Semantic properties of word prompts shape design outcomes: understanding the influence of semantic richness and similarity

Ananya Nandy[1], Monica Van[2], Jonathan Li[2], Kosa Goucher-Lambert[1], Matthew Klenk[2] and Shabnam Hakimi[2]
[1]University of California, Berkeley, [2]Toyota Research Institute
ananyan@berkeley.edu, shabnam.hakimi@tri.global

The process of design often involves translating abstract semantic information into visual artifacts. To understand how psycholinguistic properties of semantic prompts impact the design process, we conducted a behavioral study with 515 participants who were asked to design a chair that reflected a prompt word. During the task, participants made various aesthetic and functional choices to realize their design vision. We investigated how semantic richness (imageability and number of semantic neighbors) of prompts and semantic similarity between prompts impacted the characteristics of the final design. Results indicate that different outcome types (functional and geometric) are associated with semantic properties, while prompt similarity is related to outcome differentiability. The results reveal the complex relationship between semantic inputs and human-generated outputs in a creative context, with implications for how multi-modal (text, image, and 3D) systems should be built to complement design activities.

Introduction

Magical and *seamless*. *Ultimate adventure machine*. These words were used to describe the Apple Vision Pro spatial computer and Toyota Tacoma truck in their respective product launch press releases in 2023 [1, 2]. Beyond using words to describe functional performance, abstract words can give designers the power to influence product purchasing, use, and overall brand identity

Design Computing and Cognition DCC'24. J.S. Gero (ed),
pp. xx-yy.

[3]. In early stages of design, long before a product launch, such abstract words can guide inspiration for designers as they generate new ideas, which are later realized visually and/or physically. Different interpretations of these abstract words can lead to diverse outcomes, which is advantageous for spurring creative ideas and ways to embody them in design [4, 5]. Recently, it has become possible to leverage multimodal large language models, which computationally represent relationships between semantic and visual information, as a resource for creative problem solving. These models provide a means to directly generate design-related artifacts through semantic specification [6, 7]. However, different *types* of semantic information may shape the underlying psychological process that results in a design. Therefore, to understand what outputs designers expect or desire from multimodal models, it is first necessary to understand whether different types of semantic information elicit variation in human designers' outputs.

To investigate whether prompt word choice impacts design choices, we conducted a behavioral study focused on how humans translate abstract semantic information into visual artifacts during a design task. We address the question: *what is the relationship between the psycholinguistic properties of abstract words used as design prompts and their resulting outcomes?* We consider how inherent characteristics of semantic richness, 1) the ease with which a word evokes a mental image (*imageability*) and 2) the quantity of a word's contextual associations (*the number of semantic neighbors*) influence the generation of similar or dissimilar designs. Additionally, we examine 3) how the strength of words' associations with each other (*semantic similarity*) is reflected in the design output.

Related Work

The Use of Semantic and Visual Representations in Design

Connecting semantic and visual information is central to design practice. Significant effort in design research has been placed into product semantics, "the study of the symbolic qualities of man-made forms in the context of their use, and application of this knowledge to industrial design" [3]. Studies that seek to connect semantics with visual design representations often use methods such as semantic differential analysis or pairwise preference modeling, which rely on ratings and/or decisions to associate existing designs to the semantics in question [8]. At the other extreme, computational tools have been built to assist designers in turning vague ideas, or semantic intent, into visual designs, utilizing text [9, 10], images [11], or both [12].

In our study, we investigate how properties of the semantics being used for design influence outcomes of the creation process.

In addition to relying on the processing of semantic information, problem-solving involves manipulating visual material using the “visuo-spatial sketchpad” [13]. Spatial information, associated with movement or manipulation in the environment, and visual information, like color, is used to construct visual imagery. This process is particularly relevant for the creation of drawings as visual design representations [13]. For example, an analysis of sketches generated in response to a semantic keyword (a chair that portrays a “sad image”) demonstrates different paths taken to implement abstract ideas into design forms [14]. In one path, final forms arise from participants’ experiences, such as remembering body postures that express sadness. In another, participants iteratively create forms to represent sadness directly, rather than through more complex associations. When participants try to directly represent abstract concepts, mental searching can be limited to stereotypical forms, leading to outcomes that are less novel [14]. It follows, therefore, that word processing behavior may be linked different types of design outcomes.

Semantic Properties and their Impact on Cognition and Creativity

Understanding how semantic information is represented cognitively is relevant to design because words are used for both inspiring and communicating ideas (represented semantically and visually). The construct of semantic richness, consisting of dimensions such as the number of semantic neighbors, imageability, body-object interaction, and emotional valence, has been used to understand word processing behavior [15]. The properties that define such semantic richness are discussed below.

The concreteness of words impacts how they are processed. Higher concreteness (less abstract), often defined by the ability to be experienced by the senses, has been associated with enhanced recall, recognition, and comprehension [16]. A positive correlation has been found between word concreteness and imageability, which refers to the ability to generate images for the given word, making it relevant for contexts involving semantic-to-visual translation [17, 18]. Nonetheless, findings are mixed as to whether both visual and semantic information are helpful for predicting the processing of abstract words [19] or whether abstract words are implicitly visually grounded through their connections to concrete words [20]. Due to the importance of visual information in design and the open questions in this area, imageability is a word property considered in our study.

Additionally, abstract (e.g., *freedom*) vs. concrete (e.g., *bus*) words can differ in their representational structure in semantic memory. For instance, evidence shows that abstract words may be better organized by word associations compared to concrete words, which may be better organized by semantic similarity, even in the context of semantic-to-visual tasks [21, 22]. Semantic neighborhood density is suggested to impact behavior, with lexical decision reaction times found to be fastest for words with sparse neighborhoods [23]. For abstract concepts specifically, however, the existence of many semantic neighbors facilitates lower lexical decision or naming times [24]. Aside from standard word comprehension and recall, semantic richness of words may relate to creative thinking. Ideas generated when provided a semantically rich word, defined by having many semantic associations, have been found to be higher in quantity but less creative in quality [4]. Furthermore, the associative theory of creativity suggests that creative combinations can be developed by connecting concepts that are far apart, particularly through semantic memory structured by similarity [25]. At the same time, the importance of semantic memory structure may be more relevant to verbal (i.e., concerned with semantics) rather than figural (i.e., concerned with images) creativity [26]. In our study, therefore, *semantic similarity*, which is based on the embedding of words in the context of other words, is considered along with two different dimensions of *semantic richness*: imageability and the number of semantic neighbors.

Experimental Design

A behavioral study was conducted to investigate the process of translating abstract concept words into 3D designs. The study consisted of a task, where participants were presented with a word and responded with a sequence of actions, and a post-task survey. We tested two hypotheses: 1) Features of design outcomes can predict linguistic properties of prompts (imageability and the number of semantic associations), and 2) Similar prompt words result in similar designs.

Participants

Data was collected online from 540 US-based participants, recruited from Prolific [27]. Eligible participants (95%+ approval rate, 500+ previous submissions) were not required to have design experience, only an interest in the study. Pre-analysis data filtering left 515 participants for data analysis (gender identification: 317 men, 180 women, 5 non-binary or other, 13 prefer not to answer or not applicable). Participants were compensated at a rate of $15/hr for their time (M = 24.33 minutes, SD = 16.08 minutes). The methods and procedures used in this study were approved by the WCG IRB.

Design Task and Interface

During the study, each participant was presented with a word and asked to use a custom online interface to design a chair that reflected that word. Participants completed a practice task to get accustomed to the interface (*Design a creature that is cute)*. Then, they were presented with their actual task: *Design a chair that is [sleek, modern, etc.]*.

The participants first had to select one of four starting points, which were discrete combinations of what we refer to as function and geometry: static and multiple pieces (most closely reflecting a "prototypical" four-leg chairs), static and singular piece (a change in geometry), moving and singular piece (a change in function), or moving and multiple pieces (a change in function and geometry).

Once they selected the starting point, they moved to a screen where all the other possible actions were available to them. Importantly, participants were able to freely explore their options, change their design fully, or modify it slightly without any time constraints. They could make a series of discrete or continuous changes to the design, related to its function (able to move vs. static), geometry (multiple pieces vs. a singular piece, size, or shape), and perceptual (color, texture). The full task structure is shown in Fig. 1.

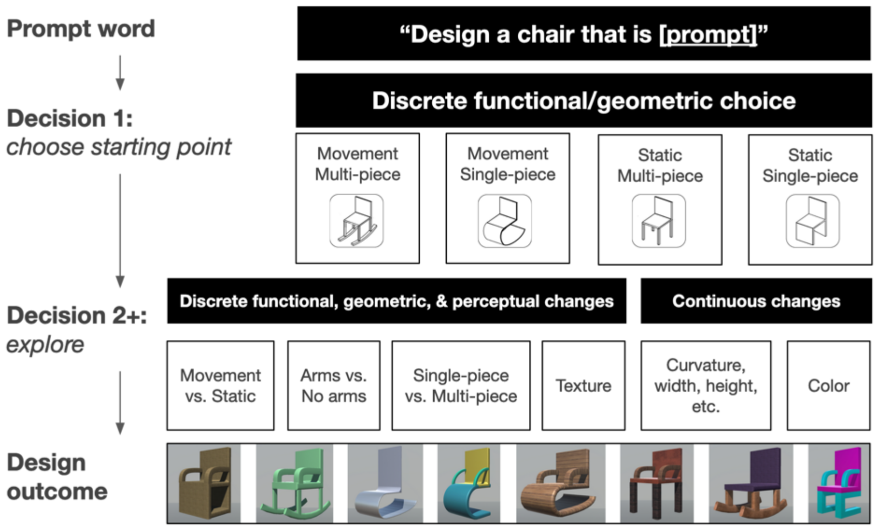


Fig. 1 Task structure and examples of outcomes from participants

Some changes, like color and texture, could be changed for the whole design or for individual parts. The goal was to provide choices that could map to higher-level intentions but also enable sufficient creative freedom. The available choices were not exhaustive but were specified based on different types of chairs available in online 3D model repositories (e.g., the chair category of the ShapeNet dataset [28]). Fig. 2 shows the interface.

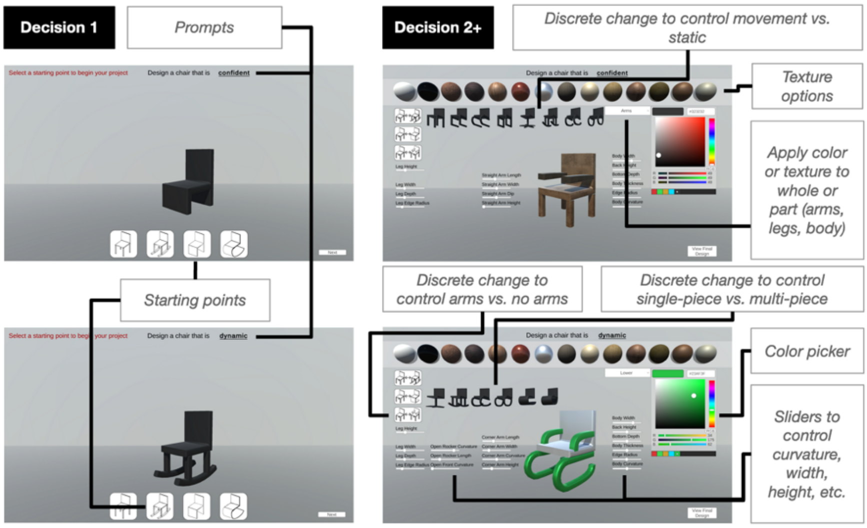


Fig. 2 Interface for completing task

Once they were finished, participants were asked to rate (from 1 to 7) and describe their satisfaction with their design, indicating both the extent to which it reflected the prompt (*How well does your final design satisfy the prompt "design a chair that is [sleek, modern, etc.]"?*) as well as their preference (*How satisfied are you with your final design?*). Participants also had the option to start the task over and create one or more additional chairs after submitting their first design. All actions and responses were recorded, as well as a constant-angle screenshot of the participants' final designs. The task was deployed via Qualtrics [29] survey with an embedded web application that was developed in Unity, using the Unity Experiment Framework [30] and the Archimatix [31] parametric modeling extension. Data from the task was sent to a database in Amazon Web Services. Participants also completed post-task questions about their engagement in the task, demographics, occupation, and creativity. Here, only outcomes (screenshots and recorded final states) from the first session (no ratings, survey responses, additional sessions, or actions) will be discussed.

Semantic Richness and Similarity of Prompt Words

The words presented as prompts in our study were primarily selected from words that had previously been associated with product semantics in a design context, for items such as furniture, household items, or even cars [32 – 38]. A subset of 12 relevant words was selected to encompass a diverse range of values for two item-level linguistic properties (imageability and the number of semantic neighbors) and a pairwise semantic similarity.

Table 1 Word prompts and linguistic properties. Properties below the median for each measure are labeled "*low*" and properties above the median are labeled "*high*."

Word	Imageability Rating	# of Semantic Neighbors	Number of Participants
Versatile	3.04 *(low)*	29 *(low)*	43
Dependable	3.33 *(low)*	0 *(low)*	42
Sleek	3.50 *(low)*	0 *(low)*	44
Modern	3.51 *(low)*	9031 *(high)*	43
Dynamic	3.73 *(low)*	5278 *(high)*	42
Confident	3.77 *(low)*	866 *(low)*	44
Unique	3.79 *(high)*	8313 *(high)*	43
Exciting	3.96 *(high)*	1922 *(high)*	42
Sophisticated	4.05 *(high)*	3999 *(high)*	41
Friendly	4.32 *(high)*	6590 *(high)*	43
Cheerful	4.58 *(high)*	0 *(low)*	44
Graceful	4.83 *(high)*	0 *(low)*	44
Median	3.78	1394	*Total* 515
Mean	3.87	3002	

Table 1 shows the prompt words and their individual properties. The item-level linguistic properties were obtained from the South Carolina Psycholinguistic Metabase (SCOPE Version 1.1) [39]. The number of semantic neighbors is calculated using a threshold distance that defines whether another word is considered a neighbor to the target word. The threshold is determined by the mean and standard deviation of a distribution of the interword distance of many randomly sampled word pairs [40, 41]. Imageability refers to the degree of effort involved in generating a mental image of a concept, ranging from 1 to 7 where lower numbers indicate words that are less imageable [42]. In the context of the full database, for example, *yellow* and *oligarchy* are words with high and low values for both properties, respectively, while *accordion* is an example with few semantic neighbors but high imageability. The word prompts in this study varied in imageability within a limited range as a minimum degree of abstractness was necessary for the words to be utilized for the design task.

In comparison to the linguistic properties, which are determined for each individual word, semantic similarity is calculated for prompt pairs. Including a range of semantic similarity values can expose how the differences in the contextual meanings of words might manifest themselves in outcomes. Fig. 3 shows the pairwise semantic similarity using the 12 possible prompt words provided to participants. Semantic similarity can be defined in various ways. In this this case, semantic similarity was calculated using a pre-trained transformer model (all-mpnet-base-v2) which creates sentence embeddings through training on large amounts of text [43]. The sentence embeddings for *"A chair that is [prompt word]"* were calculated using the *sentence transformers* library in Python [44]. The semantic similarity was then defined using the cosine similarity between embeddings.

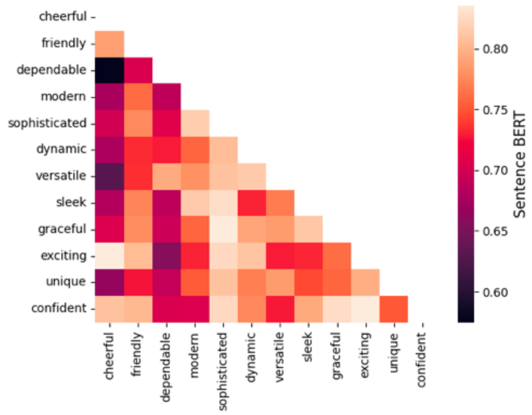


Fig. 3 Pairwise semantic similarity of prompts based on sentence embeddings

Design Outcome Characterization and Modeling Procedure

The design outcomes were parameterized using features from the recorded final states and screenshots. These outcome features were linked to each semantic property. The outcome feature extraction and analysis methods are described in the following sections and summarized in Table 2.

To properly evaluate our research questions, we first filtered our data as follows. Incomplete submissions (no completion code or missing data recorded from the interface) were removed. Then, another round of filtering was done to remove participants who had missing or improperly captured screenshots (e.g., blank images). Additionally, while most participants only completed a single session of the task, some completed multiple (i.e., chose to create another chair that reflected the same prompt). In the latter case, only the participant's first session was included in the analysis. Table 1 shows the number of participants in each condition after pre-processing.

Table 2 Summary of analyses

	Input Variables	Output Variable		Model
Richness	Style, Vibrance	Imageability	Numeric	Linear
Richness	Style, Vibrance	# of Semantic Neighbors	Numeric (count)	Zero-inflated Negative Binomial
Similarity	Image Features from ViT	Classification Accuracy	Numeric (proportion)	Support Vector Machine (SVM)
		Correlation Variables		
Similarity		Semantic Similarity	Classification Accuracy	

Relationship with Semantic Richness

Extracting Features from Design Outcomes Two features were used to model the relationship between the outcomes and individual word properties. The first is referred to as "style," addressing the discrete functional/geometric differences across outcomes that were enabled by the task structure. "Style" was modeled as a categorical variable with four levels, corresponding to the four starting points available to participants (changes to the starting point could be made in the exploration phase). The reference level for this variable is Style 2 (movement, single-piece), the most common outcome style. The second feature is vibrance, a continuous variable addressing perceptual characteristics of outcomes. Vibrance is defined by the variance of color channels of the final design screenshot. This feature was extracted from the screenshots using the *pliers* library in Python [45] and standardized. Examples of these features are shown in Fig. 4.

Connecting Design Outcomes with Dimensions of Semantic Richness Two separate models were developed, using the categorical and continuous outcome features (style and vibrance) and their interactions as predictors and the semantic properties as outcomes. Though original individual ratings were conducted on an ordinal scale, imageability is modeled using a linear model because the values were reported as norms. The number of semantic neighbors is modeled as count data, using a negative binomial model due to over-dispersion, with zero-inflation. Including zero-inflation entails

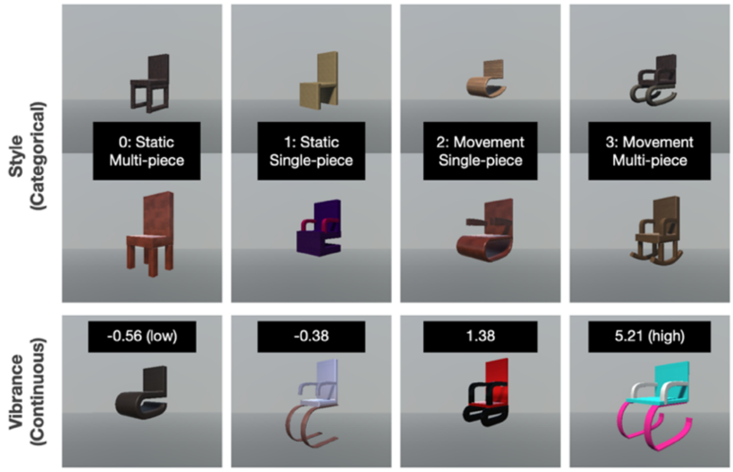


Fig. 4 Example of style and vibrance features from participants' designs

modeling excess zero counts with a logit model and the remaining zero counts with a count model. Here, several different words have zero semantic neighbors, a possible excess because of the threshold at which the neighbors were determined. Thus, zero counts of semantic neighbors may be generated by different processes [46]. The models were implemented using R and the *pscl* library [47] and are reported with no mathematical corrections.

Relationship with Semantic Similarity

Extracting Features from Design Outcomes Participants' outcome screenshots were used directly to understand the relationship between the outcomes and the semantic similarity between pairs of words. Image features were extracted from the screenshots using the Vision Transformer (ViT) used in the CLIP model [48]. The features were extracted using the *sentence transformers* library in Python (clip-ViT-B-32).

Connecting Design Outcomes with Semantic Similarity A Support Vector Machine (SVM) classifier was trained using the extracted image features after standardization in a one vs. one scheme for each pair of word prompts (83 – 88 data points per classifier). An SVM finds a hyperplane that best separates the classes in the data. For each SVM, 75% of the data was used for training and 25% was used for testing, with stratification of the two classes. A grid search cross-validation was used on the training data to set the SVM hyperparameters (kernel, C, gamma) in each case. The models were implemented using the *scikit-learn* library in Python [49]. Then, a correlation was used to investigate the relationships between semantic similarity and classification accuracy, which represents the ease of distinguishing the outcome images that resulted from pairs of prompt words.

Results

The relationship between outcomes and semantic richness dimensions

Imageability Model The results of the model predicting imageability are shown in Table 4 and visualized in Fig. 6. At mean vibrance, each alternate style type is a significant (alpha level of 0.05) predictor of imageability compared to the most selected style, Style 2 (movement, single-piece). Style 3 (movement, multi-piece) demonstrates a positive relationship with imageability, indicating that participants more likely selected this style in response to prompts with higher imageability. On the other hand, Styles 0 (static, multi-piece) and 1 (static, single-piece) have a negative relationship with imageability, indicating that participants more likely selected these styles given less imageable prompts. There is no evidence that vibrance in the outcome images or interactions with vibrance are predictors of imageability. Overall, participants often designed chairs more similar to a "prototypical" chair – Styles 0 and 1 (both static) – when faced with less imageable prompts (e.g., *dependable*), moving away from these when imageability was higher (e.g., *cheerful*).

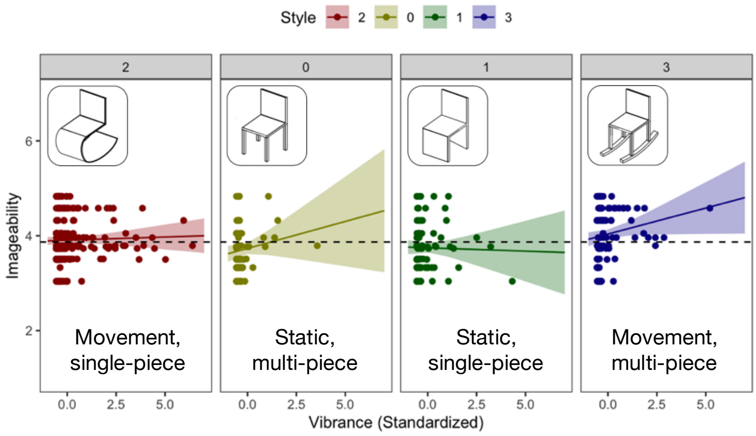


Fig. 6 Marginal effects plots for imageability model. Each style, compared to the reference Style 2, predicts imageability at mean vibrance (0). There is no evidence for the interaction between style and vibrance predicting prompt imageability.

Table 4 Parameter estimates relating outcome features and prompt imageability

	Imageability	
Predictors	*Estimates*	*p*
Intercept	**3.91 (3.84 – 3.97)**	**<0.001*****
Style [2]	*Reference*	
Style [0]	**-0.17 (-0.31 – -0.04)**	**0.011***
Style [1]	**-0.16 (-0.28 – -0.04)**	**0.007****
Style [3]	**0.13 (0.02 – 0.24)**	**0.027***
Vibrance	0.01 (-0.04 – 0.07)	0.612
Style [0] × Vibrance	0.10 (-0.08 – 0.28)	0.288
Style [1] × Vibrance	-0.03 (-0.16 – 0.11)	0.687
Style [3] × Vibrance	0.10 (-0.02 – 0.22)	0.114
Observations	515	
R^2	0.064	*$p \le .05$, ** $p \le .01$, *** $p \le .001$

Semantic Neighbors Model The results of the model predicting the number of semantic neighbors are shown in Table 5 and visualized in Fig. 7. Style 1 (static, single-piece) is a statistically significant predictor for the count

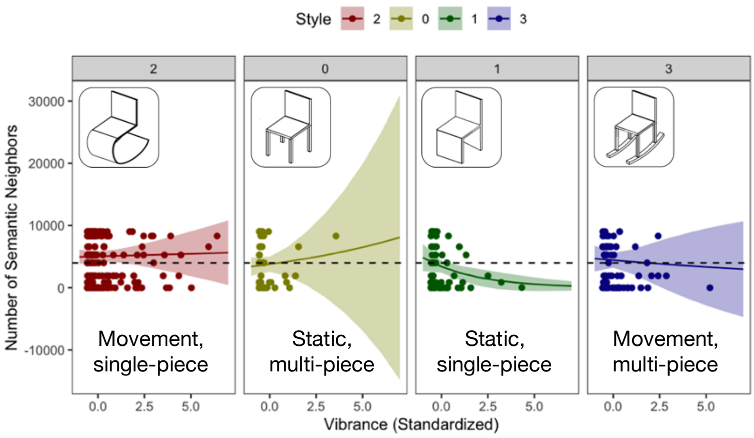


Fig. 7 Marginal effects plots for the semantic neighbors count model. Style 1, compared to the reference Style 2, predicts the number of semantic neighbors at mean vibrance (0).

portion of the model, indicating that at a mean vibrance, Style 1 outcomes tend to be associated with words with fewer semantic neighbors than Style 2 (movement, single-piece) outcomes. The intercept shows a positive effect of the reference Style 2 (movement, single-piece) on the number of semantic neighbors. The zero-inflated portion of the model distinguishes the prediction of zero vs. non-zero semantic neighbors. A Style 2 (movement, single-piece) outcome at mean vibrance (represented by the intercept) thus has a negative effect on predicting a word with zero semantic neighbors (*dependable, sleek, cheerful, graceful)*. Unlike imageability, which is related to each different style, prediction of semantic neighbor count appears to influenced primarily by two styles, which differ in "function" (there are also visual differences due to this functional change).

Table 5 Zero-inflated negative binomial model parameter estimates relating outcome features and number of semantic neighbors

	Number of Semantic Neighbors			
	Count		Zero-Inflation	
Predictors	*Estimates*	*p*	*Estimates*	*p*
Intercept	**8.53 (8.36 – 8.70)**	**<0.001***	**-0.80 (-1.07 – -0.53)**	**<0.001***
Style [2]		*Reference*		
Style [0]	-0.29 (-0.65 – 0.08)	0.120	0.20 (-0.43 – 0.82)	0.541
Style [1]	**-0.40 (-0.73 – -0.07)**	**0.019***	0.21 (-0.31 – 0.71)	0.425
Style [3]	-0.13 (-0.45 – 0.19)	0.421	0.16 (-0.34 – 0.66)	0.526
Vibrance	0.02 (-0.12 – 0.15)	0.818	-0.16 (-0.42 – 0.11)	0.252
Style [0] × Vibrance	0.09 (-0.32 – 0.50)	0.659	-0.31 (-1.34 – 0.71)	0.548
Style [1] × Vibrance	-0.37 (-0.77 – 0.02)	0.065	-0.10 (-0.77 – 0.56)	0.767
Style [3] × Vibrance	-0.07 (-0.46 – 0.31)	0.716	0.32 (-0.20 – 0.84)	0.231
Observations	515		**p ≤ .05, ** p ≤ .01, *** p ≤ .001*	

Semantic richness does not fully account for outcome variation

Although imageability and the number of semantic neighbors explore different facets of semantic richness, word meaning is defined by the interaction between these properties and others. Fig. 8 shows outcome styles and vibrances across both imageability and the number of semantic neighbors by coarsely grouping them above and below median values. Distinct words, even considering the interaction between dimensions, elicit the generation of a large variety of outcome designs despite the style-related trends observed in our study. For example, the image vibrance range is often large for high imageability words, but this is not consistent across words within the same group. Therefore, the outcome vibrance may distinguish specific words, but not *types* of words. Similarly, for some words (e.g., *graceful*), outcomes are spread evenly across all style categories, while for others (e.g., *friendly*), a single style appears to dominate. Therefore, it is evident that factors that differentiate these individual words beyond semantic richness may impact design outcomes.

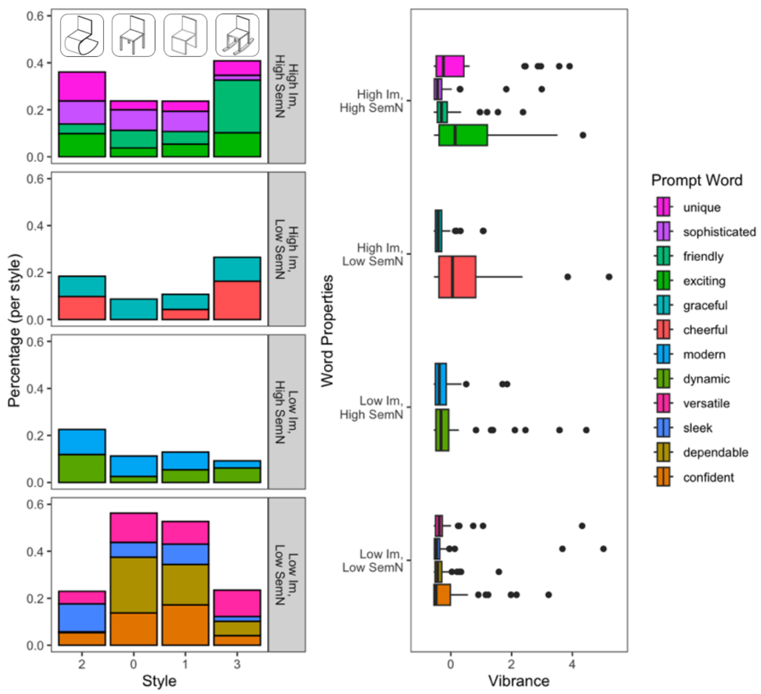
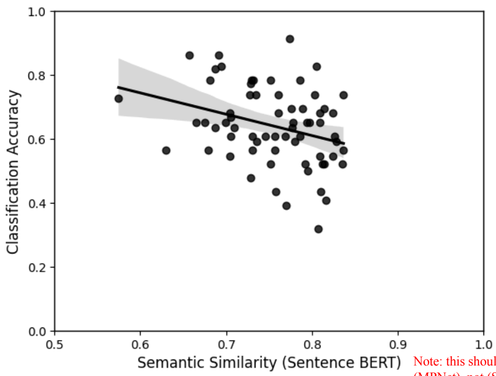


Fig. 8 Features grouped by properties below (low) or above (high) the median value.

The relationship between semantic similarity and outcome distinctiveness

Semantic similarity differentiates the meaning of words in comparison to each other, therefore providing a more holistic picture of the impact of design prompts on outcomes. Fig. 9 shows the relationship between the pairwise semantic similarity of prompts and the ability for the paired outcome images to be separated using the SVM classifier (represented by accuracy). Semantic similarity, at least as defined using the sentence embedding method, is correlated with the classification accuracy (r_p =-0.31, p = 0.011). The correlation is negative, implying that as the semantic similarity of the word prompts increases, it is more difficult to distinguish which human-generated outcomes resulted from which prompt. This result demonstrates preliminary evidence that the contextual relationship of abstract words to each other is reflected in design output, with word prompts with similar meanings more likely to result in similar designs and vice versa.



Note: this should be labeled as (MPNet), not (Sentence BERT)

Fig. 9 Prompt similarity is correlated with classifier accuracy (r_P=-0.31, p = 0.011).

Discussion

Semantic information plays a role in various design activities, from communication to inspiration, and is becoming increasingly important as the interface between humans and creativity-augmenting computational tools. This study leverages a novel task to resolve how the semantic properties of design prompts relate to outcomes. We find preliminary evidence that characteristics of the prompt can meaningfully impact outcomes.

Types of design outcomes can predict semantic richness dimensions

We found that some classes of design outcomes were relevant for predicting imageability or the number of semantic neighbors, the dimensions of semantic richness considered in our study. High semantic richness has previously been associated with lower creative quality, but better creative fluency [4]. However, this relationship may not be applicable in the context of figural creativity [26]. In our study, high imageability words were positively associated with styles that varied from a static chair with four legs (what we refer to as a "prototypical" design). Similarly, lower numbers of semantic neighbors – particularly zero – were negatively associated with styles that varied from a "prototypical" design (though in general, the number of semantic neighbors could not be easily predicted). Thus, during visual creation, a lack of associations with other words (low semantic richness) could prevent people from considering more unusual forms. Additionally, a study of keyword-driven sketches theorized that designers decompose difficult-to-visualize words into associated words that are more easily related to images [14]. Feature data considering the interaction of the two semantic richness dimensions shows that Style 2 (movement, single-piece) was a common outcome type in response to words with low imageability, but with a high number of semantic neighbors. Therefore, it is possible that in the face of particularly hard-to-visualize words, greater semantic richness can help lead people down paths away from the image of a conventional chair. Neither imageability nor the number of semantic neighbors showed a consistent relationship with vibrance, the perceptual characteristic outcomes, though variation in vibrance was inconsistent across prompt words, which merits further consideration.

Increasingly similar word prompts are associated with harder-to-distinguish design outcomes

We found that semantic similarity is correlated with how accurately a classifier can separate the design images. Increasing semantic similarity relates to worse classifier accuracy, likely induced by decreasing divergence in perceptual features of the outcome images. This result demonstrates that, in addition to linguistic properties, the meaning and relationship between prompts can be leveraged to incite varying outcomes, even when they are presented independently of each other. Additional analysis could investigate if semantically similar and dissimilar word prompts can also be distinguished by features like outcome style or vibrance.

Implications for Design and Future Directions

The results shed light on how types of abstract word prompts might elicit subtle but predictable differences for the same object (e.g., a chair). Yet, even different word prompts, if they are similar in meaning, can result in converging outputs. Collectively, these findings have implications for aligning human intentions with computational systems controlled by natural language, such as text-to-image systems, particularly for design tasks. Some types of words (e.g., *confident or dependable*), words with a low number of semantic neighbors, might easily result in expected or desired outcomes due to common understanding of how the words "look." However, diverse outputs may be expected when specifying a semantically dissimilar modifier for the same object. Further investigation into how outcomes from the human-led text-to-3D process compare to outputs from multi-modal generative models (and whether the outcome diversity is adequate) is warranted. With the rise of language-based interfaces for generating visual outputs, carefully examining how input language and expected outputs might differ for a general-purpose use compared to a creative context, if at all, is critical for allowing them to be used in designer workflows.

Limitations

The study has some limitations that should be noted. Participants were not given full freedom over chair geometry and had to use our specified decision making structure and interface, which does not reflect the open-endness of real design processes. This was purposeful to accommodate non-experts and to understand the impact of semantic prompts in a controlled way. However, due to the nature of the study (online with recruitment from a crowdsourcing website), some participants' decisions may have been impacted by the interface and motivation in ways that were difficult to capture (e.g., minimizing actions taken to finish the task or selecting arbitrarily). Given more freedom and perhaps different motivations for completing the task, experienced designers may demonstrate greater variation in response to the prompts. More broadly, design attributes likely also vary based on the use context and the designer's cultural background [32]. In this study, we did not indicate a specific context and only English-language words were considered. Using our paradigm while considering different languages might provide further insights into cultural influences on design outcomes.

Conclusion

In this study, we investigated the impact of abstract word prompts and their psycholinguistic properties on outcomes from a creative design task. During the task, participants were given a set of functional, geometric, and aesthetic choices to create a chair that reflected an abstract word. The results indicate that outcome types are sometimes predictive of prompt word types. Certain outcome styles tended to predict higher and lower imageability ratings, as well as lower numbers of semantic neighbors. Individual words were also associated with common styles and/or vibrance. Finally, there was evidence that semantic similarity of abstract prompts was related to similarity of resulting visual outputs, with similar word prompts resulting in outcomes that were harder to tell apart. Further research into the translation between abstract language and visual design artifacts can help 1) inform the types of guiding words that may elicit more or less diverse design outcomes and 2) ensure that language-controlled generative AI tools for supporting design activities provide sufficiently diverse or similar outcomes as needed.

References

1. Apple. (n.d.). https://www.apple.com/apple-vision-pro/
2. Toyota Newsroom. (n.d.). https://pressroom.toyota.com/2024-toyota-tacoma-is-the-ultimate-adventure-machine/
3. Krippendorff, K, & Butter, R. (2008). Semantics: Meanings and contexts of artifacts. Product Experience, 353-376.
4. Beaty, RE., Kenett, YN., Hass, RW, & Schacter, DL. (2023). Semantic memory and creativity: The costs and benefits of semantic memory structure in generating original ideas. Thinking & Reasoning, 29(2), 305-339.
5. Nandy, A, & Goucher-Lambert, K. (2023). Adaptive Optimization of Subjective Design Attributes: Characterizing Individual and Aggregate Perceptions. IDETC (pp. V006T06A001).
6. Liu, V., Vermeulen, J., Fitzmaurice, G., & Matejka, J. (2023). 3DALL-E: Integrating text-to-image AI in 3D design workflows. DIS (pp. 1955-1977).
7. Makatura, L, Foshey, M, Wang, B, HahnLein, F, Ma, P, Deng, B, . . . & Matusik, W. (2023). How Can Large Language Models Help Humans in Design and Manufacturing? arXiv preprint arXiv:2307.14377.
8. Crilly, N, Moultrie, J, & Clarkson, JP. (2004). Seeing things: consumer response to the visual domain in product design. Design Studies, 25(6), 547-577.
9. Zuo, H, Jing, Q, Song, T, Sun, L, Childs, P, & Chen, L. (2022). WikiLink: An Encyclopedia-Based Semantic Network for Design Creativity. Journal of Intelligence, 10(4), 103.
10. Sano, S, & Yamada, S. (2022). D-Graph: AI-Assisted Design Concept Exploration Graph. International Conference on Computational Creativity.
11. Koch, J, Lucero, A, Hegemann, L, & Oulasvirta, A. (2019). May AI? Design Ideation with Cooperative Contextual Bandits. CHI (pp. 1-12).
12. Koch, J, Taffin, N, Lucero, A, & Mackay, WE. (2020). SemanticCollage: Enriching Digital Mood Board Design with Semantic Labels. DIS (pp. 407-418).
13. Purcell, TA, & Gero, JS. (1998). Drawings and the design process: A review of protocol studies in design and other disciplines and related research in cognitive psychology. Design Studies, 19(4), 389-430.
14. Nagai, Y, & Noguchi, H. (2003). An experimental study on the design thinking

process started from difficult keywords: Modeling the thinking process of creative design. Journal of Engineering Design, 14(4), 429-437.
15. Yap, MJ, Lim, GY, & Pexman, PM. (2015). Semantic richness effects in lexical decision: The role of feedback. Memory & Cognition, 43, 1148-1167.
16. Nelson, DL, & Schreiber, TA. (1992). Word concreteness and word structure as independent determinants of recall. Journal of Memory & Language, 31(2), 237-260.
17. Christian, J, Bickley, W, Tarka, M, & Clayton, K. (1978). Measures of free recall of 900 English nouns: Correlations with imagery, concreteness, meaningfulness, and frequency. Memory & Cognition, 6, 379-390.
18. Altarriba, J, Bauer, LM, & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. Behavior Research Methods, Instruments, & Computers, 31, 578-602.
19. Gunther, F, Petilli, MA, Vergallito, A, & Marelli, M. (2022). Images of the unseen: Extrapolating visual representations for abstract and concrete words in a data-driven computational model. Psychological Research, 86(8), 2512-2532.
20. Shahmohammadi, H, Heitmeier, M, Shafaei-Bajestan, E, Lensch, H, & Baayen, H. (2022). Visual grounding of abstract and concrete words: A response to Gunther et al. (2020). arXiv preprint arXiv:2206.15381.
21. Crutch, SJ, & Warrington, EK. (2005). Abstract and concrete concepts have structurally different representational frameworks. Brain, 128(3), 615-627.
22. Dunabeitia, JA, Aviles, A, Afonso, O, Scheepers, C, & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual-world paradigm. Cognition, 110(2), 284-292.
23. Reilly, M, & Desai, RH. (2017). Effects of semantic neighborhood density in abstract and concrete words. Cognition, 169, 46-53.
24. Recchia, G, & Jones, MN. (2012). The semantic richness of abstract concepts. Frontiers in Human Neuroscience, 6(315).
25. Kenett, YN. (2019). What can quantitative measures of semantic distance tell us about creativity? Current Opinion in Behavioral Sciences, 27, 11-16.
26. He, L, Kenett, YN, Zhuang, K, Liu, C, Zeng, R, Yan, T, . . . & Qiu, J. (2020). The relation between semantic memory structure, associative abilities, and verbal and figural creativity. Thinking & Reasoning, 27(2), 268-293.
27. Prolific. (n.d.). www.prolific.com
28. Chang, AX, Funkhouser, T, Guibas, L, Hanrahan, P, Huang, Q, Li, Z, . . . Yu, F. (2015). ShapeNet: An Information-Rich 3D Model Repository. arXiv preprint arXiv: 1512.03012.
29. Qualtrics. (n.d.). https://www.qualtrics.com
30. Brookes, J, Warburton, M, Alghadier, M, Mon-Williams, M, & Mushtaq, F. (2020). Studying human behavior with virtual reality: The Unity Experiment Framework. Behavior Research Methods, 52, 455-463.
31. Archimatix. (n.d.). https://www.archimatix.com/
32. Liberman-Pincu, E., Korn, O, Grund, J., van Grondelle, E, & Oron-Gilad, T. (2023). Designing Socially Assistive Robots: Exploring Israeli and German Designers' Perceptions. arXiv preprint arXiv:2305.00419.

33. Palacios-Ibanez, A, Navarro-Martinez, R, Blasco-Esteban, J, Contero, M, & Camba, JD. (2023). On the application of extended reality technologies for the evaluation of product characteristics during the initial stages of the product development process. Computers in Industry, 144, 103780.
34. Hung, W, & Chen, LL. (2009). Exploring contradictory meanings in product semantics. DeSForM 2009.
35. Gan, Y, & Hirai, Y. (2017). Culture Reference Product Design—Case Study of Hans J. Wegner China Chair Design. Culture, 59, 30-41.
36. Yumer, ME, Chaudhuri, S, Hodgins, JK, & Kara, LB. (2015). Semantic shape editing using deformation handles. ACM TOG. 34, pp. 1-12.
37. Lin, L, Guo, G, & Xu, N. (2019). User-perceived styling experience of smart vehicles: a method to combine eye tracking with semantic differences. IET Intelligent Transport Systems. 13, pp. 72-78.
38. Khalaj, J, & Pedgley, O. (2014). Comparison of semantic intent and realization in product design: A study on high-end furniture impressions. International Journal of Design, 8(3).
39. Gao, C, Shinkareva, SV, & Desai, RH. (2023). Scope: the south carolina psycholinguistic metabase. Behavior research methods, 55(6), 2853-2884.
40. Shaoul, C, & Westbury, C. (2006). Word frequency effects in high-dimensional co-occurence models: A new approach. Behavior Research Methods, 38(2), 190-195.
41. Shaoul, C, & Westbury, C. (2010). Exploring lexical co-occurence space using HiDEx. Behavior Research Methods, 42(2), 393-413.
42. Graves, WW, Desai, RH, Humphries, C, Seidenberg, MS, & Binder, JR. (2010). Neural systems for reading aloud: a multiparametric approach. Cerebral Cortex, 20(8), 1799-1815.
43. Song, K, Tan, X, Qin, T, Lu, J, & Liu, TY. (2020). Mpnet: Masked and permuted pre-training for language understanding. Advances in Neural Information Processing Systems, 33, pp. 16857-16867.
44. Reimers, N, & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. EMNLP.
45. McNamara, Q, De La Vega, A, & Yarkoni, T. (2017). Developing a comprehensive framework for multimodal feature extraction. SIGKDD (pp. 1567-1574).
46. Young, DS, Roemmele, ES, & Yeh, P. (2022). Zero-inflated modeling part I: Traditional zero-inflated count regression models, their applications, and computational tools. Wiley Interdisciplinary Reviews: Computational Statistics, 14(1), e1541.
47. Zeileis, A, Kleiber, C, & Jackman, S. (2008). Regression models for count data in R. Journal of Statistical Software, 27(8), 1-25.
48. Radford, A, Kim, J, Hallacy, C, Ramesh, A, Goh, G, Agarwal, S, . . . Sutskever, I. (2021). Learning transferable visual models from natural language supervision. ICML (pp. 8748-8763).
49. Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, 12, 2825-2830.