UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Sketching with generative AI: verbal but not visual inspiration mitigates cognitive fixations

Permalink

https://escholarship.org/uc/item/5nq1t9q1

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors

Liu, Yaxin
Kay, Maxwell S
Green, Adam
et al.

Publication Date

2025

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Sketching with generative AI: verbal but not visual inspiration mitigates cognitive fixations

Yaxin Liu (yaxin.liu@georgetown.edu)[1]
Maxwell Kay (mk2470@georgetown.edu)[1]
Adam Green (aeg58@georgetown.edu)[1]
Roger Beaty (rebeaty@psu.edu)[2]

[1]Department of Psychology, Georgetown University, 20057
[2] Department of Psychology, Penn State University, 16802

Abstract

Symbolic visual sketching is a hallmark of human creativity, enabling the externalization of abstract concepts through figurative representations. Yet, creative expression can be constrained by pervasive conceptual associations—culturally learned mappings between abstract ideas and standard visual forms (e.g., a dove symbolizing peace). Generative AI has the potential to liberate such fixations due to AI's access to a broad range of content and ideas, but it remains unclear whether and how inspiration from verbal or visual modalities better mitigates fixations. Here, we hypothesized that the verbal modality induces greater conceptual divergence than the visual modality by bypassing perceptual constraints, whereas the visual modality may reinforce perceptually familiar mapping of visual representations. Participants generated sketches of abstract concepts (e.g., "time") before and after receiving GPT-4-generated verbal or visual inspiration. Drawings were analyzed using deep neural networks—by comparing perceptual features (VGG16-based) and semantic-perceptual content (CLIP-based)— as well as both human and GPT-4 scoring for creativity. We found that verbal inspiration significantly increased semantic distance and uniqueness, whereas visual inspiration led to minimal semantic divergence from the initial sketches. Importantly, low-level perceptual features remained unchanged across conditions, indicating that verbal prompts primarily influenced high-level conceptual framing of the sketches rather than their visual features. These findings demonstrate the effect of modality on mitigating cognitive fixations, with the verbal modality enhancing more unconventional visual sketching.

Keywords: creativity, cognitive fixation, hybrid-intelligence, large language models, GPT

Introduction

From ancient cave paintings to modern emojis, visual representations have served as a universal medium for expressing abstract concepts (Fan et al., 2018; Hoffman et al., 2015). Schematic symbols—such as mathematical notations (e.g., ∞ for infinity) or cultural icons (e.g., a dove for peace)—enable humans to externalize complex ideas through visual shorthand (Tversky, 2011). Whereas concrete symbols (e.g., an apple's outline) map directly onto real-world objects, abstract symbols (e.g., a dove for peace) rely on culturally learned associations between form and meaning (Uttal & Yuan, 2014). Although these associations facilitate efficient communication, they also contribute to cognitive fixations by reinforcing familiar visual-semantic mappings at the expense of alternative representations (Olsen & Tylén, 2023; Ward, 1994).

Cognitive fixations often arise from both perceptual and cultural learning. Repeated exposure to culturally dominant symbols (e.g., a red cross for medical service) strengthens automatic associations between concepts and their canonical visual representations (Barsalou, 1999; Wang et al., 2023). Over time, this process suppresses alternative representations, thereby making novel or unconventional mappings more challenging (Jansson & Smith, 1991). Breaking free from such fixations requires the ability to decouple semantic meaning from visual anchoring (Chrysikou et al., 2016; Chrysikou & Weisberg, 2005).

The picture-superiority effect—where images are generally better remembered than words—presents a paradox for creativity (Hockley, 2008). Images' enhanced memorability may arise from their dual encoding in both verbal and visual systems (i.e., dual-coding theory; Paivio, 1971). Furthermore, humans are more likely to label a picture than to imagine a word's referent, so pictorial stimuli often benefit from both visual and verbal memory traces (Cavazzana & Bolognesi, 2020; Glaser, 1992). Although this redundancy strengthens recall, it may also reinforce familiar associations when images depict conventional forms. In other words, while visual representations are more resistant to forgetting, they can become anchors that limit creative exploration by tethering new ideas to well-established or canonical visual forms.

Prior research has shown that creative generation is often constrained by example stimuli. For instance, Smith and colleagues (1993) demonstrated that when participants were shown pictorial examples before generating new designs, they produced fewer original ideas and their solutions closely resembled the provided examples. Studies similarly reported that exposure to pictorial examples in open-ended design tasks led participants to include more example-based elements, ultimately reducing the creativity of outputs (Chrysikou & Weisberg, 2005). Extending these findings, Chrysikou et al. (2016) examined the role of stimulus modality (pictures vs. words) and task type (open- vs. close-

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

ended) in functional fixedness. They found that when participants viewed pictures, particularly during open-ended tasks, they were more likely to adopt conventional uses, reflecting greater fixation on existing functions. By contrast, verbal prompts yielded fewer standard responses, presumably because language-based cues engage broader retrieval processes and exemplars in semantic memory and thus allow for more flexible idea generation.

Despite research on picture-superiority and stimulus modality in memory retrieval, no study has examined whether verbal versus visual prompts differentially influence the creative visual generation of abstract concepts. It is well-established that concepts are distributed across multiple representational systems, with verbal and visual modalities selectively activating different systems depending on task demands (Miyashita, 2000; Patterson et al., 2007). However, far less is known about how visual representations of concepts respond to these differing modalities—particularly when people must create new or unconventional visual forms.

Generative artificial intelligence (AI), particularly large language models (LLMs), provides a novel testbed for studying human–AI hybrid creativity (Collins et al., 2024; Rafner et al., 2023). By leveraging vast training corpora, LLMs can supply on-demand content and ideas that humans might not independently conceive, effectively serving as external cognitive aids. Whereas verbal AI prompts tend to support conceptual recombination, visual prompts inherently engage perceptual systems attuned to iconic symbols. Furthermore, LLMs can offer real-time suggestions tailored to individual inputs, thereby maximizing the effect of stimuli with specific creative tasks.

The present study leveraged AI-generated prompts to investigate how verbal versus visual modalities affect creative sketching of abstract concepts. Participants first drew an initial representation of a concept (e.g., "time"), then received tailored AI inspiration—either verbal or visual—before creating a final version. We specifically chose concepts whose well-learned associations tend to induce cognitive fixations. To ensure model consistency, the instructions to AI-generation were constant across conditions except for the presentation modality (see Methods). Finally, we analyzed the paired-drawings using both GPT-4 and two complementary deep neural networks (DDNs) to assess semantic and perceptual aspects of sketches.

We hypothesized that participants receiving verbal AI prompts would show fewer cognitive fixations—indexed by higher creativity scores—than those receiving visual prompts. Unlike the picture-superiority effect observed in memory studies, well-established visual representations can reduce creativity by anchoring participants to familiar forms. By contrast, verbal prompts may bypass perceptual constraints, resulting in greater conceptual flexibility and inducing more unconventional representations. These hypotheses were pre-registered on OSF.

Methods

Participants

An a priori power analysis determined a minimum sample size of 60 for a 2 × 2 mixed-design ANOVA (power = .80) with two between-subject conditions (verbal vs. visual AI inspiration) and two within-subject drawing stages (before vs. after AI inspiration). To account for potential exclusions, 80 participants were recruited online via Prolific (M_{age} = 28.9 years; SD = 6.0 years; 60.0% male). Recruitment targeted individuals fluent in English, possessing a high school diploma or higher, and with access to a computer mouse. All participants had normal or corrected-to-normal vision, no prior history of motor impairments, and no required artistic experience or skills. The study received approval from the university's IRB, and participants provided informed consent. Compensation was set at $8.00/hour ($2.64 for 20 minutes).

Stimuli and Procedure

Each trial consisted of three sequential phases: a pre-inspiration drawing phase, an AI-inspiration phase, and a post-inspiration drawing phase. Participants first received instructions defining visual metaphors and analogies, emphasizing the importance of conceptual creativity over artistic skill. They were then tasked with creating visual representations of abstract theme words (e.g., "time"). They were instructed to use unconventional and creative expressions by sketching and drawing on a digital canvas without using texts.

Pre-inspiration drawing phase. Three abstract themes were chosen to elicit drawings of visual metaphors: time, infinity, and trust, representing diverse conceptual domains (everyday, mathematical/physical, and mental state constructs). One of the three theme words was presented at the top of the screen. A blank digital canvas of size 600 x 400 pixels was shown below. The interface included basic tools (undo, redo, and erase) and limited participants to black line sketches. No color tools were provided. Participants were given unlimited time and were instructed to submit their drawings whenever they felt ready. Following the submission of a drawing, participants wrote a caption for their drawing. They were instructed to provide a detailed caption for the image that "communicate[s] its concept to an external observer". The caption required at minimum 20 characters. Shorter captions received a warning message.

AI inspiration. After submitting the first drawing and caption, participants were randomly assigned to either a verbal inspiration or a visual inspiration condition. In the *verbal condition,* a OpenAI GPT-4 generated full-sentence of 50 tokens or less. In the *visual condition,* a DALL-E (also from OpenAI) generated black-and white line-art style image in the size of 512 x 512 pixels.

Participants clicked a "Get AI inspiration" button and the text or image was generated via API calls to the OpenAI models

and both formats were derived from the caption of the participant's initial drawing. The instructions of the suggestion included 1) evaluating the received caption of the first drawing along with the theme word provided, 2) the evaluation of the caption in terms of conventional or unconventional, 3) providing creative suggestions based on the caption. To ensure comparability between the instructions, they were matched in length and wording as closely as possible except the instructions about modality. The full instructions to AI inspiration can be found from the OSF repository.

Post-inspiration drawing phase. Participants created a second drawing of the same theme word and captioned it following the same procedure as the pre-inspiration phase. The AI-generated text or image remained visible throughout the post-inspiration phase. Participants were instructed to decide on their own whether they use the AI-generated inspiration.

At the end of the experiment, participants rated their satisfaction with their drawings, their perceived influences from AI inspiration, and their subjective creative levels using a 10-point scale. In addition, participants completed the Object-Spatial Imagery and Verbal Questionnaire (OSIVQ) (Blazhenkova & Kozhevnikov, 2009). The analyses for individual differences can be found from the online supplemental information on OSF.

All procedures were programmed using Streamlit (Python 3.10) and deployed via Heroku. As the use of the GPT model might be time-sensitive, both the experiment and the analyses were conducted between late 2024 and January 2025.

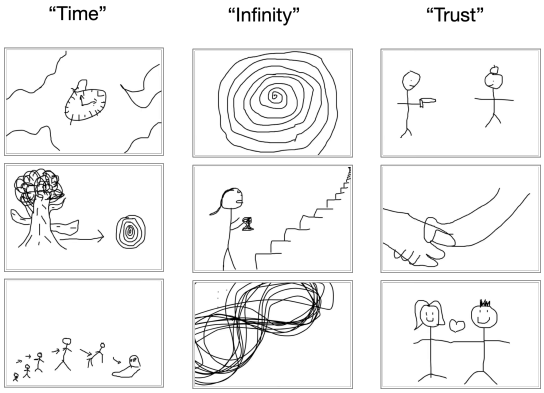


Figure 1: Sample drawings by theme topics.

Analytical Approach

Data Pre-processing. We examined the prompts generated and excluded AI-generated prompts that either contained incomprehensible information or contained incompatible information (i.e., texts in the visual condition). Six trials were excluded from analyses.

We also excluded data based on the length of image captions. Trials that contained captions that did not meet the minimum length (20 characters) were excluded. Four participants who did not provide enough captions on all three themes were excluded. The final sample consisted of 76 participants.

GPT-4 scoring of creativity. Scoring creativity with large language models has become increasingly common (DiStefano et al., 2024; Luchini et al., 2025). To assess the creativity of participants' drawings in response to task themes, GPT-4 was provided with both the theme and a short caption describing each drawing and instructed to evaluate the drawing's conceptual originality in combination with the theme provided. The model generated a single creativity score ranging from 1 (not creative) to 10 (very creative) along with a brief one-sentence rationale justifying its assessment (see Fig. 2). Specifically, the evaluation criterion focused on conceptual originality, specifically assessing whether the drawings and captions demonstrated unconventional representations of the given theme.

Human scoring. Human raters were given the same instruction as those inputted to GPT-4, with the exception of providing a one-sentence rationale. Raters were randomly presented with one drawing and its caption at the time and were asked to consider both the drawing and its caption. The theme word was provided at the top of screen and a rating scale at the bottom. The task condition and drawing stage was not available to raters. The rating task was programmed using PsychoPy3 and presented on *Pavlovia*, split into sessions (80 drawings per session). Three undergraduate students rated all sessions with a total of 480 drawings. 18 raters recruited from Prolific rated one session each (80 drawings).

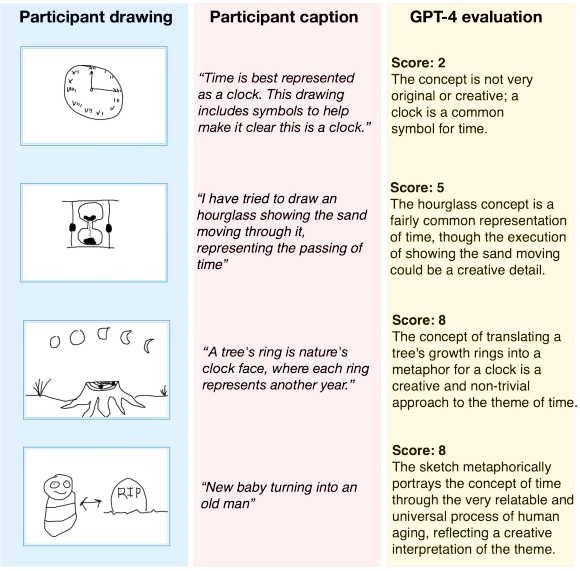


Figure 2: Sample participant drawings, captions, and GPT-4 evaluations of creativity scores (out of 10).

Visual features analysis. To capture and compare the visual features of participants' drawings from both pure visual perceptual similarities as well as the semantic relationships, we adopted two complementary models.

VGG16 (Visual Geometry Group). A truncated Visual Geometry Group 16-layer convolutional neural network (VGG16), originally trained on the ImageNet dataset, was employed to extract 4096-dimensional feature embeddings from participants' drawings (Russakovsky et al., 2014; Simonyan & Zisserman, 2014). Each image was first converted to a three-channel format, resized to 224×224 pixels, and normalized using standard ImageNet statistics. After removing the final classification layer (i.e., the softmax output) of VGG16, the resulting embedding vectors reflected low- and mid-level visual features (e.g., edges and shapes). Pairwise cosine distances were then calculated between embeddings of drawings created by the same participant under the same thematic condition; lower cosine distances indicated higher perceptual similarity among these sketches.

Contrastive Language–Image Pre-training (CLIP). CLIP, a multimodal vision–language model (Radford et al., 2021), was employed to capture the semantic content of participants' drawings. Each drawing was first preprocessed using CLIP's standard transformations (resizing and center-cropping to 224×224 pixels) and subsequently encoded into a 512-dimensional feature vector using the pretrained Vision Transformer backbone (ViT-B/32). This procedure yielded high-level, multimodal representations that emphasize conceptual meaning. Because CLIP embeddings are designed to reflect higher-level semantic information, two drawings that appear visually dissimilar may still lie close together in the embedding space if they share thematic or conceptual elements.

To compare participants' pre- and post-inspiration drawings, we calculated cosine distance between their respective CLIP embeddings, with smaller distances indicating greater semantic overlap. CLIP's advantage lies in its joint training on image–text pairs, allowing it to group semantically related images even when their perceptual features differ. This model is especially beneficial when two drawings convey a similar concept yet diverge in visual features. Thus, CLIP is a model that better captures participants' conceptual representations.

Semantic distance between caption texts. The semantic distance between drawing captions provided by participants was assessed using the all-MiniLM-L6-v2 model) from the SentenceTransformer library, a pre-trained language model optimized for computing semantic similarity between short texts. Captions were grouped by participant and thematic condition, and pairs of captions (e.g., first and second captions) within each group were analyzed. The model encoded each caption into high-dimensional word embeddings, and cosine similarity was computed to quantify the semantic similarities between paired captions. Higher similarity scores indicated greater semantic congruence between captions. For examples, see Table 1. Cosine distance was computed as one minus cosine similarity.

Table 1

Cosine similarity scores between captions

Pre-inspiration caption	Post-inspiration caption	Cosine similarity
An hour glass that signifies seasons	An hour glass that signifies change in time which brings about seasons of day and night	.89
The space journey and the endless potential which it brings	Infinite space and time in the universe	.48
A sand timer is a traditional tool to measure the passing of time.	A tree's ring is nature's clock face, where each ring represents another year.	.27

Results

Correlations between human and GPT scoring. Intraclass correlation coefficients (ICCs) were calculated using R with the psych package, based on a mean-rating ($k = 3$), absolute-agreement, two-way random-effects model (i.e., ICC [2, k]). The resulting ICC was 0.78 (95% CI [0.68, 0.85]), indicating good reliability among the three human raters' scores.

To assess the validity of GPT-rated creativity, we then correlated the mean ratings of human-rated creativity, including partial raters, and GPT-rated creativity on all participants' drawings before and after AI-inspiration. A Pearson correlation revealed significant correlations between human and GPT ratings before (r [77] = .71, t[77] = 8.8, p < .001) and after AI inspiration (r [77] = .73, t[77] = 9.23, p < .001). In the following sections, we adopted the computationally-scored creativity metrics for further analyses because human raters only evaluated partial subsets per rater. Additionally, GPT-scoring offered advantages including consistency in applying rating criteria and the elimination of rater fatigue effects.

GPT-scored creativity. Data were analyzed using a repeated-measures analysis of variance (ANOVA). The between-subject factor of task condition (verbal vs. visual) and within-subject factors drawing stage (before vs. after AI inspiration) and theme were entered to examine interaction effects on computationally-scored creativity.

A preliminary analysis of theme did not reveal significant differences in creativity scores; thus, participants' creativity scores were aggregated. The analysis revealed a significant interaction between condition and drawing stage, $F(1,74) = 43.36$, $p < .001$, $\eta^2 = .098$. Significant main effects of condition and drawing stage (before vs. after) were also found. A posthoc comparison revealed that the verbal condition showed significantly higher creativity scores after AI-generated inspiration compared to the baseline (t[74] = 9.56, $p < .001$; see Fig. 3), whereas the visual condition inspiration showed no significant differences before and after

AI inspiration (t[74] = .81, p = .849). Importantly, baseline creativity scores between verbal and visual conditions in the pre-AI inspiration stage did not differ (t[74] = -.36, p = .983), but they differed significantly after AI inspiration (t[74] = 5.17, p <.001), suggesting that the verbal modality elicited higher creativity scores than the visual modality, at least when AI prompting was closely matched.

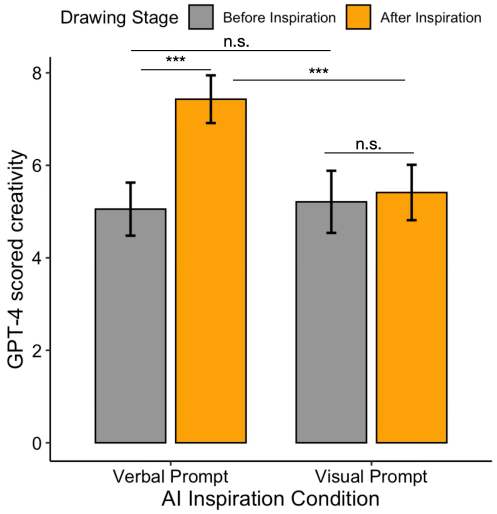


Figure 3: GPT-scored creativity scores by condition and drawing stage. Error bars represent 95% CIs. *** denotes p < .001.

Visual feature analysis. We then examined whether the type of AI inspiration (verbal vs. visual) influenced the cosine distances between participants' first and second drawings (see Methods for details).

VGG-Based Perceptual Distance. There was no significant effect of AI condition on VGG distances (F [1,74] = 1.28, p =.262, η^2=.017). In other words, participants' drawings did not differ in low-level perceptual similarity depending on whether they received a verbal or visual AI prompt (see Figure 4A).

CLIP-based distance. By contrast, condition significantly affected CLIP perceptual-semantic distance (F [1,74] = 11.45, p = .001, η^2=.13, such that distance metrics were significantly larger in the verbal condition than the visual condition (see Figure 4B). These findings suggest greater perceptual and semantic changes between participants' first and second drawings in the verbal condition compared to the visual condition.

Together, these findings demonstrate that greater changes in semantic meanings were observed between the drawings following a verbal inspiration than a visual inspiration, even though low-level perceptual features were not different between drawings following a verbal or visual prompt.

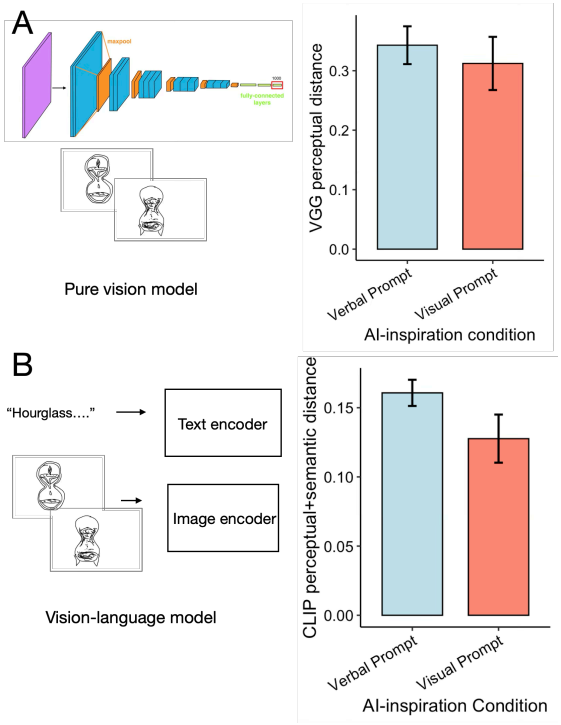


Figure 4: VGG perceptual distance metrics (A) and CLIP distance metrics (B) by AI-inspired condition. Error bars represent 95% CIs.

Semantic distance analyses. To better capture the semantic distance between the caption texts before and after AI inspiration, analyses of semantic analysis on texts only were performed without drawings.

ANOVA analyses revealed a significant main effect of condition F(1,74) = 16.42, p < .001, η^2 =.18, such that the semantic distance between pre-inspiration caption and post-inspiration caption was larger following a verbal prompt than a visual prompt (see Fig. 5).

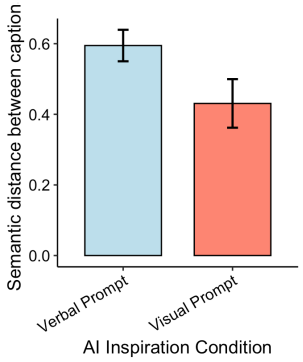


Figure 5: Semantic distance (cosine distance) by condition.

Subjective ratings. A one-way ANOVA was performed to examine the effect of AI inspiration (verbal vs. visual) on three subjective ratings: perceived AI influence, subjective creativity, and overall satisfaction.

The analysis revealed a significant main effect of AI condition on perceived AI influence, $F(1, 74) = 13.31$, $p < .001$, $\eta^2 = .152$. Participants in the verbal condition reported significantly higher perceived influence from AI compared to those in the visual condition, consistent with the higher creativity and semantic distance metrics shown above. However, the effect of AI condition on subjective creativity was not significant, $F(1, 74) = .35$, $p = .558$, $\eta^2 = .005$. Similarly, the effect of AI condition on participants' satisfaction was also nonsignificant, $F(1, 74) = 1.30$, $p = .258$, $\eta^2 = .017$.

Discussion

The present study examined the effect of verbal versus visual AI-generated inspiration on the creative sketching of abstract concepts. Our findings provide novel evidence that verbal prompts significantly enhance conceptual and semantic divergence, as reflected in higher GPT-4 creativity scores and greater CLIP-based semantic distance. Importantly, low-level perceptual features (VGG16-based) of participant sketches remained stable, suggesting that the effect of verbal prompts primarily influenced conceptual, rather than visual, transformation.

These findings support the hypothesis that verbal inspiration fosters greater conceptual flexibility, whereas visual inspiration may reinforce cognitive fixations on conventional representations. Two key mechanisms could explain this pattern: 1) Verbal prompts increase conceptual divergence by disrupting well-learned semantic-visual associations, allowing participants to generate more unconventional representations, and 2) Visual prompts reinforce existing perceptual associations, leading to greater resemblance to well-established symbolic representations.

Although these mechanisms are not mutually exclusive, our results suggest that verbal prompts operate primarily through the first mechanism. This conclusion is supported by the divergence in conceptual content (CLIP-based analysis) without corresponding differences in low-level visual features (VGG16-based analysis). That is, verbal prompts did not change the perceptual complexity of the drawings but did alter their conceptual framing.

These results extend existing research on creative cognition and human-AI hybrid creativity by demonstrating that verbal input can possibly reduce perceptual constraints imposed by well-learned semantic-visual associations. Unlike previous findings on the picture-superiority effect—which suggest that visual stimuli enhance memory encoding—our study indicates that, in creative tasks, visual representations may anchor individuals to conventional forms rather than fostering innovation. Verbal AI-generated inspiration, by contrast, appears to promote greater cognitive flexibility and conceptual recombination. Although visual artists and designers may rely on image-generation models for aiding their creativity, our findings potentially indicate that visual modality may not always be beneficial.

There are also several limitations that warrant discussion.

Variability in AI-Generated outputs. The present study relies on AI-generated prompts (OpenAI GPT-4 for text, OpenAI DALL·E for images), but AI outputs are not deterministic. The same prompt may yield different responses over time as models continue to improve their performance.

Potential limitations of AI-Generated prompts. The visual inspiration was generated in a "black-and-white line-art" style. This limitation might have led to visual prompts that were less engaging than images typically seen, thereby reinforcing conventional mappings rather than challenging them. Also, as visual prompts were generated based on LLM-derived captions, they may have had conceptual parity with verbal prompts but lacked sufficient visual novelty. In the present study, we maintained consistent instructions to AI prompts except for modality-specific instructions (i.e., word or size limit for images), and both types of prompts were generated from participant initial captions. In addition, trials where AI-generated images failed to meet criteria were excluded. Despite our procedures, future studies should evaluate whether improvements on instructions in visual AI design could yield more effective inspiration.

Potential bias in creativity evaluation. While human ratings and GPT-4 creativity scores were significantly correlated, concerns remain about whether LLMs reinforce conventional ideas (Anderson et al., 2024). When these models are used to evaluate creative outputs, they may inadvertently favor ideas that align with these prevalent patterns rather than truly novel or unconventional ones.

Together, despite potential limitations, our findings suggest that verbal AI-generated inspiration facilitates greater conceptual flexibility in creative sketching compared to visual inspiration. This effect is primarily driven by an increase in semantic divergence rather than changes in low-level perceptual features. While visual inspiration appears to reinforce existing cognitive fixations, verbal inspiration encourages unconventional representations. These insights have significant implications for human-AI collaboration in creativity and design arts, emphasizing the potential of modality-specific AI tools to scaffold human creative thinking in visual designs and beyond.

References

Anderson, B. R., Shah, J. H., & Kreminski, M. (2024). Homogenization effects of large language models on human creative ideation. *Creativity and Cognition*, 413–425.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences, 22*(4), 637–660.

Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology, 23*(5), 638–663.

Cavazzana, A., & Bolognesi, M. (2020). Uncanny resemblance: Words, pictures, and conceptual representations in the field of metaphor. *Cognitive Linguistic Studies, 7*(1), 31–57.

Chrysikou, E. G., Motyka, K., Nigro, C., Yang, S.-I., & Thompson-Schill, S. L. (2016). Functional fixedness in creative thinking tasks depends on stimulus modality. *Psychology of Aesthetics, Creativity, and the Arts, 10*(4), 425–435.

Chrysikou, E. G., & Weisberg, R. W. (2005). Following the wrong footsteps: Fixation effects of pictorial examples in a design problem-solving task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 1134–1148.

Collins, K. M., Sucholutsky, I., Bhatt, U., Chandra, K., Wong, L., Lee, M., Zhang, C. E., Zhi-Xuan, T., Ho, M., Mansinghka, V., Weller, A., Tenenbaum, J. B., & Griffiths, T. L. (2024). Building machines that learn and think with people. *Nature Human Behaviour, 8*(10), 1851–1863.

DiStefano, P. V., Patterson, J. D., & Beaty, R. E. (2024). Automatic Scoring of Metaphor Creativity with Large Language Models. *Creativity Research Journal*, 1–15. https://doi.org/10.1080/10400419.2024.2326343

Fan, J. E., Yamins, D. L. K., & Turk-Browne, N. B. (2018). Common object representations for visual production and recognition.

Glaser, W. R. (1992). Picture naming. *Cognition, 42*(1–3), 61–105.

Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & Cognition, 36*(7), 1351–1359.

Hoffman, D. D., Singh, M., & Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin & Review, 22*(6), 1480–1506.

Jansson, D. G., & Smith, S. M. (1991). Design fixation. *Design Studies, 12*(1), 3–11.

Luchini, S. A., Maliakkal, N. T., DiStefano, P. V., Laverghetta Jr*, A., Patterson, J. D.*, Beaty, R. E., & Reiter-Palmon, R. (2025). Automated scoring of creative problem solving with large language models: A comparison of originality and quality ratings. *Psychology of Aesthetics, Creativity, and the Arts*.

Miyashita, Y. (2000). Neural representation of visual objects: Encoding and top-down activation. *Current Opinion in Neurobiology, 10*(2), 187–194.

Olsen, K., & Tylén, K. (2023). On the social nature of abstraction: Cognitive implications of interaction and diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences, 378*(1870), 20210361.

Paivio, A. (1971). Imagery and language. *In Imagery* (pp. 7–32). Elsevier.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976–987.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *arXiv*.

Rafner, J., Beaty, R. E., Kaufman, J. C., Lubart, T., & Sherson, J. (2023). Creativity in the age of generative AI. *Nature Human Behaviour, 7*(11), 1836–1838.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2014). ImageNet large scale visual recognition challenge. *arXiv*.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *Proceedings of the International Conference on Learning Representations (ICLR)*.

Smith, S. M., Ward, T. B., & Schumacher, J. S. (1993). Constraining effects of examples in a creative generation task. *Memory & Cognition, 21*(6), 837–845.

Tversky, B. (2011). Visualizing thought. *Topics in Cognitive Science, 3*(3), 499–535.

Uttal, D. H., & Yuan, L. (2014). Using symbols: Developmental perspectives. *WIREs Cognitive Science, 5*(3), 295–304.

Wang, S., Okada, T., & Takagi, K. (2023). How to effectively overcome fixation: A systematic review of fixation and defixation studies. *Frontiers in Education, 8*, 1183025.

Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology, 27*(1), 1–40.