



How Familiarity and Fluency Shape Perceived Naturalness and Aesthetic Pleasure in Digital Spatial Images

Yeong Ji Hwang¹, Taig Youn Cho^{2*}

¹Major in Spatial Design, Graduate School, Ph.D. candidate, Hongik University, Seoul, Korea

²Department of Industrial Design, Professor, Ph.D., AIA, Hongik University, Seoul, Korea

Abstract

Background While creating a natural experience is a primary goal in digital environments, little research has explored the naturalness of digital spatial imagery itself. Specifically, it remains unclear whether perceived naturalness is driven by physical realism (how closely a space replicates real-world structures) or by perceptual ease, where spatial configurations are intuitively understood. To address this gap, the present study distinguishes between Physical Naturalness (familiarity-based) and Cognitive Naturalness (fluency-based) and examines their respective roles in shaping perceived naturalness and aesthetic pleasure.

Methods Physical Naturalness was operationalized as spatial imagery adhering to real-world visual norms, while Cognitive Naturalness was based on processing fluency, incorporating structured visual order and semantic clarity. Seventy-eight participants evaluated 24 artificial intelligence (AI)-generated spatial images, categorized into Physical Naturalness (PN), Physical Unnaturalness (PU), and Cognitive Naturalness (CN) conditions. Participants completed a real/virtual classification task and rated each image on familiarity, naturalness, and aesthetic pleasure. Response time and subjective ratings were analyzed using linear mixed-effects models.

Results Familiarity was the strongest predictor of perceived naturalness, supporting prior claims that familiar stimuli are perceived as more natural. However, familiarity had only a marginal effect on aesthetic pleasure. Instead, response time and perceived naturalness significantly predicted aesthetic pleasure, with faster responses associated with higher ratings. Notably, CN images, despite being physically implausible, elicited faster processing and greater aesthetic pleasure than PN images, highlighting the role of cognitive fluency in aesthetic evaluation.

Conclusions These findings suggest two cognitive pathways to aesthetic experience: familiarity enhances naturalness but influences aesthetic pleasure only indirectly, whereas fluency directly promotes aesthetic pleasure. Therefore, intuitive spatial structuring and semantic coherence may be more effective than strict adherence to physical realism in fostering positive user experiences in digital spatial design.

Keywords Digital Environment Design, AI-Generated Images, Naturalness, Aesthetic Experience, Processing Fluency

*Corresponding author: Taig Youn Cho (taigyoun@hongik.ac.kr)

Citation: Hwang, Y. J., & Cho, T. Y. (2025). How Familiarity and Fluency Shape Perceived Naturalness and Aesthetic Pleasure in Digital Spatial Images. *Archives of Design Research*, 38(3), 131-147.

<http://dx.doi.org/10.15187/adr.2025.08.38.3.131>

Received : May. 10. 2025 ; **Reviewed :** Aug. 04. 2025 ; **Accepted :** Aug. 04. 2025

pISSN 1226-8046 **eISSN** 2288-2987

Copyright : This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted educational and non-commercial use, provided the original work is properly cited.

1. Introduction

Digital spatial images are widely employed in games, animation, and metaverse applications, either by replicating real-world environments or through creative transformation. While users' perception of these spaces as natural critically shapes their experience, this perceptual naturalness has rarely been examined directly. Instead, it has been addressed indirectly through the concept of spatial presence, which is the sense of being physically located within a mediated environment. According to Steuer (1992), spatial presence depends on two factors: interactivity and vividness. Interactivity, defined as users' ability to engage with digital environments in real time, has been extensively explored. Previous studies have shown that variables such as display size, viewing angle, and controller responsiveness can significantly enhance spatial presence (Hou et al., 2012; Wu & Lin, 2018; Seibert & Shafer, 2018).

Meanwhile, vividness refers to the breadth and depth of sensory information, encompassing the number of sensory channels (e.g., vision, hearing, touch) and the quality of information within each channel (e.g., resolution or audio fidelity). Although resolution and framerate contribute to vividness, they are usually considered standard design elements rather than objects of empirical research. This is evident in the advancement of rendering techniques, such as ray tracing and global illumination, over time. However, this technically grounded view in both research and practice overlooks the cognitive mechanisms that make digital spaces feel natural.

One such mechanism may be perceptual fluency, which refers to the ease with which visual configurations are processed and understood (Reber et al., 2004). As rendering technologies advance, greater realism is generally expected to enhance perceived naturalness. This may be because visually realistic stimuli are processed more fluently, without cognitive disruption. However, perceived naturalness does not always follow this pattern. Digital spaces in films, games, and virtual environments are often creatively transformed and appear surreal or physically implausible, yet they are still perceived as natural or even aesthetically compelling. For instance, despite lacking physical realism, the spatial image generated by Midjourney in Figure 1 is described as "beautiful" or accompanied by comments like "Someone please make this in real life." The image presents colorful, fantastical desert spaces that are rendered with repeated curved lines, coherent lighting, and intuitive spatial depth. Such reactions suggest that perceived naturalness may arise not only from realism but also from perceptual ease, where spatial configurations are intuitively and fluently understood.



Figure 1 AI-generated spaces that appear implausible but are perceived as natural (Vermillion, 2023)

To clarify this distinction, we propose two types of naturalness, based on familiarity and processing fluency. Physical Naturalness (PN) arises from the familiarity of spatial features typically found in the real world. For example, spatial features like standard room layouts, realistic lighting, and correct perspective can elicit PN. Cognitive Naturalness (CN) arises from processing fluency, or the ease and speed with which a visual stimulus is understood. Even without physical realism, images can appear natural if they are visually structured and semantically clear. For example, AI-generated images, such as the one in Figure 1, may lack realistic detail, but their visual coherence and clarity can facilitate fluent processing. This ease of understanding leads viewers to perceive them as both natural and aesthetically pleasing.

This study addresses two research questions: (1) Does perceived naturalness depend more on familiarity or on fluency? and (2) How do familiarity-based (Physical Naturalness) and fluency-based (Cognitive Naturalness) spatial images differ in their effects on aesthetic pleasure? To address these questions, we operationalize Cognitive Naturalness using two core features: structured visual properties and semantic clarity. Participants evaluate AI-generated spatial images categorized according to either Physical or Cognitive Naturalness. Dependent measures include response time and aesthetic ratings, allowing us to assess the relative contributions of familiarity and fluency to the perception of naturalness and aesthetic judgment. Ultimately, this study proposes a cognitively informed and aesthetically oriented framework for understanding naturalness in digital design—one that emphasizes intuitive spatial structure and semantic coherence over physical accuracy alone.

2. Theoretical Framework

2. 1. Conceptualizing naturalness in digital spatial imagery

Although the term naturalness is commonly used, its meaning remains conceptually ambiguous in the context of digital spatial imagery. Naturalness has often been associated with physical realism, that is, the extent to which a space replicates real-world configurations. However, this perspective is insufficient to explain why spatial images that are surreal or physically implausible, such as AI-generated environments or science fiction settings, are still perceived as natural and aesthetically compelling. This suggests that perceived naturalness may also arise from perceptual ease, where spatial configurations are intuitively and fluently understood.

To address this ambiguity, the present study distinguishes between two types of naturalness: Physical Naturalness (PN) and Cognitive Naturalness (CN). PN is grounded in familiarity, shaped by repeated exposure to conventional spatial features such as standard layouts or realistic lighting. CN, by contrast, is based on processing fluency—the ease and speed with which a stimulus is cognitively processed, even in the absence of direct prior exposure. Given the aesthetic responses to implausible but intuitively structured spatial images, CN aligns with processing fluency theory, which posits that stimuli are judged more favorably when they are processed fluently (Reber et al., 2004; Winkielman et al., 2003; Unkelbach & Greifeneder, 2013).

Previous literature suggests that naturalness can stem from either familiarity or fluency. For instance, familiar elements are often perceived as natural when they align with contextual expectations (Siipi, 2008). Similarly, fluent processing enabled by prototypical or expected features can also enhance the perception of naturalness (Lee, 2021; Chen et al., 2023). However, these two mechanisms appear to differ in how they influence aesthetic responses. While the aesthetic impact of familiarity varies across domains and stimulus types (Park et al., 2010; Song et al., 2021), fluency has consistently demonstrated a robust positive effect on aesthetic pleasure (Reber et al., 2004; Winkielman et al., 2003; Unkelbach & Greifeneder, 2013). More recent work within predictive processing frameworks further suggests that aesthetic pleasure increases not only when stimuli match prior expectations, but also when they are processed more fluently than anticipated (Brouillet & Friston, 2023; Yoo et al., 2023).

Taken together, these findings imply that fluency offers a more consistent and predictive account of aesthetic pleasure than familiarity alone. Accordingly, the present study hypothesizes that Cognitive Naturalness, driven by fluent processing, plays a stronger role in eliciting aesthetic pleasure than Physical Naturalness, which is rooted in familiarity.

2. 2. Operationalization of physical and cognitive naturalness

In this study, Physical Naturalness (PN) was operationalized as spatial imagery that conforms to conventional real-world visual norms, such as accurate lighting, perspective, and object proportions, intended to evoke familiarity through physical realism. In contrast, Cognitive Naturalness (CN) was defined based on two cognitive components theorized to promote fluent processing: spatial structure and spatial semantics. These align with two core constructs in fluency theory—visual clarity and prototypicality—both enhancing perceptual

ease (Reber et al., 2004; Winkielman et al., 2006). Spatial imagery, as a form of complex visual stimulus, is also influenced by these mechanisms.

First, spatial structure refers to the formal arrangement of elements within a scene, including symmetry, repetition, and gradual transformations. These features create visual order, which facilitates rapid and effortless spatial recognition (Reber et al., 2004; Coburn et al., 2020). Predictable structures reduce cognitive load, and empirical studies show that symmetrical and regular patterns are processed more quickly (Gronchi & Sloman, 2021; Pramod & Arun, 2018; Sztuka & Kühn, 2025).

Second, spatial semantics pertains to the degree to which a space exhibits prototypical features of a known environmental category. Such features allow viewers to intuitively infer the function and identity of a scene (Rips et al., 1973; Rosch, 1975; Halberstadt, 2006). For instance, dense trees and foliage quickly signal a forest, while the combination of sand, sea, and sky readily conveys a beach (Csathó et al., 2015).

Taken together, these structured and semantically rich features promote predictability and reduce cognitive effort, thereby contributing to a stronger perception of naturalness. This study thus proposes that naturalness in digital spatial imagery does not depend solely on physical realism. Instead, Cognitive Naturalness emerges through fluency-based mechanisms grounded in visual structure and spatial meaning.

3. Method

3. 1. Research design

This study addresses two research questions:(1) Does perceived naturalness depend more on familiarity or on fluency? And (2) How do familiarity-based (Physical Naturalness, PN) and fluency-based (Cognitive Naturalness, CN) spatial images differ in their effects on aesthetic pleasure?

To investigate these questions, we designed a within-subjects experiment in which each participant experienced all stimulus types. The experiment aimed to examine how different forms of naturalness influence perceptual fluency, perceived naturalness, familiarity, and aesthetic judgment.

Three types of AI-generated spatial images were used as stimuli: (1) PN images based on spatial familiarity, (2) CN images designed to promote perceptual fluency, and (3) control images that were neither familiar nor fluent. The third category served as a contrast condition to test whether the absence of both familiarity and fluency would result in lower naturalness and aesthetic ratings.

Although image category serves as the independent variable, the primary analytical focus lies in the relationships among dependent variables. These include: (1) perceptual fluency (indexed by response time), (2) perceived naturalness, (3) spatial familiarity, and (4) aesthetic pleasure. Rather than conducting group comparisons, the analysis explores how fluency and familiarity predict naturalness and aesthetic experience. The first research question is addressed by comparing the relative contributions of fluency and familiarity to naturalness. The second is examined via a mediation analysis testing the direct and indirect effects of fluency and familiarity on aesthetic pleasure through naturalness.

The experiment consisted of two phases: a binary classification task and a rating task. Each participant completed both phases using the same set of 24 images presented in random order. Figure 2 provides an overview of the experimental sequence, stimulus types, and outcome measures.

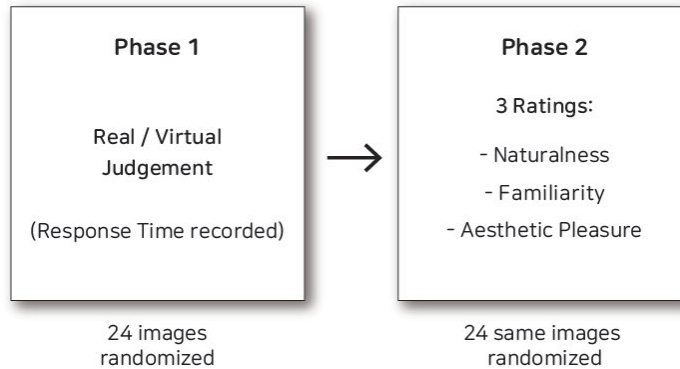


Figure 2 Overview of the experimental procedure

Prior to data collection, a power analysis based on multiple linear regression indicated that at least 77 participants would be required to detect a medium-to-large effect size, based on previous findings (Chen et al., 2023). Although we initially planned to analyze the data using multiple regression, considerable individual variability in response patterns led us to adopt a linear mixed-effects model (LMM). This approach allowed us to account for both participant- and image-level random effects, taking into account individual differences in how people perceive complex spatial images.

3. 2. Participants

A total of 106 participants with normal or corrected-to-normal vision and color perception took part in the experiment. Participants were informed that they would have the opportunity to receive monetary compensation, and 10 participants were randomly selected to receive a reward valued at 10,000 KRW each. Response time data were screened for outliers, defined as any values more than three standard deviations above the mean ($M = 3.88s$, $SD = 4.06s$). This threshold excluded responses above 16.08 seconds, likely caused by participants being interrupted during the online survey (e.g., switching tabs or leaving the screen). As a result, 27 responses (25.5%) were excluded, a relatively high rate that reflects the inherent limitations of online data collection. One additional participant was removed for providing uniform responses across all items. The final sample included 78 participants (43 females, 34 males, and one non-binary/third-gender individual; M age = 34, $SD = 7.92$), with 76 participants from Korea and two from China.

3. 3. Experimental stimuli

This study conducted an experimental investigation to examine how naturalness is evaluated in digital environments. To this end, 24 spatial images were generated using Midjourney v6.1, a text-to-image generative AI model. These images were categorized into three theoretically

defined conditions:

1. **Physically Natural (PN):** Spaces that conform to real-world visual conventions, such as natural lighting, accurate perspective, and realistic proportions.
2. **Physically Unnatural (PU):** Spaces that violate physical laws or visual norms—for example, with distorted structures or implausible light and shadow placement.
3. **Cognitively Natural (CN):** Physically implausible spaces that nonetheless convey clear spatial structure or meaning, enabling intuitive recognition through features like symmetry, repetition, or functional cues.

Eight distinct spatial scenarios were developed to ensure a systematic and balanced stimulus design, with four scenarios focused on spatial structure (e.g., symmetry, enclosure) and four on spatial meaning (e.g., implied function or environmental context). Each scenario was rendered into the three conditions above, resulting in a total of 24 images.

Table 1 Prompt structure and visual characteristics by image type

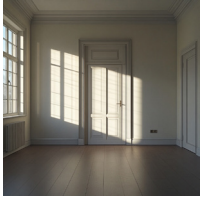
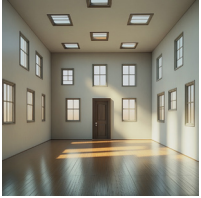





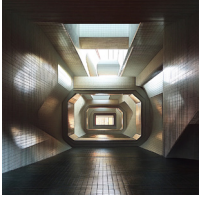

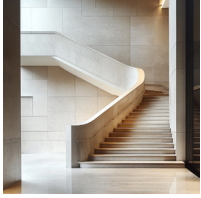





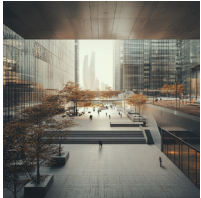


Condition	Prompt Structure	Key Visual Features
Physically Natural (PN)	“Realistic photograph of a [space type]”	Conforms to real-world norms: natural lighting, accurate perspective, realistic proportions
Physically Unnatural (PU)	PN prompt + “with impossible structures”	Violates physical norms: misaligned structures, inconsistent shadows, perceptual ambiguity
Cognitively Natural (CN)	PN + select distortions from PU + perceptual fluency cues	Structured yet implausible: symmetry, repetition, coherent depth, implied function

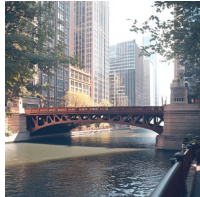


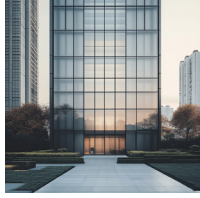

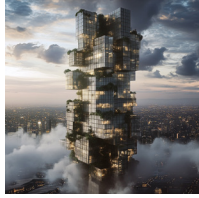
Table 1 summarizes the prompt structure and visual characteristics used to generate the three types of images. The PU images were created by modifying the PN prompts to introduce visual or structural violations that disrupt realism and increase perceptual ambiguity—for example, through conflicting shadow directions or spatial distortions. CN images were generated using the same PN base, with selected distortions from PU retained or exaggerated. Crucially, CN images included additional cues to promote perceptual fluency through structured composition and meaningful spatial cues.

For instance, in a window-based scenario, PN images depict conventionally arranged windows with balanced lighting and realistic proportions. PU images display irregular arrangements and shadow mismatches that disrupt coherence. CN images retain structural irregularities but introduce intuitive spatial depth through directional lighting or suggestive views, allowing the space to be processed more fluently.

All images underwent post-processing using Adobe Photoshop (2024) to ensure consistency in both composition (e.g., the inclusion of doors, windows, and sky views) and visual properties, such as tone and brightness, across conditions. Table 2 presents the full list of spatial scenarios and their corresponding PN, PU, and CN image versions.

Table 2 Image variants (PN, PU, CN) across eight spatial scenarios

Scenario	PN (Physically Natural)	PU (Physically Unnatural)	CN (Cognitively Natural)
Windows/ Doors (Structure)			
	Typical interior with windows/ doors	irregular arrangements and shadows	Irregular windows emphasizing depth through view
Arches (Structure)			
	Arched retail hall	Flat repetitive arches	Depth-creating arches
Tunnel (Structure)			
	Open-ended tunnel with stairs	Functionally ambiguous tunnel	Tunnel with rhythmic surfaces
Staircase (Structure)			
	Standard linear staircase	Disjointed and illogical stairs	Structured staircase with flowing form
Living Room (Meaning)			
	Realistic living room with nature view	Disjointed living room layout	Room merging with landscape, inviting rest
Square (Meaning)			
	Open and accessible square	Fragmented and contextless square	Waterside square with resting function

Bridge (Meaning)			
	Typical urban bridge	Irrationally structured bridges	Functional bridges with novel forms
Building (Meaning)			
	Steel-glass modern building	Contextless irregular building	Distorted building with urban cues

Note. Each scenario is classified as either spatial structure-focused or spatial meaning-focused, as indicated in parentheses.

3. 4. Materials and procedure

This study adhered to the ethical guidelines for human-subject research. Before participating, all respondents read and agreed to a digital consent form that outlined the study's purpose, procedures, and data handling. Only those who gave informed consent could proceed. According to the institution's guidelines for low-risk online research, this study did not fall under IRB review.

The experiment consisted of two phases. Phase 1 was a binary classification task in which participants viewed 24 AI-generated spatial images and judged whether each represented a real or virtual space. This task was designed to assess perceptual fluency, the ease and speed with which spatial characteristics were processed. The method builds on prior fluency research using lexical categorization tasks (Whittlesea & Williams, 1998), adapted here for spatial imagery. Because the boundary between real and artificial is often ambiguous in digital environments—for example, stylized real-world architecture may appear artificial, while photorealistic virtual spaces may seem real—this task was designed not to evaluate classification accuracy, but to assess how intuitively spatial features are processed. Thus, response time is treated as a proxy for fluency: faster responses indicate more intuitive and fluent processing, while slower responses suggest perceptual ambiguity or cognitive hesitation.

Participants completed this task on Qualtrics, an online survey platform. One image was presented per page, accompanied by two response buttons ("real" or "virtual"). The page submission time was used as the primary measure, as it captured the full decision-making process, including any hesitation or changes of mind.

Phase 2 asked participants to rate the same 24 images on three dimensions using 5-point Likert scales:

- Naturalness: “How natural do you find this space?” (1 = certainly not natural, 5 = certainly natural)
- Familiarity: “How familiar do you find this space? In other words, how common or usual does the space look?” (1 = certainly not familiar, 5 = certainly familiar)
- Aesthetic pleasure: “How visually pleasing do you find this space? In other words, how good or beautiful does the space look?” (1 = certainly not pleasing, 5 = certainly pleasing)

These questions were adapted from Chen et al. (2023) and were presented in Korean. To reduce order effects, the sequence of image presentation was randomized in both phases, and no time limit was imposed.

Before the task, participants were informed that the spaces had been selected from designer portfolios featured on architecture and design-related websites. This framing was intended to direct participants’ attention to the spatial qualities of the environments rather than the authenticity of the images. Following the experiment, participants were fully debriefed and informed that all images were AI-generated. The debriefing included a clear explanation of the study’s purpose, procedures, and data usage to ensure ethical transparency.

4. Results

4. 1. Descriptive statistics

The study analyzed descriptive statistics for response time, naturalness, familiarity, and aesthetic pleasure using a two-tailed approach with an alpha level of 0.05.

Regarding response time, PN images had the slowest average response ($M = 4.77s$, $SD = 2.35s$), while CN images had the quickest ($M = 3.15s$, $SD = 1.49s$). Details on average response time are provided in Table 2.

Table 3 Descriptive statistics of response time by image type

Image Type	Mean (s)	SD (s)	Skewness	Kurtosis
PN	4.77	2.35	1.29	2.25
PU	3.72	2.16	1.98	6.12
CN	3.15	1.49	1.18	1.78

Note. PN = Physically Natural; PU = Physically Unnatural; CN = Cognitively Natural.

For naturalness, the PN image group received the highest average rating ($M = 4.26$, $SD = .629$). CN images were rated slightly higher in naturalness compared to PU ($M = 2.28$, $SD = .628$ vs. $M = 2.11$, $SD = .572$). Similarly, familiarity ratings were highest for PN images ($M = 4.16$, $SD = .575$), while both PU and CN images received lower familiarity scores ($M = 2.01$, $SD = .493$ and $M = 2.07$, $SD = .685$, respectively). Aesthetic pleasure was rated highest for CN ($M = 3.41$, $SD = .564$), followed by PN ($M = 3.33$, $SD = .564$) and PU ($M = 2.69$, $SD = .693$). Additional details on average scores and standard deviations are provided in Table 3.

Table 4 Descriptive statistics of aesthetic ratings by image type

Measure	Image Type	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
Naturalness	PN	4.26	0.63	-1.008	0.770
	PU	2.11	0.57	0.789	0.845
	CN	2.28	0.63	0.765	0.158
Familiarity	PN	4.16	0.58	-0.931	1.483
	PU	2.07	0.69	0.977	0.970
	CN	2.01	0.49	0.532	0.207
Aesthetic pleasure	PN	3.33	0.56	-0.415	-0.023
	PU	2.69	0.69	0.469	-0.055
	CN	3.41	0.68	0.045	-0.849

Note. PN = Physically Natural; PU = Physically Unnatural; CN = Cognitively Natural.

4. 2. Inferential analyses

Before conducting the study, we initially planned to analyze the data using multiple linear regression. However, upon inspecting the completed dataset, we observed substantial individual variability in both response times and aesthetic ratings—even within the same image categories. Some participants showed unusually slow responses, likely due to external distractions during the online survey. Given this heterogeneity, simple averaging could obscure meaningful patterns in the data.

To address this, we employed a linear mixed-effects model (LMM), which accounts for both fixed effects (e.g., image type) and random effects (e.g., participant- and image-specific variance). Although the spatial images were carefully designed to represent distinct theoretical conditions (i.e., Physical Naturalness, Physical Unnaturalness, and Perceptual Naturalness), their rich visual and semantic complexity may have led to subtle differences in individual interpretation. By treating both participants and images as random intercepts, the LMM captured baseline variation across individuals and images. All analyses were conducted using the Mixed Models module in Jamovi (Version 2.3.28).

4. 2. 1. Predictors of perceived naturalness

A linear mixed-effects model was conducted to investigate the cognitive underpinnings of perceived naturalness in digital spatial environments. Familiarity, response time, and image type were treated as fixed effects to examine their influence on perceived naturalness. Random intercepts were included for participants ($ICC = .096$) and images ($ICC = .032$) to account for the nested data structure. The model explained 52% of the variance in naturalness ratings through fixed effects (marginal $R^2 = .520$), and 57.8% when random effects were included (conditional $R^2 = .578$). Visual inspection of the Q–Q plot and histogram of residuals indicated approximate normality, with only minor deviations at the tails. Thus, model assumptions were considered sufficiently met. Alternative response time filters (e.g., <10s, 12s) yielded similar results, indicating that the model findings were not sensitive to outlier definitions.

The model revealed significant effects of familiarity ($F(1, 1842.7) = 348.51, p < .001$), image type ($F(2, 26.2) = 63.21, p < .001$), and response time ($F(1, 1726.5) = 4.75, p = .029$). Familiarity emerged as the strongest predictor of naturalness ($\beta = 0.427, SE = 0.02, t = 18.67, p < .001$). Response time showed a modest but significant positive relationship with naturalness ratings ($\beta = 0.021, SE = 0.01, t = 2.18, p = .029$).

Regarding image type, Physically Unnatural (PU) images received significantly lower naturalness ratings compared to Physically Natural (PN) images ($\beta = -1.23$, $SE = 0.11$, $t = -10.76$, $p < .001$). Cognitively Natural (CN) images were also rated lower than PN ($\beta = -1.03$, $SE = 0.11$, $t = -9.00$, $p < .001$). These results confirm that PN images are perceived as the most natural among the three types, as displayed in Figure 3.

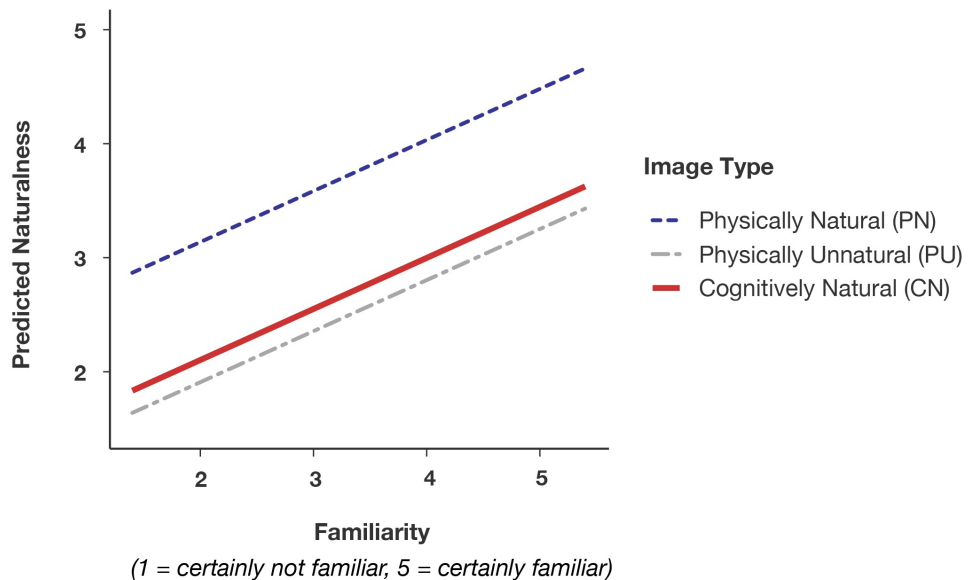


Figure 3 Predicted naturalness as a function of familiarity and image type

4. 2. 2. Predictors of aesthetic pleasure

The second model examined the predictors of aesthetic pleasure. Naturalness, familiarity, response time, and image type were entered as fixed effects. Random intercepts were included for participants ($ICC = .106$) and images ($ICC = .074$) to account for individual and stimulus-level variability. The fixed effects accounted for 11.5% of the variance in aesthetic pleasure (marginal $R^2 = .115$), and the full model explained 26.2% of the variance (conditional $R^2 = .262$). Model assumptions were checked visually via residual plots and were deemed acceptable. Robustness checks using different response time filters confirmed that key findings remained stable.

Results revealed significant effects of naturalness ($F(1,1838.1) = 67.46$, $p < .001$), response time ($F(1,1762.2) = 13.44$, $p < .001$), and image type ($F(2,24.5) = 9.39$, $p < .001$), with familiarity showing a marginal effect ($F(1,1837.1) = 3.86$, $p = .050$). Naturalness positively predicted aesthetic pleasure ($\beta = 0.215$, $SE = 0.03$, $t = 8.21$, $p < .001$), while response time was negatively associated ($\beta = -0.040$, $SE = 0.01$, $t = -3.87$, $p < .001$), indicating that faster responses led to higher aesthetic ratings. There was no significant difference in aesthetic ratings between PU and PN images ($\beta = -0.11$, $SE = 0.18$, $t = -0.60$, $p = .549$). However, CN images were rated significantly higher than PN images ($\beta = 0.57$, $SE = 0.18$, $t = 3.20$, $p = .003$), as shown in Figure 4.

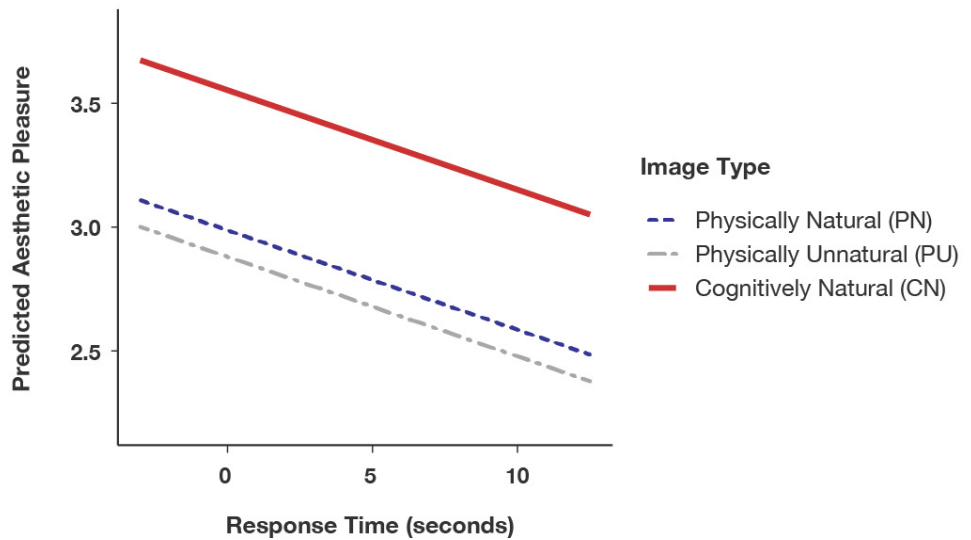


Figure 4 Predicted aesthetic pleasure as a function of response time and image type

4. 3. Mediation analysis

To further clarify the relationship between perceived naturalness and aesthetic pleasure, two mediation analyses were conducted. We tested whether the effect of perceived naturalness on aesthetic pleasure was mediated by processing fluency (indexed by response time) or by familiarity. Although the main analyses employed linear mixed-effects models, the mediation analyses were conducted using linear regression due to methodological limitations in applying standard mediation procedures to LMMs. Indirect effects were estimated via nonparametric bootstrapping (5,000 simulations), a widely accepted and robust approach in mediation research (Preacher & Hayes, 2008).

The first analysis tested response time as a mediator. Results indicated a significant but weak negative mediation effect ($ACME = -0.019, p < .001$), suggesting that increased naturalness was associated with slightly longer response times, which in turn predicted lower aesthetic ratings. While the total effect of naturalness on aesthetic pleasure remained strongly positive ($Total\ effect = 0.217, p < .001$), the proportion mediated was small (-8.6%), indicating that fluency plays only a modest and suppressive role in this relationship.

In contrast, the second analysis tested familiarity as a mediator and found no significant mediation effect ($ACME = -0.0007, p = .96$). The indirect pathway through familiarity was negligible, and the relationship between naturalness and aesthetic pleasure was primarily driven by a direct effect.

These results suggest that while perceived naturalness robustly predicts aesthetic pleasure, this relationship is largely direct. Processing fluency plays a minor and suppressive role, and familiarity does not serve as a significant mediator.

5. Discussion

This study addressed two primary research questions concerning the cognitive underpinnings of perceived naturalness and its relationship to aesthetic pleasure in digital spatial imagery.

First, we asked whether perceived naturalness depends more on familiarity or on fluency. Statistical modeling revealed that familiarity was the strongest predictor of perceived naturalness. Response time, a proxy for fluency, also showed a small but significant positive effect on naturalness. Although this seems counterintuitive—since longer times typically suggest lower fluency—this result likely reflects the specific task demands: evaluating whether a space was real or virtual may have required more deliberate scrutiny for realistic (PN) images. Thus, in this context, longer processing may signal more effortful but confident categorization, not disfluency. Future studies could better isolate intuitive fluency by employing tasks that assess semantic or functional recognition.

Second, we examined how familiarity-based (PN) and fluency-based (CN) spatial images influence aesthetic pleasure. Two levels of analysis provide insight. First, CN images—designed to enhance fluency through visual structure and semantic clarity—received the highest aesthetic ratings, significantly outperforming PN and PU images. Second, in the mixed-effects model, naturalness emerged as the strongest positive predictor of aesthetic pleasure, followed by response time, which showed a negative effect: faster responses predicted higher aesthetic ratings. Familiarity had only a marginal effect.

To further clarify these relationships, mediation analyses were conducted. Results indicated that response time partially and negatively mediated the relationship between naturalness and aesthetic pleasure. This suggests that, although naturalness tends to increase aesthetic ratings, this effect is somewhat weakened when naturalness is accompanied by slower processing. In contrast, familiarity did not significantly mediate the relationship between naturalness and aesthetic pleasure, implying that its contribution is limited to shaping naturalness judgments.

These findings align with previous research on processing fluency theory, which posits that more fluently processed stimuli elicit stronger aesthetic responses (Reber et al., 2004). By contrast, the literature on familiarity offers mixed results: while some studies report positive effects (Zajonc, 1968; Cutting, 2003), this is not always the case across different contexts (Hekkert et al., 2003; Park et al., 2010; Song et al., 2021). In our model, familiarity did not exert a direct effect on aesthetic pleasure. However, given that perceived naturalness—strongly influenced by familiarity—was the most robust predictor of aesthetic ratings, prior reports of familiarity's positive aesthetic effects may be explained through this indirect pathway.

Taken together, these findings point to a dual-pathway model of aesthetic experience. Familiarity enhances perceived naturalness (as seen in PN images) but contributes only marginally to aesthetic pleasure. Fluency, on the other hand, exerts a more direct and consistent influence on aesthetic pleasure, as reflected in both the high ratings and fast processing of CN images. Despite being less familiar and physically implausible, CN images were intuitively grasped and highly aesthetically rated, demonstrating the central role of perceptual fluency.

In conclusion, while Physical Naturalness enhances naturalness through familiarity, it is Cognitive Naturalness—structured for fluent processing—that most powerfully supports aesthetic pleasure. These insights provide theoretical clarity on the perceptual mechanisms of naturalness and offer practical guidance for digital spatial design. Visually intuitive composition and semantic coherence may be more effective than strict physical realism in evoking natural and pleasing spatial experiences.

6. Conclusion

This study examined how perceived naturalness and aesthetic pleasure are shaped by familiarity and fluency in digital spatial images. The findings show that while familiarity significantly enhances perceptions of naturalness, it exerts only a marginal influence on aesthetic pleasure. In contrast, fluency—measured through faster response times—exhibited a direct and robust effect on aesthetic evaluation. These effects were most evident in Cognitively Natural (CN) images, which, despite their physical implausibility, were rated as more aesthetically pleasing than Physically Natural (PN) images.

This study offers a novel operational distinction between Physical Naturalness (driven by familiarity) and Cognitive Naturalness (driven by fluency), addressing a gap in the literature on the perception of digital environments. While previous research has focused primarily on technological realism or spatial presence, the current findings underscore the importance of cognitive-level processes in shaping naturalness. Two key insights emerge: first, that perceived naturalness is strongly associated with familiarity; and second, that even unfamiliar and creatively constructed digital environments can be perceived as natural when processed fluently, thereby enhancing aesthetic pleasure.

These findings also suggest practical directions for design practices involving generative AI. Replicating real-world environments in full detail is not essential to convey naturalness. While visual familiarity—achieved through realistic modelling, materials, and lighting—can enhance perceived naturalness, it does not consistently improve aesthetic pleasure. In contrast, perceptual fluency, supported by clear visual structure and spatial meaning, contributes more directly to aesthetic responses. As generative AI tools already perform well in producing visual realism, future design strategies may benefit more from prioritizing structured visual composition and semantic coherence. This suggests a shift in focus from visual plausibility (i.e., how closely images resemble the physical world) to how spatial compositions are creatively structured and intuitively interpreted.

This study has several limitations that warrant consideration. First, response time was collected via an online platform (Qualtrics), which enabled a larger and more diverse sample for this exploratory investigation. However, online data collection is inherently subject to variability in devices, settings, and participant attention. Although a substantial proportion of outliers was removed and robustness checks confirmed the stability of the results, uncontrolled conditions may have introduced noise into the response time data. Second, the use of single-item measures for key constructs such as naturalness and aesthetic pleasure may limit measurement precision. Third, the study relied exclusively on subjective self-reports, which—while useful for capturing conscious evaluations—may not reflect embodied

or spontaneous responses, highlighting a limitation in ecological validity. Fourth, the generalizability of the findings is limited by the use of a fixed set of image categories (PN, PU, CN) without independent validation. While these categories were theory-driven, their external credibility remains to be tested. Nonetheless, low image-level variance in our mixed-effects models suggests that participants responded more to category-level distinctions than to individual image differences.

Future studies should consider more controlled experimental platforms (e.g., PsychoPy or lab-based settings) and employ multi-item validated scales to improve measurement reliability. Combining subjective ratings with physiological measures—such as eye-tracking, EEG, or facial EMG—may also provide deeper insight into perceptual and affective responses. In addition, immersive technologies such as virtual reality could be used to explore whether the observed relationships between naturalness, fluency, and aesthetic pleasure extend beyond static imagery to more interactive and embodied digital environments.

Overall, this study addresses a key gap in prior research by introducing a cognitive account of naturalness in digital environments. Moving beyond technical realism, we distinguish between physical naturalness, based on familiarity, and cognitive naturalness, grounded in perceptual fluency. Our findings show that even physically implausible but fluently processed images can feel natural and enhance aesthetic pleasure. By articulating this distinction, the study not only contributes to a theoretical understanding of aesthetic experience but also provides actionable guidance: designers of AI-generated environments may benefit more from emphasizing spatial clarity and semantic coherence than from pursuing visual realism alone.

References

1. Adobe. (2024). Adobe Photoshop (Version 2024) [Computer software]. Adobe. <https://www.adobe.com/>
2. Brouillet, D., & Friston, K. (2023). Relative fluency (unfelt vs felt) in active inference. *Consciousness and Cognition*, 115, 103579. <https://doi.org/10.1016/j.concog.2023.103579>
3. Chen, Y., Pollick, F., & Lu, H. (2023). Aesthetic preferences for prototypical movements in human actions. *Cognitive Research Principles and Implications*, 8(1). <https://doi.org/10.1186/s41235-023-00510-0>
4. Coburn, A., Vartanian, O., Kenett, Y. N., Nadal, M., Hartung, F., Hayn-Leichsenring, G., ... & Chatterjee, A. (2020). Psychological and neural responses to architectural interiors. *Cortex*, 126, 217–241.
5. Csathó, Á., van der Linden, D., & Gács, B. (2015). Natural scene recognition with increasing time-on-task: the role of typicality and global image properties. *Quarterly Journal of Experimental Psychology*, 68(4), 814–828.
6. Cutting, J. E. (2003). Gustave Caillebotte, French impressionism, and mere exposure. *Psychonomic Bulletin & Review*, 10(2), 319–343.
7. Gronchi, G., & Sloman, S. A. (2021). Regular and random judgements are not two sides of the same coin: Both representativeness and encoding play a role in randomness perception. *Psychonomic Bulletin & Review*, 28(5), 1707–1714.
8. Halberstadt, J. (2006). The generality and ultimate origins of the attractiveness of prototypes. *Personality and Social Psychology Review*, 10(2), 166–183.
9. Hekkert, P., Snelders, D., & Van Wieringen, P. C. (2003). 'Most advanced, yet acceptable': Typicality and novelty as joint predictors of aesthetic preference in industrial design. *British Journal of Psychology*, 94(1), 111–124.

10. Hou, J., Nam, Y., Peng, W., & Lee, K. M. (2012). Effects of screen size, viewing angle, and players' immersion tendencies on game experience. *Computers in Human Behavior*, 28(2), 617–623.
11. Lee, D. (2021). Exploring the concept of naturalness. *The Korean Journal of Animation*, 17(2), 86–108. <https://doi.org/10.51467/asko.2021.06.17.2.86>
12. Park, J., Shimojo, E., & Shimojo, S. (2009). Roles of familiarity and novelty in visual preference judgments are segregated across object categories. *Proceedings of the National Academy of Sciences*, 107(33), 14552–14555. <https://doi.org/10.1073/pnas.1004374107>
13. Pramod, R. T., & Arun, S. P. (2018). Symmetric objects become special in perception because of generic computations in neurons. *Psychological science*, 29(1), 95–109.
14. Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods*, 40(3), 879–891.
15. Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4), 364–382. https://doi.org/10.1207/s15327957pspr0804_3
16. Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of verbal learning and verbal behavior*, 12(1), 1–20.
17. Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of experimental psychology: General*, 104(3), 192.
18. Seibert, J., & Shafer, D. M. (2018). Control mapping in virtual reality: effects on spatial presence and controller naturalness. *Virtual Reality*, 22(1), 79–88.
19. Song, J., Kwak, Y., & Kim, C. (2021). Familiarity and novelty in aesthetic preference: the effects of the properties of the artwork and the beholder. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.694927>
20. Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, 42(4), 73–93.
21. Sztuka, I. M., & Kühn, S. (2025). Neurocognitive dynamics and behavioral differences of symmetry and asymmetry processing in working memory: insights from fNIRS. *Scientific Reports*, 15(1), 4740.
22. Unkelbach, C., & Greifeneder, R. (2013). A general model of fluency effects in judgment and decision making. In *The experience of thinking* (pp. 11–32). Psychology Press.
23. Vermillion, J. [@joshuavermillion]. (2023, June 25). *Desert Paperscapes: AI-rendered imaginaries from MidJourney* [Image carousel]. Instagram. <https://www.instagram.com/p/Ct6Z6Wij3JZ/>
24. Whittlesea, B. W., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy–attribution account of feelings of familiarity. *Acta Psychologica*, 98(2–3), 141–165. [https://doi.org/10.1016/s0001-6918\(97\)00040-1](https://doi.org/10.1016/s0001-6918(97)00040-1)
25. Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science*, 17(9), 799–806. <https://doi.org/10.1111/j.1467-9280.2006.01785.x>
26. Wu, D. Y., & Lin, J. H. T. (2018). Ways of seeing matter: the impact of a naturally mapped perceptual system on the persuasive effects of immersive virtual reality advertising. *Communication Research Reports*, 35(5), 434–444.
27. Yoo, J., Jasko, K., & Winkielman, P. (2023). Fluency, prediction and motivation: how processing dynamics, expectations and epistemic goals shape aesthetic judgements. *Philosophical Transactions of the Royal Society B Biological Sciences*, 379(1895). <https://doi.org/10.1098/rstb.2023.0326>
28. Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology*, 9(2p2), 1.