Psychology of Aesthetics, Creativity, and the Arts
2009, Vol. 3, No. 1, 52–56

1931-3896/09/$12.00 DOI: 10.1037/a0014835

Natural Scenes Are Indeed Preferred, but Image Quality Might Have the Last Word

Pablo P. L. Tinio and Helmut Leder
University of Vienna

Numerous studies have shown that people prefer natural scenes over human-made scenes. Evolutionarily and neurologically based explanations for this preference have been proposed. This study examined the impact that image quality has on the preference for natural scenes. The authors also explored the influence of image quality on familiarity ratings and on how scenes are classified in a rapid, two-alternative, forced-choice task. Finally, the authors propose a framework for conceptualizing the nature of different image manipulation procedures, and how they influence aesthetic experiences.

Keywords: aesthetics, image quality, image manipulation, natural scenes, photography

Research in environmental psychology and visual aesthetics have repeatedly shown that people prefer images of natural scenes over images of human-made scenes (e.g., Biederman & Vessel, 2006; Kaplan, 1992; Kaplan & Kaplan, 1989; Kaplan, Kaplan, & Wendt, 1972; Ulrich, 1981). Three explanations have been proposed to account for this preference. First, natural scenes are the type of environment that humans evolved in and that they are well adapted to (e.g., Kaplan & Kaplan, 1989; Orians & Heerwagen, 1992). Second, being in or looking at nature has restorative effects on physical and mental health (e.g., Hartig, Mang, & Evans, 1991; Ulrich, 1984; van den Berg, Hartig, & Staats, 2007). Third, natural scenes activate association areas in the brain that have a high concentration of mu-opioid receptors, which have been linked to the experience of pleasure (Biederman & Vessel, 2006; Yue, Vessel, & Biederman, 2007).

In this paper, we report on a study that directly and systematically examined the impact that image quality has on the robust preference for natural scenes. As the studies cited above suggest, there has been a considerable amount of research directed toward examining why people prefer natural scenes. As many of those studies employed photographic images, consideration of image quality as a methodological element is crucial. We expected to support previous findings (e.g., Biederman & Vessel, 2006) that people prefer natural scenes over human-made scenes. In addition, we predicted image quality to have an impact on the preference for natural scenes. Specifically, and of main interest to our research, was whether human-made scenes of high image quality would be preferred over natural scenes of low image quality.

Why would image quality influence scene preference? One explanation is based on the concept of *perceptual fluency*, wherein stimuli that are high in perceptual fluency are liked more. Image quality (e.g., based on contrast) has previously been shown to influence perceptual fluency (e.g., Reber, Schwarz, & Winkielman, 2004). Another possible explanation is that image quality corresponds to photographic quality. In this sense, people are aware that a photograph of a scene is a mere graphic depiction of the scene. Thus, photographs of scenes with high image quality could be liked more because they possess high photographic (and artistic) value.

We also explored whether people exhibit greater familiarity for a particular type of scene. Monin (2003) proposed that stimuli that are liked are perceived as more familiar, a concept he referred to as the *warm glow heuristic*. Thus, high-quality images may be higher in perceptual fluency, which could lead to higher judgments of liking (e.g., Reber et al., 2004), which in turn could induce the warm glow heuristic. As a result, high-quality images may be judged as more familiar.

If natural scenes are indeed preferred because they are familiar in an evolutionary sense, or because they promote well-being, or because they elicit pleasure, then it is appropriate to expect that natural scenes would be faster and more accurate to classify than human-made scenes. To examine this prediction, we used a rapid, two-alternative, forced-choice classification task wherein participants had to classify each image into natural or human-made. A short presentation duration of 100 ms was used in order to promote spontaneous responses. It has been shown that people are able to grasp the "gist" of a complex scene quickly (e.g., Oliva & Torralba, 2006) even with short presentation durations (e.g., Potter, 1975; Thorpe, Fize, & Marlot, 1996). In addition, affective evaluation of environmental scenes occurs rapidly (e.g., Hietanen, Klemettilä, Kettunen, & Korpela, 2007). Because of the short presentation duration used, careful attention was placed on creating sets of stimuli that clearly belonged to the two scene categories.

Image Manipulation Procedure and Category Validation Pre-Study

The original set of stimuli consisted of 50 high-quality color photographs of natural scenes and 50 high-quality color photo-

Pablo P. L. Tinio and Helmut Leder, Department of Psychology, University of Vienna, Vienna, Austria.

We thank Sofia Tidman for her valuable comments regarding this project, and for proofreading the manuscript. We would also like to thank Martin Wieser for his help in preparing the stimuli.

Correspondence concerning this article should be addressed to Pablo P. L. Tinio, Department of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria. E-mail: pablo.tinio@univie.ac.at

graphs of human-made scenes. The natural scene photographs depicted scenes such as mountain ranges, prairie fields, desert landscapes, canyons, and forests. The human-made scene photographs depicted scenes such as city skylines, bridges, road systems, and housing complexes. The latter set of scenes did not include any indoor scenes; it has been shown that different cortical areas of the brain are activated in response to outdoor and indoor scenes (Henderson & Larson, 2007). The presence and visual dominance of other aspects of the environment (e.g., water) that may be present in both types of scenes were roughly equated.

Using Adobe Photoshop CS2 (version 9.0, www.adobe.com), a degraded version of each photograph was produced by manipulations on the following image properties: sharpness (decrease); noise and grain (increase); contrast (decrease); color fidelity (toward less accurate); and color saturation (decrease). Figure 1 shows an example of the results of the image manipulation procedure. All of the manipulations were performed globally—linearly, across the entire image area—for each photograph, with the restriction that the changes be moderate enough so that the content of the photograph was retained, thus avoiding unnatural or artificial looking images. The final stimulus set consisted of 50 natural-normal scenes, 50 natural-degraded scenes, 50 human-made-normal scenes, and 50 human-made-degraded scenes.

An initial category validation prestudy involving four participants (2 females, 2 males; mean age of 24.00) was conducted to validate the authors' initial classification of the scenes into the natural and human-made categories. The scenes (approx. 8.5in. × 11in. in size) were presented in random order with a response-dependent presentation duration. This self-paced procedure was employed in order to simulate a manual sorting study.

The results of the prestudy indicated that the initial classification of the scenes was accurate. The performances of the participants in percent accuracy were as follows: 100%, 96%, 100%, and 99%. It was important that there was no particular scene for which the participants consistently made errors regarding its classification into its respective category.

Method

Participants

Sixteen psychology students (14 females, 2 males; mean age of 21.05) from the University of Vienna participated in the main experiment. None of these participants took part in the category validation prestudy.

Materials

The stimuli consisted of the same four sets used in the prestudy.

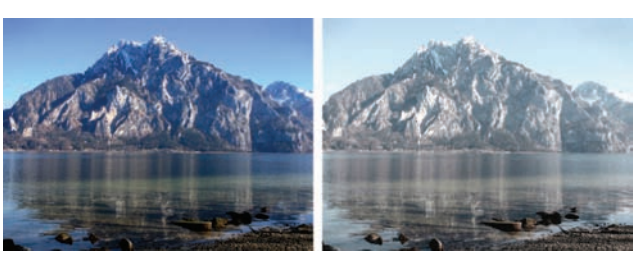

Figure 1. Illustration of the image manipulation procedure, with original (left) and degraded (right) versions.

Procedure

The experiment consisted of three blocks. The first block involved liking ratings of the stimuli using a Likert-type scale with 1 indicating "dislike" and 7 indicating "like." The second block consisted of familiarity ratings also using a Likert-type scale with 1 indicating "unfamiliar" and 7 indicating "familiar." The third block consisted of a forced-choice classification of the stimuli as either natural or human-made. However, for this block, a short presentation duration was used in order to elicit spontaneous responses. All participants completed the liking block first, as this was the dependent measure that was most closely tied to the main research question. The order of the familiarity and classification blocks was counterbalanced across participants.

Trials for the three blocks consisted of the following sequence of stimulus events: a fixation cross for 200 ms; the stimulus for 1500 ms (100 ms for the classification block); a cue for 2000 ms; and an intertrial interval for 2000 ms. Participants provided their responses while the cue was on the screen, and were instructed to provide their responses as quickly as possible based on their initial "gut" feeling. For each block, participants completed 12 practice trials (using three of each type of stimuli not included in the main trials), which were identical in structure to the main trials. Each participant viewed only one version of each scene, either normal or degraded, counterbalanced across participants. The presentation order of the stimuli was fully randomized.

Results

Liking

See Figure 2 for graphic depictions of the results from the three blocks. The mean liking ratings for the scenes sampled across participants for each scene type were: natural-normal, 5.47 (SD = 0.67); natural-degraded, 3.64 (SD = 0.88); human-made normal, 4.82 (SD = 0.86); and human-made degraded, 3.09 (SD = 0.66). An analysis of variance with scene type (natural and human-made) and image quality (normal and degraded) as within-subjects factors yielded significant main effects of: scene type, $F(1, 15) = 10.66$, $p < .01$, $\eta_p^2 = 0.42$; and image quality, $F(1, 15) = 89.80$, $p < .001$, $\eta_p^2 = 0.86$. The interaction between scene type and image quality was not significant ($p = .46$). These results confirm previous findings that, in general, natural scenes are liked more than human-made scenes. Of primary interest to our study was whether image quality could influence the relationship between scene type and preference judgments. The planned comparison of human-made normal versus natural-degraded was significant, with human-made normal scenes being liked more than natural-degraded scenes, $t(15) = 3.74$, $p < .01$.

In order to test whether perceptual fluency (e.g., Reber et al., 2004) was involved in the liking judgments, response latencies were analyzed. An analysis of variance with scene type and image quality as the within-subjects factors and response latency for the liking ratings as the dependent variable yielded a significant main effect of scene type, $F(1, 15) = 7.52$, $p < .05$, $\eta_p^2 = 0.33$. In general, latencies for human-made scenes were higher than natural scenes. There were no other significant effects. The paired comparison between response latencies for natural-degraded (2357.20 ms) and human-made normal (2407.95 ms) scenes was not significant ($p = .07$).

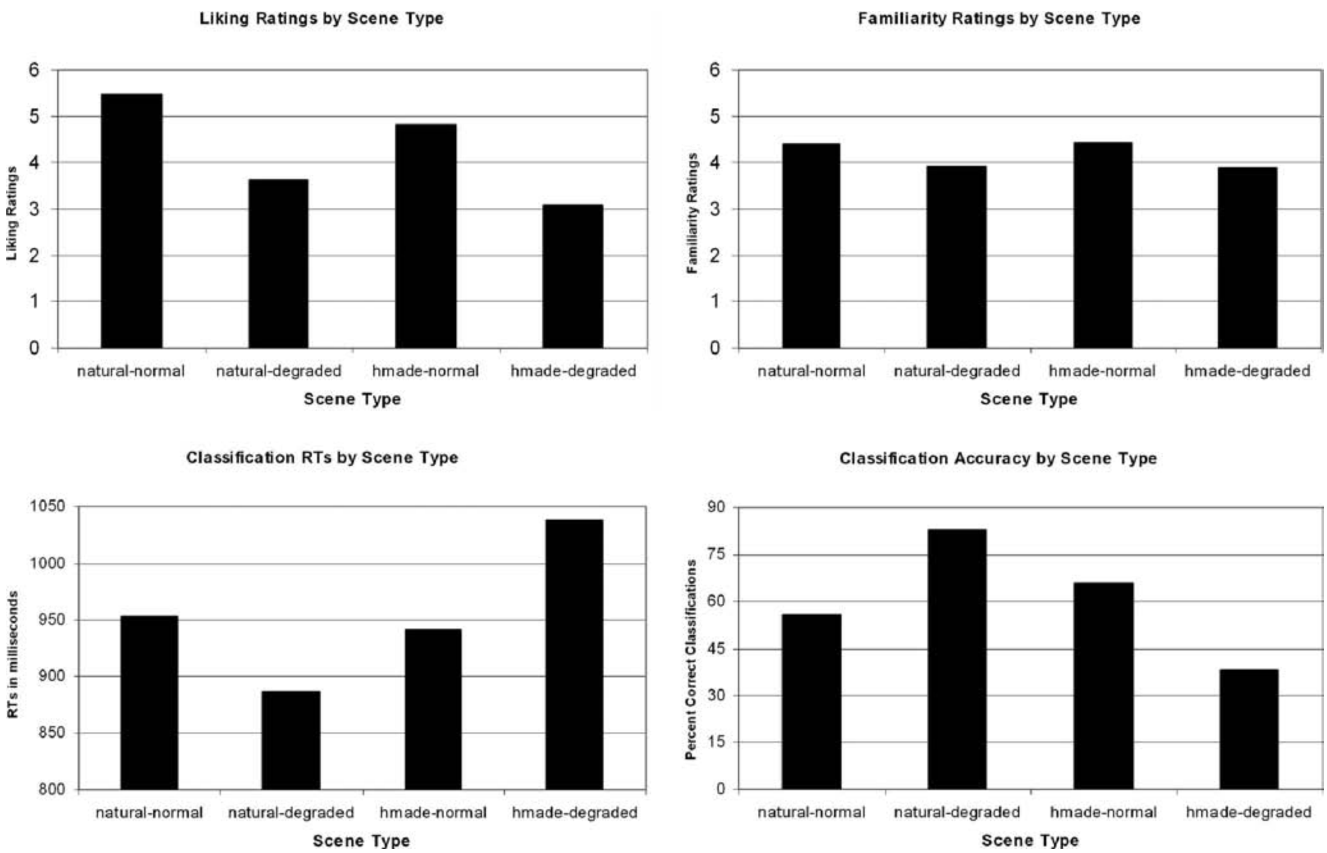

Figure 2. Mean liking ratings, familiarity ratings, classification response times, and classification accuracy by scene type.

Familiarity

As with the liking ratings, the mean familiarity ratings for the scenes were sampled across participants for each scene type and were: natural-normal, 4.42 (SD = 0.94); natural-degraded, 3.94 (SD = 0.76); human-made normal, 4.44 (SD = 1.00); and human-made degraded, 3.90 (SD = 0.79). An analysis of variance with scene type (natural and human-made) and image quality (normal and degraded) as within-subjects factors revealed a significant main effect of image quality, $F(1, 15) = 5.91$, $p < .05$, $\eta_p^2 = 0.28$. There were no other significant effects. Overall, the high-quality scenes were judged as more familiar than the degraded scenes.

Classification Into Natural or Human-Made

The principle measures of interest in this block were reaction time and response accuracy *for correct classifications*. The mean reaction times sampled across participants for each scene type were as follows: natural-normal, 953.63 ms (SD = 154.69); natural-degraded, 886.55 ms (SD = 157.08); human-made normal, 941.14 (SD = 226.69); and human-made degraded, 1038.04 (SD = 232.37). The main effects of scene type and image quality were not significant, $p = .07$ and $p = .58$, respectively. An analysis of variance with scene type (natural and human-made) and image quality (normal and degraded) as within-subjects factors revealed a significant interaction, $F(1, 15) = 10.73$, $p < .01$, $\eta_p^2 = 0.42$. For the natural scenes, classification times were significantly higher for normal than for degraded versions, $t(15) = 3.27$, $p < .01$. For the human-made scenes, the difference in classification times between degraded and normal images only approached significance ($p = .06$). For nondegraded scenes, there was no difference in reaction times between natural and human-made scenes ($p = .78$). However, for degraded scenes, classification times for human-made scenes were significantly higher than for natural scenes, $t(15) = 3.45$, $p < .01$.

A notable finding in this task was that the classification accuracy was highest for the natural-degraded scenes (83%), followed in decreasing order by human-made normal (66%), natural-normal (56%), and human-made degraded (38%) scenes. These findings are especially noteworthy in light of the results of the prestudy, which showed that participants were very good at classifying images when the task was self-paced in both viewing of the scenes and in providing responses. Such high error rates make the interpretation of the reaction times complex.

Discussion

Previous studies would undoubtedly have attempted to control for the quality of the stimuli; however, the extent of such control is not clear. While these results confirmed previous findings (e.g., Biederman & Vessel, 2006) that in general, natural scenes are liked more than human-made scenes, the results also showed that the image quality of the scenes could influence preference judgments. Specifically, human-made scenes of high image quality were liked more than natural scenes of low image quality. Furthermore, perceptual fluency (e.g., Reber et al., 2004) did not seem to influence the results. Thus, it seems that the photographic (or artistic) value of high-quality images could explain why high-quality images are preferred. Several studies designed to further test this idea are planned.

In addition to the liking ratings—the traditional dependent measure for scene preference studies—we examined familiarity ratings. Overall, participants found the original high-quality images more familiar than the degraded images. Participants also generally liked high-quality images more than degraded images. Thus, a possible interpretation of these effects is that people judged what they liked as more familiar, which is consistent with Monin's (2003) proposal of the warm glow heuristic.

The classification task was aimed at examining the prediction that people should be faster and more accurate at classifying natural scenes because doing so would be adaptive, as natural environments possess elements that are necessary for survival. The results indicated an interaction between scene type and image quality with classification times for human-made scenes significantly higher than natural scenes, but only for degraded images. At first glance, it may seem that people's ability to classify scenes is more tolerant to degradation for natural than for human-made scenes. However, this interpretation could be disputed by the fact that for the natural scene category, classification times were significantly higher for the normal than for the degraded images. Furthermore, accuracy was negatively related to response times. Specifically, classification was fastest for natural-degraded scenes, which were the scenes most correctly classified; classification was slowest for human-made degraded scenes, which were the scenes most incorrectly classified. The complex data that resulted from the classification task is difficult to interpret, although upon closer examination, the results indicated that participants may have been biased toward judging scenes as belonging to the natural scene category when the presentation duration was short. The question that beckons is why were the natural-degraded scenes the fastest and most accurate to classify? Further studies are required to examine this specific question and to explore the relationship between reaction times and classification accuracy when judging scenes.

While it is true that we tend to like trees and forests more than concrete roads, the quality of the image is important. Image quality also has an influence on what scenes we find familiar, and how those are classified. Our findings reiterate the importance of having a controlled set of stimuli, especially for studies fundamentally based on evaluations of visual representations. If image quality is not systematically controlled, it could become an artifact of the experimental context.

Taxonomy of Image Manipulation Procedures

In light of the results of this study, we would like to present a taxonomy of image manipulation procedures that is especially relevant for aesthetics research. In proposing this taxonomy, we do not intend to oversimplify the many variations and combinations of image manipulation procedures. Instead, we hope that it could serve as a starting point for conceptualizing how different techniques vary and are related to each other, and how they subsequently influence aesthetic experiences.

For this study, the manipulations performed on the original set of images were basic (albeit powerful) *surface-level manipulations*. These consisted of transformations in terms of sharpness, noise, grain, contrast, color fidelity, and color saturation. Manipulations were performed globally for each image. In general, surface-level manipulations could be performed easily, efficiently, and in a direct manner, thus possessing *high manipulability*. It is interesting that such manipulations are often subtle enough that people cannot readily identify that certain image properties were changed. In terms of experimental studies in psychological aesthetics, surface-level changes lend themselves well to studies such as those reported in this paper—those that involve short presentation durations, fast judgment, and dependent variables such as response latencies.

Surface-level manipulations are in stark contrast to *composition-level manipulations*, wherein changes are made on structural aspects of images. An example of such a manipulation is a change in the cropping of an image. In cropping, the external borders of the image are manipulated, which results in changes in the image's complexity, symmetry, and balance. Compared to surface-level manipulations, composition-level manipulations are less direct, and typically involve both local and global changes. Thus, they possess *moderate manipulability*.

Finally, *semantic-level manipulations* consist of changes to local image areas. Typically, such manipulations require direct manipulations of image content (the actual referent) such as adding or removing elements in an image, or altering the positions of different elements. Semantic-level manipulations are not as easy to perform as surface-or composition-level manipulations; the former type of manipulations generally involve local changes. Thus, semantic-level manipulations have *low manipulability*. Changes to the image using semantic-level manipulations could be more easily identified by perceivers than surface level or composition level manipulations.

In terms of aesthetic appreciation, this taxonomy is congruent with Leder, Belke, Oeberst, and Augustin's (2004) model of aesthetic appreciation. The influence of surface- and composition-level manipulations occurs during the initial perceptual analysis of an image and during subsequent implicit integration into memory, which both occur automatically. In contrast, semantic-level manipulations are directed toward the image content, and are therefore closely related to the generally deliberate explicit classification stage as outlined in the model of aesthetic appreciation. Simultaneous consideration of both the taxonomy presented here and Leder et al.'s model is an approach that could help researchers to conceptualize how stimulus manipulations performed on photographic images impact how the images are perceived aesthetically.

References

Biederman, I., & Vessel, E. A. (2006). Perceptual pleasure and the brain. *American Scientist, 94,* 249–255.

Hartig, T., Mang, M., & Evans, G. (1991). Restorative effects of natural environment experiences. *Environment and Behavior, 23,* 3–26.

Henderson, J. M., & Larson, C. L. (2007). Cortical activation to indoor versus outdoor scenes: An fMRI study. *Experimental Brain Research, 179,* 75–84.

Hietanen, J. K., Klemettilä, T., Kettunen, J. E., & Korpela, K. M. (2007). What is a nice smile like that doing in a place like this? Automatic affective responses to environments influence the recognition of facial expressions. *Psychological Research, 71,* 539–552.

Kaplan, S. (1992). Environmental preference in a knowledge-seeking, knowledge-using organism. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 581–598). Oxford, United Kingdom: Oxford University Press.

Kaplan, S., & Kaplan, R. (1989). *The experience of nature: A psychological perspective.* New York: Cambridge University Press.

Kaplan, S., Kaplan, R., & Wendt, J. S. (1972). Rated preference and complexity for natural and urban visual material. *Perception & Psychophysics, 12,* 354–356.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology, 95,* 489–508.

Monin, B. (2003). The warm glow heuristic: When liking leads to familiarity. *Journal of Personality and Social Psychology, 85,* 1035–1048.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23–36.

Orians, G. H., & Heerwagen, J. H. (1992). Evolved responses to landscapes. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 555–579). Oxford, United Kingdom: Oxford University Press.

Potter, M. C. (1975). Meaning in visual search. *Science, 187,* 965–966.

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,* 364–382.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520–522.

Ulrich, R. S. (1981). Natural versus urban scenes. *Environment and Behavior, 13,* 523–556.

Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science, 224,* 420–421.

Van den Berg, A. E., Hartig, T., & Staats, H. (2007). Preference for nature in urbanized societies: Stress, restoration, and the pursuit of sustainability. *Journal of Social Issues, 63,* 79–96.

Yue, X., Vessel, E. A., & Biederman, I. (2007). The neural basis of scene preferences. *NeuroReport, 18,* 525–529.

Received October 31, 2008
Revision received November 20, 2008
Accepted November 21, 2008 ■