## 1 INTRODUCTION

Providing designers with external sources of inspiration can help promote novel and innovative ideation by serving as cues to retrieve relevant concepts from long-term memory to aid conceptual design (Sio et al., 2015). Inspirational stimuli may also lead to undesirable outcomes such as design fixation (Jansson and Smith, 1991), thus requiring the thoughtful selection of stimuli presented to designers. Consideration of stimulus selection is especially relevant when derived from large datasets, increasingly used when applying data-driven design methods (Jiang et al., 2022). In a recent review on data-driven design-by-analogy (DbA), Jiang et al. identify *analogical distance* between source and target domains as the most widely studied factor impacting this process. To retrieve stimuli with varying analogical distances to a given design problem or designer specified input, similarity relationships need to be computed. For instance, text-based processing has been used by Fu et al. (2013) to define contextual similarity between patents. Semantic networks include the Technology Semantic Network (TechNet), which uses unsupervised learning and natural-language-processing (NLP) techniques to retrieve and relate technology-based knowledge from patent texts (Sarica et al., 2021).

However, beyond utilizing stimuli represented by textual information, there is increasing interest in using 2D-image and 3D-model datasets to support *multi-modal analogy* for design inspiration (Jiang et al., 2022). Zhang and Jin (2021) used an unsupervised deep-learning model to construct a latent space for a dataset of sketches. Jiang et al. (2021) constructed a convolutional neural network-based model to derive a vector space where feature vectors embed visual and technology-related information from patent images. Kim and Maher (2023) developed a co-creative artificial intelligence (AI) partner that provides inspirational sketches related by visual and conceptual similarity to designers' sketches. The effects of providing sketches with varying levels of visual and conceptual similarity to the designer's sketch were investigated (Kim and Maher, 2023). Visually similar sketches with low conceptual similarity were associated with higher quality ideation. However, visual similarity was not examined in isolation from conceptual similarity. In prior work from our team, deep neural networks modeling visual and functional relationships between 3D-model parts were used to develop a multi-modal search platform for inspiration discovery (Kwon et al., 2022a). Findings across two subsequent user studies involving this system revealed that, when searching for stimuli in terms of appearance and function-based similarities to a specified input, designers frequently encountered results they did not expect (Kwon et al., 2023).

Insights from this prior work motivate the aim of the present study. Specifically, to deepen our understanding of the definition and evaluation of non-text-based measures of similarity, which have not been widely studied in interactive settings. Currently, there is a gap in knowledge regarding how relationships defined in terms of non-textual properties of inspirational stimuli align with human perceptions and expectations. Increased availability, interest, and use of 2D-image and 3D-model datasets support the development of tools enabling discovery of design stimuli related to an input by *non-text-based features* rather than semantic distances. Evaluations of human and AI-based representations of visual and functional similarity are compared in this work through the two synergystic research questions:

- (RQ1) How do evaluations of human and AI-based representations of non-text-based similarities of inspirational stimuli compare?
- (RQ2) How does the evaluation of non-text-based similarities compare when considering visual vs. functional attributes of inspirational stimuli?

These research questions are studied within the context of using 3D-model parts as a source of design relevant inspiration, which contain both visual and functional information. Findings from a categorization task are analyzed in which stimuli are organized based on visual and functional similarity (n=36). This study uses the categorization of stimuli to evaluate similarities computed in terms of human vs. AI-based representations and visual vs. functional similarity. Low levels of similarity are expected to align with different-group categorization and vice versa. Leveraging the notion of "near" and "far", typically attributed to conceptual distances (e.g., (Fu et al., 2013)), "near" distances are associated with same-group categorization and "far" with different-group categorization. Understanding how similarity in terms of non-textual information is assessed can support the retrieval of relevant and impactful sources of inspiration for designers.

## 2 METHODS

This work investigates how humans evaluate visual and functional similarities between 3D-model parts. Similarity is represented by distances between stimuli in embedding spaces derived using two approaches. The first approach uses deep learning to construct neural networks to model these relationships, resulting in *computational embedding spaces* for a large dataset of 3D-model parts (developed in prior work by Kwon et al. (2022a)). 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 stimuli. In Sec. 2.1, the study conducted and steps taken in order to develop psychological embedding spaces for both visual and functional similarity are outlined. The analytical approach presented in Sec. 2.2 is followed to quantify how the perception of similarity between stimuli in terms of their mutual categorization can be related to measures of similarity, as defined within these two embedding spaces.

### 2.1 Experimental design

A human subjects study (n=36) was conducted, consisting of two main tasks: a triplet rating task and a categorization task, each completed twice (for visual and functional similarity, separately). 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 presentation was counterbalanced across participants. 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.

#### *2.1.1 Participants*

For this study, 36 participants (13 female, 22 male, 1 non-binary) were recruited including 14 graduate students, 12 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) (Kwon et al., 2022b). No particular level of engineering design knowledge or experience was required. Participants were recruited via email from among current students in Mechanical Engineering and those who previously completed research studies related to engineering design. Compensation of $10 was provided for completion of the 30 min. study. This human subjects research study has been approved by the Institutional Review Board at the University of California, Berkeley.

#### *2.1.2 Computationally derived stimulus sets*

The stimulus sets provided to participants in the triplet rating and categorization tasks 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 similarity of 3D-model parts by *visual appearance*, 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). The development of these neural networks is fully described in our past work (Kwon et al., 2022a). While components of common objects may provide limited inspiration, prior work utilizing these stimuli in a 3D-modeling design task demonstrated that they could be useful for uncovering new and unexpected features to designers' ideas (Kwon et al., 2022b, 2023).

Euclidean distances between parts in the computational embedding spaces are used to represent how similar (low distance) or dissimilar (high distance) parts are. 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 (Mo et al., 2019)), resulting in 2043 possible parts. This constraint was imposed 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. The full 16-part stimulus sets selected to present in the triplet rating and categorization tasks are shown in Fig. 1. While distance between neighbors is 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 contain a diversity of distances, where all parts belong to both low and high distance pairs.

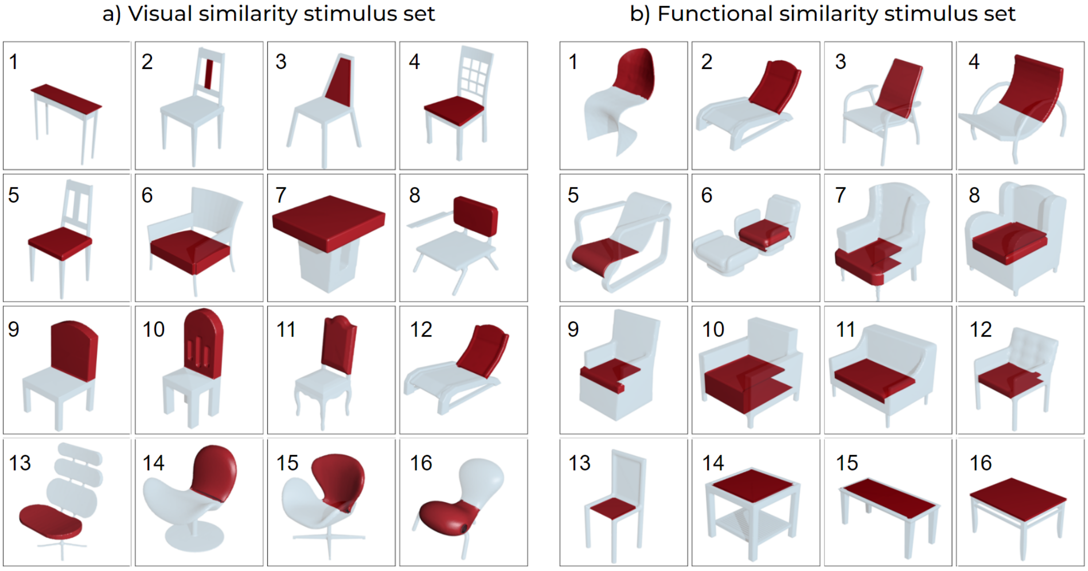


*Figure 1. Stimulus set used when assessing **(a)** visual similarity and **(b)** functional similarity between parts in triplet rating and categorization tasks.*

#### *2.1.3 Human-evaluated similarities of stimuli*

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 and Goucher-Lambert (2022) and Ahmed et al. (2019) have used triplet similarity ratings to generate embedding spaces for human representations of design stimuli. 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.

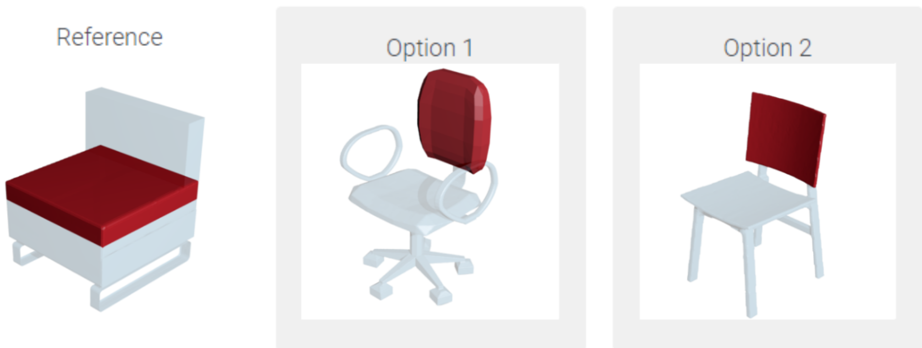


*Figure 2. Example triplet of 3D-model parts shown to participants.*

For the number of parts in each stimulus set (16), a total of 1680 unique triplet trials are possible. Ahmed et al. (2019) recommend that a minimum of 30% of the full stimulus set is needed to construct a robust embedding space of human representations. 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 similarity and 826 trials for functional similarity were included, constituting 48% and 49% of all potential trials.

#### 2.1.4 Categorization of stimuli

Following the triplet rating task, participants provided a short written response to describe the criteria used to evaluate visual or functional similarity between parts. A categorization task was next conducted for the same stimulus set (presented unordered). Rather than allow participants to freely group parts, two different criteria were specified to consider for each similarity type. Participants were instructed to examine the (1) **shape** (e.g., geometry) and (2) **size** (e.g., thickness) of a part when categorizing parts by visual similarity. The (1) **object** the reference part belongs to and (2) **neighboring parts** to the reference part were criteria specified when categorizing parts by functional similarity. These criteria were selected based on knowledge of the part features learned by the computationally derived neural networks as well as the similarity evaluation criteria participants provided in pilot testing. For each stimulus set, participants constructed two sets of 3 or 4 categories for the specified criteria. By associating computed similarities derived by the previously specified methods and criteria with categorization outcomes, insight into how similarities are evaluated can be gained.

### 2.2 Definition and evaluation of similarities between stimuli

Using the similarity assessments obtained in the triplet rating tasks, a psychological embedding space was constructed for each stimulus set to model human representations of parts, as described in Sec. 2.2.1. Similarity between parts is represented by their embedding-space distances. The relationship between similarity and categorization was then used to gain insight into how computed similarities between stimuli are evaluated, as described in Sec. 2.2.2.

#### 2.2.1 Construction of psychological embedding spaces

Using the triplet rating task observations, psychological embedding spaces were constructed, which each 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 representations (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 (Roads and Mozer, 2019). The number of dimensions to use for each model was determined by training models with dimensions between two and ten and selecting the highest value at which validation set (10% of trials) losses stopped improving. The final psychological embedding spaces for both visual and functional similarity are two dimensional.

#### 2.2.2 Relationship between similarity and categorization of parts

To represent similarity between pairs, the process conceptualized in the example in Fig. 3 was followed. First, the Euclidean distances between all 120 pairs of parts in both the psychological and computationally derived embedding spaces were computed and ordered by decreasing distance. Measuring similarity in terms of Euclidean distances implies symmetry in pairwise relationships that may not always be appropriate to maintain, as explored by Chaudhari et al. (2019). However, in the context of 3D-shape retrieval for the properties of similarity explored, symmetric distance-based measures are considered suitable (Tangelder and Veltkamp, 2004). Lower distance between parts is used to represent higher similarity, and vice versa. According to pairwise embedding-space distances, pairs were assigned a similarity level between 1 and 5, where each level contains 24 pairs. Lower levels were assigned to high distance, and thus low similarity, pairs (such as 5 and 15 in the example) and higher levels to low distance, high similarity pairs (such as 5 and 9 in the example).

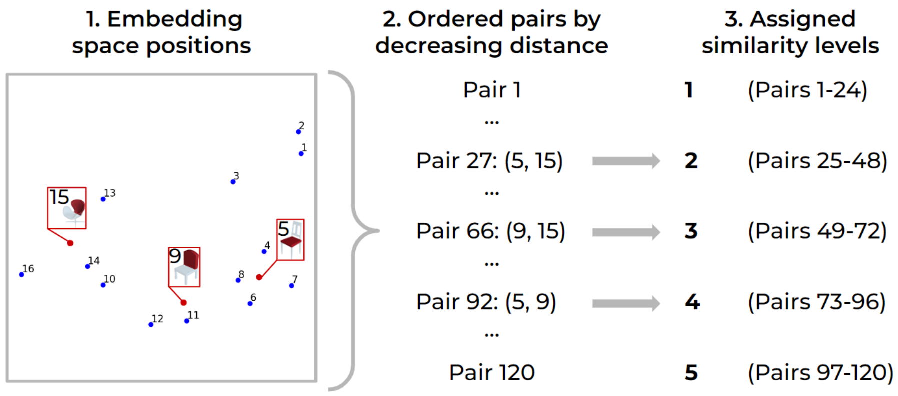


*Figure 3. Conceptual overview of process used to assign similarity levels to pairs of parts by ordered pairwise distances. Each similarity level (1-5) was assigned to 24/120 part pairs.*

As described in Sec. 2.1.4, each participant created three to four categories to organize stimuli in terms of given criteria. Based on these categories, every unique pair of parts (120 in total) was associated with an outcome of either being grouped together or separately. Each pair of parts was also assigned to a similarity level (1-5) consisting of 24 pairs in total. For each participant, the proportion of pairs in each similarity level categorized into different groups was computed. The mean proportion across participants was then found for each similarity level. Significantly above chance group means (where the lower limit of the 95% confidence interval is greater than 50%) indicate different-group categorization of pairs of parts with the specified similarity level. Using the notion of "near" and "far" prevalent in DbA (e.g., Chan et al. (2015)), **above chance different-group categorization** is used to distinguish the boundary between similarity levels at which stimuli may be evaluated as "too far". Analogously, stimuli with a similarity level associated with **below chance different-group categorization** (i.e., same-group categorization) may be "too near" to be relevant. We propose that in between these similarity levels lies the "sweet spot" referred to by Fu et al. (2013) in inspirational stimuli avoiding these extremes. A similar approach was used by Cooke et al. (2007) to model the relationship between similarity and categorization of 3D objects. The boundary separating parts that are "too far" is expected to be observed at low similarity levels, corresponding to parts separated by greater distances within the embedding spaces, while stimuli perceived as "too near" are expected to be related by high levels of similarity.

## 3 RESULTS

Using the approach outlined in Sec. 2.2.2, the relationship between the similarity and categorization of parts is explored. RQ1 and RQ2 are addressed by comparing how the categorization of parts with varying similarities differs when similarities are obtained from computational or psychological embedding spaces and define visual or functional similarity. Represented in Fig. 4, mean proportions of pairs categorized in different groups are shown for each level of visual and functional similarity between parts. The computational and psychological similarity levels of a pair of parts are plotted separately and are assigned based on the pairwise distances between the parts in each embedding space. High-distance pairs are associated with low similarity levels (1) while low-distance pairs have high similarity levels (5). Mean proportion values and associated 95% confidence intervals are also detailed in Table 1. Above chance (50%) average proportions represent that participants did not categorize pairs with the specified level of similarity together. Proportions are plotted with error bars of 95% confidence intervals to determine whether same or different group categorization of pairs with the specified similarity level is statistically significant. Confidence interval limits that do not cross the 50% threshold are bolded in Table 1 to indicate significant proportions of same or different group categorization.

### 3.1 Categorization of visually similar inspirational stimuli

The categorization of stimuli related by visual similarity is first examined. When evaluating based on **shape** (Fig. 4a), parts with similarity levels up to 2 and 3 in terms of psychological and computational embedding space distances, respectively, are categorized in different groups above chance. Psychological boundaries in Fig. 4a show humans may consider pairs with similarity levels up to 2 to be "too far".

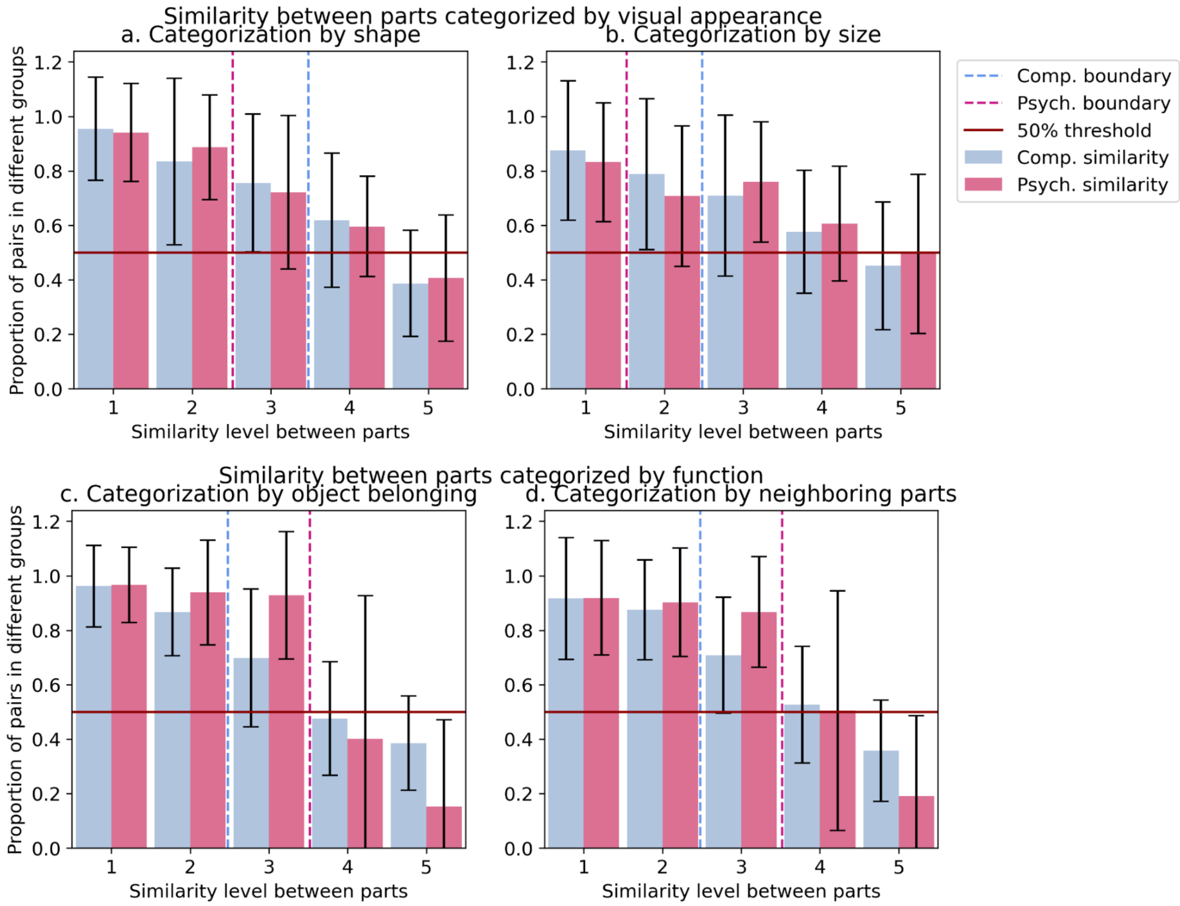


*Figure 4. Mean proportions (with 95% confidence intervals) of pairs in different groups when categorizing by visual similarity based on **(a)** shape and **(b)** size and by functional similarity based on **(c)** object belonging and **(d)** neighboring parts.*

*Table 1. Proportions of pairs categorized in different groups across similarity levels*

| Embedding space | Similarity level | Categorization criteria by visual similarity | | | | Categorization criteria by functional similarity | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *Shape* | | *Size* | | *Object* | | *Neighbors* | |
| | | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI |
| Computational | 1 | .95 | (**.77**, 1.1) | .88 | (**.62**, 1.1) | .96 | (**.81**, 1.1) | .92 | (**.69**, 1.1) |
| | 2 | .83 | (**.53**, 1.1) | .79 | (**.51**, 1.1) | .87 | (**.71**, 1.0) | .88 | (**.69**, 1.1) |
| | 3 | .76 | (**.50**, 1.0) | .71 | (.41, 1.0) | .70 | (.44, .95) | .71 | (.49, .92) |
| | 4 | .62 | (.37, .87) | .58 | (.35, .80) | .48 | (.27, .68) | .53 | (.31, .74) |
| | 5 | .39 | (.19, .58) | .45 | (.22, .69) | .39 | (.21, .56) | .36 | (.17, .54) |
| Psychological | 1 | .94 | (**.76**, 1.1) | .83 | (**.61**, 1.1) | .97 | (**.83**, 1.1) | .92 | (**.71**, 1.1) |
| | 2 | .89 | (**.69**, 1.1) | .71 | (.45, .97) | .94 | (**.75**, 1.1) | .90 | (**.70**, 1.1) |
| | 3 | .72 | (.44, 1.0) | .76 | (.54, .98) | .93 | (**.69**, 1.2) | .87 | (**.66**, 1.1) |
| | 4 | .59 | (.41, .78) | .61 | (.39, .82) | .40 | (-.12, .93) | .51 | (.07, .95) |
| | 5 | .41 | (.17, .64) | .50 | (.20, .79) | .15 | (-.17, **.47**) | .19 | (-.10, **.49**) |

AI makes this distinction for pairs with similarity levels 1-3. For parts categorized based on **size** (Fig. 4b), the same difference between human and AI-based representations is found where pairs with a similarity level of 1 in the psychological embedding space and up to level 2 in the computational embedding space may be “too far”. Differences between human and AI evaluations of similarity are discussed in Sec. 4.1. The “farther” boundary in size vs. shape suggests that size is a less discriminating factor when evaluating visual similarity. Combining both categorization criteria, stimuli sharing similarity up to level 3 are not grouped together and are thus associated with being "too far". These stimuli represent 20-60% of all pairs that may be discounted as being too visually dissimilar to be relevant, with humans associating greater visual dissimilarity (lower similarity levels) with different group categorization.

### 3.2 Categorization of functionally similar inspirational stimuli

The relationship between categorization and functional similarity is also determined. Across both criteria used to categorize functional similarity, the most similar pairs in the psychological embedding space (level 5) are placed in different categories below chance (or, in the same categories above chance). High similarity pairs in this context are evaluated as "too near" or too obviously related to be relevant. This finding, which is not observed with high levels of computationally determined similarity, may reveal that these criteria align with how participants made their similarity judgments. For categories made based on the **object** a part belongs to (Fig. 4c), differences were observed based on how similarities between parts were derived. For computationally derived similarities, pairs with similarity levels of 1 and 2, the 40% least similar pairs, are considered "too far". Based on psychological embedding space distances, pairs of stimuli with similarity levels up to 3 are "too far", including 60% of all pairs. Also notable is the steep dropoff between the mean proportion of pairs that are grouped in different categories with psychological similarity levels of 3 (0.93) compared to 4 (0.40). The boundary separating parts that are "too far" to categorize together by **neighboring parts** (Fig. 4d) is between similarity levels 2-3 and 3-4 in computational and psychological embedding space distances, respectively. Across both categorization criteria, stimuli that are "too far" to be grouped together constitute 40-60% of pairs. These boundaries are "nearer" than observed when categorizing by visual similarity. Parts do not need to be as far in functional as in visual similarity to be divided into separate groups. The opposite relationship between human and AI evaluations of functional vs. visual similarity is found where more similar pairs in terms of human than AI-based representations are categorized separately. Differences in human vs. AI evaluations and visual vs. functional similarity are further discussed in Sec. 4.1 and 4.2.

## 4 DISCUSSION

In general, the results presented demonstrate that across definitions of pairwise similarities (human-evaluated and computationally derived) and criteria for categorization (form and function-based features), increasing similarity between pairs is consistently associated with decreasing proportions of pairs categorized in different groups. "Near" and "far" are used to describe similarity levels associated with stimuli that are categorized together and separately. Based on the present findings, the evaluations of different measures of similarities are impacted by various factors discussed in this section.

### 4.1 Comparing evaluation of human and computational representations of similarity

Through RQ1, a comparison of the perception of computationally derived and human-evaluated similarities was sought. Results from four categorization tasks were used to determine whether the placement of stimuli into the same or different groups aligned with these similarity measures. Boundaries between similarity levels of pairs of stimuli associated with different-group categorization were observed. Comparing human and AI-evaluated similarities, AI-based boundaries were "nearer" when categorizing by visual similarity, but "farther" when categorizing by functional similarity. Parts need to be *less visually similar* according to psychological than computational similarities and *less functionally similar* according to computational than psychological similarities to be categorized separately. Visual features considered by AI may be less obvious to humans (e.g., shapes at different angles), which can explain the increased separation of parts with higher levels of computationally derived similarities. Regarding functional similarity, this difference reflects that in the triplet rating tasks participants may have rated parts as more functionally similar if they belonged to the same object type. This finding is in line with the presence of three distinct object categories (4/16 chair backs, 9/16 chair seats, and 4/16 tabletops) in the functional similarity stimulus set. By contrast, the computationally derived functional embedding space does not incorporate semantic knowledge of individual parts and their associated object assemblies, thus similar objects may share low functional similarity. It is possible that by introducing more complex and diverse stimuli, the use of object-related criteria to categorize stimuli by functional similarity may be less relevant to investigate and further differences may be observed when other criteria are considered. This emphasizes that it is important for computational methods to develop relationships between inspirational stimuli that agree with human representations. These findings suggest that these representations are dependent on the stimuli considered (e.g., object diversity within the stimulus set).

### 4.2 Comparing evaluation of visual and functional similarity

Addressing RQ2, the evaluations of visual and functional similarity are compared. Distinct boundaries dividing stimuli that are "too far" were observed, where boundaries identified for functional similarity are "nearer" than those found for visual similarity and include more similar parts as "too far". When considering visual similarity between parts, the top 20-60% least similar pairs are found to be "too far" to categorize together, while these boundaries lie between the top 40-60% of least similar pairs when categorizing by functional similarity. One reason for this difference may be that, when evaluating visual similarity, multiple visual features may be easily analyzed simultaneously (e.g., parts should have the same shape and thickness). If multiple criteria are used to determine visual similarity through triplet comparisons, but categorization is performed for one criteria at a time (shape or size), less similar pairs (as assessed in the triplet rating task) may be similar enough to group together. It is notable that under evaluation of these non-text-based similarities, up to 60% of stimulus pairs are found to be "too far". Definitions of "far" applied in prior work to retrieve conceptually related analogies tend to be more extreme, constituting crowd-sourced ideas occurring once or examples in the 10th percentile of text-based similarities to a design challenge (Goucher-Lambert and Cagan, 2019; Chan et al., 2015). Taken together, one level of similarity at which stimuli are "too far" may not exist for all sets of design stimuli or definitions of similarity.

### 4.3 Limitations and future work

It is important to note that the interpretation of these results is limited to the specific stimulus sets presented and the similarity types assessed by participants. While selected to reduce the difficulty of this study and present exploratory findings on evaluation of non-text-based similarities, the reduced diversity and complexity of stimuli chosen limit the current results. 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. 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 effectively engaged with and utilized by designers.

## 5 CONCLUSION

This work investigates similarities based on non-textual features of inspirational stimuli, represented by distances in computational and psychological embedding spaces derived using deep learning and human-evaluated similarity assessments. Through a categorization task, similarity levels between stimuli that were categorized together or separated were associated with distances considered "near" and "far". Boundaries separating stimuli that are "too far" were observed, which include stimuli that were not grouped together at above chance levels. For similarities derived from psychological and computational embedding spaces, differences in the boundaries separating stimuli that are "too far" were observed. Human vs. AI-based boundaries are "farther" in visual similarity but "nearer" in functional similarity. Comparing across similarity types, the boundaries for stimuli that are "too far" in functional similarity are "nearer" than for visual similarity, implying that functionally related stimuli do not need to be as dissimilar as visually related stimuli to be "too far". Thus, depending on the type of similarity criteria involved, computed distances between stimuli have different implications with respect to perceptions of "near" and "far". Overall, this work presents an exploration of less explored, increasingly relevant, definitions of similarity of inspirational stimuli based on form and function-based attributes.

## ACKNOWLEDGMENTS

The authors thank Dr. Forrest Huang for the computational models used in this work. This work was supported by the National Science Foundation under grant 2145432 - CAREER.

## REFERENCES


Ahmed, F., Ramachandran, S., Fuge, M., Hunter, S. and Miller, S. (2019), "Interpreting Idea Maps: Pairwise Comparisons Reveal What Makes Ideas Novel", *J. Mech. Des.*, Vol. 141 No. 2, p. 021102, 10.1115/1.4041856.

Chan, J., Dow, S.P. and Schunn, C. (2015), "Do the Best Design Ideas (Really) Come from Conceptually Distant Sources of Inspiration?", *Des. Stud.*, Vol. 36, pp. 31–58, 10.1016/j.destud.2014.08.001.

Chaudhari, A.M., Bilionis, I. and Panchal, J.H. (2019), "Similarity in Engineering Design: A Knowledge-Based Approach", in: *Proc. of ASME IDETC/CIE*, Anaheim, CA, USA, p. V007T06A045, 10.1115/DETC2019-98272.

Cooke, T., Jäkel, F., Wallraven, C. and Bülthoff, H.H. (2007), "Multimodal similarity and categorization of novel, three-dimensional objects", *Neuropsychologia*, Vol. 45 No. 3, pp. 484–495, 10.1016/j.neuropsychologia.2006.02.009.

Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C. and Wood, K. (2013), "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, p. 021007, 10.1115/1.4023158.

Goucher-Lambert, K. and Cagan, J. (2019), "Crowdsourcing Inspiration: Using Crowd Generated Inspirational Stimuli to Support Designer Ideation", *Des. Stud.*, Vol. 61, pp. 1–29, 10.1016/j.destud.2019.01.001.

Jansson, D. and Smith, S. (1991), "Design fixation", *Des. Stud.*, Vol. 12 No. 21, pp. 3–11, 10.1016/0142-694X(91)90003-F.

Jiang, S., Hu, J., Wood, K.L. and Luo, J. (2022), "Data-Driven Design-By-Analogy: State-of-the-Art and Future Directions", *J. Mech. Des.*, Vol. 144 No. 2, p. 020801, 10.1115/1.4051681.

Jiang, S., Luo, J., Ruiz-Pava, G., Hu, J. and Magee, C.L. (2021), "Deriving Design Feature Vectors for Patent Images Using Convolutional Neural Networks", *J. Mech. Des*, Vol. 143 No. 6, p. 061405, 10.1115/1.4049214.

Kim, J. and Maher, M.L. (2023), "The effect of AI-based inspiration on human design ideation", *Int. J. Des. Creat. Innov.*, pp. 1–18, 10.1080/21650349.2023.2167124.

Kwon, E., Huang, F. and Goucher-Lambert, K. (2022a), "Enabling multi-modal search for inspirational design stimuli using deep learning", *AI EDAM*, Vol. 36, p. e22, 10.1017/S0890060422000130.

Kwon, E., Rao, V. and Goucher-Lambert, K. (2022b), "Investigating the roles of expertise and modality in designers' search for inspirational stimuli", in: *Proc. of ASME IDETC/CIE*, St. Louis, MO, USA, p. V006T06A015, 10.1115/DETC2022-90638.

Kwon, E., Rao, V. and Goucher-Lambert, K. (2023), "Exploring designers' encounters with unexpected inspirational stimuli", in: J.S. Gero (Editor), *Design Computing and Cognition'22*, Springer International Publishing, Cham, pp. 397–408, 10.1007/978-3-031-20418-0$_2$4.

Mo, K., Zhu, S., Chang, A.X., Yi, L., Tripathi, S., Guibas, L.J. and Su, H. (2019), "Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding", in: *2019 IEEE/CVF Conference on CVPR*, p. 909–918, 10.1109/CVPR.2019.00100.

Nandy, A. and Goucher-Lambert, K. (2022), "Do Human and Computational Evaluations of Similarity Align? An Empirical Study of Product Function", *J. Mech. Des.*, Vol. 144 No. 4, p. 041404, 10.1115/1.4053858.

Roads, B. and Mozer, M. (2019), "Obtaining psychological embeddings through joint kernel and metric learning", *Behav. Res.*, Vol. 51, pp. 2180–2193, 10.3758/s13428-019-01285-3.

Sarica, S., Song, B., Luo, J. and Wood, K. (2021), "Idea generation with technology semantic network", *AI EDAM*, Vol. 35, pp. 265–283, 10.1017/S0890060421000020.

Sio, U.N., Kotovsky, K. and Cagan, J. (2015), "Fixation or inspiration? a meta-analytic review of the role of examples on design processes", *Des. Stud.*, Vol. 39, pp. 70–99, 10.1016/j.destud.2015.04.004.

Tangelder, J. and Veltkamp, R. (2004), "A survey of content based 3d shape retrieval methods", in: *Proceedings Shape Modeling Applications, 2004.*, pp. 145–156, 10.1109/SMI.2004.1314502.

Zhang, Z. and Jin, Y. (2021), "Toward Computer Aided Visual Analogy Support (CAVAS): Augment Designers through Deep Learning", in: *Proc. of ASME IDETC/CIE*, Virtual, Online, p. V006T06A057, 10.1115/detc2021-70961.