Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Controlling Visual Stimuli in Digital Tourism Experiments: A Replicable Framework

Lucília Cardoso
Universidade de Leiria e Oeste
Esaú Ventura Pupo Sirius
Universidade Federal do ABC
Arthur Araújo
Universidade Lusófona
Pablo Henrique Moreira
pablohom@id.uff.br
Universidade Federal de Minas Gerais
Sérgio Silva
University of Trás-os-Montes and Alto Douro
Rui Cardoso
University of Cumbria
Carla Fraga
Universidade Federal de Juiz de Fora
Stanislav Ivanov
Varna University of Management

Research Article

Keywords: visual stimuli, naturalistic images, digital tourism, stimulus control, experimental design, replicability

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11307132/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Controlling Visual Stimuli in Digital Tourism Experiments: A Replicable Framework

Abstract

Digital tourism research increasingly uses visual stimuli to investigate attention, memory, evaluation, and decision-making, often within technology-mediated environments in which visual content forms part of the experimental setting**.** Yet stimulus-selection and control procedures are often insufficiently documented, and uncontrolled affective, perceptual, semantic, or technical variation may provide alternative explanations for observed effects**.** This paper presents a replicable framework for multidimensional control of naturalistic visual stimuli in digital tourism experiments while balancing ecological realism and experimental control. The framework was developed from the stimulus-control requirements of a digital tourism experimental design and was applied to 1,356 Nencki Affective Picture System (NAPS) images using normative metadata filtering, AI-supported semantic classification, rule-based decision criteria, technical normalisation, perceptual screening, expert validation, experimental allocation, and set-level comparison**.** The workflow produced 54 tourism images evenly distributed across Gastronomy, Nature, and Events and allocated to three category-balanced, non-overlapping experimental sets. These sets showed no statistically significant differences in the affective and low-level visual characteristics examined. The contribution lies in a transparent and auditable decision process integrating affective, perceptual, semantic, and technical stimulus control, providing a transferable methodological framework for strengthening the validity, comparability, and replicability of technology-mediated tourism experiments.

Keywords*: visual stimuli; naturalistic images; digital tourism; stimulus control; experimental design; replicability.*

1. Introduction

Images are among the most influential elements of digital tourism communication. Destination websites, booking platforms, social media campaigns, and digital tourism interfaces rely extensively on visual content to attract attention, communicate destination attributes, shape expectations, and influence behavioural intentions (MacKay and Fesenmaier 1997; Stepchenkova and Zhan 2013; Molinillo et al. 2018). This visual dimension has become increasingly important as tourism information and destination representations have expanded across online and image-based platforms, where photographs contribute to how destinations are presented and how tourists form visual impressions before travel (Nixon 2024; Höpken et al. 2026). Tourism photographs contribute to destination image formation and affect how potential visitors process information and evaluate tourism experiences (Li et al. 2016; Lourenção et al. 2020). Evidence from eye-tracking research further demonstrates that even apparently secondary image characteristics, such as the presence of textual information, can alter patterns of visual attention and information processing (Li et al. 2016). Consequently, visual stimuli have become central not only to tourism practice but also to tourism research.

At the same time, the growing use of digital tourism environments means that such information is encountered through technological devices and interfaces. Device use can affect when, how, and in what form travel information is consumed, while content may be displayed differently across devices

(Zarezadeh et al. 2023). These technologically mediated conditions are therefore relevant to research that examines how tourists interact with and process digital tourism content.

As tourism researchers increasingly investigate attention, memory, perception, destination evaluation, website usability, digital experiences, and online consumer behaviour, photographs are frequently used as experimental stimuli (Deng and Li 2018; Molinillo et al. 2018). However, photographs cannot be viewed as neutral carriers of information. Naturalistic images differ in emotional content, visual complexity, familiarity, semantic richness, brightness, contrast, and salience, all of which may influence cognitive and behavioural outcomes independently of the experimental manipulation (Clarke and Ludington 2018; Lang et al. 1993; Marchewka et al. 2014). If these characteristics are not adequately controlled, researchers may attribute observed effects to theoretical constructs of interest when they are partly produced by unintended differences between the visual materials themselves (Lakens et al. 2013; Mather 2007).

This methodological concern becomes especially important when photographs are used to investigate responses to digitally presented tourism content. Studies examining destination websites, tourism advertising, visual communication, digital interfaces, and online destination image increasingly depend on photographic materials to understand how individuals attend to, interpret, and remember tourism content (Deng and Li 2018; Molinillo et al. 2018; Stepchenkova and Zhan 2013). In these settings, visual elements such as background, spatial arrangement, attractiveness, and composition require explicit methodological consideration (Liu et al. 2023). If image characteristics vary unintentionally within or across experimental conditions, they may introduce alternative explanations for observed responses alongside the digital content or manipulation under investigation. Inadequate stimulus control may therefore undermine conclusions regarding how tourists perceive, evaluate, remember, and respond to digital tourism content.

The challenge becomes even more pronounced when researchers seek to employ naturalistic images. Naturalistic photographs preserve much of the perceptual, semantic, and contextual richness of real-world environments and therefore provide stronger ecological validity than artificial laboratory stimuli (Sonkusare et al. 2019). However, this realism comes at a methodological cost. Naturalistic stimuli vary simultaneously across affective, perceptual, semantic, and technical dimensions, creating multiple pathways through which they may influence attention and memory processes (Clarke and Ludington 2018; Marchewka et al. 2014). A photograph may be affectively appropriate but semantically ambiguous; semantically relevant but perceptually dominant; or technically standardised while still attracting disproportionate visual attention. Consequently, selecting tourism images for experimental purposes requires more than choosing visually attractive or thematically relevant content.

Existing approaches address only part of this problem. Normative image databases provide information regarding affective characteristics such as valence and arousal, and physical properties such as luminance, contrast, and entropy (Marchewka et al. 2014). Research in visual cognition demonstrates that familiarity, semantic clarity, category agreement, and visual complexity can also influence recognition and interpretation (Clarke and Ludington 2018). Furthermore, low-level perceptual characteristics such as brightness may influence affective evaluations (Lakens et al. 2013), while emotionally salient content may alter memory performance independently of experimental manipulations (Mather 2007). Together, this evidence shows that different properties of visual stimuli require different forms of methodological control. Affective comparability, for example, does not establish semantic adequacy, while technical standardisation does not eliminate differences in perceptual salience or meaning. As a result, no single control procedure is sufficient to ensure that naturalistic stimuli are simultaneously relevant, comparable, and experimentally appropriate.

The present paper emerged from the development of a digital tourism experimental protocol requiring controlled visual stimuli within predefined semantic categories. The framework emerged from the

stimulus-construction requirements of an experimental design using a simulated tourism webpage composed of photographic cards representing tourism-related information. Because the study manipulates agency through temporal contingencies while holding visual content constant, the construction of the image set became a critical methodological challenge. The visual stimuli therefore had to be controlled carefully so that differences in attention and memory could be examined without introducing systematic variation in the materials presented across the experimental conditions. The objective was not simply to identify appealing tourism photographs but to develop a controlled, category-balanced, and replicable stimulus pool capable of supporting experimental investigations of attention and memory without introducing unintended affective, perceptual, semantic, or technical confounds.

Beyond its methodological contribution, this study responds to a broader challenge facing experimental tourism research. Destination websites, digital platforms, tourism advertising, and online communication increasingly depend on visual content. Yet the properties and selection procedures of visual stimuli remain inconsistently documented in experimental research, including tourism studies using eye-tracking and related visual methods (Hong 2026; Lin et al. 2023). This limits the transparency with which researchers can evaluate how experimental materials were constructed and whether uncontrolled stimulus characteristics may have affected the reported outcomes.

Accordingly, the objective of this paper is to present a replicable framework for selecting, controlling, and validating naturalistic visual stimuli in digital tourism experiments. The proposed framework integrates normative image databases, metadata filtering, artificial-intelligence-supported semantic classification, symbolic decision rules, technical normalisation, perceptual screening, and expert validation within a single transparent workflow. By making stimulus-selection decisions explicit and auditable, the framework helps researchers distinguish the effects of tourism content from the effects of uncontrolled visual properties. Its contribution lies in integrating these complementary control procedures within a single transparent and auditable process, rather than in any individual computational component or in the particular stimulus set produced. In doing so, it contributes to improving the validity, comparability, and replicability of tourism experiments while also generating methodological guidance for research on destination communication, tourism advertising, digital platforms, website design, and online tourism experiences.

2. Methodological foundations of naturalistic visual stimulus control

2.1 Visual stimuli, attention and episodic memory: The methodological challenge

Visual stimuli occupy a central role in tourism research because photographs are widely used to represent destinations, experiences, products, attractions, and digital environments. Studies examining destination image, tourism advertising, website design, online platforms, and visitor experiences increasingly rely on visual materials to investigate how tourists allocate attention, interpret information, evaluate destinations, and form behavioural intentions (Li et al. 2016; Lourenção et al. 2020; MacKay and Fesenmaier 1997; Molinillo et al. 2018; Stepchenkova and Zhan 2013). When photographs are employed as experimental stimuli, however, they cannot be treated as neutral illustrations. Their affective, perceptual, and semantic properties may influence participants' responses independently of the experimental manipulation, thereby affecting attention, memory, recognition, and evaluation processes (Clarke and Ludington 2018; Lang et al. 1993; Marchewka et al. 2014; Mather 2007).

This issue is particularly relevant in digital tourism experiments that seek to understand how tourists process and remember online content. Episodic memory involves not only recognising previously

encountered information but also retrieving the context in which that information was experienced (Davachi 2006; Tulving 2002). Source-memory research further demonstrates that individuals may successfully recognise an item while failing to recover the contextual details associated with it (Johnson et al. 1993; Yonelinas 2002). Consequently, if some images are more distinctive, familiar, emotionally salient, or visually dominant than others, differences in recognition or recall may reflect characteristics of the stimuli rather than the mechanisms under investigation (Clarke and Ludington 2018; Marchewka et al. 2014; Mather 2007).

Attention represents one of the principal pathways through which these effects emerge. Eye-tracking research indicates that gaze allocation reflects both stimulus-driven and goal-directed processes, shaped by factors such as colour, contrast, brightness, visual organisation, task relevance, and expectations (Corbetta and Shulman 2002; Rayner 1998; Tatler et al. 2011; Wolfe 1994). Because naturalistic images vary substantially across these dimensions, they may attract attention in different ways even under identical experimental conditions. Consequently, differences in the affective, perceptual, and semantic properties of naturalistic images may influence the same cognitive processes that tourism experiments seek to measure.

2.2 The methodological tension between ecological validity and experimental control

One of the central methodological challenges in tourism experiments concerns balancing ecological validity and experimental control. Tourism experiences are inherently visual and contextual, and researchers frequently rely on photographs to represent destinations, accommodation, gastronomy, events, and digital tourism environments (MacKay and Fesenmaier 1997; Molinillo et al. 2018; Stepchenkova and Zhan 2013). Naturalistic photographs increase realism and strengthen the correspondence between experimental settings and actual tourism encounters (Sonkusare et al. 2019). However, the same characteristics that enhance realism may also introduce uncontrolled sources of variation. In digital tourism experiments, this means preserving the realism of naturalistic visual content while ensuring that stimuli remain sufficiently comparable across experimental conditions.

Research in visual perception demonstrates that observers do not process all images equally. Features such as colour, brightness, contrast, complexity, figure-ground organisation, and spatial composition may influence visual processing and attentional allocation before conscious interpretation occurs (Treisman and Gelade 1980; Wagemans et al. 2012; Wolfe 1994). Perceptual characteristics may also affect affective evaluations, even when the underlying content remains unchanged (Lakens et al. 2013). In tourism contexts, these differences are particularly relevant because destination photographs often vary considerably in visual composition, emotional appeal, distinctiveness, and familiarity (Li et al. 2016; Lourenção et al. 2020).

The challenge extends beyond perceptual factors. Tourism images differ in recognisability, semantic clarity, familiarity, cultural meaning, and contextual associations (Clarke and Ludington 2018). Iconic landmarks, distinctive gastronomic experiences, or highly recognisable events may draw disproportionate attention and receive different evaluations than less distinctive stimuli. As a result, differences between images may introduce alternative explanations for experimental outcomes, making it difficult to distinguish the effects of the intended manipulation from the effects of the stimuli themselves.

The objective is therefore not to eliminate the richness of naturalistic tourism imagery but to ensure that this richness does not create systematic differences between experimental conditions. Achieving this objective requires a multidimensional approach to stimulus control. Technical standardisation alone cannot guarantee equivalence because images may remain different in emotional tone, semantic clarity, familiarity, visual complexity, or salience even when displayed under identical conditions (Clarke and Ludington 2018; Marchewka et al. 2014**).** Technical consistency is therefore necessary but not sufficient, because affective, semantic, and perceptual differences may remain even after

images have been standardised for presentation**.** Likewise, affective comparability does not necessarily imply semantic comparability, and semantic appropriateness does not guarantee perceptual comparability.

For tourism experiments using visual stimuli, the methodological challenge is therefore not merely selecting relevant photographs but constructing image sets that preserve ecological realism while reducing competing explanations for observed outcomes. This requires researchers to consider affective, perceptual, semantic, and technical characteristics simultaneously as part of experimental design.

2.3 Requirements for a replicable stimulus-control procedure

The issues discussed in the previous sections suggest that image selection should be treated as a formal component of experimental design. Rather than relying on intuitive or *ad hoc* decisions, tourism researchers require procedures capable of documenting, controlling, and validating the visual materials used in experiments. This is particularly important in studies examining destination image, visual communication, digital experiences, tourism advertising, and consumer responses to online content, where photographs frequently serve as the primary experimental stimuli (Deng and Li 2018; Li et al. 2016; Molinillo et al. 2018).

A replicable stimulus-control procedure should ensure that visual stimuli remain ecologically meaningful while minimising unintended differences that may influence participants' responses. The methodological requirements identified above can be organised around four complementary dimensions of control. Affective control concerns properties such as valence and arousal, which may influence responses to visual material independently of the intended manipulation (Lang et al. 1993; Marchewka et al. 2014; Mather 2007). Perceptual control concerns characteristics such as brightness, contrast, visual organisation, complexity, and salience, which may affect visual processing, attention, or evaluation (Lakens et al. 2013; Tatler et al. 2011; Treisman and Gelade 1980; Wagemans et al. 2012; Wolfe 1994). Semantic control concerns familiarity, recognisability, category agreement, and semantic clarity, which are not captured by affective or perceptual measures alone (Clarke and Ludington 2018). Technical control concerns the standardisation of presentation-related characteristics required to prepare the images for consistent use within the digital experimental environment.

These dimensions address different sources of variation and are therefore complementary rather than interchangeable. Technical consistency does not remove affective, perceptual, or semantic differences, just as affective comparability does not establish semantic or perceptual comparability. Together, they provide the operational structure for the proposed stimulus-selection framework.

In practice, this requires documenting the origin of the images, establishing explicit inclusion and exclusion criteria, assessing semantic appropriateness for the categories under investigation, and standardising technical characteristics that may affect presentation conditions. Transparency and auditability apply across these decisions. This is important because incomplete reporting of visual-stimulus properties limits the evaluation and replication of experimental procedures (Lin et al. 2023), while methodological reporting remains a concern in tourism research using visual methods such as eye-tracking (Hong 2026).

Replicability depends not only on reporting the final stimulus set but also on documenting the decisions through which that set was constructed. Recent discussions on stimulus sampling highlight the risks of drawing conclusions from a limited or idiosyncratic collection of stimuli that may not adequately represent the broader category under investigation (Simonsohn et al. 2025). In tourism research, where conclusions are often intended to generalise beyond individual photographs to broader categories such as destinations, experiences, gastronomy, or events, transparent selection procedures and the use of multiple exemplars strengthen both methodological rigour and the

credibility of the resulting findings. Explicit validation of the selected materials is also important rather than assuming their suitability for the intended research purpose (Hogenboom et al. 2024).
The framework proposed in this paper was developed in response to these requirements. Rather than treating image selection as a separate technical task, it integrates affective, perceptual, semantic, technical, and expert-validation procedures within a single auditable workflow. The four control dimensions provide its conceptual structure, while transparency and auditability apply across the selection, standardisation, validation, and allocation decisions through which the final stimulus set is constructed. The objective is to provide tourism researchers with a transparent and transferable approach for constructing image sets that support experimental validity, comparability, and replication across different tourism contexts.

3. Experimental context and stimulus requirements: illustrative application in a digital tourism experiment

The approach presented in this paper was developed within the digital tourism experiment *From Control to Memory: How Digital Interactivity Improves Tourism Experiences*. The framework was developed in response to the stimulus-control requirements of a digital tourism experimental design involving a simulated tourism webpage composed of photographic cards. The webpage provided the technological environment for which the visual materials were constructed. Because differences between images could influence attention, interpretation, or memory independently of the intended manipulation, stimulus construction was treated as part of the experimental design rather than as a preparatory task (Lakens et al. 2013; Marchewka et al. 2014). The stimulus-construction procedure therefore formed part of the experimental design rather than a separate preparatory task.
The experimental design incorporated a recognition-memory paradigm with immediate and delayed testing. Previously presented images were designated as OLD items, whereas previously unseen images used to assess false recognition and discrimination were designated as LURE items. The experimental protocol comprises three phases: an Encoding phase, an immediate memory test (T1), and a delayed memory test (T2) administered 24-48 hours later. To support this design, the stimulus pool was organised into three non-overlapping sets comprising 18 images each: an OLD set presented during Encoding and repeated during both memory tests, a LURE L1 set introduced only during T1, and a LURE L2 set introduced only during T2. Each set contained six Gastronomy, six Nature and six Events images, resulting in a final pool of 54 unique stimuli. Table 1 summarises the composition and experimental function of the three stimulus sets.

Table 1. Composition and experimental function of the stimulus sets

Stimulus set	Experimental function	Number of images	Category distribution	Main constraint
OLD	Encoding	18	6 Gastronomy, 6 Nature, 6 Events	Presented during digital tourism task
LURE L1	Immediate memory test	18	6 Gastronomy, 6 Nature, 6 Events	New exemplars, not shown at encoding
LURE L2	Delayed memory test	18	6 Gastronomy, 6 Nature, 6 Events	New exemplars, not used in OLD or L1

Total	Full stimulus pool	54	18 per category	No overlap between sets

The Gastronomy, Nature and Events categories were selected because they represent common themes used in destination communication, tourism marketing and digital tourism platforms (Molinillo et al. 2018; Stepchenkova and Zhan 2013). At the same time, the content depicted in each image needed to be clearly identifiable, since ambiguity in category membership may influence image interpretation, recognition and memory performance (Clarke and Ludington 2018). The selected images needed to support recognition without being so distinctive, emotionally salient or familiar that they could artificially facilitate memory performance. This requirement reinforced the need for balanced allocation across categories and stimulus sets (Mather 2007). Figure 1 illustrates how this balanced allocation structure was implemented across Encoding, T1 and T2, distinguishing the repeated OLD stimuli from the phase-specific LURE sets.

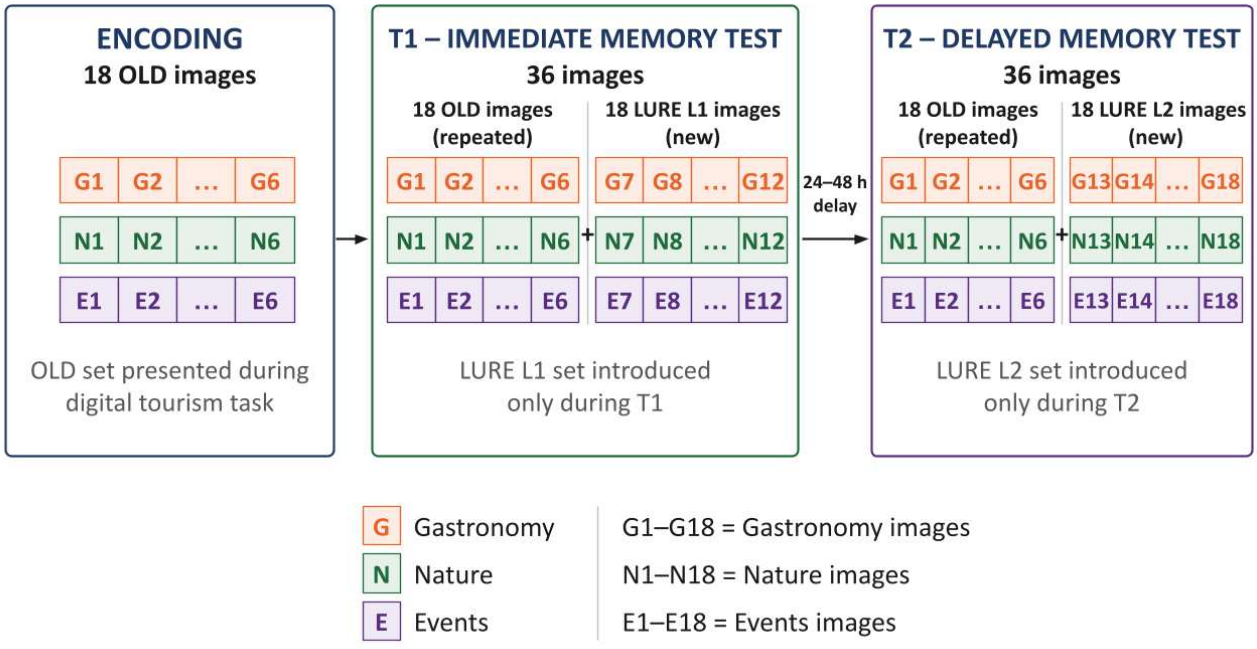


Fig. 1 Experimental deployment of OLD and LURE stimulus sets across Encoding, T1 and T2

The stimulus requirements were therefore strict. The final pool needed to contain 54 unique images distributed equally across the three semantic categories and across the OLD, LURE L1 and LURE L2 sets. Consistent with the four control dimensions defined in Section 2.3, the stimuli also had to satisfy affective, perceptual, semantic, and technical requirements**.** Valence and arousal were considered because of their influence on emotional processing and memory (Marchewka et al. 2014; Mather 2007), while perceptual attributes such as luminance, contrast and entropy may affect attention independently of the intended manipulation (Lakens et al. 2013; Marchewka et al. 2014). Familiarity and semantic clarity also required consideration because they can influence both image recognition and contextual retrieval (Clarke and Ludington 2018). Technical requirements concerned the

consistent preparation of the images for presentation within the simulated webpage, including orientation, image dimensions, and file format.

This structure made intuitive image selection inadequate. The stimulus pool had to satisfy affective, perceptual, semantic and technical requirements simultaneously. Technical standardisation alone could not guarantee comparability, since images with identical dimensions and presentation formats may still differ substantially in emotional content, familiarity, visual complexity or meaning (Clarke and Ludington 2018; Lakens et al. 2013; Marchewka et al. 2014). Conversely, images that were affectively, perceptually, and semantically suitable still required consistent technical preparation for use within the digital experimental environment. The following sections describe how these requirements were addressed through a layered and replicable stimulus-selection workflow.

4. Exploratory normative filtering and semantic insufficiency

The first stage of the workflow aimed to determine whether a normatively documented image database could provide a sufficiently large and category-balanced pool of tourism-relevant stimuli while maintaining control over affective and physical image characteristics. For this purpose, the Nencki Affective Picture System (NAPS) was selected as the starting point because it offers a large collection of naturalistic photographs accompanied by normative information on valence, arousal, approach-avoidance tendencies, luminance, contrast, and entropy (Marchewka et al. 2014). Among the image databases considered, NAPS provided the largest pool of photographs with potential relevance to the tourism categories required by the experimental design. Access to the database was formally obtained from the Nencki Institute of Experimental Biology for academic, non-commercial research purposes.

The initial selection strategy relied on the assumption that normative affective and physical metadata could serve as a first layer of stimulus control. Particular attention was given to valence and arousal because affective characteristics of images may influence emotional responses and memory independently of the experimental manipulation (Marchewka et al. 2014; Mather 2007). Physical descriptors such as luminance, contrast, and entropy were also examined because low-level visual characteristics can influence responses to pictures independently of their semantic content, as demonstrated, for example, for brightness and affective evaluation (Lakens et al. 2013), while these properties are explicitly documented within NAPS (Marchewka et al. 2014). The objective of this stage was therefore to identify images that were affectively moderate, visually comparable, and potentially suitable for subsequent analysis.

To support this exploration, a filtering tool was developed to apply sequential selection criteria to the NAPS metadata and generate auditable subsets of candidate images. The tool allowed systematic filtering of the database according to affective, physical, and orientation-related criteria while documenting the records retained at each stage. Additional implementation details and a screenshot of the NAPS-Filter Tool interface are provided in Appendix A (Fig. 2).

The filtering procedure retained images with mean valence scores between 5.5 and 7.0 and mean arousal scores between 3.5 and 5.5. Horizontal orientation was then required to ensure compatibility with the experimental interface. Additional screening retained images with approach-avoidance, luminance, contrast, LAB colour and entropy values between the 10th and 90th percentiles of the 305

horizontal candidates**.** From the original pool of 1,356 images, 417 satisfied the affective criteria, 305 remained after orientation filtering, and 74 images met the final affective and physical requirements. Detailed results of the exploratory normative and physical screening are provided in Appendix B (Table 7, Table 8, Table 9 and Table 10)**.**

Although this process successfully identified a subset of images meeting the predefined normative screening criteria, it also revealed a critical limitation. NAPS was developed as a general-purpose affective image database rather than as a tourism-specific resource. Consequently, normative metadata could indicate whether an image met affective or physical requirements, but they could not determine whether the image clearly represented the tourism categories required by the experiment. Semantic dimensions such as familiarity, category agreement, recognisability, and semantic clarity are not captured directly by affective ratings or low-level visual descriptors (Clarke and Ludington 2018). An image could therefore be normatively suitable while remaining semantically inappropriate for a tourism website simulation.

To assess this issue, the 74-image subset was examined for correspondence with the three target categories of Gastronomy, Nature, and Events. The detailed assessment is presented in Table 2.

Table 2. Summary of the semantic assessment of the filtered NAPS images

Assessment category	NAPS content represented	n	Tourism relevance	Decision	Methodological interpretation
Clear Gastronomy candidates	Meals, fruit, food stalls	8	High	Retain for further evaluation	Clearly identifiable food-related content
Potential Nature candidates	Wildlife, landscapes, leaves	11	Moderate	Visual review required	Natural content; tourism fit requires confirmation
Potential Events candidates	Play, stall, jogging, painting	4	Low	Insufficient without further review	No clear event or organised activity
Tourism-related content outside target categories	Buildings, sculptures, transport, fountains	12	Low	Exclude	Tourism-related but outside target categories
Face- or person-dominant content	Children, adults, couples, groups	26	Low	Exclude	Potential attentional dominance
Domestic animals	Cats, dogs	5	Low	Exclude	Poor fit with intended Nature category
Semantically irrelevant objects and scenes	Household objects, coins, toys, damaged buildings	8	Low	Exclude	No fit with target categories

As Table 2 shows, only eight images could be clearly identified as Gastronomy candidates, eleven showed potential relevance to Nature, and four could be provisionally associated with Events. Even these counts overestimated the number of usable images because several candidates still required visual inspection to assess tourism relevance, perceptual salience, and semantic clarity. The remaining

images primarily depicted faces, generic social scenes, domestic animals, transport-related objects, buildings, or other content that did not align with the experimental requirements.
The analysis demonstrated that normative filtering alone could not generate the category-balanced stimulus pool required by the experimental design. Most notably, the Events category was severely underrepresented, highlighting the difficulty of deriving tourism-specific categories from a database that was not originally organised around tourism concepts. The problem was therefore not a lack of affective or physical information but a mismatch between normative image classification and the semantic requirements of the research context.
This outcome provided an important methodological insight. Normative filtering proved valuable as an initial mechanism for controlling affective and physical characteristics and for documenting a transparent screening process. However, it was insufficient for identifying tourism-relevant image categories and constructing the 54-image pool required by the experiment. Rather than proceeding with the reduced subset, the workflow returned to the complete NAPS database for a subsequent stage of domain-specific semantic classification. The exploratory filtering stage was therefore retained not as the definitive selection route but as evidence that affective and physical control alone cannot resolve the semantic requirements of naturalistic visual stimulus control in tourism experiments.

5. Proposed multidimensional framework for visual stimulus selection and control

5.1 Semantic classification and decision rules

The limitations identified during exploratory metadata filtering highlighted the need for a semantic classification procedure capable of identifying tourism-relevant content across the complete NAPS database. To address this challenge, the research team developed the NAPS Augmented Classifier, a Python-based application that combines CLIP image-text similarity with explicit decision rules. The prompts used during classification are reported in Table 3, while additional implementation details and decision thresholds are provided in Appendix C.
The classifier was applied to the complete set of 1,356 NAPS images. Using Contrastive Language-Image Pre-training (CLIP; Radford et al. 2021), images were compared with researcher-defined prompts representing the target categories of Nature, Gastronomy and Events, as well as exclusion categories associated with unsuitable content. Rather than relying on predefined labels, CLIP estimates the semantic similarity between an image and textual descriptions, allowing categories to be operationalised according to the requirements of the experiment. This stage operationalised the semantic-control requirement identified in Section 2.3 by screening the database against tourism-specific category descriptions rather than relying solely on the original NAPS organisation. This distinction is relevant because general-purpose visual classifications do not necessarily capture tourism-specific semantic attributes (Nixon 2024).
To make the classification decisions explicit and traceable, classification outputs were subjected to explicit symbolic rules. An initial 80% classification-score threshold was used with eight prompts. When the inclusion set was reduced from five to three categories, the threshold was lowered to 75% to reflect the smaller number of competing categories rather than the observed classification results. A 10-percentage-point margin between the highest and second-highest scores was also specified. This

rule did not exclude any images because, with six softmax-normalised prompts, a leading score of ≥75% necessarily produced a margin of at least 50 percentage points; the smallest observed margin was 51.6. Images reaching the ≥80% blocking threshold for an exclusion prompt were automatically removed. Images failing the classification-score or margin thresholds were classified as uncertain and retained for potential expert evaluation.

Application of this procedure generated category-specific candidate pools while preserving a large number of uncertain cases rather than forcing artificial classifications. This conservative strategy prioritised semantic clarity over category coverage and reinforced the role of subsequent expert evaluation. Table 3 presents the prompts used during classification. The symbolic thresholds and decision rules are reported in Appendix C to facilitate replication.

Table 3. Textual prompts used in the CLIP-based classification

Role	Category label	Textual prompt
Inclusion	Nature	Natural landscapes, mountains, canyons, forests, wild scenery, wild animals, natural parks and preserves.
Inclusion	Events	Meeting spots, concerts, theatres, shows and venues, parks, outdoor and cultural activities, leisure and nightlife.
Inclusion	Gastronomy	Cooked food, restaurants, traditional dishes, wine, cocktails, sweets, desserts and high gastronomy.
Exclusion	Medical_Veterinary	A medical, surgical or veterinary scene, dissection or surgical tools.
Exclusion	Violence_Weapons	Violence, firearms, handguns, rifles or military conflict.
Exclusion	Neutral_Boring	A plain everyday object, dull texture or functional item.

The classification procedure identified 108 Nature candidates, 96 Gastronomy candidates and 52 Events candidates, while 851 images remained uncertain and 249 were blocked. The large uncertain category reflected the conservative nature of the decision rules and reinforced the need for expert review**.** The asymmetry among categories also confirmed the uneven representation of tourism-related content within the original database, particularly for Events.

5.2 Technical and perceptual control

Following semantic classification, candidate images were subjected to technical and perceptual control procedures designed to reduce non-experimental sources of variation. Of the 256 classified candidates, 49 were removed by the orientation criterion; the remaining images were standardised in dimensions and file format to support consistent presentation within the simulated tourism interface. These technical specifications were implementation choices for the present experiment rather than general standards for digital tourism research.

Technical normalisation alone, however, does not guarantee perceptual equivalence. Therefore, an additional screening stage was implemented to identify elements likely to generate disproportionate bottom-up attentional capture. Images dominated by close-up human faces or highly salient foreground animals were excluded, while images depicting people or animals as integrated

components of broader scenes could be retained. Researcher pre-screening removed 39 images. One image falling just below the classification-score threshold was nevertheless retained for expert review, resulting in 169 images entering the initial expert-validation round.

Together, technical and perceptual control procedures contributed to improving the comparability of the candidate stimulus pool while preserving the ecological realism required for a digital tourism environment. Technical control addressed consistency of experimental presentation, whereas perceptual screening addressed visual characteristics that could make particular stimuli disproportionately salient. The technical-normalisation specifications applied to the images are summarised in Appendix C.

5.3 Expert validation of domain and experimental adequacy

The final stage of the workflow involved expert validation of the candidate images. Although the previous stages addressed perceptual, semantic, and technical control, human evaluation remained necessary to assess whether the images were appropriate for the specific tourism categories and experimental objectives. This validation complemented the preceding control stages by assessing category and experimental suitability that could not be established from affective, physical, or computational information alone (Clarke and Ludington 2018; Hogenboom et al. 2024).

The stimulus set was evaluated by a four-member expert panel, including tourism researchers and one specialist with neurocognitive training. Candidate images were assessed using a structured evaluation instrument containing image identifiers, individual ratings, category rankings, inclusion decisions, and qualitative observations. The evaluation instrument used by the expert panel is reproduced in Appendix D.1 (Table 11 and Table 12). Images were rated on a five-point scale, where 1 indicated exclusion and 5 indicated high-confidence inclusion. Images with average ratings of 2.5 or above were considered eligible for selection.

The expert-validation pool comprised 172 candidate images, including 33 Events images, 66 Gastronomy images, and 73 Nature images. Of these, 169 images were independently evaluated by all four experts, while three additional Events images were introduced later during the consensus phase. The process resulted in the selection of 54 unique images, with 18 images retained in each tourism category. Table 4 summarises the number of candidates assessed, descriptive statistics for the selected images, their NAPS source categories, and the final selection decisions. Detailed validation records, rankings, and consensus decisions are reported in Appendix D.2.

Across the three categories, Nature received the highest evaluations (M = 4.79), followed by Gastronomy (M = 4.13) and Events (M = 3.78). While quantitative ratings provided the primary basis for selection, final decisions also considered category quotas, semantic appropriateness, visual diversity, and the avoidance of redundant imagery.

Table 4. Expert-validation outcomes and final image selection by category

Category	Pool	Rated	Selected	Mean	Range	NAPS source categories of selected images	Decision
Events	33	30	18	3.78	2.50–5.00	Faces, People and Objects	Three images with mean ratings below 2.5 were excluded and replaced by three alternative images through expert consensus.
Gastronomy	66	66	18	4.13	3.50–5.00	Objects and People	The final selection followed the category ranking and the requirement of eighteen images.
Nature	73	73	18	4.79	4.50–5.00	Landscapes, People and Animals	Expert consensus was used to resolve the tie at the final selection boundary and to avoid visual redundancy.
Total	172	169	54	—	—	NAPS naturalistic photographs	The final pool met the predefined requirement of 54 stimuli.

Note. For Events, the mean and range refer to the 15 selected images that received formal expert ratings; the three additional images selected during the consensus phase were not part of the initial rating round. NAPS source categories are reported at category level only. Detailed image-level validation outcomes are reported in Appendix D.2.

Most decisions followed directly from the ranking procedure. However, several situations required expert consensus. In the Events category, three initially evaluated images received mean ratings below the inclusion threshold and were excluded. Because the remaining ranked images were insufficient to complete the category, three alternative images were introduced through consensus agreement. In the Nature category, consensus was used to resolve ties at the final selection boundary and ensure visual diversity among the retained scenes. Gastronomy did not require additional consensus decisions, as the final eighteen images emerged directly from the ranking process.

The outcome of this stage was a validated pool of 54 tourism images distributed equally across Gastronomy, Nature, and Events. All ratings, rankings, observations, inclusion decisions, and consensus outcomes were retained in the validation records, ensuring a transparent audit trail of the final stimulus-selection process. These records are summarised in Appendix D.2.

Following validation, the 54 selected images were randomly allocated within each tourism category to the three experimental sets defined in Section 3. Each set contained six Gastronomy, six Nature, and six Events images, preserving category balance across the Encoding, T1, and T2 phases of the experiment. The complete allocation of images to the OLD, LURE L1, and LURE L2 sets is reported in Appendix E (Table 13).

5.4 Comparability of final stimulus sets

Following the allocation of the 54 images, the three experimental sets were compared on the affective and low-level visual characteristics relevant to stimulus control. Valence and arousal represented the normative affective properties of the images, whereas luminance, contrast and entropy were used as low-level visual descriptors relevant to stimulus comparability. Descriptive statistics were calculated for each set. Welch's one-way analyses of variance were used to examine differences across OLD, LURE L1 and LURE L2. The descriptive statistics and analysis-of-variance results are presented in Table 5.

Table 5. Comparison of affective and physical properties across the final stimulus sets

Characteristic	Encoding M (SD)	T1 M (SD)	T2 M (SD)	Welch's F	*p*
Valence	7.04 (0.82)	6.74 (1.10)	7.00 (0.76)	F(2, 33.33) = 0.46	.633
Arousal	4.59 (0.92)	4.51 (1.18)	4.82 (0.90)	F(2, 33.57) = 0.47	.629
Luminance	128.26 (29.45)	119.32 (34.16)	113.70 (39.43)	F(2, 33.53) = 0.84	.439
Contrast	64.28 (12.93)	64.00 (12.56)	68.01 (15.76)	F(2, 33.71) = 0.40	.671
Entropy	7.63 (0.27)	7.47 (0.41)	7.28 (0.77)	F(2, 30.27) = 2.17	.131

Note. Each set comprised 18 unique images, including six Gastronomy, six Nature and six Events stimuli. OLD refers to the images presented during the initial interactive task and repeated during both memory tests. LURE L1 and LURE L2 refer to the new images used in the immediate and delayed memory tests, respectively. Valence and arousal are the overall normative mean ratings reported in NAPS, with higher values indicating greater pleasantness and affective activation, respectively. Luminance represents mean image brightness, contrast represents variation in pixel intensity, and entropy represents the amount of visual information or complexity within an image. M = mean; SD = standard deviation. The allocation structure underlying these comparisons is provided in Appendix E.

These results do not demonstrate strict statistical equivalence between the three sets. They nevertheless provide quantitative support for the final selection and allocation decisions by showing that OLD, LURE L1 and LURE L2 did not differ significantly in the affective and low-level visual characteristics examined. Together with the equal category distribution, the expert-validation results and the absence of overlap between sets, this analysis indicates that the workflow produced three stimulus sets with no statistically significant differences in the examined affective and low-level visual characteristics.

6. The Final Replicable Workflow

The final approach can be described as a layered workflow moving from a broad, normatively documented image database to a reduced, controlled and expert-validated stimulus set. The workflow was not entirely linear. An exploratory metadata-filtering stage first produced a reduced subset, but its semantic insufficiency led the selection process to return to the complete NAPS database before semantic classification was conducted. Across its 11 phases, the workflow integrates successive

functions of stimulus sourcing and exploration, semantic control, technical and perceptual control, expert validation, experimental allocation, and final set-level comparison. These stages operationalise the affective, perceptual, semantic and technical control requirements defined earlier within a single documented process. The main contribution, summarised in Table 6, is the transparent sequence of decisions that produced the final set of images, rather than the set itself. This emphasis responds to concerns about incomplete reporting of visual-stimulus characteristics and methodological procedures (Lin et al. 2023; Hong 2026), while explicit stimulus-selection and sampling decisions are also relevant to the interpretation and generalisation of experimental findings (Simonsohn et al. 2025).

Table 6. Final replicable workflow for naturalistic visual stimulus control

Phase	Problem addressed	Methodological operation	Output
1. Normative image base	Avoid arbitrary image sourcing	Start from NAPS, a normatively documented database	Initial pool of 1,356 images
2. Metadata exploration	Characterise affective and physical properties	Review normative affective and physical metadata	Normative reference for stimulus control
3. Exploratory metadata filtering	Assess whether normative and physical filtering could produce a sufficient domain-relevant pool	Apply sequential multivariate filters through the NAPS-Filter Tool	Auditable subset of 74 images and evidence of semantic insufficiency
4. Return to the complete image database	Avoid excluding potentially relevant images before semantic assessment	Reintroduce the complete set of 1,356 NAPS images as the input for semantic classification	Complete image pool available for domain-specific classification
5. Semantic classification	Identify domain-relevant categories	CLIP-based image–text similarity using explicit prompts	Softmax-normalised similarity scores across the inclusion and exclusion prompts
6. Symbolic rules	Reduce ambiguity and exclude unsuitable content	Classification-score, margin and blocking thresholds	Classified, uncertain or blocked images
7. Technical normalisation	Control image format, dimensions and presentation consistency	Horizontal orientation, 500 × 375 px, JPEG and approximate file-size control	Technically standardised stimuli
8. Perceptual filtering	Reduce bottom-up attentional confounds	Exclusion of close-up faces and foreground animals	Perceptually more comparable candidates
9. Expert validation	Ensure experimental and domain adequacy	Four-expert rating, mean threshold, ranking and consensus	Final validated pool of 54 images, with 18 images per tourism category
10. Experimental allocation	Create category-balanced, non-	Randomly allocate images within each tourism category to OLD, LURE L1 and LURE	Three category-balanced sets of 18

	overlapping stimulus sets	L2, maintaining six images per category in each set	images with no overlap between sets
11. Set-level comparison	Check whether the final allocation introduced systematic affective or low-level visual differences across experimental sets	Compare OLD, LURE L1 and LURE L2 on valence, arousal, luminance, contrast and entropy using descriptive statistics and inferential tests	Quantitative evidence that no statistically significant differences were detected across the three sets in the properties examined

This workflow preserves ecological validity because it retains naturalistic photographic stimuli rather than reducing the experiment to abstract laboratory displays. At the same time, naturalistic stimuli introduce greater variability across affective, perceptual and semantic characteristics, creating a methodological tension between realism and experimental control (Sonkusare et al. 2019; Clarke and Ludington 2018; Marchewka et al. 2014). The workflow addresses this tension by applying different forms of control at successive stages rather than relying on a single standardisation procedure. The approach also strengthens auditability by ensuring that each stage leaves an inspectable record of the decisions through which the final stimulus set was constructed, consistent with broader concerns regarding stimulus reporting, methodological transparency and stimulus sampling (Lin et al. 2023; Hong 2026; Simonsohn et al. 2025). These records include database identifiers, normative metadata, exploratory filtering decisions, softmax-normalised similarity scores, category margins, blocking decisions, technical transformations, perceptual exclusions, expert ratings, final inclusion decisions and experimental-set allocation.

Replicability emerges from the explicitness of the process. Other researchers can use the same database, adapt the prompts, modify thresholds according to their domain, report the distribution of classified and uncertain images, document technical normalisation and structure expert validation. The exact final images may differ across studies, but the decision logic can be repeated, evaluated and adapted. In this sense, replicability concerns the transparency and repeatability of the decision process rather than reproduction of an identical final stimulus set (Hong 2026; Simonsohn et al. 2025). To facilitate replication of the computational stages of the workflow, the NAPS-Filter Tool and the NAPS Augmented Classifier used for CLIP-based semantic classification will be made available for academic and non-commercial research upon reasonable request.

7. Discussion

7.1 Methodological Implications and Transferability

The primary contribution of the proposed framework is a multidimensional and auditable approach to visual stimulus control in technology-mediated tourism experiments. It integrates affective, perceptual, semantic, and technical control within a single documented workflow, addressing the fact that tourism information is increasingly encountered through digital environments in which the mode of technological presentation forms part of how content is accessed and consumed (Zarezadeh et al. 2023). Visual stimuli are central to contemporary tourism studies because destinations, experiences,

tourism products, and services are increasingly communicated through digital images distributed via websites, social media, online travel platforms, and destination marketing campaigns (Molinillo et al. 2018; Nixon 2024; Höpken et al. 2026). As a result, conclusions regarding tourists' attention, perceptions, preferences, memory, and behavioural intentions often depend directly on the characteristics of the images used in experimental settings. Yet the procedures used to select and control those images are frequently underreported. The framework proposed here demonstrates that image selection should be regarded as an explicit component of experimental design rather than as a preparatory stage conducted before data collection.

This contribution is particularly relevant for research on destination image formation and destination communication. Tourism studies frequently seek to understand how visual representations influence the way destinations are perceived, remembered, and evaluated. However, comparisons between destination photographs can be distorted when differences in emotional salience, visual complexity, familiarity, or semantic clarity are not adequately controlled. By providing a structured process for selecting and validating images, the proposed workflow improves researchers' ability to isolate the effects of destination attributes from the effects of uncontrolled visual characteristics. This, in turn, supports more robust evidence regarding the relationship between projected and perceived destination images (Deng and Li 2018; Stepchenkova and Zhan 2013).

The framework also has implications for research on tourism advertising and digital persuasion. Experimental studies increasingly examine how visual content influences attention allocation, emotional responses, advertising effectiveness, and behavioural intentions (Li et al. 2016; Lourenção et al. 2020). Because attention and affective responses can also vary with characteristics of the visual stimulus itself, including familiarity and other image-related features (Liu et al. 2023), differences in image composition, recognisability, familiarity, or perceptual salience may affect outcomes independently of the marketing message under investigation. The workflow presented here offers researchers a practical approach for reducing these sources of bias while preserving the ecological realism that characterises tourism communication.

The relevance of stimulus control becomes particularly clear in research on digital interfaces and platforms. Online visual media play an increasingly important role in destination representation and destination-image formation (Nixon 2024), while image-based platforms allow tourists to form visual impressions and destinations to present themselves to potential visitors (Höpken et al. 2026). At the same time, tourism information may be presented and consumed differently across technological devices and interfaces (Zarezadeh et al. 2023). In experimental studies of websites, platforms, online interaction, or digitally mediated tourism experiences, visual content therefore forms part of the material through which those environments are evaluated. Controlling that content helps reduce alternative explanations when researchers seek to attribute observed responses to interface design, interaction, engagement, or other digital manipulations.

The computational component of the workflow should be interpreted within this wider methodological structure. AI-supported semantic classification was used as one operational stage to screen the complete image database against researcher-defined semantic categories at scale. CLIP provided the image-text similarity mechanism used for this purpose (Radford et al. 2021), while explicit decision rules and subsequent validation determined how the resulting classifications entered the selection process. The methodological contribution therefore does not lie in AI-assisted classification itself, but in its integration with normative filtering, technical standardisation, perceptual screening, expert validation, experimental allocation, and set-level comparison within a transparent workflow.

Although developed within a study of visual attention and memory, the framework is transferable to a wide range of tourism contexts. Its most direct applications are technology-mediated settings in which visual content forms part of the experimental environment, including destination websites, booking and online travel platforms, digital advertising, social media, and virtual or immersive tourism experiences. It may also support research on destination branding, gastronomy tourism, event marketing, accommodation marketing, digital storytelling, and visitor engagement with destination content. The present application also showed that tourism categories were not equally straightforward to operationalise. Nature and Gastronomy generated relatively robust candidate pools, whereas Events proved substantially more challenging and required additional expert intervention during the validation stage. This illustrates that implementation of the framework may need to be adapted to the semantic characteristics and representation of the tourism domain under investigation.

The specific categories, prompts, thresholds, and validation procedures may vary according to the research question, but the underlying logic remains the same: images should be selected through transparent procedures that balance ecological realism with methodological control.

Ultimately, the contribution of the framework extends beyond the construction of a particular stimulus set. By making the selection, validation, and allocation of tourism images explicit and auditable, the workflow contributes to improving the validity, comparability, and replicability of experimental tourism research. Its methodological value lies in integrating previously separate control and validation procedures into a traceable decision process that can be adapted to different technology-mediated tourism settings. More reliable experimental evidence, in turn, can support stronger theoretical development and provide destination managers, marketers, platform designers, and tourism organisations with more trustworthy knowledge regarding how tourists perceive, remember, and evaluate visual tourism content.

7.2 Limitations and future applications

Several limitations should be considered when interpreting the proposed framework and its application within the present study.

First, the NAPS was not developed specifically for tourism research. Although it provides a robust normative foundation and extensive affective and perceptual metadata (Marchewka et al. 2014), the database was not designed around tourism constructs. Consequently, some tourism-relevant categories were more readily represented than others. Nature, for example, was relatively easy to populate, whereas Events proved considerably more difficult because of its greater semantic heterogeneity. This limitation reflects a broader challenge facing tourism researchers who rely on general-purpose image databases to construct domain-specific experimental stimuli. While the exploratory filtering process demonstrated the value of normative databases for affective and perceptual control, it also highlighted the importance of additional semantic-validation procedures when the objective is to represent tourism-related concepts accurately.

Second, the framework was validated using three broad tourism categories: Nature, Gastronomy, and Events. These categories were selected because they are commonly represented across destination websites, digital platforms, and tourism-promotion materials. Nevertheless, they do not encompass the full diversity of tourism content. Future applications should examine whether the same workflow can be adapted to other tourism domains and visual categories rather than assuming that the present operational criteria transfer unchanged.

Third, the semantic-classification stage depended on researcher-defined prompts and the performance of the CLIP model. The categories investigated in this study do not exist as predefined experimental

constructs within the model itself; rather, they are operationalised through textual descriptions supplied by the researchers (Radford et al. 2021). Different prompt formulations may therefore produce different candidate pools. The resulting classifications should consequently be interpreted as outputs of the specific semantic operationalisation used in this study, rather than as fixed or universally valid category assignments.

Fourth, the symbolic thresholds used during classification represented study-specific methodological decisions. More restrictive thresholds would increase semantic purity but reduce the number of available candidates, whereas more permissive thresholds would expand the candidate pool at the cost of greater ambiguity. Accordingly, the values adopted in this study should be understood as justified design choices rather than universal standards. Future applications may adapt these thresholds to their own categories and experimental requirements, provided that the criteria and resulting decisions are reported transparently.

Fifth, technical normalisation reduced but did not eliminate all potential sources of presentation variability. Standardising image orientation, dimensions, and approximate file size helped minimise unintended technical differences, but loading times may still vary depending on hardware, browser implementation, platform architecture, and user environments. More broadly, the framework was developed within one simulated digital tourism environment, and its technical implementation choices should not be assumed to transfer unchanged across devices, interfaces, screen conditions, or platforms. This caution is particularly relevant because tourism information may be presented and consumed differently across technological devices (Zarezadeh et al. 2023).

Sixth, expert validation inevitably involves human judgement. The expert panel was used to assess domain and experimental adequacy that could not be established from normative metadata or computational classification alone**.** Nevertheless, the present study relied on a relatively small panel of specialists. Future applications may use larger or differently constituted panels where the research context requires broader domain, cultural, or contextual expertise.

Finally, the present study establishes methodological validation of the stimulus-construction process but does not provide participant-level evidence regarding the behavioural performance of the resulting image sets. Subsequent research may examine whether the selected stimuli produce comparable distributions of visual attention, recognition, source memory, behavioural responses, and subjective evaluations. Such analyses would help determine whether any remaining differences between images or categories influence participant responses in meaningful ways.

Beyond these limitations, the framework opens several avenues for future tourism research. Priority applications include destination websites, booking and online travel platforms, digital advertising, social media, and virtual or immersive tourism environments, followed by other contexts such as destination branding, gastronomy, events, accommodation, and digital storytelling. More broadly, the framework offers a practical foundation for improving the transparency, comparability, and replicability of visual experiments in tourism. As digital communication continues to shape how destinations are presented, perceived, and remembered, rigorous approaches to visual stimulus control will become increasingly important for both tourism scholarship and evidence-based tourism management.

8. Conclusion

Visual content has become a central component of digital tourism communication. Destination websites, online platforms, tourism advertising, and social media increasingly rely on photographs to influence how destinations and experiences are perceived, remembered, and evaluated. At the same

time, tourism information is increasingly encountered through technologically mediated environments, making the control of visual stimuli an important component of experimental validity. As a result, the quality of experimental evidence in tourism research depends not only on the manipulation being tested but also on the characteristics of the images used to represent tourism content. Uncontrolled differences in visual stimuli may provide alternative explanations for observed effects and reduce the validity, comparability, and replicability of research findings.

Using a digital tourism experiment as an illustrative application, this paper presented a replicable framework for selecting, controlling, and validating naturalistic images. The proposed workflow integrates exploratory normative filtering, semantic classification, rule-based decision criteria, technical normalisation, perceptual screening, expert validation, experimental allocation, and final set-level comparison within a transparent and auditable process. Across these stages, the framework operationalises four complementary dimensions of stimulus control: affective, perceptual, semantic, and technical. Rather than treating image selection as a preparatory or aesthetic task, the framework positions it as an integral component of experimental design.

The principal contribution of this study lies in the development of a transferable methodology that can be adapted to a wide range of tourism research contexts, rather than in the construction of a specific stimulus set. Its contribution therefore resides in the transparent decision process through which stimuli are sourced, controlled, validated, and allocated, rather than in any particular computational component or final image pool. The framework provides researchers with a structured approach for balancing ecological validity and experimental control when using visual materials to study destination communication, tourism advertising, digital experiences, website interaction, attention, memory, and decision-making.

More broadly, the framework contributes to improving the quality of experimental research in tourism by making stimulus-selection decisions explicit, transparent, and replicable. As tourism becomes increasingly digital and visually mediated, rigorous approaches to visual stimulus control will be essential for producing more credible evidence about how tourists attend to, remember, evaluate, and respond to tourism content. By supporting transparent and multidimensional stimulus control, the framework can strengthen the validity, comparability, and replicability of experiments conducted through destination websites, digital platforms, and other technology-mediated tourism environments.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this paper, the authors used ChatGPT (OpenAI) to assist with textual revision, language refinement, and translation. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the published article.

References

Clarke AJB, Ludington JD (2018) Thai norms for name, image, and category agreement, object familiarity, visual complexity, manipulability, and age of acquisition for 480 color photographic objects. J Psycholinguist Res 47:607–626. https://doi.org/10.1007/s10936-017-9544-5

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201–215. https://doi.org/10.1038/nrn755

Davachi L (2006) Item, context and relational episodic encoding in humans. Curr Opin Neurobiol 16:693–700. https://doi.org/10.1016/j.conb.2006.10.012

Deng N, Li XR (2018) Feeling a destination through the "right" photos: A machine learning model for DMOs' photo selection. Tour Manag 65:267–278. https://doi.org/10.1016/j.tourman.2017.09.010

Hogenboom SAM, Schulz K, van Maanen L (2024) Implicit association tests: Stimuli validation from participant responses. Br J Soc Psychol 63:975–1002. https://doi.org/10.1111/bjso.12688

Hong WCH (2026) Towards rigorous eye-tracking methodology in interdisciplinary fields: Insights from and recommendations for tourism research. J Eye Mov Res 19:31. https://doi.org/10.3390/jemr19020031

Höpken W, Wang Z, Schwarzenbacher S, Fuchs M (2026) Comparison of tourists' perceptions and destinations' self-presentation through image recognition. Inf Technol Tourism 28:16. https://doi.org/10.1007/s40558-025-00355-x

Johnson MK, Hashtroudi S, Lindsay DS (1993) Source monitoring. Psychol Bull 114:3–28. https://doi.org/10.1037/0033-2909.114.1.3

Lakens D, Fockenberg DA, Lemmens KPH, Ham J, Midden CJH (2013) Brightness differences influence the evaluation of affective pictures. Cogn Emot 27:1225–1246. https://doi.org/10.1080/02699931.2013.781501

Lang PJ, Greenwald MK, Bradley MM, Hamm AO (1993) Looking at pictures: Affective, facial, visceral, and behavioral reactions. Psychophysiology 30:261–273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

Li Q, Huang ZJ, Christianson K (2016) Visual attention toward tourism photographs with text: An eye-tracking study. Tour Manag 54:243–258. https://doi.org/10.1016/j.tourman.2015.11.017

Lin Z, Ma Q, Huang X, Wu X, Zhang Y (2023) Pervasive failure to report properties of visual stimuli in experimental research in psychology and neuroscience: Two metascientific studies. Psychol Bull 149:487–505. https://doi.org/10.1037/bul0000399

Liu B, Moyle B, Kralj A, Li Y (2023) Celebrity endorsement in tourism: Attention, emotional arousal and familiarity. Tour Manag 98:104750. https://doi.org/10.1016/j.tourman.2023.104750

Lourenção MTA, Giraldi JME, de Oliveira JHC (2020) Destination advertisement semiotic signs: Analysing tourists' visual attention and perceived ad effectiveness. Ann Tour Res 84:103001. https://doi.org/10.1016/j.annals.2020.103001

MacKay KJ, Fesenmaier DR (1997) Pictorial element of destination in image formation. Ann Tour Res 24:537–565. https://doi.org/10.1016/S0160-7383(97)00011-X

Marchewka A, Żurawski Ł, Jednoróg K, Grabowska A (2014) The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. Behav Res Methods 46:596–610. https://doi.org/10.3758/s13428-013-0379-1

Mather M (2007) Emotional arousal and memory binding: An object-based framework. Perspect Psychol Sci 2:33–52. https://doi.org/10.1111/j.1745-6916.2007.00028.x

Molinillo S, Liébana-Cabanillas F, Anaya-Sánchez R, Buhalis D (2018) DMO online platforms: Image and intention to visit. Tour Manag 65:116–130. https://doi.org/10.1016/j.tourman.2017.09.021

Nixon LJB (2024) Do deep learning models accurately measure visual destination image? A comparison of a fine-tuned model to past work. Inf Technol Tourism 26:377–406. https://doi.org/10.1007/s40558-024-00293-0

Radford A, Kim JW, 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. In: Meila M, Zhang T (eds) Proceedings of the 38th International Conference on Machine Learning. Proc Mach Learn Res 139:8748–8763. https://proceedings.mlr.press/v139/radford21a.html

Rayner K (1998) Eye movements in reading and information processing: 20 years of research. Psychol Bull 124:372–422. https://doi.org/10.1037/0033-2909.124.3.372

Simonsohn U, Montealegre A, Evangelidis I (2025) Stimulus sampling reimagined: Designing experiments with mix-and-match, analyzing results with stimulus plots. J Pers Soc Psychol 129:71–90. https://doi.org/10.1037/pspa0000449

Sonkusare S, Breakspear M, Guo CC (2019) Naturalistic stimuli in neuroscience: Critically acclaimed. Trends Cogn Sci 23:699–714. https://doi.org/10.1016/j.tics.2019.05.004

Stepchenkova S, Zhan F (2013) Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. Tour Manag 36:590–601. https://doi.org/10.1016/j.tourman.2012.08.006

Tatler BW, Hayhoe MM, Land MF, Ballard DH (2011) Eye guidance in natural vision: Reinterpreting salience. J Vis 11:5. https://doi.org/10.1167/11.5.5

Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cogn Psychol 12:97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Tulving E (2002) Episodic memory: From mind to brain. Annu Rev Psychol 53:1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Wagemans J, Elder JH, Kubovy M, Palmer SE, Peterson MA, Singh M, von der Heydt R (2012) A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. Psychol Bull 138:1172–1217. https://doi.org/10.1037/a0029333

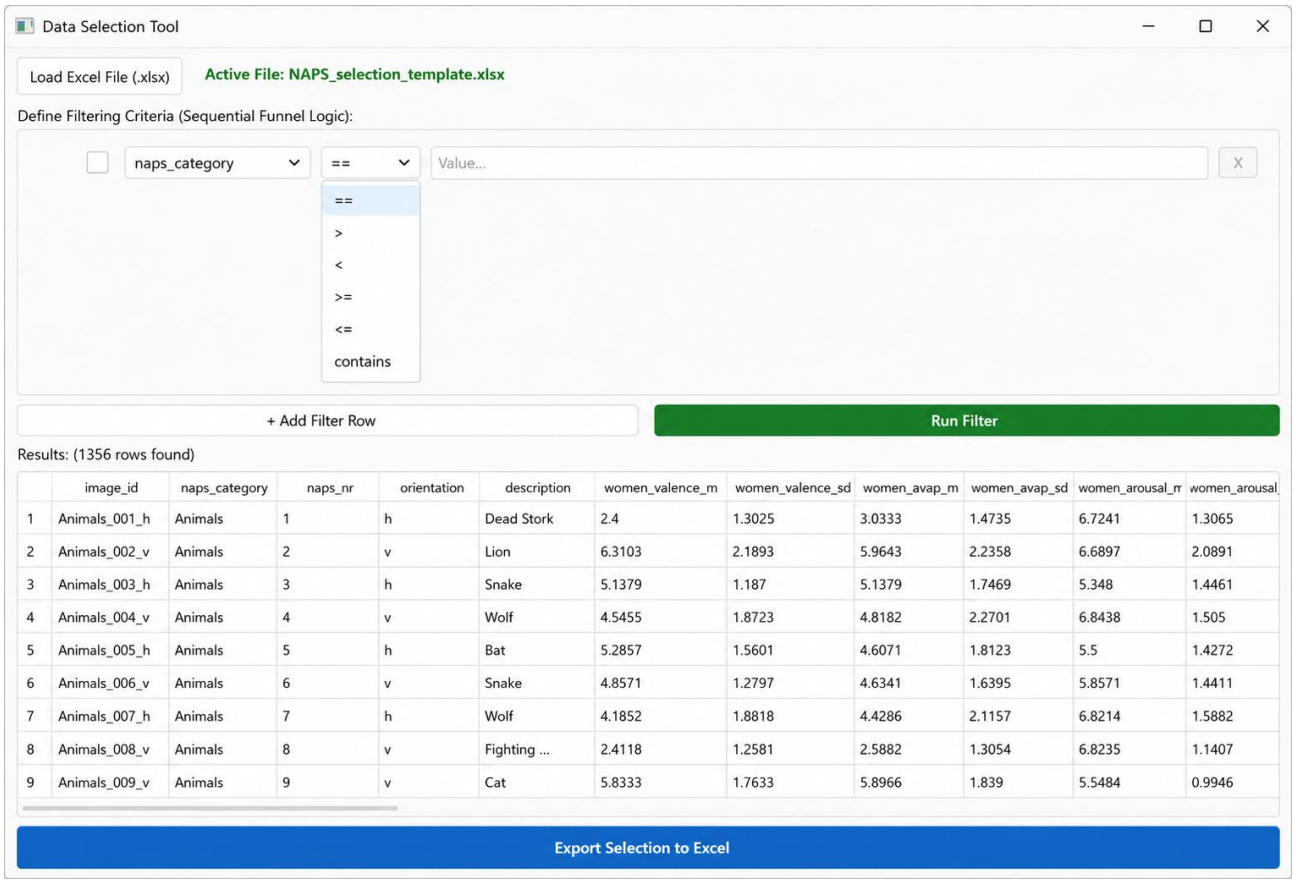
Wolfe JM (1994) Guided Search 2.0: A revised model of visual search. Psychon Bull Rev 1:202–238. https://doi.org/10.3758/BF03200774

Yonelinas AP (2002) The nature of recollection and familiarity: A review of 30 years of research. J Mem Lang 46:441–517. https://doi.org/10.1006/jmla.2002.2864

Zarezadeh Z, Benckendorff P, Gretzel U (2023) Explicating affordances in travel information search: Investigating device use in relation to goals and personal characteristics. Inf Technol Tourism 25:235–257. https://doi.org/10.1007/s40558-023-00250-3

Appendix A: NAPS Filter Tool

Fig. 2: NAPS-Filter Tool interface used for exploratory metadata filtering



	image_id	naps_category	naps_nr	orientation	description	women_valence_m	women_valence_sd	women_avap_m	women_avap_sd	women_arousal_m	women_arousal
1	Animals_001_h	Animals	1	h	Dead Stork	2.4	1.3025	3.0333	1.4735	6.7241	1.3065
2	Animals_002_v	Animals	2	v	Lion	6.3103	2.1893	5.9643	2.2358	6.6897	2.0891
3	Animals_003_h	Animals	3	h	Snake	5.1379	1.187	5.1379	1.7469	5.348	1.4461
4	Animals_004_v	Animals	4	v	Wolf	4.5455	1.8723	4.8182	2.2701	6.8438	1.505
5	Animals_005_h	Animals	5	h	Bat	5.2857	1.5601	4.6071	1.8123	5.5	1.4272
6	Animals_006_v	Animals	6	v	Snake	4.8571	1.2797	4.6341	1.6395	5.8571	1.4411
7	Animals_007_h	Animals	7	h	Wolf	4.1852	1.8818	4.4286	2.1157	6.8214	1.5882
8	Animals_008_v	Animals	8	v	Fighting ...	2.4118	1.2581	2.5882	1.3054	6.8235	1.1407
9	Animals_009_v	Animals	9	v	Cat	5.8333	1.7633	5.8966	1.839	5.5484	0.9946

The NAPS-Filter Tool was developed to apply sequential inclusion and exclusion criteria to NAPS metadata and document the images retained at each stage. In this study, it was used only for exploratory metadata screening before the semantic-classification stage described in Section 5.

Appendix B. Exploratory normative and physical screening of NAPS photographs

Screening criteria. Photographs were retained when they met the study-specific valence range of 5.5–7.0, arousal range of 3.5–5.5, and horizontal-orientation requirement. Approach–avoidance, luminance, contrast, LAB colour components and entropy were then restricted to the 10th–90th percentile range of the 305 affectively eligible horizontal photographs, excluding the most extreme 10% at each tail.

Interpretive note. The categories reported here are the original NAPS categories. Assignment to Gastronomy, Nature and Events required separate semantic classification and visual review. These results therefore describe the exploratory normative and physical screening stage, not the final experimental stimulus pool. Table 7 presents the distribution of the eligible photographs across the original NAPS categories.

Table 7. Distribution of eligible photographs by original NAPS category

NAPS category	Eligible photographs
Animals	13
Faces	23
Landscapes	4
Objects	23
People	11
Total	**74**

Table 8 summarises the sequential screening stages, operational criteria, and number of photographs retained at each stage.

Table 8. Sequential results and operational criteria of the exploratory screening

Screening stage	Operational criterion	Input set (n)	Retained (n)
Affective screening	Valence 5.5–7.0 and arousal 3.5–5.5	1,356	417
Orientation screening	Horizontal photographs only	417	305
Physical screening	Approach–avoidance, luminance, contrast, LAB L, LAB A, LAB B and entropy within the 10th–90th percentiles	305	74

Table 9 reports the descriptive affective and physical characteristics of the 74 photographs retained after exploratory screening.

Table 9. Descriptive characteristics of the 74 photographs retained after exploratory screening, by original NAPS category

NAPS category	n	Valence M (SD)	Arousal M (SD)	Approach–avoidance M (SD)	Luminance M (SD)	Contrast M (SD)	Entropy M (SD)
Animals	13	6.38 (0.42)	4.55 (0.48)	6.20 (0.33)	111.70 (18.04)	65.58 (6.27)	7.72 (0.10)
Faces	23	6.47 (0.41)	4.61 (0.35)	6.07 (0.35)	109.96 (16.50)	65.94 (8.62)	7.68 (0.12)
Landscapes	4	6.25 (0.39)	4.49 (0.21)	6.27 (0.27)	129.30 (5.05)	65.43 (10.61)	7.77 (0.07)

NAPS category	n	Valence M (SD)	Arousal M (SD)	Approach–avoidance M (SD)	Luminance M (SD)	Contrast M (SD)	Entropy M (SD)
Objects	23	6.22 (0.34)	4.64 (0.33)	6.18 (0.28)	112.96 (15.60)	66.50 (7.66)	7.75 (0.10)
People	11	6.36 (0.22)	4.46 (0.41)	6.13 (0.27)	104.25 (14.85)	68.65 (8.96)	7.70 (0.12)
Total	**74**	**6.35 (0.37)**	**4.58 (0.37)**	**6.15 (0.31)**	**111.40 (16.34)**	**66.43 (7.95)**	**7.72 (0.11)**

Table 10 reports the LAB colour characteristics of the 74 retained photographs by original NAPS category.

Table 10. LAB colour characteristics of the 74 photographs retained after exploratory screening, by original NAPS category

NAPS category	n	LAB L M (SD)	LAB A M (SD)	LAB B M (SD)
Animals	13	45.96 (7.21)	1.98 (3.28)	6.14 (4.85)
Faces	23	45.62 (6.63)	1.72 (3.21)	4.93 (3.51)
Landscapes	4	53.31 (1.95)	0.81 (2.28)	7.24 (9.73)
Objects	23	46.68 (6.19)	2.10 (3.16)	4.53 (4.27)
People	11	43.37 (5.96)	2.76 (3.20)	4.59 (4.99)
Total	**74**	**46.09 (6.52)**	**1.99 (3.11)**	**5.09 (4.58)**

Access note. In accordance with the NAPS Terms of Use, original photographs, image identifiers, descriptions and image-level normative metadata are not reproduced or publicly distributed. The complete screening record is retained internally for audit purposes and reporting to the NAPS Authors where required.

Calculations performed. Category frequencies, arithmetic means and sample standard deviations (SD; denominator n−1) were calculated for the 74 retained photographs by original NAPS category and overall. Values are reported as M (SD) and rounded to two decimal places. No inferential tests were conducted because these tables describe the exploratory screening output rather than differences between NAPS categories

Appendix C. CLIP classification rules and implementation details

The rules reported below correspond to the final classification configuration used after the inclusion set was reduced to the three target tourism categories.

1. **Classification-score threshold:** images were assigned to a category only when the highest softmax-normalised similarity score reached or exceeded 75%.
2. **Margin threshold:** the difference between the highest and second-highest similarity scores had to be at least 10 percentage points.
3. **Blocking threshold:** images were automatically excluded when similarity to an exclusion category reached or exceeded 80%.

Images meeting the inclusion rules were classified into their dominant category. Images meeting an exclusion rule were blocked. Cases failing the classification-score or margin thresholds were classified as uncertain and retained for possible expert review.

Following classification and pre-validation screening, 169 images entered the initial independent expert-validation round. Three additional Events images were introduced during consensus, bringing the total considered across validation to 172.

Candidate images retained for the experimental stimulus pool were standardised to horizontal orientation, 500 × 375 px and JPEG format, with approximate file-size control to reduce avoidable presentation variability.

Appendix D. Expert validation and stimulus selection

Appendix D1: Expert Validation Instrument

The final validation stage combined quantitative ratings and expert judgement to assess candidate images for the experimental objectives and tourism categories.
Four experts participated, including tourism researchers and one specialist with neurocognitive expertise.
Each of the 169 images submitted to the initial expert-validation round was independently rated using the instrument below. Table 11 presents the structure of the expert-evaluation instrument used during the validation stage.

Table 11. Structure of the expert-evaluation instrument

Variable	Description
Image ID	Unique identifier from the NAPS database
Proposed category	Gastronomy, Nature or Events
Adequacy rating	Score from 1 to 5
Inclusion recommendation	Include / Exclude
Ranking position	Relative position within category
Comments	Optional qualitative observations

The five-point rating scale applied in the expert evaluation is presented in Table 12.

Table 12. Rating Scale

Score	Interpretation
1	Clear exclusion
2	Low adequacy
3	Moderate adequacy
4	Good adequacy
5	High-confidence inclusion

Images with mean ratings of 2.5 or above were considered eligible for selection. Final inclusion decisions also considered category balance, visual diversity, semantic appropriateness, and the avoidance of redundant imagery.
Disagreements were resolved through expert consensus where necessary.

Appendix D2: Detailed Expert Validation Outcomes

For Events, three images received average ratings below the predefined inclusion threshold and were excluded. Because the remaining ranked candidates were insufficient to complete the category quota, three alternative images were selected through expert consensus.
For Nature, consensus resolved ties at the final selection boundary while preserving visual diversity and reducing redundancy among landscape scenes.

For Gastronomy, the final selection followed the ranking procedure directly and required no consensus replacements.

Validation records

The complete validation file retained:

- Individual expert ratings
- Mean category ratings
- Ranking positions
- Inclusion and exclusion decisions
- Qualitative observations
- Consensus outcomes

These records constitute the audit trail for the final 54-image stimulus pool.

Appendix E: Final Stimulus Allocation

Following validation, the 54 selected images were randomly allocated within each tourism category to the three experimental sets. Each set contained six Gastronomy, six Nature and six Events images, with no overlap between sets. Table 13 summarises the allocation structure of the final OLD, LURE L1, and LURE L2 stimulus sets.

Table 13. Allocation structure of the stimulus sets

Experimental set	Gastronomy	Nature	Events
OLD (Encoding)	6 images	6 images	6 images
LURE L1 (Immediate test)	6 images	6 images	6 images
LURE L2 (Delayed test)	6 images	6 images	6 images

Definition of the sets

OLD: Images presented during the Encoding phase and repeated during both memory tests.
LURE L1: Previously unseen images introduced during the immediate memory test (T1).
LURE L2: Previously unseen images introduced during the delayed memory test (T2).

Allocation principles

The allocation procedure followed four principles:

1. Equal representation of the three tourism categories within each set.
2. No repetition of images across sets.
3. Subsequent comparison of the three sets on affective and low-level visual characteristics.
4. Equal semantic-category representation across the three experimental sets.

The complete allocation record is retained by the authors but not publicly reproduced in accordance with the NAPS Terms of Use.