Article:

Van der Jagt, APN, Craig, T, Anable, J et al. (2 more authors) (2014) Unearthing the picturesque: The validity of the preference matrix as a measure of landscape aesthetics. Landscape and Urban Planning, 124. pp. 1-13. ISSN 0169-2046

https://doi.org/10.1016/j.landurbplan.2013.12.006

Published in: Landscape and Urban Planning, 124:pp1-13

Final draft post refereeing version

Unearthing the picturesque: The validity of the preference matrix as a measure of landscape aesthetics

Alexander P. N. Van der Jagt,[1,2] Tony Craig,[2] Jillian Anable,[1] Mark J. Brewer,[3] and David G. Pearson[1]

[1] University of Aberdeen, Aberdeen, United Kingdom

[2] The James Hutton Institute, Aberdeen, United Kingdom

[3] Biomathematics and Statistics Scotland, Aberdeen, United Kingdom

Authors Note

Alexander P. N. van der Jagt, School of Psychology, University of Aberdeen, Aberdeen, United Kingdom; Social, Economic and Geographical Sciences, The James Hutton Institute, Aberdeen, United Kingdom.

Tony Craig, Social, Economic and Geographical Sciences, The James Hutton Institute, Aberdeen, United Kingdom.

Jillian Anable, School of Geosciences, University of Aberdeen, Aberdeen, United Kingdom.

Mark J. Brewer, Biomathematics and Statistics Scotland, Aberdeen, United Kingdom.

David G. Pearson, School of Psychology, University of Aberdeen, Aberdeen, United Kingdom.

Address correspondence to David Pearson, School of Psychology, William Guild Building, King's College, University of Aberdeen, Aberdeen, AB24 2UB, United Kingdom. E-mail: d.g.pearson@abdn.ac.uk, Telephone: +44 (0)1224 273940, Fax: +44 (0)1224 273426.

1 Introduction

Despite a long and rich history of enquiry into landscape aesthetics, and its purported role in influencing both levels of stress and attentional functioning (R. Kaplan & Kaplan, 1989; Ulrich, 1983), a consensus on its explanatory attributes is lacking (Lothian, 1999). For instance, following a meta-analysis encompassing studies attesting the most influential model on landscape aesthetics – the preference matrix – it was concluded that: “the postulated theory has not generated reproducible results” (Stamps, 2004, p. 14).

Although this could imply that the preference matrix is simply invalid as a theory of landscape aesthetics, it could alternatively be that: (1) the measures of the informational qualities have been unreliable, (2) confounding variables have influenced how the informational variables load on scenic quality, (3) specific scene content of images have influenced the type or direction of the relationship between the informational variables and scenic quality or (4) the relationships could have been better mapped by nonlinear polynomials.

To address these alternative explanations, the methodological approach of the present study diverged from that of previous research with regard to: (1) item definitions, (2) control for confounding variables, (3) variety of stimulus material, and (4) presupposed type of relationship between predictor and target variables. We present evidence showing support for each of the variables in the preference matrix following a series of methodological improvements addressing these limitations.

1.1 The Preference Matrix

The preference matrix by Kaplan and Kaplan (1989) is an evolutionary theory which is based on the assumption that the ability for aesthetic appraisal has evolved to encourage adaptive habitat selection. It coincides with other evolutionary theories (Appleton, 1975; Orians &

Heerwagen, 1992; Ulrich, 1983), which all have been popular to account for the strong cross-cultural similarities in preferences for particular configurations of landscapes and the elements therein (Parsons & Daniel, 2002). Kaplan and Kaplan (1989) reason that a good Understanding (i.e., having a valid mental map) of the physical environment is crucial to human survival (also, see S. Kaplan, 1987). For that reason, they postulate that humans are attracted to landscapes that provided a sense of order. Furthermore, they argue that ongoing exploration of new habitat conveyed adaptive benefits as well. Hence, environments that incite further Exploration – due to high levels of complexity and/or mystery – will also be experienced as attractive. The four variables of the preference matrix – Coherence, Complexity, Legibility, and Mystery – are defined by crossing the two needs of Understanding and Exploration with a time perspective (immediate or inferred/predicted; see Figure 1 & Table 1).

The preference matrix is an example of a perception-based approach to explaining landscape aesthetics. This implies that the authors of this theory consider the aesthetic response to originate from the interplay between objective, quantifiable landscape features and the subjective appraisal of these attributes (Daniel & Vining, 1983; Daniel, 2001). The Understanding and Exploration vector of the preference matrix can be regarded as experiential conceptualizations of objective attributes such as: "uniformity and variety" as well as "order and complexity", which have been contemplated as predictors of landscape aesthetics by philosophers for centuries (Lothian, 1999). It has been argued that such informational are experienced as attractive because these enticed our ancestors to continuously build upon and extend their mental map of the environment, yet prevented them from wandering off to potentially unsafe settings for which such an overview could not be readily achieved (Kaplan, 1987; Kaplan & Kaplan, 1989).

The time perspective vector was, however, a relatively new addition within the preference matrix by Kaplan and Kaplan (1989, but see Woodcock, 1984). It was introduced to account for the high preference of natural scenes which included an element of Mystery such as a path disappearing around a bend or a partly precluded clearing within a forest (Kaplan, 1987). The authors noted that this informational quality does not have a one-to-one relationship with the visual features of an environment; it requires a process of cognitive inference or prediction to be coded. This is unlike the informational qualities which are immediately available (e.g., Complexity). At first sight, such inferential processing seems to run counter to the evolutionary backbone of the model, based on which we would expect affective responses to spatial qualities to be intuitive and automatic. However, the authors make explicit that the cognitive operations required for making predictions about functioning do not require any conscious processing and therefore are made very rapidly. In agreement with this contention, recent research in visual cognition has shown that the scene exposure time that is required to detect the navigability of a scene – a concept related to the inferred Legibility construct of the preference matrix – at a 75% accuracy threshold is very low (i.e., 35-45 ms) and alike to that required for detecting qualities of the "immediate" environment such as openness and concealment (Greene & Oliva, 2009a).

Given the rootedness of the preference matrix in a long-lasting research tradition on landscape aesthetics and recent empirical support for the ability to derive both immediate and inferred informational qualities rapidly and automatically, we wanted to address the current status quo whereby conclusive support exists for neither one of the informational qualities of the preference matrix as predictors of scenic quality (Stamps, 2004). To this end, methodological limitations of previous research which could have contributed to inconsistencies between findings regarding this theoretical model need to be addressed.

1.2 Methodological Considerations

1.2.1 Item definitions

Stamps (2004), when discussing the findings of his meta-analysis, touches upon the high variability between previous studies with regard to the size and direction of reported correlations between each of the variables in the preference matrix and scenic quality. He then goes on to suggest that a replacement of questionnaire items tapping on the variables from the model by objective measurements (e.g., estimates of visible area from GIS maps as indicator of Mystery). Although we concur with the contention that measurement error might have been introduced in previous research, we are less convinced by the suggestion that this is addressed most effectively by downgrading the preference matrix to a mere objective landscape aesthetics paradigm. Instead, we reconsidered the standard definition of the variables in the preference matrix. To this end, we conducted two pilot studies to measure participant understanding of item definitions in relation to a set of 20 images highly variable in terms of scene content (see Table 1).

Participants in the pilot studies rated all images on the variables of the preference matrix, which were operationally defined in line with previously used definitions in the literature (e.g., Stamps, 2004). Subsequently, participants indicated their level of comprehension of each of the items on a scale from 1 (very low) to 7 (very high). Additionally, participants were invited to comment on those definitions for which comprehension was low. An analysis of these comments showed that the ease of rating items varied between different images. For instance, the item definition of Legibility ("It would be easy to find my way around the environment depicted") is derived based on the assumption that the environment affords locomotion. Rating the legibility of a scene, however, proved to be challenging with regard to images depicting inaccessible ground surface like rugged mountaintops or seascapes. The standard definition of Mystery ("The setting promises more to be seen if you could walk deeper into it") obviously brings about similar limitations. We found this surprising as Kaplan claims that: "The variables in the matrix apply to a large variety of environments and situations" (S. Kaplan, 1987, p. 11). We therefore employed alternative definitions with could also be interpreted to imply visual exploration. The definition of Mystery; "This would be an interesting scene to explore further", was adopted from Van den Berg, Vlek and Coeterier (1998). The definition of Legibility, was rephrased as: "It would be easy to orient myself around the depicted scene".

Whereas comprehension of the definition of Complexity ("The scene contains diverse elements and features") was rated very favourably, a subset of participants reported low comprehension of the definition of Coherence ("The visual elements of the scene fit together well"). That is, the "fit together" construct between different environmental features can be judged differently depending on how one conceptualizes this item. If one regards a conception of coherence along the lines of typicality one might judge an abstract sculpture placed in a forest as incoherent. However, from an artistic, compositional point of view, the same scene can be judged to 'hang together' perfectly well. Along similar lines, a racing track in a forest might be judged as coherent by an automobile fan and as incoherent by an ecologist. We believe that statistical control for scene familiarity will to an extent account for these influences of user background on Coherence ratings. Hence, we employed the standard definition of Coherence ("The scene 'hangs together'; it is easy to organize and structure") in conjunction with statistical control for familiarity (also, see section on: control for confounding variables).

A final concern regarding item definitions involved the assessment of the target variable; scenic quality or aesthetic response. In previous research, we both find instances of scenic quality being measured by environmental ratings on preference ("I like this scene") and beauty

("This is a beautiful scene"). This follows from the assumption that each of these constructs can be deemed valid as a measurement of scenic quality (Daniel, Boster, & Forest, 1976; Purcell, 1987). Perhaps not surprisingly, the preference matrix has been predominantly tested with preference as target variable. It should, however, be noted that the origins of the dimensions underlying the preference matrix can be traced back to centuries-old philosophical studies specifically aimed at understanding landscape beauty (Lothian, 1999). In addition, it has been argued that using beauty statements helps one to deal more effectively with inadvertent influences by user perspective – shaped by pertinent goals and intentions – on the measure of scenic quality (De Vries, de Groot, & Boers, 2012; Han, 2010). For these reasons, it was decided to test the preference matrix in the context of scenic beauty. This approach is consistent with previous attempts by Real, Arce and Sabucedo (2000) and Van den Berg et al. (1998).

1.2.2 Control for confounding variables

Both the scene content of an image and the user experience of the observer are known to influence the appraisal of scenic quality (e.g., Bishop & Hulse, 1994; De Groot & Van den Born, 2003; Howley, 2011; R. Kaplan & Kaplan, 1989; Purcell, Peron, & Berto, 2001). In addition, ratings of the informational qualities are also likely to be confounded to an extent by user experience. Nasar (1994), for example, reasoned that those environments which are discrepant with the expectations that have become associated with that structure or category through experience are likely to be experienced as lower in coherence and higher in complexity. Similarly, Coeterier (1996) described that the complexity and mystery of a landscape are perceived differently by locals and outsiders due to a discrepancy in the level of knowledge regarding the setting.

To account for the effects of both scene content and user experience on ratings of the items in the preference matrix and beauty, we opted for a multiple regression methodology that enabled the confounding influence of these constructs to be factored out statistically, as described later. Inclusion of these variables will likely bring down the size of the error term in the model (as the model is then using these confounding variables to account for extra uncertainty), resulting in an improved estimate of the effect size for each of the preference matrix variables as represented by the regression weights (by accounting for potential confounders).

To assess scene content, we incorporated questionnaire items measuring natural and built character of scene content. Natural and built character were measured individually because these constructs are not mutually exclusive (e.g., a ruin could be rated both high on natural and built character; Coeterier, 1996). User experience was assessed by an item measuring the familiarity of each scene ("I am familiar with this type of scene"), alongside an item measuring rural experience in general ("What number of years have you spent living in a village, hamlet, and/or farm?").

In the present study the same observers who provided ratings of their perception of informational variables in the preference matrix also rated scenic quality. Previously, researchers addressing the preference matrix have condemned the use of multiple ratings as such practice might confound the correlations between the respective constructs (e.g., Herzog, Kaplan, & Kaplan, 1982; Nasar, 1994). It is important to stress that within our multiple regression framework we are able to fully control the effect of each predictor on the target variable for those of all other variables in the equation (cf. type-III sums of squares in ANOVA; Draper & Smith, 1981), and even if such confounding is present, the validity of significant effects in other variables still holds. The outcome of our analysis thus more nearly reflects the unique contribution by each of the variables in the preference matrix to the aesthetic response.

1.2.3 Variety of stimulus material

Previous studies on the preference matrix have typically employed small to moderately sized image databases varying from 14 to 191 images, typically capturing a confined range of scene content (Stamps, 2004). Whilst this practice has clear practical advantages, the use of small image sets imposes the risk of failing to sample the full spectrum of each of the informal qualities. The importance of employing a high variety of stimulus material is illustrated by the finding that the strength of the relationship between the variables from the preference matrix and the aesthetic evaluation of an environment shifts alongside the type of scene content under consideration (Herzog & Bosley, 1992; Herzog & Leverich, 2003). For example, it has been shown that legibility is a stronger predictor of preference when using a set of forest scenes, as compared to a combined set of forest and field scenes (Herzog & Leverich, 2003). This can be accounted for as in dense forest environments high legibility hints at the availability of paths or vantage points, which aid in scene understanding and exploration. On the contrary, high legibility within the context of open and exposed field settings suggests a lack of navigational elements such as distinctive landmarks; interfering with scene understanding (Herzog & Kutzli, 2002; R. Kaplan & Kaplan, 1989). For that reason, failing to sample a wide variety of images is likely to lead to inconsistencies between studies in how the informational qualities relate to scenic quality.

To effectively deal with the limitations of using a low variety of stimulus material we employed a substantially sized database comprised of 1600 images portraying a wide variety of natural, built, and "mixed" scene content. All images depicted Scottish scene content in order to minimize variability in terms of participant familiarity. To deal with the logistics of having such sizeable database we showed each participant a subset of 80 randomly selected images, which they rated on all items in the questionnaire. Each participant thus rated a unique set of images (and conversely, each image was not viewed by every participant). This, however, poses the risk that individual differences in rating behaviour will impact on the outcome of the regression equation (i.e., the observations are not independent). To deal with this potential limitation we used a regression model with random effects (i.e., a mixed effects model) to account for between-image and between-participant variability, thereby allowing for observations that are not independent. This random effects analysis also enabled us to make generalizations from findings obtained with a random sample of participants and images to the population and environment as a whole (Baayen, Davidson, & Bates, 2008).

1.2.4 Presupposed type of relationship between predictor and target variables

Studies on the preference matrix have in the vast majority tested solely for linear relationships between the predictor variables and scenic quality. With regard to the type of relationship between the variables of the preference matrix and scenic quality, Kaplan and Kaplan (1989, p. 58), however, note the following: “A lack of Coherence makes it difficult to understand what is before one; a lack of Complexity diminishes one’s likelihood of becoming engaged in viewing. It is not necessarily the case, however, that preference is enhanced by having increasing amounts of these informational factors. For the two factors that rely on greater inference, however, there is an implied suggestion along the lines of “the more the merrier”. With more Legibility, confidence is enhanced that the setting will continue to be understandable. More Mystery entices one to further exploration.” The contention that intermediate levels of Coherence and Complexity have the most positive impact on the aesthetic response follows up on seminal research showing an inverted-U shaped relationship between complexity and preference for both nonsense stimuli and real-world environments (Berlyne, 1971; Wohlwill, 1974). For that reason, we contend with the assertion of Nasar (1994) that a lack of testing for nonlinear relationships may account for inconsistent support for complexity as predictor of landscape aesthetics.

It could be hypothesized that nonlinear relationships between the informational variables and scenic quality are not just confined to the inverted-U shaped connection for complexity. For instance, several studies have shown a surprising negative relationship between Mystery and preference regarding (forest) scenes with low levels of visual access (Herzog & Kutzli, 2002; Herzog & Kropscott, 2004; Herzog & Kirk, 2005). In addition, Legibility has been found to be a more effective predictor of scenic quality for image sets incorporating densely vegetated scenes in comparison to open field settings (Herzog & Leverich, 2003; Herzog & Kropscott, 2004). For that reason, we performed formal tests to assess the type of relationship between the variables of the preference matrix and scenic quality (for a detailed description of the statistical procedure, see Appendix A)

In addition to taking nonlinear relationships between the items of the preference matrix and scenic quality into consideration, the authors of the preference matrix also propose an interactive association between Coherence and Complexity (R. Kaplan & Kaplan, 1989). As mentioned above, this assumption dates back to centuries-old philosophical reasoning and is shared amongst a variety of theories of landscape aesthetics (Appleton, 1975; Lothian, 1999; Ulrich, 1983; Wohlwill, 1983). The centrality of the interplay between these concepts of unity and diversity to aesthetic responses has recently been corroborated by research showing that scenes with fractal geometry – a combination of complex, yet coherent (i.e. self-repetitive) structure – are consistently the most preferred (see Taylor, Spehar, Van Donkelaar, & Hagerhall, 2011, for an overview). To our knowledge, however, it has not been formally tested for an interaction between Coherence and Complexity. For that reason, we chose to incorporate an interactive relationship between Coherence and Complexity in our statistical model.

1.3 The present study

Here we set out to test the validity of the variables in the preference matrix as predictors of aesthetic evaluation of a scene. Informational qualities are gauged with a set of item definitions that has been piloted to ensure high participant comprehension. This practice increases the reliability of our measurement. In addition, we take into account the unsystematic variance associated with user experience and scene content, which improves the estimate of effect size associated with the informational variables. As the predictors of the preference matrix may show a different relationship to scenic quality depending on the type of content that is presented, we employed a substantial image set with a high variety of natural, built and "mixed" scene content. Finally, we also test if the validity of the preference matrix can be improved by allowing for the possibility of a nonlinear relationship with scenic quality.

It is hypothesized that, following implementation of the set of the aforementioned methodological amendments, the variables Coherence, Complexity, Legibility, and Mystery will emerge as independent predictors of scenic quality. We further predict an interactive relationship between Coherence and Complexity to appear. In addition, we hypothesize that rated Natural Character and Familiarity contribute positively to the aesthetic evaluation of a scene, whereas Built Character will abate the attractiveness of an environment.

2 Methods

2.1 Participants

A total of 100 participants (71 female) participated in the study. The ages of the participants varied from 18 to 51 years with a median age of 19 years old. All participants were undergraduate students from the University of Aberdeen with normal or corrected-to-normal vision who were rewarded by course credit. The majority of participants had not lived in a village, hamlet, and/or farm at any stage in their life (N = 61). A total of 31 participants spent five or more years living in a rural setting.

2.2 Stimuli

A total of 1600 high quality images were selected from two online image banks; SCRAN (Scottish Cultural Resources Access Network, http://www.scran.ac.uk) and Flickr (http://www.flickr.com; using the Creative Commons search option). Images were selected to represent a wide variety of natural, built and “mixed” content environments. Natural scenes ranged from rugged mountain peaks to well-maintained gardens and built scenes varied from modern city panorama to ruined castle. All scenes were captured in daylight conditions. Some photographers use certain filters offered by graphics editing software to change the appearance of their pictures. We made sure not to select any images for which we could identify that such manipulations had been applied. Care was taken to select only images shot in Scotland by using search terms referring to the country as a whole (‘Scotland’) or to a part of the country (e.g., ‘Aberdeenshire’). We excluded those images portraying commercial messages or well-known landmarks and images shot at unusual viewpoints (e.g. aerial photographs, macro photographs, photographs with a low depth of field). All images were cropped to measure 668 x 501 pixels (501 x 668 when vertically oriented) and were presented on a 19-inch flat panel monitor (Dell

Inc., Round Rock, TX) with screen resolution set at 1280 x 1024 pixels.. Details of the database can be provided by the corresponding author upon request.

2.3 Instruments

The computerized questionnaire was designed using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). All statistical analyses were run in the statistical package R (R Development Core Team, 2011). Initial data exploration was done with a generalized additive model (GAM; Beck & Jackman, 1998) and a linear mixed model (LMM) using the function lmer (lme4 package: Bates, Maechler, & Bolker, 2011). The GAM was used to derive plots which helped to investigate the nature of the relationships between the explanatory variables and beauty, whereas the LMM was used for initial informal testing for significance. In the final step of the analysis, the model with the highest goodness of fit was estimated with an ordinal mixed model (OMM) using the function MCMCglmm (MCMCglmm package; Hadfield, 2010), which uses a Bayesian approach to model the ordinal response on an ordinal scale, rather than rely on an arbitrary assignment to a numeric scale (for a more detailed description of the statistical procedure, see Appendix A).

2.4 Procedure

Each participant individually completed the questionnaire on a computer situated in a PC lab. Following a brief on-screen introduction to the task, participants indicated their Age, Gender, and Rural Experience ("What number of years have you spent living in a village, hamlet, and/or farm?"). Subsequently, participants rated 80 randomly pre-selected images on all items in the questionnaire using a Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). Items were always presented in the same sequence (i.e., Natural Character → Built Character → Beauty → Familiarity → Coherence → Complexity → Legibility → Mystery). An image remained displayed on the screen until each of the items – presented one-by-one – had been scaled. A group of 20 participants (1600 / 80 = 20) was required for each of the 1600 images rated once on all items.

Coherence was defined as the following: “The scene ‘hangs together’; it is easy to organize and structure”. Complexity was defined as: “The scene contains diverse elements and features”. The definition of Mystery read: “This would be an interesting scene to explore further”. The fourth informational variable – Legibility – was defined as: “It would be easy to orient myself around the depicted scene”. Natural Character and Built Character were assessed by the phrases: “This is a natural scene”, and: “This is a built scene” respectively. Finally, Familiarity was measured using the definition: “I am familiar with this type of scene”, whereas the definition for Beauty read: “This is a beautiful scene”.

3 Results

In order to investigate whether observer background had an effect on image ratings, the variable Rural Experience was divided by the age of the participant. One (non-native) participant expressed difficulties grasping the meaning of the concepts “natural” and “built” and reported having shifted interpretation during the experiment. The data for this participant were therefore excluded from the analysis.

Cronbach’s alpha coefficients for scene ratings were calculated for Beauty ($\alpha = 0.853$), Coherence ($\alpha = 0.242$), Complexity ($\alpha = 0.562$), Legibility ($\alpha = 0.253$), and Mystery ($\alpha = 0.763$). However, it was not an objective of the present study to perform a strong test of the reliability of matrix quality measurement. While the alpha coefficients provide some indication of internal consistency of ratings, it should be noted that observers were randomly allocated across every scene. Furthermore, the random effects analysis already takes account of the differences between individual observers when establishing the significance of relationships between each of the predictor variables in the regression modelling.

In step one of the data analysis, we analysed the correlations between each of the variables sampled in this experiment (see Table 2). We interpreted correlations between 0.1 and 0.3 as small, between 0.3 and 0.5 as moderate and greater than 0.5 as large or strong (Cohen, 1988). As a result of the large sample size (df = 7918), each of the correlations between the explanatory variables and beauty was significant; although markedly different in size. Amongst the explanatory variables, Mystery and Natural Character showed the largest correlation with scenic quality. Built Character was also strongly, albeit negatively related to ratings of beauty. Whereas the correlation between Complexity and scenic quality was moderate, the correlations of both Legibility and Coherence with aesthetic judgement were relatively small. Finally, the variables capturing the user experience construct (i.e., Familiarity & Rural Experience) showed small correlations with scenic beauty as well.

Table 2 indicates that the majority of predictor variables were inter-correlated as well. In accordance with predictions by the preference matrix a moderate correlation was established between the Complexity and Mystery variables, which are both covering the ‘exploration’ component of the model. Similarly, Coherence and Legibility – both touching upon the ‘understanding’ component – were also positively correlated with a moderate sized correlation. With regard to the content-related variables – Natural and Built Character – we established a strong inverse interrelationship. Furthermore, natural and built scene content ratings were each moderately correlated with the Mystery and Complexity variables. However, ratings of scene content were unrelated to Legibility and only very weakly related to Coherence (built scene content only). With regard to the experience-related variables, a small correlation could be observed between Familiarity and Rural Experience. In addition, we established moderately sized positive correlations between Familiarity and both Legibility and Coherence. The remaining correlations between the predictor variables were all either of small size or not significant (see Table 2).

In the next step of our analysis we aimed to establish the type of relationship between each of the predictor variables and beauty (see Appendix A for a step-by-step description of the rationale for this statistical procedure). To gain more insight into this aspect of our data, we fitted simple generalized additive models (GAMs) to the raw data and plotted the resulting curves (solid lines in Figures 2A-D & 3A-D) with uncertainty ranges (dashed lines). The individual GAMs provide an informal screening for the possible types of relationships (i.e., linear, quadratic or cubic) we might want to fit. Since overly complex multiple regression linear mixed models (LMMs; e.g. fitting every single predictor variable with a cubic equation) can lead to instability of model fitting in the software, the GAMs served the purpose of reducing the complexity of the initial models that were fitted. For example, if a relationship is obviously linear, fitting first order (i.e., linear) polynomials instead of third order (i.e., cubic) polynomials could simplify the LMMs and ordinal mixed models (OMMs). The lack of straight curves within the GAM plots highlighted that the relationships between each of the explanatory variables and beauty were not necessarily linear, thereby indicating the need to formally test for quadratic and cubic, in addition to linear relationships. Similarly, there were no obvious u- or n-shaped curves indicative of a quadratic relationship.

To formally test for (the shape of) the relationships, an LMM was fitted including all predictor variables as fixed factors and participant number and image identification number (ID) as random factors. All relationships were fitted with a cubic polynomial initially as the GAM plots did not indicate any obvious linear or quadratic relationships. Note that due to employing a (student) participant group the variation in terms of Age was limited; hence, this variable was not incorporated in the regression model.

The LMM provided evidence for nonlinear relationships between a subset of the predictor variables and beauty. The relationships of the variables Coherence and Mystery with beauty were best estimated by a third order (i.e., cubic) polynomial. For Complexity we found a first order (i.e., linear) polynomial to provide the highest goodness of fit. The relationship between Legibility and beauty was found to be most adequately mapped after converting the original seven-point (Likert) scale into a two-level nominal factor discriminating between low (Legibility = 1-3) and high legibility (Legibility = 4-7). That is, contributions of Legibility levelled off at the top-end of the scale. For the remaining variables in the equation we also established various types of relationships. Amongst the content variables, Natural Character mapped best on beauty ratings using a cubic polynomial; whereas for Built Character a linear polynomial sufficed. With regard to the experience-related variables we established a linear polynomial to generate the most effective prediction of the association between Rural Experience and beauty. The relationship between Familiarity and beauty was again represented best by a two-level nominal factor discriminating between low (Familiarity = 1-2) and high familiarity (Familiarity = 3-7).

The interaction between Coherence and Complexity showed the highest goodness of fit when specifying a second-order (i.e., quadratic) polynomial for Coherence in combination with a nominal factor for Coherence. This entailed that the LMM estimated the independent contribution of Complexity to beauty with a quadratic function as well.

Amongst the random effects in the final model, participant had a variance of 0.13 (SD = 0.35) whereas image ID had a variance of 0.16 (SD = 0.40). A relatively large variance of 1.04 (SD = 1.02) was left unexplained when between participant and between image variability was taken into account, showing a high degree of unsystematic variability associated with individual participants (or images). This suggests it is important to have each participant view a sufficient number of images and for each image to be rated by a sufficient number of participants, as this will help counter the largest source of residual variability.

In the third step of our analysis we entered the variables from the LMM into an ordinal mixed model (OMM) for hypothesis testing. The rationale for doing so is that an OMM better represents the (ordinal) structure of the response data than the commonly used linear regression analysis (for a more detailed rationale, see Appendix A). In contrast to the lmer procedure, the MCMCglmm software produces non-sequential output, which means that the effect of each variable on the target variable is conditional on (i.e., controlled for) that of all other variables in the equation (cf. type-III sums of squares). The MCMCglmm package uses a Bayesian method to estimate the model parameters via the method of "Markov chain Monte Carlo" (Hadfield, 2010). We stress that the interpretation of Bayesian output is slightly different from frequentist output (e.g., Aspinall, 2012). Firstly, the "CI" in Table 3 refers to Credible (rather than "Confidence") Interval; providing a measure of the uncertainty range in which the true value lies with a 95 percent probability. This is unlike the confidence interval – common in frequentist statistics – which gives an indication of the percentage of samples in which the mean would occur in a certain range. The Bayesian p-values in Table 3 should also be interpreted slightly differently. That is, a regression coefficient with a Bayesian p-value of 0.05, for example, indicates that the probability of the corresponding variable not being a useful predictor of beauty is five percent. In frequentist statistics, however, this particular p-value would indicate that there is a five percent probability of obtaining a similarly sized Beta-weight in a replication of the experiment on the premise that the null hypothesis (of there being no relationship) is true.

The OMM provided support for each of the four informational variables of the preference matrix as predictors of beauty (see Table 3). Figure 4, however, depicts a strong variability in the type, size and direction of the relationships between each of these predictor variables and beauty. Most notable amongst the predictor variables is Mystery due to its strong positive association with beauty; mostly at the extremes of the scale. The curve for Coherence similarly shows strongest contributions of this attribute to scenic quality at the extremes of the scale. The quadratic curve for Complexity can be interpreted to suggest that the independent contribution by this particular attribute to beauty was somewhat more pervasive at the top-end of the scale. For Legibility we observe that low levels of this attribute affected scenic beauty negatively, whereas it did not make an independent contribution to ratings of beauty for scenes rated in the mid- top top-end of the scale.

With regard to the remaining variables (see Figure 5), it can be observed that Natural Character served as an important positive predictor of scenic quality with the strength of its contribution gradually increasing towards the top-end of the scale. Built Character had a modestly sized negative effect on beauty ratings which was best estimated with a linear function. We further show that low image rankings on the construct of Familiarity were detrimental to beauty ratings. However, for those scenes ranked intermediate to high on this particular construct, Familiarity was insignificant as independent predictor of beauty. The variables Gender and Rural Experience did not have an independent contribution towards predicting scenic quality.

In addition, our findings also support an independent contribution by the interactive relationship between Complexity and Coherence on appraisals of scenic beauty (see Table 3). It can be observed in Figure 6 that the positive contribution of Complexity to scenic beauty is always more positive for high, as opposed to low or intermediate levels of Coherence. However, the inverted-U shaped function of Complexity and beauty for high levels of Coherence suggests that scenes with an intermediate level of Complexity – combined with a high level of Coherence – have a higher level of beauty than scenes with an abundance of these attributes.

Finally, two post-hoc OMMs were conducted to further elucidate the relationship between Mystery and beauty given a high correlation. Firstly, we investigated whether exclusion of Mystery as predictor variable would render any of the other predictors in the model non-significant, which proved not to be the case. Next, we ran an OMM in which all variables, supplemented by beauty, were entered as predictors of the target variable Mystery. In this model, both content variables and Coherence failed to reach significance, confirming that beauty and Mystery do not share a single underlying construct.

4 Discussion

The aim of the present study was to investigate the validity of the preference matrix in predicting aesthetic responses to physical environments. This endeavour was motivated on the finding that the preference matrix – despite its rootedness in a long tradition of enquiry into landscape aesthetics – lacks conclusive empirical support (Stamps, 2004). We reasoned that the inconsistencies between previous research findings could be accounted for by the following methodological limitations: Low comprehension of item definitions, poor control for confounding variables, low variety of stimulus material and purported linear type of relationship between predictor and target variables. To deal with such shortcoming we piloted understanding of item definitions, controlled for confounding effects of variables related to both scene content and user experience, employed a large and highly variable image database and tested for nonlinear relationships. Doing so, we provide convergent evidence for our hypothesis that all informational variables of the preference matrix, as well as a variety of variables related to scene content and user experience, are independently predictive of aesthetic judgements.

4.1 The Informational Variables

All four variables of the preference matrix – Coherence, Complexity, Legibility and Mystery – were found to be predictive of aesthetic appraisal as measured by beauty. Previous research on the preference matrix has shown the variable Mystery to be one of the most potent predictor variables (R. Kaplan & Kaplan, 1989). The present findings are in line with this contention as Mystery had a strong correlation with beauty and emerged as the most significant predictor in the regression model after controlling for other variables. Similarly-sized correlations between Mystery and scenic quality have been reported before; however, in a subset of previous studies a significant correlation failed to emerge (Stamps, 2004). It has, however, been reasoned that such null-effects with regard to Mystery can be explained by a misinterpretation of the item definition in combination with an image set low in variability (Herzog & Bryce, 2007). We find that the effect of Mystery on the aesthetic appraisal is best described as a cubic relationship. That is, the contribution of Mystery is less pronounced for advancements in the mid-level of the scale as compared to the bottom- and top-ends of the scale.

Despite a very substantial correlation with Mystery we established an independent contribution of the variable Complexity to ratings of beauty; described best as a linear positive relationship. This finding might be somewhat surprising given research showing that scenes with intermediate levels of complexity have the highest hedonic value (Berlyne, 1971; Wohlwill, 1974). However, this finding should not be interpreted to imply that an inverted-U relationship could not exist between a statistically uncontrolled version of the Complexity variable and scenic quality (although the GAM plot in Figure 2B does not suggest so).

Notwithstanding the small correlation with beauty, the variable Coherence emerged as an independent predictor of scenic quality. That can in part be explained by the very small correlations between Coherence and the variables Mystery, Natural Character and Built Character. With regard to Coherence we show – similar to Mystery – a positive relationship that is best represented by a cubic function. This signals a particularly adverse influence of scenes with low levels of Coherence on scenic quality, as opposed to a particularly beneficial influence of this attribute for scenes deemed to have high levels of this attribute.

In previous research, Legibility has been found to be a relatively ineffective predictor of the aesthetic response to landscapes. This entails that Legibility was either a relatively weak predictor, or not predictive at all, of scenic quality (Herzog & Leverich, 2003; R. Kaplan & Kaplan, 1989; Stamps, 2004). However, the outcome of the present study counteracts such previous studies. That is, we found a significant and positive correlation between Legibility and beauty. Furthermore, we showed that, despite a strong correlation with Familiarity, the Legibility construct uniquely contributed to predicting aesthetic evaluations. The relationship between Legibility and beauty was most effectively represented by condensing the original 7-factor measurement scale to a nominal scale discriminating between high and low levels of this attribute. We find that low levels of legibility are detrimental to scenic beauty whereas high levels of Legibility leave the aesthetic judgement unaffected. This is convergent with studies showing legibility to be a more effective predictor in densely vegetated forest settings than in open field settings (Herzog & Leverich, 2003; Herzog & Kropscott, 2004), as well as with other studies finding that densely vegetated scenes lacking in open views can compromise hedonic value (Hammitt, Patterson, & Noe, 1994; Herzog & Kutzli, 2002; Staats, Gatersleben, & Hartig, 1997).

Contrary to the common practice of investigating the contribution of Coherence and Complexity to scene aesthetics in isolation, we tested for an interactive relationship between both variables in the present study. In line with our expectation, such interaction was established independently of all other predictor variables. This showed that Complexity contributed more strongly towards predicting the quality of scenes that were also judged as highly coherent. Importantly, however, the combination of high Coherence with intermediate levels of Complexity was deemed as more attractive than that of high Coherence with either low or high levels of Complexity. A potential explanation for this finding is that scenes combining high Coherence with high Complexity relatively often represented pristine wilderness environments or natural environments with ruins or other overgrown man-made structures. These categories of scenes might be responded to with fear by some participants, which could have compromised judgements of scene attractiveness. To illustrate, research has shown that participants of a wilderness program, despite its overall positive impact on psychological well-being, almost all initially experienced a degree of fear and discomfort during their visit (Kaplan & Talbot, 1983). Another study showed that wilderness environments prime death-related thoughts to a higher extent than cultivated environments (Koole & Van den Berg, 2005). Overall, our findings are convergent with seminal theories in landscape aesthetics conceptualizing the interplay between order and diversity (or similar constructs) in accounting for scenic quality (e.g., Appleton, 1975; R. Kaplan & Kaplan, 1989).

4.2 Scene Content and User Experience

In line with previous research, we demonstrated that naturalness is a powerful and positive predictor of scenic quality (Herzog, 1989; R. Kaplan & Kaplan, 1989; Ode, Fry, Tveit, Messager, & Miller, 2009; Real et al., 2000; Ulrich, 1983). Interestingly, the relationship between Natural Character and landscape beauty was best represented by a quadratic curve suggesting that increments at the top-end of the scale have a stronger beneficial effect when compared to the bottom-end of the scale. This suggests that there is a level of naturalness that needs to be achieved before the strongest progression in scenic quality associated with improved Natural Character can be discerned.

In contrast to Natural Character, the variable Built Character emerged as a monotonous negative predictor of beauty. This corroborates the findings of previous studies demonstrating a detrimental effect of built artefacts on scenic quality (Arriaza, Canas-Ortega, Canas-Madueno, & Ruiz-Aviles, 2004; De Vries et al., 2012; Molnarova et al., 2012; White et al., 2010). It is an interesting observation that, despite a substantial negative correlation between the respective constructs, both Natural and Built Character independently contributed towards predicting ratings of beauty. This confirms our notion that natural and built scene content cannot be regarded as the opposite directions along a single dimension. Participants thus likely have varied in the degree to which they conceptualized Natural and Built Character alongside ecological or perceptually derived criteria (i.e., spatial information such as availability of sharp edges; Gobster, 1999; Wohlwill, 1983).

Notwithstanding the weak correlations between each of the experience-related variables (i.e., Familiarity and Rural Experience) and beauty, we demonstrated an independent contribution of Familiarity to predicting scenic quality. The unique relationship between both constructs was best captured by a nominal function as low levels of Familiarity affected landscape attractiveness negatively whereas for medium to high levels of this attribute no effect could be observed. Our findings are in line with that of Peron and colleagues (2002) who suggested that high familiarity is less effective in furthering scenic quality than low familiarity is in hampering it. This contrasts with findings of previous research showing that observer experience is an important positive predictor of beauty or preference as well (e.g., Buijs, Elands, & Langers, 2009; Howley, 2011; Van den Berg & Koole, 2006). In the present study the influence of Familiarity might have been underrepresented due to the relatively homogenous participants of undergraduate students of similar age. It should also be taken into account that Familiarity showed medium correlations with the informational variables Coherence and Legibility. Factoring out the influence of these variables in the regression equation is likely to have diminished the predictive value of Familiarity. Although alternative definitions of this item could therefore be considered, the present definition (i.e., "I am familiar with this type of scene") was chosen based on the data of a pilot study, which revealed the importance of specifying the type of familiarity; is one primarily interested in the participant's acquaintance with the specific type of scene (e.g., busy city street) or with the specific content of the image (e.g., The Eiffel Tower in Paris)?

Finally, we failed to find support for Rural Experience as an independent predictor of beauty. It should, however, be noted that our measure of the number of years lived in a rural setting gives at most a rather crude approximation of exposure to nature; one might spend most time indoors despite living in a rural setting or vice versa. As Kaplan and Kaplan (1989, p. 74) put it: "The very categorization of rural, urban, and suburban may not usefully parallel the way people experience different kinds of natural settings”. Furthermore, variation on this measure was very low as most participants had an urban background.

4.3 Is Beauty Confounded With Mystery?

Based on the strong correlation between Mystery and beauty, one might argue that participants have used their ratings of beauty as a proxy for assessments of the mystery of a scene. The finding, however, that when controlling for Mystery each of the informational and content variables predicted the scenic judgement significantly, suggests that participants have not done so. For that reason, two post-hoc OMMs were fitted in which it was investigated whether the same variables can be used interchangeably as predictor and target variables. This proved not to be the case, which argues against the idea that participants collectively failed to discriminate between their ratings of Mystery and scenic beauty.

4.4 Limitations and Suggestions for Future Research

Despite the considerable care that has been taken in finding a justifiable methodology for testing the preference matrix, we would like to outline several limitations with regard to the present study. Firstly, although we incorporated content- and experience-related measures to account for unsystematic variation in our data, other variables with potential confounding effects on (the predictors of) scenic beauty have not been taken into account. Amongst such attributes outlined in the literature are: Historicity, openness, weather and seasonal influences, maintenance, style, intensity of use, surprise and colour (Berlyne, 1971; Nasar, 1994; Tveit, Ode, & Fry, 2006). In addition, we can also outline specific types of scene content which have been considered to elicit particularly strong affective responses. Within the category of natural environments, for example, it has been shown that water and trees are particularly strong positive predictors of scenic quality; unlike dense forests, rocks, and modern agriculture (Hammitt et al., 1994; Herzog, 1989; Kaltenborn & Bjerke, 2002; Real et al., 2000; Ulrich, 1983; White et al., 2010; Yang & Brown, 1992). Similarly, we could have included additional variables predictive of scene typicality; ethnicity, demographic background, user group and environmental knowledge or interest (Buijs et al., 2009; Gobster, Nassauer, Daniel, & Fry, 2007; R. Kaplan & Kaplan, 1989; Van den Berg et al., 1998). Not surprisingly, we lacked sufficient scope to incorporate such large variety of variables in the present experiment.

Although further research is required to better establish the extent of the confounding influence of such attributes as mentioned above on the variables of the preference matrix, we believe that the measures in our study provided at least partial control for this myriad of potentially confounding variables. That is, we may expect strong correlations of such other variables with the measures in the present experiment (e.g., historicity-natural character; maintenance-coherence; modern agriculture-built character; user group-familiarity). In addition, our study design, which accommodated the presentations of ever-changing sets of highly variable scenes to different participants, has likely diminished any systematic effects by scene content or typicality. Finally, the use of random effects in our regression model factored out systematic variation associated with individual participants and images which would perhaps have otherwise been controlled for by incorporating such variables as mentioned above.

In the present study we piloted understanding of the questionnaire items as conventional definitions have been criticized on grounds of ambiguity (Stamps, 2004). Based on the outcome of these studies we applied subtle modifications to a subset of item definitions, thereby heeding the theoretical conceptualizations of the respective variables. Nonetheless, we cannot completely exclude the possibility that participants have diverged in their interpretation of item definitions or that interpretation of the item definitions has been inconsistent across images. That is, previous efforts have shown that it is not an easy task to standardize item interpretation across different (sets of) participants and images (Herzog & Leverich, 2003; Herzog & Bryce, 2007).

In future research, an alternative solution for dealing with measurement error resulting from high variability in item interpretation might be to limit the exposure time to scenes. To illustrate, research on visual perception has shown that a mere 35 ms of exposure time is sufficient for reliable judgements of scene affordances to be made (Fei-Fei, Iyer, Koch, & Perona, 2007; Greene & Oliva, 2009a). We also know from research in perception that awareness of basic-level scene category (e.g., forest, field) is still very limited following such ultra-rapid scene presentations (Loschky & Larson, 2010). It follows from this that limiting exposure time to such levels at which structural scene properties and gross content categories (i.e., natural, built) can be extracted while awareness of specific land use is compromised, is likely to diminish the extent to which content- and experience-related attributes serve as a confound on item ratings.

Our use of random effects analysis has enabled us to make generalizations from findings obtained with a random sample of participants and images to the population and environment as a whole (Baayen, Davidson, & Bates, 2008). However, all of the images used in our study were nonetheless still shot in Scotland. This was a deliberate constraint introduced in order to minimize variability in terms of participant familiarity, but it may also have limited the absolute generality of our findings. It is possible that relationships described in our study depend on qualities of Scottish scenes that we are not currently aware of. This potential limitation could be addressed by future replications of our paradigm that select scenes using different environmental contexts.

Finally, we would like to highlight the need for research on the mechanism(s) that drive the evaluation of the attributes described in the preference matrix. For instance, to what extent does our personal experience influence our experience of landscape coherence and complexity? Furthermore, future research is required for showing if, and illuminating how, the psychological indicators of the preference matrix are represented in the content and spatial information of a scene. We believe that conducive to such efforts will be a consideration of developments in related fields of scientific enquiry. For instance, Oliva and Torralba (2001) have presented a computational model which shows that the perception of depth is correlated with image statistics indicative of the presence of long, vanishing lines. In the foreseeable future, a methodological approach to computationally map functional scene properties – informational variables that may manifest themselves in a variety of ways (e.g., mystery) – might be uncovered as well (Greene & Oliva, 2009b).

5 Conclusions

In the present study we aimed to test the validity of the preference matrix in response to inconclusive support from previous research. To this end, a series of methodological refinements were made concerning: (1) item definitions, (2) control for confounding variables, (3) variety of stimulus material and (4) presupposed type of relationship between predictor and target variables. Doing so, we provide convergent evidence for our hypothesis that each of the predictor variables of the preference matrix independently contributes to predicting scenic quality evaluations. Secondly, we show a significant interaction between the constructs of Coherence and Complexity in predicting landscape aesthetics. Finally, we demonstrate how the hedonic value of scenes can increase in association with ratings of naturalness, whereas built character and low familiarity exert a negative effect. In line with expectations, our analysis shows substantial variation in the form and strength of the relationships between each of these constructs and scene attractiveness. We advocate for further research to elucidate if and how the psychological indicators of the preference matrix can be translated into specific design recommendations.

References

1. Appleton, J. (1975). The experience of landscape. London: Wiley.
2. Arriaza, M., Canas-Ortega, J., Canas-Madueno, J., & Ruiz-Aviles, P. (2004). Assessing the visual quality of rural landscapes. Landscape and Urban Planning, 69(1), 115-125. doi: 10.1016/j.landurbplan.2003.10.029
3. Aspinall, P. A. (2012). On avoiding research into the blindingly obvious. Architectoni.Ca, 1, 74-82.
4. Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 59(4), 390-412. doi: 10.1016/j.jml.2007.12.005
5. Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes (R package version 0.999375-42) [Software]. Vienna, Austria: R Foundation for Statistical Computing. Available from http://CRAN.R-project.org/package=lme4
6. Beck, N., & Jackman, S. (1998). Beyond linearity by default: Generalized additive models. American Journal of Political Science, 42(2), 596-627. doi: 10.2307/2991772
7. Berlyne, D. E. (1971). Aesthetics and psychobiology. New York: Appleton-Century-Crofts. Bishop, I. D., & Hulse, D. W. (1994). Prediction of scenic beauty using mapped data and geographic information-systems. Landscape and Urban Planning, 30(1-2), 59-70. doi: 10.1016/0169-2046(94)90067-1
8. Buijs, A. E., Elands, B. H. M., & Langers, F. (2009). No wilderness for immigrants: Cultural differences in images of nature and landscape preferences. Landscape and Urban Planning, 91(3), 113-123. doi: 10.1016/j.landurbplan.2008.12.003
9. Coeterier, J. (1996). Dominant attributes in the perception and evaluation of the Dutch landscape. Landscape and Urban Planning, 34(1), 27-44. doi: 10.1016/0169-2046(95)00204-9
10. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.
11. Daniel, T. C. (2001). Whither scenic beauty? Visual landscape quality assessment in the 21st century. Landscape and Urban Planning, 54(1-4), 267-281. doi: 10.1016/S0169-2046(01)00141-4
12. Daniel, T. C., Boster, R. S., & Forest, R. M. (1976). Measuring landscape esthetics: The scenic beauty estimation method. (USDA Forest Service Research Paper RM-167). Rocky Mountain Forest and Range Experiment Station Fort Collins, CO.
13. Daniel, T. C., & Vining, J. (1983). Methodological issues in the assessment of landscape quality. In I. Altman, & J. F. Wohlwill (Eds.), Human behavior and environment: Advances in theory and research, Vol. 6 (pp. 39-84). New York: Plenum Press.
14. De Groot, W. T., & van den Born, R. J. G. (2003). Visions of nature and landscape type preferences: An exploration in The Netherlands. Landscape and Urban Planning, 63(3), 127-138. doi: 10.1016/S0169-2046(02)00184-6
15. De Vries, S., de Groot, M., & Boers, J. (2012). Eyesores in sight: Quantifying the impact of man-made elements on the scenic beauty of Dutch landscapes. Landscape and Urban Planning, 105(1-2), 118-127. doi: 10.1016/j.landurbplan.2011.12.005
16. Draper, N. R., & Smith, H. (1981). Applied Regression Analysis (2nd ed.). New York: Wiley.

17. Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? Journal of Vision, 7(1), 1-29. doi: 10.1167/7.1.10
18. Gobster, P. H. (1999). An ecological aesthetic for forest landscape management. Landscape Journal, 18(1), 54-64. doi: 10.3368/lj.18.1.54
19. Gobster, P. H., Nassauer, J. I., Daniel, T. C., & Fry, G. (2007). The shared landscape: What does aesthetics have to do with ecology? Landscape Ecology, 22(7), 959-972. doi: 10.1007/s10980-007-9110-x
20. Greene, M. R., & Oliva, A. (2009a). The briefest of glances: The time course of natural scene understanding. Psychological Science, 20(4), 464-472. doi: 10.1111/j.1467-9280.2009.02316.x
21. Greene, M. R., & Oliva, A. (2009b). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. Cognitive Psychology, 58(2), 137-176. doi: 10.1016/j.cogpsych.2008.06.001
22. Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. Journal of Statistical Software, 33(2), 1-22. Available from http://www.jstatsoft.org/v33/i02/
23. Hammitt, W. E., Patterson, M. E., & Noe, F. P. (1994). Identifying and predicting visual preference of Southern Appalachian forest recreation vistas. Landscape and Urban Planning, 29(2-3), 171-183. doi: 10.1016/0169-2046(94)90026-4
24. Han, K. -T. (2010). An exploration of relationships among the responses to natural scenes: Scenic beauty, preference, and restoration. Environment and Behavior, 42(2), 243-270. doi: 10.1177/0013916509333875
25. Herzog, T. R. (1989). A cognitive analysis of preference for urban nature. Journal of Environmental Psychology, 9(1), 27-43. doi: 10.1016/S0272-4944(89)80024-6
26. Herzog, T. R., & Bosley, P. J. (1992). Tranquillity and preference as affective qualities of natural environments. Journal of Environmental Psychology, 12(2), 115-127. doi: 10.1016/S0272-4944(05)80064-7
27. Herzog, T. R., & Bryce, A. G. (2007). Mystery and preference in within-forest setting. Environment and Behavior, 39(6), 779-796. doi: 10.1177/0013916506298796
28. Herzog, T. R., Kaplan, S., & Kaplan, R. (1982). The prediction of preference for unfamiliar urban places. Population and Environment, 5(1), 43-59. doi: 10.1007/BF01359051
29. Herzog, T. R., & Kirk, K. M. (2005). Pathway curvature and border visibility as predictors of preference and danger in forest settings. Environment and Behavior, 37(5) doi: 10.1177/0013916505275306
30. Herzog, T. R., & Kropscott, L. S. (2004). Legibility, mystery, and visual access as predictors of preference and perceived danger in forest settings without pathways. Environment and Behavior, 36(5), 659-677. doi: 10.1177/0013916504264138
31. Herzog, T. R., & Kutzli, G. E. (2002). Preference and perceived danger in field/forest settings. Environment and Behavior, 34(6), 819-835. doi: 10.1177/001391602237250
32. Herzog, T. R., & Leverich, O. L. (2003). Searching for legibility. Environment and Behavior, 35(4), 459-477. doi: 10.1177/0013916503251455
33. Howley, P. (2011). Landscape aesthetics: Assessing the general publics' preferences towards rural landscapes. Ecological Economics, 72, 161-169. doi: 10.1016/j.ecolecon.2011.09.026

34. Kaltenborn, B. P., & Bjerke, T. (2002). Associations between landscape preferences and place attachments: A study in Røros, Southern Norway. Landscape Research, 27(4), 381-396. doi: 10.1080/01426390220000023943
35. Kaplan, R., & Kaplan, S. (1989). The experience of nature: A psychological perspective. New York: Cambridge University Press.
36. Kaplan, S. (1987). Aesthetics, affect, and cognition: Environmental preference from an evolutionary perspective. Environment and Behavior, 19(1), 3-32. doi: 10.1177/0013916587191001
37. Kaplan, S., & Talbot, J. F. (1983). Psychological benefits of a wilderness experience. In I. Altman, & J. F. Wohlwill (Eds.), Human behavior and environment: Advances in theory and research, Vol. 6 (pp. 163–203). New York: Plenum Press.
38. Koole, S. L., & Van den Berg, A. E. (2005). Lost in the wilderness: Terror management, action control, and evaluations of nature. Journal of Personality and Social Psychology, 88(6), 1014-1028. doi: 0.1037/0022-3514.88.6.1014
39. Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. Visual Cognition, 18(4), 513-536. doi: 10.1080/13506280902937606
40. Lothian, A. (1999). Landscape and the philosophy of aesthetics: Is landscape quality inherent in the landscape or in the eye of the beholder? Landscape and Urban Planning, 44(4), 177-198. doi: 10.1016/S0169-2046(99)00019-5
41. Molnarova, K., Sklenicka, P., Stiborek, J., Svobodova, K., Salek, M., & Brabec, E. (2012). Visual preferences for wind turbines: Location, numbers and respondent characteristics. Applied Energy, 92, 269-278. doi: 10.1016/j.apenergy.2011.11.001
42. Nasar, J. L. (1994). Urban design aesthetics - the evaluative qualities of building exteriors. Environment and Behavior, 26(3), 377-401. doi: 10.1177/001391659402600305
43. Ode, A., Fry, G., Tveit, M. S., Messager, P., & Miller, D. (2009). Indicators of perceived naturalness as drivers of landscape preference. Journal of Environmental Management, 90(1), 375-383. doi: 10.1016/j.jenvman.2007.10.013
44. Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision, 42(3), 145-175. doi: 10.1023/A:1011139631724
45. Orians, G. H., & Heerwagen, J. H. (1992). Evolved responses to landscapes. In J. Barkow, L. Cosmides & J. Tooby (Eds.), The adapted mind: Evolutionary psychology and the generation of culture (pp. 555-579). New York: Oxford University Press.
46. Parsons, R., & Daniel, T. C. (2002). Good looking: In defense of scenic landscape aesthetics. Landscape and Urban Planning, 60(1), 43-56. doi: 10.1016/S0169-2046(02)00051-8
47. Peron, E., Berto, R., & Purcell, T. (2002). Restorativeness, preference and the perceived naturalness of places. Medio Ambiente y Comportamiento Humano, 3(1), 19-34.
48. Purcell, A. T. (1987). Landscape perception, preference, and schema discrepancy. Environment and Planning B-Planning & Design, 14(1), 67-92. doi: 10.1068/b140067
49. Purcell, A. T., Peron, E., & Berto, R. (2001). Why do preferences differ between scene types? Environment and Behavior, 33(1), 93-106. doi: 10.1177/00139160121972882

50. R Development Core Team (2011). R: A language and environment for statistical computing (Version 2.15.0) [Software]. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org
51. Real, E., Arce, C., & Sabucedo, J. M. (2000). Classification of landscapes using quantitative and categorical data, and prediction of their scenic beauty in North-Western Spain. Journal of Environmental Psychology, 20(4), 355-373. doi: 10.1006/jevp.2000.0184
52. Staats, H., Gatersleben, B., & Hartig, T. (1997). Change in mood as a function of environmental design: Arousal and pleasure on a simulated forest hike. Journal of Environmental Psychology, 17(4) doi: 10.1006/jevp.1997.0069
53. Stamps, A. (2004). Mystery, complexity, legibility and coherence: A meta-analysis. Journal of Environmental Psychology, 24(1), 1-16. doi: 10.1016/S0272-4944(03)00023-9
54. Taylor, R. P., Spehar, B., Van Donkelaar, P., & Hagerhall, C. M. (2011). Perceptual and physiological responses to Jackson Pollock's fractals. Frontiers in Human Neuroscience, 5, 60.
55. Tveit, M., Ode, A., & Fry, G. (2006). Key concepts in a framework for analysing visual landscape character. Landscape Research, 31(3), 229-255. doi: 10.1080/01426390600783269
56. Ulrich, R. S. (1983). Aesthetic and affective response to natural environment. In I. Altman, & J. F. Wohlwill (Eds.), Human behavior and environment: Advances in theory and research, Vol. 6 (pp. 85-125). New York: Plenum Press.
57. Van den Berg, A. E., & Koole, S. L. (2006). New wilderness in The Netherlands: An investigation of visual preferences for nature development landscapes. Landscape and Urban Planning, 78(4), 362-372. doi: 10.1016/j.landurbplan.2005.11.006
58. Van den Berg, A. E., Vlek, C. A. J., & Coeterier, J. F. (1998). Group differences in the aesthetic evaluation of nature development plans: A multilevel approach. Journal of Environmental Psychology, 18(2), 141-157. doi: 10.1006/jevp.1998.0080
59. White, M., Smith, A., Humphryes, K., Pahl, S., Snelling, D., & Depledge, M. (2010). Blue space: The importance of water for preference, affect, and restorativeness ratings of natural and built scenes. Journal of Environmental Psychology, 30(4), 482-493. doi: 10.1016/j.jenvp.2010.04.004
60. Wohlwill, J. F. (1974). Human adaption to levels of environmental stimulation. Human Ecology, 2, 127-147. doi: 10.1007/BF01558117
61. Wohlwill, J. F. (1983). The concept of nature: A psychologist's view. In I. Altman, & J. F. Wohlwill (Eds.), Human behavior and environment: Advances in theory and research, Vol. 6 (pp. 5-38). New York: Plenum Press.
62. Woodcock, D. M. (1984). A functionalist approach to landscape preference. Landscape Research, 9(2), 24-27. doi: 10.1080/01426398408706109
63. Yang, B. E., & Brown, T. J. (1992). A cross-cultural-comparison of preferences for landscape styles and landscape elements. Environment and Behavior, 24(4), 471-507. doi: 10.1177/0013916592244003
64. 3924(82)90009-0

List of Tables

Table 1.
Mean Comprehension Scores of the Informational Variables in the Pilot Studies and the Standard Definitions

Table 2.
Pearson Correlations Between the Questionnaire Items

Table 3.
The Order of Polynomial and Parameter Estimates of the Variables in the Bayesian MCMC Model With Best Penalized Fit

Individual Tables

Table 1.
Mean Comprehension Scores of the Informational Variables in the Pilot Studies and the Standard Definitions

Variable	Item	Definition	M (SD)
Coherence	Pilot [b]	The visual elements of the scene fit together well	5.10 (1.91)
	Standard	How well does the scene ‘hang together’? How easy is it to organize and structure the scene?	-
Complexity	Pilot [a]	The scene contains diverse elements and features	6.18 (0.88)
	Standard	How much is going on in the scene? How much is there to look at? If the scene contains a lot of elements of different kinds, rate it high in complexity	-
Legibility	Pilot [b]	It would be easy to find my way around the environment depicted	5.40 (1.90)
	Standard	How easy would it be to find your way around in the setting? How easy would it be to figure out where you are at any given moment or to find your way back to any given point in the setting?	-
Mystery	Pilot [b]	This would be an interesting scene to explore further	5.90 (1.73)
	Standard	How much does the setting promise more to be seen if you could walk deeper into it? Does the setting seem to invite you to enter more deeply into it and thereby learn more?	-

Note. [a] Item was piloted in Study 1 (N = 17), [b] Item was piloted in Study 2 (N = 10)

Table 2.
Pearson Correlations Between the Questionnaire Items

Variable	1.	2.	3.	4.	5.	6.	7.	8.
1.Beauty	-	-	-	-	-	-	-	-
2.Complexity	0.38 [a]	-	-	-	-	-	-	-
3.Coherence	0.11 [a]	-0.02 [c]	-	-	-	-	-	-
4.Legibility	0.15 [a]	0.09 [a]	0.41 [a]	-	-	-	-	-
5.Mystery	0.75 [a]	0.48 [a]	0.07 [a]	0.17 [a]	-	-	-	-
6.Natural	0.63 [a]	0.25 [a]	-0.01	0.01	0.50	-	-	-
7.Built	-0.55 [a]	-0.17 [a]	0.03 [c]	0.01	-0.44 [a]	-0.86 [a]	-	-
8.Familiarity	0.08 [a]	0.03 [c]	0.27 [a]	0.35 [a]	0.05 [a]	0.06 [a]	-0.06 [a]	-
9.Experience	0.03 [b]	0.05 [a]	0.01	0.06 [a]	0.03 [c]	0.00	0.01	0.04 [b]

Note. [a] Correlation is significant at the 0.001 level (two-tailed), [b] Correlation is significant at the 0.01 level (two-tailed), [c] Correlation is significant at the 0.05 level (two-tailed).

Table 3.
The Order of Polynomial and Parameter Estimates of the Variables in the Bayesian MCMC Model With Best Penalized Fit

Parameter	Order of polynomial	Beta *(β)*	95% CI lower limit	95% CI upper limit	Bayesian p-value
Intercept	-	3.95	3.41	4.44	< 0.001
Natural	linear	72.6	65.4	80.2	< 0.001
Natural	quadratic	11.8	7.69	15.6	< 0.001
Built	linear	-16.3	-23.3	-9.32	< 0.001
Familiarity < 2.5	<factor>	-0.26	-0.38	-0.14	< 0.001
Rural Experience	linear	6.04	-4.50	16.7	0.25
Gender (male)	<factor>	-0.15	-0.40	0.11	0.26
Mystery	linear	137	131.4	142.8	< 0.001
Mystery	quadratic	1.85	-2.16	6.08	0.37
Mystery	cubic	22.9	19.4	26.5	< 0.001
Complexity	linear	10.7	5.76	15.6	< 0.001
Complexity	quadratic	1.46	-3.38	6.37	0.56
Legibility < 3.5	<factor>	-0.23	-0.33	-0.13	< 0.001
Coherence	linear	19.2	15.1	23.1	< 0.001
Coherence	quadratic	6.76	2.78	10.8	< 0.01
Coherence	cubic	7.62	3.99	11.2	< 0.001
Complexity * Coherence > 5.5	linear * <factor>	1.65	-5.68	8.95	0.65
Complexity * Coherence > 5.5	quadratic * <factor>	-14.7	-22.3	-7.09	< 0.001

List of Figures

List of Appendices

Appendix A.

Statistical Procedure for the Determination of the Type of Relationship Between the Predictor Variables and Beauty.

Initial data exploration was done using a generalized additive model (GAM: Beck & Jackman, 1998) and a standard linear mixed model (LMM); the latter with the lmer function in R. Both methods allow for fast analysis of the data but incorrectly assume that the Likert response is continuous. We chose to perform initial exploratory analyses with these methods as they are more efficient (i.e., less resource intensive) than fitting the full Bayesian ordinal mixed model using Markov chain Monte Carlo (MCMC) algorithms.

The GAM plots were fitted to investigate if there were any grounds for assuming a complex (i.e., nonlinear) relationship between the predictor variables and Beauty. The actual relationship, when we do a multiple regression, is unlikely to be more complex, but quite possibly it may be less complex because of collinearity (correlation with other covariates).

In the next step, we fitted a series of LMMs to test the actual relationship of variables with Beauty. The functional form of the explanatory variables was selected using a carefully planned stepwise procedure. First, cubic relationships, corresponding to third order polynomials between the predictor and target variable, were investigated.[1] If a cubic association between the variables could not be established, the cubic polynomial was discarded and it was tested for a quadratic relationship, corresponding to a second order polynomial. If no quadratic association existed either, the quadratic polynomial was discarded and a first order polynomial, testing for a linear relationship, was fitted between the explanatory variable and beauty.

The next step was to scrutinize the GAM plots for relationships between each of the explanatory variables and the target variable beauty that signalled tendencies to the minimum or maximum (i.e., floor and ceiling effects). Subsequently, for the variables showing any such tendency it was investigated whether the model could be simplified by using a nominal (i.e., two-level factor) instead of a seven-level factor coding with the cut-off point between the factors set in accordance with the point at which saturation was achieved. Such practice reduces the degrees of freedom, thereby increasing the likelihood of finding an effect. Note, however, that whenever variables have been reduced to a two-level factor scale this should not be interpreted as evidence for the respective construct being more effectively measured on a dichotomous scale. That is, because such an amendment is likely to result in a different kind of rating behaviour. The data was also visually inspected to identify possible interactions between informational variables. The goodness of fit of each newly fitted model – penalized for the number of parameters – was scrutinized using the Akaike Information Criterion (AIC; Kuha, 2004). The following criteria for model acceptance were applied: Regression models with an AIC difference > 2 were regarded as significantly different in terms of goodness of fit and that model with the smallest AIC-value was adhered to.

Finally, the model with the best relative fit was estimated using an MCMC procedure (the MCMCglmm function in R) to obtain variable regression weights. The main rationale for shifting from a linear to an ordinal mixed effects model at this stage is that assignment of a simple numerical value to each point on the Likert scale, which is an assumption of linear models, is

[1] Note that an equation with a cubic polynomial automatically includes a quadratic and linear polynomial as well. An equation with a quadratic polynomial includes a linear polynomial, but not a cubic polynomial. An equation with a linear polynomial does include neither a cubic nor a quadratic polynomial (see Table 3).

essentially an arbitrary choice; one that may affect the outcome of the analysis (for example, see Chapter 7 in Agresti, 1984). Such difficulties can, for example, come about when participants do not experience the difference between each of the successive points on the Likert scale as being equidistant (McCullagh, 1980). Another argument for choosing an ordinal mixed model is its robustness against non-normality and skewness. While it is true that linear models have been found to be robust to deviations from normality (Norman, 2010), and that their use in analysing ordinal response data is common, developments in methodology and software now enable us to fit more natural models for this type of data; making simplifying assumptions unnecessary. Ordinal models of this type are essentially straightforward extensions of models for nominal data, such as logistic regression.

References

Agresti, A. (1984). Analysis of ordinal categorical data. New York: John Wiley and Sons.

Beck, N., & Jackman, S. (1998). Beyond linearity by default: Generalized additive models. American Journal of Political Science, 42(2), 596-627. doi: 10.2307/2991772

Kuha, J. (2004). AIC and BIC - comparisons of assumptions and performance. Sociological Methods & Research, 33(2), 188-229. doi: 10.1177/0049124103262065

McCullagh, P. (1980). Regression-models for ordinal data. Journal of the Royal Statistical Society Series B-Methodological, 42(2), 109-142.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. Advances in Health Sciences Education, 15(5), 625-632. doi: 10.1007/s10459-010-9222-y