV, no pairs in terms of function are considered more distant in the psychological than computational embedding spaces. This suggests that, across all pairs, pairs were considered more functionally similar by humans. Parts 5-13 appear to be closely related by humans in functional similarity, as reflected by the presence of many lightly colored, high similarity, pairs in heatmap-PF. Inspecting Fig. 1b, these parts correspond to chair seats, regardless of the type of chair the seats belong to (e.g., 1, 2, 4-legged). In heatmap-CF, a range of distances is observed between parts 5-13, which is a consequence of how the stimuli were selected. The human criteria used to assess functional similarity between these parts may be less nuanced, and consider less information available in the shown stimuli. Evaluation criteria used by humans to make functional similarity judgments is detailed in Sec. 4.3.2.

4.2.2 Alignment across Similarity Levels. In addition to the analysis of pairwise similarities, agreement between human and computational evaluations of similarities is also examined at a higher level. Rather than compare all pairwise similarities, the agreement of pairs assigned to a range of levels of similarity is considered by using measures of percent agreement. The process of defining similarity levels is conceptualized in the example in Fig. 6.
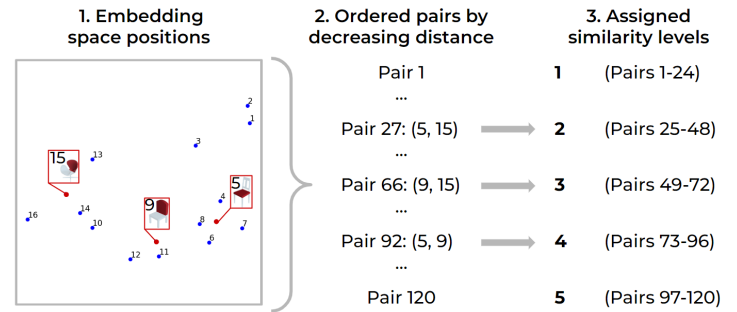

FIGURE 6: CONCEPTUAL OVERVIEW OF PROCESS USED TO ASSIGN SIMILARITY LEVELS TO PAIRS OF PARTS BY ORDERED PAIRWISE DISTANCES

First, Euclidean distances between all 120 pairs of parts in both the psychological and computational embedding spaces are computed and ordered by decreasing distance. Lower distance between parts represents higher similarity, and vice versa. According to pairwise embedding-space distances, pairs are assigned a similarity level between 1 and 5, where each level contains 24 pairs. Lower levels are assigned to high distance, and thus low similarity, pairs (e.g., 5 and 15 in Fig. 6) and higher levels to low distance, high similarity pairs (e.g., 5 and 9).

Evaluating computed distances in terms of low to high levels of similarity (i.e., 1-5) may provide more generalizable insights beyond specific distances computed in this work. In studies investigating stimuli used for design-by-analogy, for example, retrieval criteria for "near" or "far" analogies are often expressed in terms of percentiles of similarity (e.g., in [2, 38]). Insights pertaining to high and low levels of similarity, rather than specific distances, may be relevant toward identifying "near" and "far" stimuli in terms of non-text-based relationships, different from more commonly explored text-based analogies.

In Fig. 4, the relationship between pairwise distances and assigned similarity levels are shown, where overlapping pairs assigned to the same psychological and computational similarity levels are plotted darker. Percent agreement measures overlap in the number of pairs assigned based on both embedding space distances to the same similarity levels. Relatively low values for percent agreement of 44% (53/120 pairs) and 43% (51/120 pairs) are observed for similarity levels assigned in terms of visual and functional similarity, respectively. This definition can be broadened to also include adjacent levels such that a pair assigned to level 2 based on distance in one embedding space is considered to agree when assigned to level 1 or 3 based on distance in the other. Accounting for adjacent levels is a popular modification to percent agreement, e.g., for ratings assigned on a 1-7 scale [37]). Adjusting for this modification, there is an 85% (102/120) overlap of pairs assigned to the same visual similarity levels, and 82% (98/120 pairs) overlap in functional similarity levels. Alignment of pair assignment to similarity levels improves considerably when adjacent levels are included, indicating that the low percent agreement is not due to large discrepancies between embedding space distances. Computationally derived measures may therefore sufficiently represent human-evaluated similarities at a coarser view, i.e., when identifying near vs. far or high vs. low similarity between stimuli. However, misalignment is apparent at a more granular level, such as across five levels of similarity, which can be impactful if retrieval of stimuli at varying distances from an input is desired.

In Fig. 7, the percent agreement of pairs assigned to similarity levels in terms of psychological or computational embedding space distances is shown across similarity levels and by criteria represented (visual or functional). Adding insight to the low percent agreement observed, there appears to be variation across similarity levels. As indicated using Chi-square tests, the difference in percent agreement across similarity levels is observed to be statistically significant for both visual similarity (χ^2(4, N=120)=12.03, p=0.017) and functional similarity (χ^2(4, N=120)=12.07, p=0.017). The largest contribution to this difference appears to be due to the high overlap of pairs identified as sharing low similarity by both human and computational representations. These results demonstrate there is improved alignment, particularly for pairs sharing low similarity, suggesting that higher similarity between pairs may be driven by different factors considered by humans and AI.

Overall, high agreement between human and AI-based representations of visual and functional similarity was found, but not across all analyses. Specifically, our findings demonstrate high consistency in defining pairwise embedding space distances and high alignment in assigning pairs to coarsely defined levels of low to high similarity. These results support the notion that existing AI-based models (e.g., those used in this work) represent human perspectives of visual and functional similarities effectively overall. However, successful retrieval of inspirational stimuli may rely on alignment of similarities across more granularly defined levels not currently achieved. Observed differences and areas of misalignment therefore encourage further examination of stimulus features that may not currently be considered by computational methods. Qualitative post-task responses are analyzed in
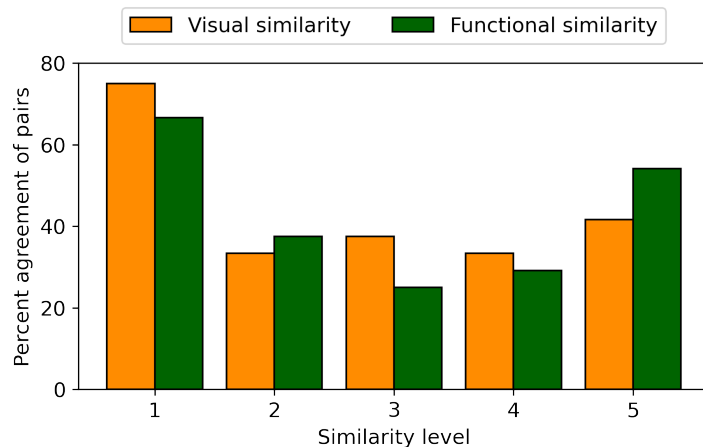

FIGURE 7: PERCENT AGREEMENT OF PAIRS ACCORDING TO HUMAN AND AI-BASED VISUAL AND FUNCTIONAL SIMILARITIES ACROSS SIMILARITY LEVELS

the following subsection to identify features of stimuli underlying human similarity judgments of non-text-based information.

4.3 Exploring Human Criteria for Similarity Assessments

Supporting the main findings of this work measuring the agreement between human and computational representations of non-text-based characteristics of stimuli, an examination of human evaluation criteria is additionally conducted through qualitative analysis. This analysis can reveal areas where computational methods may improve to further align with human evaluation criteria, specifically when considering visual and functional similarity. The evaluation criteria discussed are obtained through the process described in Sec. 3.3.

4.3.1 Evaluation Criteria for Visual Similarity. Following a qualitative inductive category formation procedure, eight criteria for evaluating visual similarity were identified from 35 collected responses (one response was missing in data collection). Participants' responses were assigned to an average of 2 codes each. These criteria are listed in Table 1, where counts refer to the number of participants whose responses were coded into the relevant criteria. Multiple components of a response coded into the same criteria were counted only once. Most frequently referenced (by 34 participants) was the shape of the part, which included references to specific shape geometries (e.g., rectangle, triangle, circle, etc.), curvature or straightness, angularity (sharpness or roundedness), etc. Size was also highly referenced, by 24 participants, which mostly considered thickness (or flatness), dimensions (length, width, height, volume), and proportions between dimensions. More unique responses made reference to non-visual features, including the part's function, "how it interacts with the body" (coded as 'user interaction'), or "what part of the chair it was on" (coded as 'object', referring to the part's placement within the whole object). These criteria were provided from both participants who completed either the visual or functional similarity trials first.

As referenced in Sec. 4.2.1, parts 10 and 11 and 11 and 13 in the visual similarity stimulus set (Fig. 1a) are farther in distances according to human than AI-based representations. Differences

TABLE 1: HUMAN EVALUATION CRITERIA OF VISUAL SIMILARITY

Criteria: Description	Count
Shape: Geometry, curvature, edges, angles	34
Size: Thickness, dimensions, proportions, volume	24
Style: Distinctive features, aesthetics	3
Surface Area: Presence of gaps, holes	2
Orientation: Plane of part (vertical/horizontal)	2
Function: Function of part	2
User Interaction: How it interacts with body	1
Object: Placement of part within object	1

in curvature of edges, angularity of corners, and continuity of surface area in these examples may contribute to greater perceived dissimilarities by humans. When modeling visual features using deep-neural networks, since multiple random perspectives of parts are considered, similarities in e.g., edge thickness, may be emphasized. Though the same criteria may be used (e.g., curvature or thickness) by humans and AI, when applied to multiple perspectives compared to one isometric view, differences may be observed. The role of these criteria and information seen and emphasized by humans vs. AI in contributing to their representations of visual features of stimuli is discussed in Sec. 5.1.

4.3.2 Evaluation Criteria for Functional Similarity. To assess functional similarity between parts, responses provided by participants were coded into nine different criteria, shown in Table 2. Compared to criteria used to evaluate visual similarity, responses were more variable across 36 participants. The most frequently appearing criteria, referenced by 18 participants, was coded broadly as 'interaction' and included both what might interact with the part and how. For example, body parts or objects that might be supported by the part were considered as well as the type of support provided (e.g., for vertical or horizontal loads). Interestingly, some responses were explicitly human-centered when assessing part function in terms of interactions, including "I imagined how I would most often interact with the part" or "I categorized based on how a person would use it". Other criteria were more objective regarding the identity of the part (e.g., chair seat), the whole object (e.g., chair), or the primary use and function of the part (e.g., "for human seating"), aligning with the clustering of chair seats, chair backs, and tabletops in Fig. 1b and of pairwise distances in heatmap-PF (Fig. 5).

Visual attributes were referenced including size and shape, and were acknowledged by some participants as useful if others were exhausted, with one participant stating that if other criteria did not decide the selection "the choice was mostly arbitrary and based on shape matching". More participants who completed functional similarity trials first referred to criteria based on visual features, suggesting that the ordering of tasks may have had the opposite effect than expected. Participants completing visual similarity trials first may have known not to rely on these features when assessing functional similarity. Parts may share high functional similarity in the psychological embedding space due to non-function-based features. As a result, when forming categories, high-similarity parts (as determined by outcomes of the triplet rating task) may then be categorized separately.

TABLE 2: HUMAN EVALUATION CRITERIA OF FUNCTIONAL SIMILARITY

Criteria: Description	Count
Interaction: How and what objects/body parts interact with part under use	18
Function: Main use/purpose function of part	12
Position: Location of part within object	11
Shape: Geometry or curvature	8
Size: Thickness or flatness	8
Type: Comfortable/lounge or structural/rigid	6
Object: Identity of whole object	5
Material: Stiffness, softness, stress fields	5
Neighboring parts: Adjacent parts in object	3

Physical attributes were also considered, classified under 'material', including properties such as stiffness, stress fields, and softness or hardness. One participant noted that these physical qualities "could influence how the user would feel using the object". A related sentiment was expressed by several responses categorized broadly as 'type' in Table 2 to correspond to criteria based on whether the part appeared comfortable, provided cushioning, or in one example, "was more of a lounge type for fitting surface or if it was more of an upright type sitting on surface". These types of surfaces were in contrast to those that appeared rigid and were more structural. Abstract criteria such as perceived comfort are impactful in the evaluation of the overall function served by the object part, but may be difficult to capture using AI, since corresponding visual attributes may not be obvious. Further considerations of representing abstract features of inspirational stimuli are discussed in Sec. 5.1.

5. DISCUSSION

This work investigates how non-text-based attributes of inspirational stimuli are represented by humans and AI. Human representations of visual and functional similarity modelled based on triplet ratings of 3D-model parts collected in a human subjects study were compared to AI-based representations derived using deep learning. Computed similarities between stimuli were found to be consistent across both representations, with greater agreement found when assessing low similarities in appearance and function. Qualitative data regarding how humans formed similarity judgments was analyzed, furthering insight into human representations of the examined non-text-based relationships. Based on these findings, implications for representing inspirational stimuli in terms of non-textual information are discussed.

5.1 Implications for Representing Inspirational Stimuli by Non-Text-Based Attributes

In this work, 3D-model parts are considered as a source of inspirational stimuli that contain multi-modal information, including visual features and hierarchical relationships to other parts. By exploring methods to computationally define non-text-based relationships between these stimuli and then compare these representations to human evaluations, insight can be gained into how to support the retrieval and use of inspirational stimuli in a design context. Toward this aim, two implications are suggested for

representing and defining similarity between inspirational stimuli based on, specifically, visual and functional attributes of stimuli.

5.1.1 Framing of Similarity Assessments.
The first implication of representing non-text-based information of inspirational stimuli is the framing used when making similarity assessments. Humans assessed visual features of stimuli by interacting with 2D images of 3D-model parts taken from one isometric view. Instead, neural networks were trained on multiple images taken at random angles of each 3D-model part. Therefore, neural networks may equally represent similarity between geometries and shapes from less obvious perspectives (e.g., the side edge of a chair seat) to the most common or meaningful views, from the human perspective. As noted in Table 1, the plane or orientation of the part in the shown image influenced participants' perception of visual similarity. Instead, the equal weighting of all perspectives in the neural networks may explain differences in pairwise computational and psychological embedding space distances between stimuli. For example, parts 10 and 11 in Fig. 1a are considered more similar by AI than parts 10 and 14, but the opposite relationships are true based on human representations (as shown in heatmap-PV in Fig. 5). While humans may have emphasized the rounded top edge in parts 10 and 14 as the most influential criteria determining their similarity, the AI-determined relationships also consider the straight edges seen from side views of each part. This retrieval of stimuli based on less obvious features of parts can lead to discovery of seemingly distant inspiration, which may be helpful to designers, but may also be distracting if too unexpected.

The issue of framing is also present in the representation of functional relationships. As presented in Sec. 4.3.2, multiple perspectives may be relevant to consider such as the interaction and relationship of the part with other parts, objects, or humans. Notably absent from the AI-based representation of functional similarity is the human-centered framing and identification of intended and afforded interactions with parts referenced by participants in this study. Instead, functional relationships are derived based on relationships to other parts within whole object assemblies. It is therefore suggested that data-driven methods should account for the framing of representation of inspirational stimuli that is most impactful or appropriate for the type of similarity modeled.

5.1.2 Capturing Information at Varying Levels of Abstraction.
A second implication of representing non-semantic attributes of inspirational stimuli is to capture information at varying levels of abstraction. In the example of 3D-model parts, several more concrete features of stimuli were represented by both humans and AI. These features included the identity of the object the part belonged to and neighboring parts within the same object assembly. While the neural networks used did not explicitly input semantic labels of parts or objects, these relationships were inferred through hierarchical information. When representing function, this concrete information regarding part and object identity was meaningful across both human and computational representations. Not examined in the current work is the role of product domain on human vs. AI representations, reflecting a higher level of abstraction than individual object

parts. Exploring stimuli from diverse domains may impact the present findings, particularly the observation that parts were more functionally similar in AI than human representations. Introducing a greater variety of stimuli may inform how humans assess similarity between stimuli across domains differently from AI, which in our present implementation is unaware of the semantic groupings parts belong to.

As revealed through qualitative insights in Sec. 4.3.2, more abstract information was also relevant. For example, participants referenced a product's style or its type in terms of level of comfort or use for lounging. This criteria incidentally aligned with computational embedding space distances since comfortable chairs (e.g., parts 6-8 in Fig. 1b) share visual attributes, which the function-based neural networks also incorporate. Prior work has relatedly employed visual information through shape grammars and 3D geometries of products to assess overall similarities in product style [39, 40]. For humans, visual style was found to be associated with a more conceptual meaning (i.e., appearance of cushioned chairs with comfort), influencing representation of functional relationships. Insights from this study encourage further understanding of the relationship between visual attributes and function and the use of AI to computationally define abstract, conceptual features of stimuli toward improved alignment with human representations.

5.2 Limitations and Future Work
This work presents a comparison of human and AI-based representations of visual and functional similarity between 3D-model parts. We acknowledge the potential limitation of the present findings to the specific stimulus sets presented and similarity types assessed by participants. In this work, a limited set of stimuli was utilized in order to reduce the significant complexity of this study and to make the task of assessing similarity of non-textual information tractable for humans. Future work might explore the generalizability of these findings to additional examples and contexts. Several features of stimuli influencing similarity assessments, e.g., the number of different objects presented or perception of comfort, may be specific to the types of objects presented. As well, although the tasks conducted were not explicitly design relevant, design experience of participants may impact their judgments of relationships between the shown stimuli. Despite providing definitions for functional relationships at the study's onset, participants' knowledge of other definitions of function in engineering design may have impacted their similarity assessments. Furthering this study, future work is encouraged to investigate additional sources of inspirational stimuli containing multi-modal information from which to extract and define non-text-based similarities and study in a design context. By gaining more knowledge regarding how these similarities are perceived and evaluated, new sources of inspiration can be more effectively engaged with and utilized by designers. Towards this aim, efforts to define a more holistic definition of similarity are encouraged, which appropriately account for varied features of stimuli (semantic and non-semantic) and components of similarity. The retrieval of and interaction with inspirational stimuli across various forms can be enabled by computational platforms developed by the broader design research community.

6. CONCLUSION

The growing interest in representing inspirational stimuli across multiple modalities, in contrast to by text-based labels or descriptions only, motivates the present work. The aim of this work was to understand how human and AI-based similarity assessments of non-text-based features of inspirational stimuli compare. Using measures of consistency and alignment, high agreement between humans and AI was found for representing both visual and functional similarities, at a coarse level. In both cases, this agreement was highest for identifying pairs of stimuli sharing low similarity, suggesting that humans and AI agree on identifying obvious differences, but less on features driving increased similarity between pairs. The framing of how humans and AI assess features of parts as well as the representation of abstract information are proposed as factors that may need further consideration in modeling visual and functional similarities using computational methods. Findings from this study encourage further research on representing multi-modal information of various sources of inspirational stimuli to better understand and support effective inspiration representation and retrieval for designers.

ACKNOWLEDGMENTS

REFERENCES

[1] Sio, Ut Na, Kotovsky, Kenneth and Cagan, Jonathan. "Fixation or inspiration? A meta-analytic review of the role of examples on design processes." *Des. Stud.* Vol. 39 (2015): pp. 70–99.

[2] Chan, Joel, Fu, Katherine, Schunn, Christian, Cagan, Jonathan, Wood, Kristin and Kotovsky, Kenneth. "On the Benefits and Pitfalls of Analogies for Innovative Design: Ideation Performance Based on Analogical Distance, Commonness, and Modality of Examples." *J. Mech. Des.* Vol. 133 No. 8 (2011): p. 081004.

[3] Fu, Katherine, Chan, Joel, Cagan, Jonathan, Kotovsky, Kenneth, Schunn, Christian and Wood, Kristin. "The Meaning of "Near" and "Far": The Impact of Structuring Design Databases and the Effect of Distance of Analogy on Design Output." *J. Mech. Des.* Vol. 135 No. 2 (2013): p. 021007.

[4] Goucher-Lambert, Kosa, Gyory, Joshua T., Kotovsky, Kenneth and Cagan, Jonathan. "Adaptive Inspirational Design Stimuli: Using Design Output to Computationally Search for Stimuli that Impact Concept Generation." *J. Mech. Des.* Vol. 142 No. 9 (2020): p. 091401.

[5] Jiang, Shuo, Hu, Jie, Wood, Kristin L. and Luo, Jianxi. "Data-Driven Design-By-Analogy: State-of-the-Art and Future Directions." *J. Mech. Des.* Vol. 144 No. 2 (2022): p. 020801.

[6] Kwon, E., Huang, F. and Goucher-Lambert, K. "Enabling multi-modal search for inspirational design stimuli using deep learning." *AI EDAM* Vol. 36 (2022): p. e22.

[7] Kwon, Elisa, Rao, Vivek and Goucher-Lambert, Kosa. "Exploring Designers' Encounters with Unexpected Inspirational Stimuli." Gero, John S. (ed.). *Design Computing and Cognition'22*: pp. 397–408. 2023.

[8] Borgianni, Yuri, Rotini, Federico and Tomassini, Marco. "Fostering Ideation in the Very Early Design Phases: How Textual, Pictorial and Combined Stimuli Affect Creativity." *Proc. of the 21st International Conference on Engineering Design*: pp. 139–148. 2017. The Design Society.

[9] Yuri Borgianni, Lorenzo Fiorineschi, Lorenzo Maccioni and Rotini, Federico. "Forms of Stimuli and their Effects on Idea Generation in terms of Creativity Metrics and Non-Obviousness." *International Journal of Design Creativity and Innovation* Vol. 8 No. 3 (2020): pp. 147–164.

[10] Linsey, J. S., Clauss, E. F., Kurtoglu, T., Murphy, J. T., Wood, K. L. and Markman, A. B. "An Experimental Study of Group Idea Generation Techniques: Understanding the Roles of Idea Representation and Viewing Methods." *J. Mech. Des.* Vol. 133 No. 3 (2011). 031008.

[11] Gonçalves, Milene, Cardoso, Carlos and Badke-Schaub, Petra. "What Inspires Designers? Preferences on Inspirational Approaches During Idea Generation." *Des. Stud.* Vol. 35 No. 1 (2014): pp. 29–53.

[12] Linsey, J.S., Wood, K.L. and Markman, A.B. "Modality and Representation in Analogy." *AI EDAM* Vol. 22 No. 2 (2008): p. 85–100.

[13] Han, J., Shi, F., Chen, L. and Childs, P. R. "The Combinator–A Computer-Based Tool for Creative Idea Generation Based on a Simulation Approach." *Des. Sci.* Vol. 4 No. 11 (2018).

[14] Malaga, Ross A. "The Effect of Stimulus Modes and Associative Distance in Individual Creativity Support Systems." *Decision Support Systems* Vol. 29 (2000): pp. 125–141.

[15] Menezes, Alexandre and Lawson, Bryan R. "How designers perceive sketches." *Des. Stud.* Vol. 27 No. 5 (2006): pp. 571–585.

[16] Murphy, Jeremy, Fu, Katherine, Otto, Kevin, Yang, Maria, Jensen, Dan and Wood, Kristin. "Function Based Design-by-Analogy: A Functional Vector Approach to Analogical Search." *J. Mech. Des.* Vol. 136 No. 10 (2014): p. 101102.

[17] Fu, Katherine, Cagan, Jonathan, Kotovsky, Kenneth and Wood, Kristin. "Discovering Structure in Design Databases Through Functional and Surface Based Mapping." *J. Mech. Des.* Vol. 135 No. 3 (2013): p. 031006.

[18] Song, Hyeonik and Fu, Katherine. "An Exploration-Based Approach to Computationally Supported Design-by-Analogy using D3." *AI EDAM* Vol. 34 No. 4 (2020): pp. 444–457.

[19] Song, Hyeonik and Fu, Katherine. "Design-by-Analogy: Effects of Exploration-Based Approach on Analogical Retrievals and Design Outcomes." *J. Mech. Des.* Vol. 144 No. 6 (2022): pp. 1–15.

[20] Ji Han, Feng Shi Jianxi Luo, Serhad Sarica. "Semantic Networks for Engineering Design: State of the Art and Future Directions." *J. Mech. Des.* Vol. 144 No. 2 (2022): pp. 1–11.

[21] Sarica, S., Song, B., Luo, J. and Wood, K.L. "Idea generation with technology semantic network." *AI EDAM* Vol. 35 (2021): pp. 265–283.

[22] Zhang, Zijian and Jin, Yan. "An Unsupervised Deep Learning Model to Discover Visual Similarity Between Sketches for Visual Analogy Support." *Proc. ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*: p. V008T08A003. 2020.

[23] Zhang, Zijian and Jin, Yan. "Toward Computer Aided Visual Analogy Support (CAVAS): Augment Designers through Deep Learning." *Proc. ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*: p. V006T06A057. 2021.

[24] Jiang, Shuo, Luo, Jianxi, Ruiz-Pava, Guillermo, Hu, Jie and Magee, Christopher L. "A Convolutional Neural Network-Based Patent Image Retrieval Method for Design Ideation." *Proc. ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*: p. V009T09A039. 2020.

[25] Jiang, Shuo, Luo, Jianxi, Ruiz-Pava, Guillermo, Hu, Jie and Magee, Christopher L. "Deriving Design Feature Vectors for Patent Images Using Convolutional Neural Networks." *J. Mech. Des* Vol. 143 No. 6 (2021). 061405.

[26] Kazi, Rubaiat Habib, Grossman, Tovi, Cheong, Hyunmin, Hashemi, Ali and Fitzmaurice, George. "DreamSketch: Early Stage 3D Design Explorations with Sketching and Generative Design." *Proc. of the 30th Annual ACM Symposium on User Interface Software and Technology*: p. 401–414. 2017.

[27] Arora, R., Darolia, I., Namboodiri, V. P., Singh, K. and Bousseau, A. "SketchSoup: Exploratory Ideation Using Design Sketches." *Computer Graphics Forum* Vol. 36 No. 8 (2017): pp. 302–312.

[28] Chaudhuri, Siddhartha and Koltun, Vladlen. "Data-Driven Suggestions for Creativity Support in 3D Modeling." *ACM Transactions on Graphics* Vol. 29 No. 6 (2010): pp. 1–10.

[29] Kim, Jingoog, Maher, Mary Lou and Siddiqui, Safat. "Collaborative Ideation Partner: Design Ideation in Human-AI Co-creativity." *Proceedings of the 5th International Conference on Computer-Human Interaction Research and Applications (CHIRA)*. 2021.

[30] Kim, Jingoog and Maher, Mary Lou. "The effect of AI-based inspiration on human design ideation." *International Journal of Design Creativity and Innovation* (2023): pp. 1–18.

[31] Kwon, Elisa, Rao, Vivek and Goucher-Lambert, Kosa. "Investigating the Roles of Expertise and Modality in Designers' Search for Inspirational Stimuli." *Proc. ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. 2022. V006T06A015.

[32] Mo, Kaichun, Zhu, Shilin, Chang, Angel X., Yi, Li, Tripathi, Subarna, Guibas, Leonidas J. and Su, Hao. "PartNet: A Large-scale Benchmark for Fine-grained and Hierarchical Part-level 3D Object Understanding." (2018).

[33] Nandy, A. and Goucher-Lambert, K. "Do Human and Computational Evaluations of Similarity Align? An Empirical Study of Product Function." *J. Mech. Des.* Vol. 144 No. 4 (2022): p. 041404.

[34] Ahmed, F., Ramachandran, S.K., Fuge, M., Hunter, S. and Miller, S. "Interpreting Idea Maps: Pairwise Comparisons Reveal What Makes Ideas Novel." *J. Mech. Des.* Vol. 141 No. 2 (2019): p. 021102.

[35] Roads, B.D. and Mozer, M.C. "Obtaining psychological embeddings through joint kernel and metric learning." *Behav. Res.* Vol. 51 (2019): pp. 2180–2193.

[36] Mayring, P. "Qualitative content analysis." Sage, Great Britain (2004): pp. 266–269.

[37] Stemler, Steven E. "A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Inter-rater Reliability." *PARE* Vol. 9 No. 4 (2004).

[38] Goucher-Lambert, Kosa and Cagan, Jonathan. "Crowdsourcing Inspiration: Using Crowd Generated Inspirational Stimuli to Support Designer Ideation." *Des. Stud.* Vol. 61 (2019): pp. 1–29.

[39] Culbertson, Timothy D. and Simpson, Timothy W. "Using Shape Grammars to Identify Salient Features in Support of Product Family Design." *Proc. ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. 2014. V007T07A036.

[40] Ranscombe, Kinsella Philip, Charlie and Blijlevens, Janneke. "Data-Driven Styling: Augmenting Intuition in the Product Design Process Using Holistic Styling Analysis." *J. Mech. Des.* Vol. 139 No. 11 (2017): pp. 1–11.