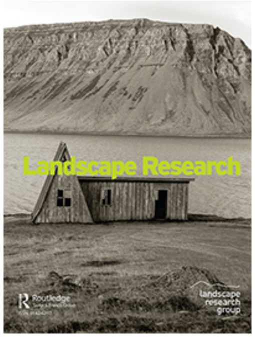

Landscape Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/clar20

Explaining landscape preference heterogeneity using machine learning-based survey analysis

Xiaozi Liu, Endre Tvinnereim, Kristine M. Grimsrud, Henrik Lindhjem, Liv Guri Velle, Heidi Iren Saure & Hanna Lee

To cite this article: Xiaozi Liu, Endre Tvinnereim, Kristine M. Grimsrud, Henrik Lindhjem, Liv Guri Velle, Heidi Iren Saure & Hanna Lee (2021) Explaining landscape preference heterogeneity using machine learning-based survey analysis, Landscape Research, 46:3, 417-434, DOI: 10.1080/01426397.2020.1867713

To link to this article: https://doi.org/10.1080/01426397.2020.1867713

Published online: 19 Jan 2021.

LANDSCAPE RESEARCH
2021, VOL. 46, NO. 3, 417–434
https://doi.org/10.1080/01426397.2020.1867713

landscape research group
Routledge Taylor & Francis Group

Explaining landscape preference heterogeneity using machine learning-based survey analysis

Xiaozi Liu [a], Endre Tvinnereim [a,b], Kristine M. Grimsrud [c], Henrik Lindhjem [d,e], Liv Guri Velle [f], Heidi Iren Saure [g] and Hanna Lee [h]

[a]Social science department, NORCE Norwegian Research Centre, Bergen, Norway; [b]Department of Administration and Organization Theory, University of Bergen, Bergen, Norway; [c]Research department, Statistics Norway, Oslo, Norway; [d]Menon Centre for Environmental and Resource Economics, Oslo, Norway; [e]Norwegian Institute for Nature Research (NINA), Oslo, Norway; [f]Møreforsking, Alesund, Norway; [g]Department of Teacher Education, NLA Høgskolen, Bergen, Norway; [h]Climate department, NORCE Norwegian Research Centre, Bergen, Norway

ABSTRACT
We conducted a national survey on a high-quality internet panel to study landscape preferences in Norway, using photos as stimuli. We examined preference heterogeneity with respect to socio-demographic characteristics and latent topics brought up by the respondents, using ordinal logistic regression and structural topic modelling (STM), a machine learning-based analysis. We found that pasture landscapes are the most favoured (55%), while densely planted spruce forests are the least favoured (8%). The contrast was particularly strong between eastern and western Norway, between men and women, and between young and old. STM revealed that the choices were mainly driven by the preference for landscape openness, especially by women. Other important drivers were concerns regarding reforestation of former farmlands, aesthetic properties, forest management, biodiversity issues, and cultural values. Our results suggest that landscape policies may clash with socio-cultural preferences, and failure to account for these may undermine the success of a policy.

KEYWORDS
Landscape preference; preference heterogeneity; ordinal logit regression; machine learning; structural topic modelling

1. Introduction

Landscapes are not only the natural heritage of a society but also a part of the cultural heritage, due to the intimate and evolving relationships between humans and their natural environments. There is a consensus among theorists that landscape aesthetics have two important components, a biological one and a cultural one (Bourassa, 1992; Tveit et al., 2006; Zube et al., 1982). The biological landscape preference is genetic, triggered by the common biological need of humans to adapt and survive in a natural environment (Arnberger & Eder, 2011). This may explain a general preference for specific landscapes across population groups, e.g., a common desire for a landscape containing water. The cultural mode of landscape aesthetics is a learned factor, shaped as it is by the social and cultural traits of individuals and their past experiences, and is hence more varied. Under the circumstances, landscape sometimes becomes a symbol or identity for cultural groups (Jones, 2008). When a particular landscape is perceived as cultural or place identity, any threat to that landscape will be seen as a threat to the associated identity (Devine-Wright, 2009). Rapid economic development and urbanisation profoundly impact landscape aesthetics and functionality. If a

CONTACT Xiaozi Liu Xiaozi.liu@hi.no

landscape becomes modified in a way that disrupts cultural or place identity, mounting social dismay and discontent may result.

The ongoing debate on how to manage the vanishing pasture landscapes in Norway reflects this concern. About one-third (8500 km^2) of the former farmlands in Norway are currently been abandoned, and the Norwegian government has proposed planting trees in the abandoned areas in order to meet carbon emission reduction targets (Haugland et al., 2013). The afforestation plan involves extensive planting of spruce forest along the coastline of western Norway (Søgaard et al., 2019). The decision has met with resistance and sparked heated debate (Iversen et al., 2019). For many people from western Norway, the open and almost treeless landscape, developed and maintained through grazing by domestic animals and burning of heathland since ancient times (Hjelle et al., 2010), carries unique cultural connotations and defines the region. The afforestation plan may cause a permanent change in their cultural landscape. The example of a landscape policy clashing with traditional socio-cultural values is not unique to Norway. Unwanted landscape changes may take several forms, driven by anthropogenic or natural causes, including reforested nature due to the progressive decline and abandonment of traditional farming practices (MacDonald et al., 2000) and the intrusions of silviculture, forest harvesting (Hull et al., 2001) and energy infrastructures (Jefferson, 2018) into the rural landscapes.

How do citizens relate to changes in the landscape arising from economic development, population growth, and climate change? How can the potential negative effects be mitigated? To answer these questions, it is necessary to map public landscape preferences and identify the potential conflicts. The goal of this paper is twofold: to elicit the landscape preferences of the Norwegian public in connection to the controversial afforestation plan, and to investigate the potential sources of preference heterogeneity. We achieve the goal by applying a new approach that combines a survey experiment with a machine learning-based textual data analysis. While the contribution of our study to the landscape literature is mainly methodological, knowledge of preference heterogeneity also has relevance for further discussion of afforestation policies.

Landscape planning and management have traditionally been expert driven, with trained experts interpreting the landscape (Daniel, 2001; Panagopoulos, 2009). There are concerns that expert approaches may be suitable for preserving landscapes with remarkable features, but less so for preserving ordinary landscapes, such as cultural landscapes (Scott, 2002), because the landscape preferences of expert groups and the general public may differ (Tveit, 2009). Although new approaches, such as the public participation GIS approach, as demonstrated by Brown and Hausner (2017) and Zoderer et al. (2019), have been developed to integrate public opinions into GIS-based landscape planning, the scope of public outreach for this method remains limited.

By contrast, research on landscape quality assessment has been dominated by perception-based approaches that emphasise the cultural component of human views (Daniel, 2001). One strand of the literature adopts empirical approaches to examine visual landscape preferences across different socio-economic groups, e.g., Sayadi et al. (2009) and Van Zanten et al. (2016) on agricultural landscapes, Gundersen et al. (2016, a review) on boreal forests, and Iversen et al. (2019) on climate forest. These studies use survey instruments (e.g., choice experiments) and discrete choice models (e.g., mixed logit models) as analytical tools. These models may identify heterogeneity arising from known (observed) classes, such as gender, education and other demographic variables, but not the heterogeneity of unobserved traits (e.g., personal identity). Latent class analysis (LCA) has been proposed to resolve this limitation (Arnberger & Eder, 2011; Häfner et al., 2018). LCA detects the extent to which respondents agree. LCA models capture preference heterogeneity by probabilistically assigning individual respondents to behavioural groups or latent (i.e., unobserved) segments as a function of the respondent's characteristics (Colombo et al., 2009). The use of LCA in the landscape preference literature is, to our knowledge, still limited.

Our study demonstrates an alternative method, termed structural topic modelling (STM), a machine learning-based qualitative and quantitative analysis of textual data (Roberts et al., 2016), designed to identify heterogeneous landscape preferences. Respondents are classified into different topic or preference groups on the basis of their text responses and socio-demographic characteristics. STM, which originated in the political science literature (Roberts et al., 2014), has been expanded to other research fields including survey experiments (e.g., Tvinnereim & Fløttum, 2015). We demonstrate how open-ended questions can be integrated into a landscape survey, and how qualitative textual data can be analysed quantitatively (and qualitatively) to reveal further insights about landscape preferences[1]. To our knowledge, this is the first study that has used STM for landscape preferences.

Landscape preferences in Norway have previously been examined at the local level (e.g., Strumse, 1996). In contrast to our study, which explores the preference heterogeneity of individual respondents arising from known or unknown socio-demographic traits, previous studies have focused on specific factors that affect landscape preferences, including land tenure or ownership (Hausner et al., 2015), environmental value orientation (Kaltenborn & Bjerke, 2002) and landscape composition (Dramstad et al., 2006). Our study utilised a national web-based https://www.uib.no/en/digsscore/122111/norwegian-citizen-panel Norwegian Citizen Panel (Norsk Medborgerpanel) to conduct the survey. The non-profit panel is run by the University of Bergen for the purpose of studying public opinion. This national study complements existing studies well, allowing us to compare results derived from different settings.

2. Materials and methods

2.1. Survey design

We presented the respondents three sets of landscape photos, each of which represented a specific strategy to manage abandoned agricultural lands (see Table 1). Alternative *A* represents a business-as-usual scenario, where the abandoned farmlands will be gradually reforested by nature; *B* requires effort to maintain the productivity of the existing farmlands; alternative *C* corresponds to the proposed afforestation plan for climate change mitigation (Haugland et al., 2013).

Table 1. Management alternatives and the represented landscapes.

Management alternatives	Landscapes	
	Photo ID *	Label
(A) Natural succession of abandoned agricultural lands	1	Grassland
	2	Heathland
(B) Continued management through grazing	3	Young mixed
	4	Old mixed
(C) Planting of spruce forest	5	Young planted
	6	Old planted

Notes: "*" Photos are shown in Figure 1.

Each landscape was represented by two photos depicting alternative subtypes: heathland vs. grassland in the case of open landscapes, and young vs. old forest for the forest landscapes. Care was taken to make the photos as similar as possible in dimensions other than the type of landscape itself. In particular, we sought photos with a minimal presence of water or grazing animals, taken in the same season and under the same weather conditions. At the same time, there was no way to keep lighting constant, as a dense spruce forest will necessarily let in less daylight than a pasture. Thus, while seeking to minimise confounding variables, our use of natural and non-manipulated landscapes makes our study quasi-experimental rather than fully experimental.

There were in total eight combinations of photo pairs, and each respondent was randomly assigned three alternatives. Respondents were shown one pair of photos at a time without receiving any additional information. Specifically, they did not observe the descriptions about management alternatives nor photo labels in Table 1. The photos were accompanied by the

Figure 1. The landscape photos used in the survey.

following two simple survey questions presented in Norwegian (translated here to English). The first question is:

1. Which of the two landscapes in the photos below do you like best:
- Like landscape *a* best
- Like landscape *b* best
- Neutral/don't know

The second question is an open-ended question. The generated textual data will be analysed by structural topic modelling (STM):

2. You can give the reasons for your choice here if you wish?

2.2. Survey panel and data

Our survey was distributed to the Norwegian Citizen Panel (NCP), an online, probability-sample infrastructure for studying public opinion in Norway. The non-profit panel is run by the University of Bergen. Owing to the high response rate and fully randomised recruitment of the respondents[2], many researchers considered NCP as high-quality survey panel. However, NCP shares a problem with other panels in Norway, that is, individuals with higher education are somewhat over-represented (Skjervheim et al., 2018).

The data for the current study were taken from Wave 12 of the NCP panel, collected in June 2018 (Ivarsflaten et al., 2018). The survey response rate, defined by total responses divided by active panel members, was 70.4%. Of 1,452 respondents selected for our landscape survey, only four did not answer. The sampling frame of the survey was the Norwegian population above the age of 18. The names of potential participants were drawn at random from the Norwegian population registry, and they received a log-in code via the mail. Respondents may choose to answer the survey using a computer, a tablet, or a mobile phone.

The survey generated a total of 4,344 answers (three per respondent). Nearly 60% of the respondents also provided text to explain the reason for their choice at least once. Descriptive statistics for the dataset are shown in Table 2. The respondents' profiles, including those responsive and unresponsive to the open-ended text question, are summarised in the appendix Table A1. An interesting observation is that those who have answered the open-ended text question tended to be female, older, with higher education, lower reported income, and more often living in northern and western Norway.

We collected additional county-specific data on the area covered by spruce plantations in the period 1957–1967 in order to investigate the potential linkages between the historical plantation programme and region-specific landscape preferences. During this period, approximately 19,000 ha of spontaneous pine in the coastal counties of Norway were converted into timber-producing spruce. The spruce species planted include Norway spruce (*Picea abies*) and Sitka spruce (*P. sitchensis*) (Statistics Norway, 1971). This conversion was not made in eastern Norway. Existing evidence suggests that past negative experiences with Sitka spruce that spread seeds beyond the plantation area may affect overall preferences for planted spruce (Nygård & Øyen, 2017; Saure et al., 2013).

2.3. Quantitative and qualitative data analysis

The analysis involved applying an ordinal logistic regression model to the landscape photo choices made by the respondents, followed by structural topic modelling (STM) of the responses to the open-ended text questions in order to reveal their underlying motives.

Table 2. Descriptive statistics.

	Dislike	Neutral	Like	Total
Photo.a[a]				
Grassland	149 (8.3%)	163 (18.7%)	412 (24.6%)	724 (16.7%)
Heathland	130 (7.2%)	176 (20.2%)	371 (22.1%)	677 (15.6%)
Old mixed	201 (11.2%)	182 (20.9%)	340 (20.3%)	723 (16.6%)
Old planted	285 (15.9%)	139 (15.9%)	327 (19.5%)	751 (17.3%)
Young mixed	402 (22.4%)	136 (15.6%)	169 (10.1%)	707 (16.3%)
Young planted	628 (35.0%)	76 (8.7%)	58 (3.5%)	762 (17.5%)
N (observations)	1795	872	1677	4344
Photo.b[a]				
Grassland	128 (7.6%)	164 (18.8%)	432 (24.1%)	724 (16.7%)
Heathland	158 (9.4%)	166 (19.0%)	447 (24.9%)	771 (17.7%)
Old mixed	202 (12.0%)	151 (17.3%)	372 (20.7%)	725 (16.7%)
Old planted	254 (15.1%)	160 (18.3%)	283 (15.8%)	697 (16.0%)
Young mixed	397 (23.7%)	144 (16.5%)	200 (11.1%)	741 (17.1%)
Young planted	538 (32.1%)	87 (10.0%)	61 (3.4%)	686 (15.8%)
N (observations)	1677	872	1795	4344
Spruce share[b]				
Mean (SD)	0.060 (0.092)	0.049 (0.087)	0.055 (0.090)	0.056 (0.090)
Range	0.000–0.308	0.000–0.308	0.000–0.308	0.000–0.308

[a]Photo.a and photo.b are presented simultaneously to respondents.
[b]The share of spruce is region specific, indicating the proportion of nation-wide spruce planting that took place in a region in the 1950s-1960s.

2.3.1. Ordinal logistic regression

The dependent variable of the survey is ordered at three levels: dislike, neutral and like. This calls for an ordinal logit model. We used a mixed-effect model to account for two sources of randomness: (1) Respondents made three repeated choices (one from each photo pair). The choices made by the same individual might be correlated; (2) the pairing of photos is random, but preferring a landscape photo over another photo also depends on what is shown in the other photo. We, therefore, decided to treat both respondent ID and the paired photo (i.e. photo.b) as random variables.

We fitted the data using R with 'clmm2' function in the ordinal package (Christensen, 2019), which supports both an ordinal dependent variable and random effects. The function is based on cumulative link mixed models (CLMM) and parameters are derived from maximum likelihood estimations. The error distribution of the mixed-effect model is complex; we derived confidence intervals for the estimates using the bootstrap re-sampling technique with 1000 replications. We selected the best-fit model using likelihood ratio tests. Detailed model selection procedure is listed in Table A2 in the appendix.

2.3.2. Textual data analysis

Before analysing the textual data from the open-ended question, we processed the data as follows: (i) textual responses made by the same respondents were combined; (ii) text in Nynorsk was translated into Bokmål.[3] The Nynorsk–Bokmål language pair of Apertium (Forcada et al., 2011; Velldal et al., 2017), an open-source machine translation system, was used for the conversion; (iii) the Bokmål language model of UDPipe (Straka et al., 2016) and the R package quanteda (Benoit et al., 2018) were used for tokenisation (lexical analysis) and lemmatisation[4]; the latter was verified manually. The final dataset contained a text corpus of 805 documents (i.e. text responses) and 1159 unique words/terms.

Structural topic modelling (STM) was then used to identify the most prevalent topics, which are interpreted as factors of concern for respondents in this study. STM is unsupervised machine learning that uses statistical models of text for measurement and inference. STM outperforms many other topic models by incorporating document-level metadata, in our case information about respondents, into the topic model (Roberts et al., 2016).

Researchers need to define a data generation process for STM (e.g., the number of topics and covariates) and use the text data to find the most likely values for model parameters. An STM model has two key conceptual components: a topic prevalence model that assigns words to topics as a function of covariates, and a topical content model that controls word frequency in each topic as a function of covariates. Each textual response is composed of multiple topics and the sum of topic proportions across all topics is one (Roberts et al., 2016).

We formulated an STM model that allows interactions between content covariates and latent topics. Following statistical relationships shown in the best-fit ordinal logistic model (see Section 2.3), we treated age category, region, and gender as covariates for the STM. The topic prevalence model used in estimating the STM was[5]:

$$Prevalence \sim region + age + region \times age + gender. \tag{1}$$

We used the function *searchK*, a data-driven STM approach, to select the potential number of topics. The final decision was based on (a) the trade-offs between semantic coherence and the exclusivity of a word to topics; (b) qualitative readings of words and example documents that were estimated by STM to be strongly associated with the topics (see online supplement). Using this evaluation, we selected the STM model with six topics as the final model for subsequent analyses and assigned a text label to each topic. Finally, we estimated the statistical relationship between topic prevalence (proportions) and covariates using a similar relationship as shown in Equation 1.

3. Results

We report the results based on the model best fit for ordinal logistic model (see Table A2 in the appendix for model selection procedures). The best-fit model shows that region, gender and age are the statically significant covariates, whereas education and income are not. These statistical relationships are utilised in the STM analysis to derive the most prevalent topics.

3.1. Rating landscape photos

The overall result of our photo-based survey is that the Norwegian public has clear landscape preferences. Table 3 shows the predicted probability of each landscape being favoured or disfavoured in relation to any other landscape photos. The order of preference from high to low is ranked as follows: heathland/grassland, old mixed, old planted, young mixed, and young planted. The difference between heathland and grassland is statistically insignificant (Table A3 in the appendix), indicating that the Norwegian public is indifferent with respect to these two.

The probability gap between the most favoured and the least favoured landscape photo is large: the 'dislike' probability is 80% versus 20% for pasture landscapes and young planted spruce; the 'like' probability between the two is 55% versus 8%.

3.2. Preference heterogeneity

Our analysis reveals a clear heterogeneity in landscape preferences across certain socio-demographic groups. As suggested by the likelihood ratio tests (reported in Table A2 in Section 2), the variations of income, education, and current residence (rural or city) do not bear much explanatory power. By contrast, demographic variables such as region, age, and gender are statistically significant. We summarise these effects under two headings: regional disparity and generational shift (Figure 2).

Regional disparity is particularly evident in the two pastoral landscapes (grassland and heathland, photos 3 & 4 in Figure 1), which were more favoured in the western than in the eastern regions. The effect is consistent across all age cohorts but more pronounced among the older and female cohorts; for instance, western Norwegian women over the age of 60 are 10% more likely to choose pasture photos than their counterparts from eastern Norway[6]. Regional disparity is also evident in the preference for the old spruce photo (photo 6 in Figure 1). The probability of disliking old spruce is higher than the probability of liking it in western and northern Norway, but the opposite is true in eastern Norway. This is probably because spruce is native or naturally spread in eastern Norway but considered introduced and partly invasive in western Norway (further discussion in the next section).

The generational shift refers to changes in landscape preferences within the same region. While the probability of older generations (both male and female) from western Norway who dislike the old spruce (48%) is higher than the probability of their liking it (29%) (Figure 2(a)), the preference is the reverse for the younger respondents. For the youngest cohort ($<$ 30 years) from the western regions, the mean probability of their favouring old spruce is 16 percentage points greater than their

Table 3. Predicted probability of landscape photo preference.

Photo ID	Landscape	Probability		
		Dislike	Neutral	Like
3	Grassland	0.206	0.241	0.553
4	Heathland	0.210	0.243	0.547
2	Old mixed	0.274	0.267	0.459
6	Old planted	0.321	0.275	0.404
1	Young mixed	0.506	0.256	0.238
5	Young planted	0.789	0.132	0.079

Note: The prediction is based on the statistical outputs of model 1 in Table A2.

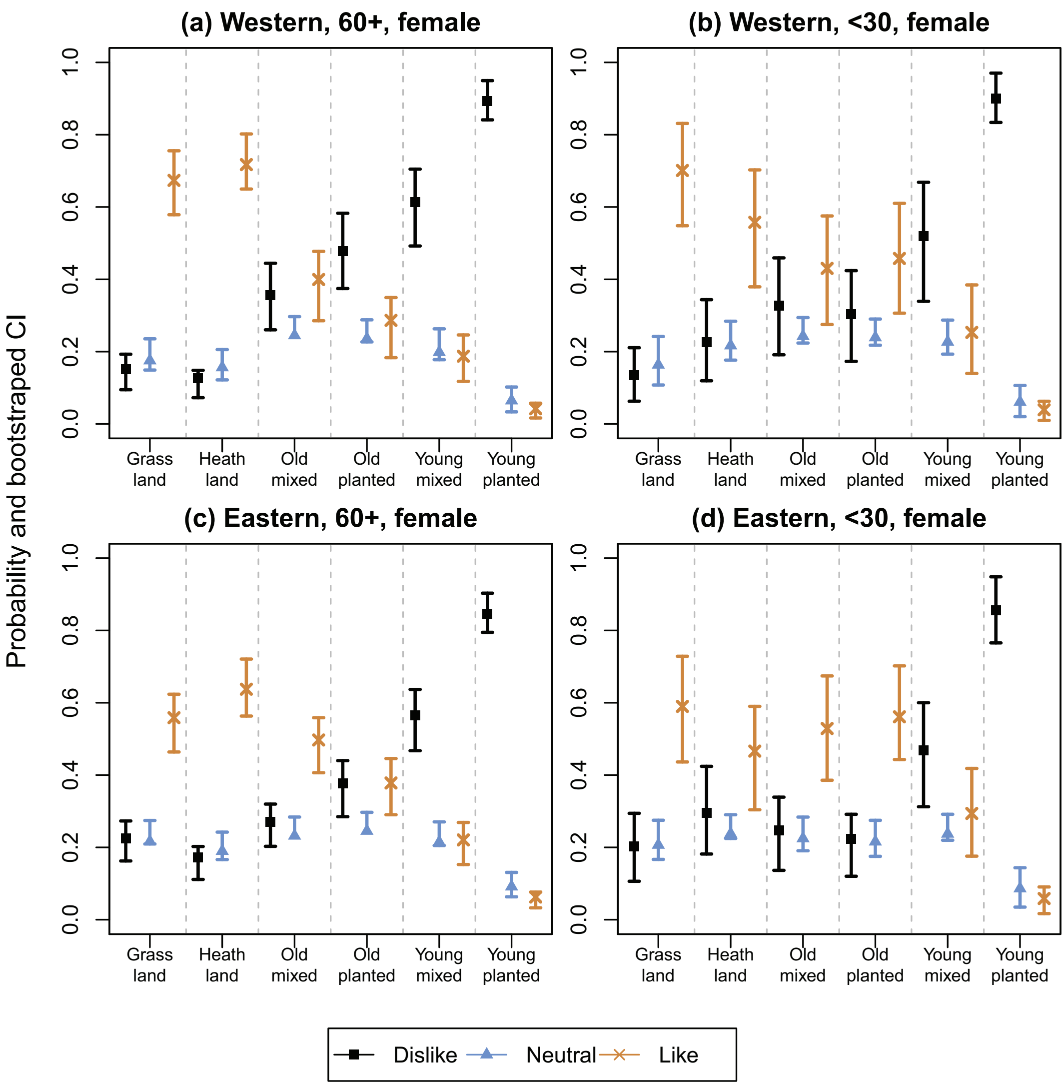

Figure 2. Preference heterogeneity by region, age and gender. The mean probabilities are predicted by the full model (i.e. model 8 in Table A2). Error bars refer to 95% bootstrapped confidence intervals.

disfavouring it (46% like vs. 30% dislike in Figure 2(b)), but is still lower than for their eastern counterparts. The eastern region is under a similar transformation: whereas the 60+ age cohort from eastern Norway shares a neutral view of old spruce (evidenced by the overlapped error bars for dislike and like in Figure 2(c)), the youngest cohort (< 30 years) shows a clear preference for old spruce (mean probability of 56% like versus 22% dislike, and statistically significant; Figure 2(d)). This suggests that landscape preferences across the country are becoming more homogeneous with time.

3.3. The spruce effect

To shed further light on the observed regional disparity and generational shift of landscape preferences, we replaced the covariate 'region' with region-specific shares of historical spruce plantation in the 1950s–1960s in the best-fit model (model 8 in Table A2). Note that large-scale

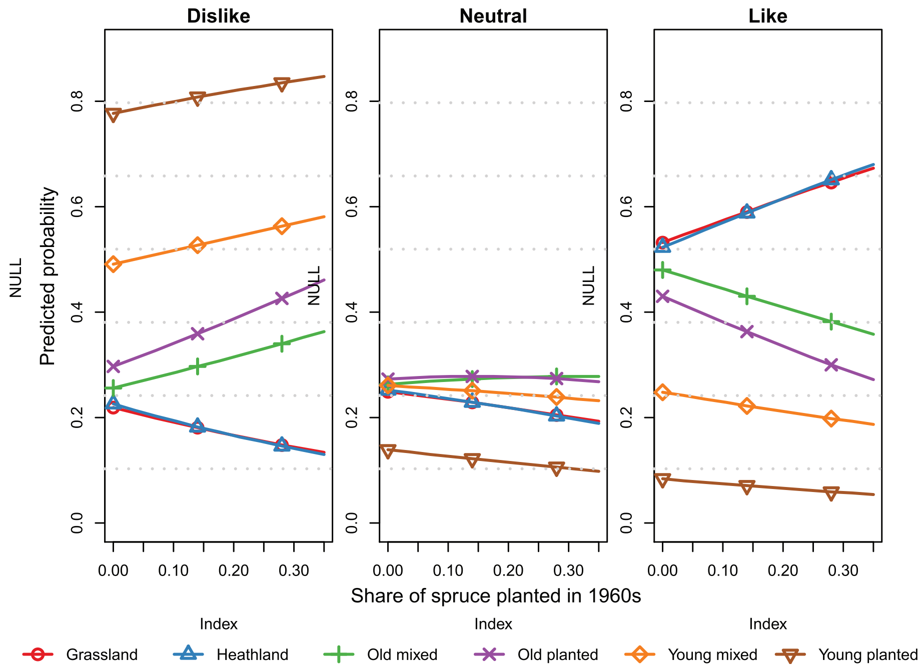

Figure 3. Evidence that historical spruce plantations in the 1950s–1960s may affect the landscape preferences of the respondents. Probabilities are predicted by the model that includes spruce share, photo and their interactions as independent variables and the pair photo as a random effect.

spruce planting took place only in the western part of Norway. The results are displayed in Figure 3: the share of spruce planted in a region is positively correlated (Table A3 in the appendix) with the probability of pasture landscape photos being selected as favourite. However, it reduces the probability of any other landscapes being selected, and the photo representing the old spruce forest experienced the sharpest decline. The spruce effect on the preference for the young planted spruce photo is also statistically significant, albeit to a lesser extent. This indicates that other factors such as the light in the photo or landscape may play a role in respondents' landscape preferences.

3.4. Evidence from textual data

The open-ended text responses allow us to clarify further what drives the respondents' preferences. We show (Table A4 in the appendix) that the words most frequently brought up are related to the openness of the landscape (e.g., *open, bright*), cultural values (e.g., *cultural landscape, graze*), aesthetic characteristics (e.g., *see, pretty, beautiful, lush, green*), and biodiversity (e.g., *diversity, vary, animal, tree, and forest floor*). The STM model has induced the six most prevalent topics, or factors of concern to the Norwegian public, that are strongly associated with these frequently occurring words in Table 4.

We label these topics in descending order of prevalence as follows (topic proportions are in parentheses): openness (0.30), overgrowth (0.17), aesthetics (0.15), biodiversity (0.14), forest (0.13), and culture (0.12). Topic labels are described below:

- Openness (Topic 4): emphasis on topographic openness and other related characteristics such as being airy, bright and good for viewing.
- Aesthetics (Topic 3): on the aesthetic properties of a landscape, such as the general impressions related to compositions and colours, and the feelings it has triggered (e.g., refreshing, lively and natural).

Table 4. Words with highest probability by induced topic with suggested labels and topic proportion. English translation in parentheses.

Topic	Words with highest prob.	Label	Proportion
1	kulturlandskap/viktig/vegetasjon (cultural landscape/important/vegetation)	Culture	0.12
2	liten/mangfold/lett (a little/diversity/easy)	Biodiversity	0.14
3	fin/frisk/tre (nice/fresh/tree)	Aesthetics	0.15
4	åpen/grønn/lys (open/green/bright)	Openness	0.30
5	frodig/natur/gjengro (lush/nature/overgrown)	Overgrowth	0.17
6	skog/naturlig/vakker (forest/natural/beautiful)	Forest	0.13

Topic 4 is negatively correlated with topic 1 (−0.7) and topic 6 (−0.8); Topic 1 is positively correlated with topic 2 (0.6) and 6 (0.5). These correlations have been accounted in the analysis.

- Overgrowth (Topic 5): concerning the natural succession and reforestation of the former farmlands.
- Forest (Topic 6): on the forest in general and the need for forest management and interventions; e.g., maintaining a certain distance between the forest and houses, pruning and thinning of branches, and reducing the density of planted forest.
- Biodiversity (Topic 2): on the species richness of a landscape, e.g., the complaints about mono-culture planting and a lack of life on the forest floor, and the comments on the diverse species and habitats of traditional grazing lands.
- Culture (Topic 1): emphasis on the cultural values of a landscape; some referred to spruce forest as cultural landscape, some attached cultural value to pastures maintained by domestic animal grazing.

To arrive at an answer as to how the prevalence of these topics varies with the socio-demographic group, we examined topic-specific marginal effects of covariates, namely gender, region and age (Figure 4). Gender has the most pronounced effect: women placed more emphasis than men on the openness and aesthetics of the landscapes, and men placed more emphasis on the need for forest management, the problem of the overgrowing of the former farmlands and cultural landscapes (Figure 4(a)). The result is consistent across regions and age cohorts (see Table A5 in the appendix). The effects of region and age only become statistically significant when the two variables interact; e.g., the factors of concern to women aged under 60 (Figure 4(b)), are similar in both coastal and non-coastal regions. For women from coastal areas (Figure 4(c)), the older population (aged 60 and above) is more concerned about biodiversity, whereas the younger cohorts (under 60) are more concerned about the problem of overgrown nature.

4. Discussion and conclusions

We applied a new approach to study landscape preference heterogeneity of the Norwegian public. We found pastoral landscapes to be the most favoured and densely planted spruce forests the least favoured. The contrasts between eastern and western Norway, between men and women, and between young and old, were particularly strong.

This study confirms some old wisdom but also reveals some new insights. The negative sentiments of the elderly from the coastal areas towards spruce landscapes appeared to be associated with historical plantations of spruce, suggesting that a preference can be shaped by the 'reputation' of a species (in this case the invasive Sitka spruce) (Kueffer & Kull, 2017). However, with the passage of time, this sentiment is gradually softening. The heterogeneity of the preferences between young and old respondents reflects the evolving nature of the cultural landscape (Dien, 2000). Some researchers advocated for targeted education programmes to raise awareness of historical landscapes (Tempesta,

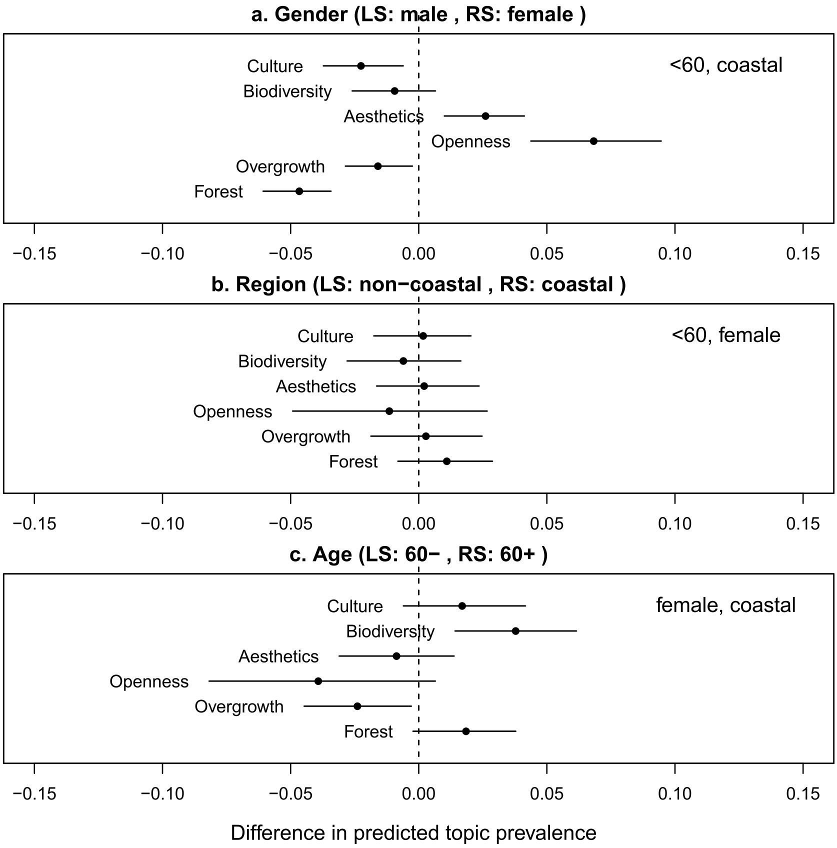

Figure 4. Graphical display of topical prevalence contrast. The text labels refer to the six most prevalent topics. Topics on the right of the zero line (RS) are more likely to be brought up by women (a) or those from coastal areas (b), or those aged 60 and above (c). The levels of other co-variates are indicated at the top right corner of each panel. LS and RS refer to left- and right- sides, respectively, of the zero line. Bars indicate the 95% confidence intervals.

2010). The structural topic modelling analysis applied in this study has revealed new insights about gender differences: female respondents were more likely to bring up topics such as the topographic openness and aesthetic properties of a landscape, while men were more likely to emphasise the need for forest management, reforestation of the former farmlands and the cultural values of the landscape.

Our findings are broadly in line with the earlier studies. Culture differences between Norway and Poland explained differences in management preferences for protected areas (Brown et al., 2015). The forest landscape preferences among Nordic people were closely linked to place identity and stewardship (Gundersen et al., 2016), and afforestation on the former agricultural lands was viewed more negatively than the afforestation within established forests (Gundersen & Frivold, 2008). With respect to coastal landscapes, old Norwegians were more attached to traditional grazing practices than the younger ones (Strumse, 1996), and Norwegian residents valued the open landscapes more than foreign tourists (Jacobsen & Tømmervik, 2016). This raises a question for management authorities: whom should we manage the landscapes for?

As a policy implication, we emphasise the importance of integrating public preferences into landscape planning and management. Previous studies have acknowledged that including citizens' opinion in landscape policy-making can enhance their trust in authorities as well as improve policy implementation (Jones, 2007). In this study, we demonstrated how public preferences can be mapped via a national landscape survey. Moreover, our study has revealed clear heterogeneity of landscape preferences in Norway. This information is valuable. Extensive tree planting on abandoned farmlands, as has been proposed by the Norwegian government with a view to reduce greenhouse gases, may have starkly contrasting local receptions. Failure to take account of preference differences may lead to a clash between climate policy and traditional socio-cultural values.

Several related questions following this study are worthy further investigation: (i) how are the past policies and practices influencing current landscape preferences? (ii) to what extent are Norwegian citizens willing to accept a trade-off between planting spruce to offset carbon emissions and preserving traditional agricultural landscapes, and (iii) how do stated landscape preferences change when key elements are altered in a fully experimental setting? The answers to these questions are critical for designing sustainable afforestation policies that achieve climate targets.

Notes

1. For a complete list of STM publications, please refer to https://www.structuraltopicmodel.com/.
2. DIGSSCORE status report, https://www.uib.no/en/digsscore/132816/digsscore-status-report
3. The textual data consist of a mixture of Nynorsk and Bokmål, the two official standards for written Norwegian. It is reported that approximately 8% of the population use Nynorsk as the main form of Norwegian.
4. Lemmatisation removes inflectional endings only and returns the base or dictionary form of a word, which is known as the lemma.
5. Some simplifications were made to enhance model performance. This includes reducing the number of interaction terms and the number of categorical levels for age and region.
6. The mean probabilities are 12% for grassland and 8% for heathland, respectively.

Acknowledgments

The manuscript has benefited from comments by Ole Martin Lægreid, Elisabeth Ivarsflaten, Mikko Heino and Knut Grove on various occasions. We would like to thank Emma Waterton, the editor of *Landscape Research*, and two anonymous reviewers for their suggestions and comments. We also acknowledge the language editing by Beverley Wahl.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Research Council of Norway [268243 and 280393].

Notes on contributors

Xiaozi Liu is an environmental and resource economist at NORCE Norwegian Research Centre and Institute of Marine Research in Norway. She is specialized in applied game theory, fisheries economics and discrete choice analysis.

Endre Tvinnereim is Associate Professor of political science at the University of Bergen and Research Professor at NORCE Norwegian Research Centre. His main research interests are comparative public opinion on climate policies, quantitative text analysis in survey research and evidence-based evaluations of cap-and-trade and other CO_2 pricing mechanisms.

Kristine M. Grimsrud is a senior economist at Statistics Norway, specialized in environmental economics and stated preference research.

Henrik Lindhjem is the head of research at Menon Centre for Environmental and Resource Economics and scientific advisor at the Norwegian Institute for Nature Research, with extensive experience in quantitative survey method in environmental economics.

Liv Guri Velle has a PhD in vegetation ecology, and is a researcher and research manager at Møreforsking. Her research interests are related to ecological dynamics and sustainable management regimes within the frames of land-use and climate change.

Heidi Iren Saure is Associate Professor at NLA College Bergen, Department of Teacher Education. Saure's research interests are within the fields of vegetation ecology and science education.

Hanna Lee is Research Professor at NORCE. She is a terrestrial ecosystem ecologist and climate modeller interested in the feedback cycles of ecosystem and climate change.

ORCID

Xiaozi Liu http://orcid.org/0000-0002-2299-8612
Endre Tvinnereim http://orcid.org/0000-0001-8891-2503
Kristine M. Grimsrud http://orcid.org/0000-0002-6580-1031
Henrik Lindhjem http://orcid.org/0000-0003-2206-3603
Liv Guri Velle http://orcid.org/0000-0003-0433-5337
Heidi Iren Saure http://orcid.org/0000-0002-3232-7564
Hanna Lee http://orcid.org/0000-0002-2003-4377

References

Arnberger, A., & Eder, R. (2011). Exploring the heterogeneity of rural landscape preferences: An image-based latent class approach. *Landscape Research*, *36*(1), 19–40. https://doi.org/10.1080/01426397.2010.536204

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Bourassa, S. C. (1992). The aesthetics of landscape. *Journal of Aesthetics and Art Criticism*, *50*(4), 343–345. https://doi.org/10.2307/431417

Brown, G., & Hausner, V. H. (2017, June). An empirical analysis of cultural ecosystem values in coastal landscapes. *Ocean & Coastal Management*, *142*, 49–60. https://doi.org/10.1016/j.ocecoaman.2017.03.019

Brown, G., Hausner, V. H., Grodzińska-Jurczak, M., Pietrzyk-Kaszyńska, A., Olszańska, A., Peek, B., Rechciński, M., & Lægreid, E. (2015, November). Cross-cultural values and management preferences in protected areas of Norway and Poland. *Journal for Nature Conservation*, *28*, 89–104. https://doi.org/10.1016/j.jnc.2015.09.006

Christensen, R. H. B. (2019). *Cumulative link models for ordinal regression with the R package ordinal*. https://cran.r-project.org/web/packages/ordinal/

Colombo, S., Hanley, N., & Louviere, J. (2009). Modeling preference heterogeneity in stated choice data: An analysis for public goods generated by agriculture. *Agricultural Economics*, *40*(3), 307–322. https://doi.org/10.1111/j.1574-0862.2009.00377.x

Daniel, T. C. (2001). Whither scenic beauty? Visual landscape quality assessment in the 21st century. *Landscape and Urban Planning*, *54*(1), 267–281. https://doi.org/10.1016/S0169-2046(01)00141-4

Devine-Wright, P. (2009). Rethinking NIMBYism: The role of place attachment and place identity in explaining place-protective action. *Journal of Community & Applied Social Psychology*, *19*(6), 426–441. https://doi.org/10.1002/casp.1004

Dien, D. S.-F. (2000). The evolving nature of self-identity across four levels of history. *Human Development*, *43*(1), 1–18. https://doi.org/10.1159/000022650

Dramstad, W. E., Tveit, M. S., Fjellstad, W., & Fry, G. L. A. (2006). Relationships between visual landscape preferences and map-based indicators of landscape structure. *Landscape and Urban Planning*, *78*(4), 465–474. https://doi.org/10.1016/j.landurbplan.2005.12.006

Forcada, M. L., Ginestí–Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J. A., Sánchez-Martínez, F., Ramírez-Sánchez, G., & Tyers, F. M. (2011). Apertium: A free/open-source platform for rule-based machine translation. *Machine Translation*, *25*(2), 127–144. https://doi.org/10.1007/s10590-011-9090-0

Gundersen, V., Clarke, N., Dramstad, W., & Fjellstad, W. (2016). Effects of bioenergy extraction on visual preferences in boreal forests: A review of surveys from Finland, Sweden and Norway. *Scandinavian Journal of Forest Research*, *31*(3), 323–334. https://doi.org/10.1080/02827581.2015.1099725

Gundersen, V. S., & Frivold, L. H. (2008). Public preferences for forest structures: A review of quantitative surveys from Finland, Norway and Sweden. *Urban Forestry & Urban Greening*, *7*(4), 241–258. https://doi.org/10.1016/j.ufug.2008.05.001

Häfner, K., Zasada, I., Zanten, B. T. V., Ungaro, F., Koetse, M., & Piorr, A. (2018). Assessing landscape preferences: A visual choice experiment in the agricultural region of Märkische Schweiz, Germany. *Landscape Research*, *43*(6), 846–861. https://doi.org/10.1080/01426397.2017.1386289

Haugland, H., Anfinnsen, B., Aasen, H., Løbersli, E., Selboe, O-K., Terum, T., Lileng, J., Granhus, A., Kjersti Holt Hanssen, K. S. (2013). *Planting forests on new areas as climate measures: Suitable areas and environmental criteria (in Norwegian)* (Technical Report M26-2013). Norwegian Environment Agency.

Hausner, V. H., Brown, G., & Lægreid, E. (2015, December). Effects of land tenure and protected areas on ecosystem services and land use preferences in Norway. *Land Use Policy*, *49*, 446–461. https://doi.org/10.1016/j.landusepol.2015.08.018

Hjelle, K. L., Halvorsen, L. S., & Overland, A. (2010). Heathland development and relationship between humans and environment along the coast of western Norway through time. *Quaternary International*, *220*(1), 133–146. https://doi.org/10.1016/j.quaint.2009.09.023

Hull, R. B., Robertson, D. P., & Kendra, A. (2001). Public understandings of nature: A case study of local knowledge about "natural" forest conditions. *Society & Natural Resources*, *14*(4), 325–340. https://doi.org/10.1080/08941920151080273

Ivarsflaten, E., Dahlberg, E., Arnesen, S. (2018). *Norwegian citizen panel 2018: Study documentation* (Technical Report NMP-18-2). Ideas2Evidence and University of Bergen.

Iversen, E. K., Lindhjem, H., Jacobsen, J. B., & Grimsrud, K. (2019). Moving (back) to greener pastures? Social benefits and costs of climate forest planting in norway. *Land Use Policy*, page 104390. https://doi.org/10.1016/j.landusepol.2019.104390

Jacobsen, J. K. S., & Tømmervik, H. (2016, July). Leisure traveller perceptions of iconic coastal and fjord countryside areas: Lush naturalness or remembrance of agricultural times past? *Land Use Policy*, *54*, 38–46. https://doi.org/10.1016/j.landusepol.2016.01.013

Jefferson, M. (2018, March). Safeguarding rural landscapes in the new era of energy transition to a low carbon future. *Energy Research & Social Science*, *37*, 191–197. https://doi.org/10.1016/j.erss.2017.10.005

Jones, M. (2007). The European landscape convention and the question of public participation. *Landscape Research*, *32*(5), 613–633. https://doi.org/10.1080/01426390701552753

Jones, M. (2008). The "two landscapes" of north Norway and the "cultural landscape" of the south. In M. Jones & O. R. Kenneth Eds., *Nordic landscapes: Region and belonging on the Northern Edge of Europe* (pp. 283–300). University of Minnesota Press. chapter 4.

Kaltenborn, B. P., & Bjerke, T. (2002). Associations between environmental value orientations and landscape preferences. *Landscape and Urban Planning*, *59*(1), 1–11. https://doi.org/10.1016/S0169-2046(01)00243-2

Kueffer, C., & Kull, C. A. (2017). Non-native species and the aesthetics of nature. *In Impact of Biological Invasions on Ecosystem Services*, Springer, *12*, 311–324. https://doi.org/10.1007/978-3-319-45121-3_20

MacDonald, D., Crabtree, J. R., Wiesinger, G., Dax, T., Stamou, N., Fleury, P., Gutierrez Lazpita, J., & Gibon, A. (2000). Agricultural abandonment in mountain areas of Europe: Environmental consequences and policy response. *Journal of Environmental Management*, *59*(1), 47–69. https://doi.org/10.1006/jema.1999.0335

Norway, S. (1971). *Census of forestry 1967* (Technical report).

Nygård, P. H., & Øyen, B.-H. (2017). Spread of the introduced Sitka spruce (Picea sitchensis) in coastal norway. *Forests*, *8*(1), 24. https://doi.org/10.3390/f8010024

Panagopoulos, T. (2009). Linking forestry, sustainability and aesthetics. *Ecological Economics*, *68*(10), 2485–2489. https://doi.org/10.1016/j.ecolecon.2009.05.006

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, *111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Saure, H. I., Vandvik, V., Hassel, K., & Vetaas, O. R. (2013). Effects of invasion by introduced versus native conifers on coastal heathland vegetation. *Journal of Vegetation Science*, *24*(4), 744–754. https://doi.org/10.1111/jvs.12010

Sayadi, S., González-Roa, M. C., & Calatrava-Requena, J. (2009). Public preferences for landscape features: The case of agricultural landscape in mountainous Mediterranean areas. *Land Use Policy*, *26*(2), 334–344. https://doi.org/10.1016/j.landusepol.2008.04.003

Scott, A. (2002). Assessing public perception of landscape: The LANDMAP experience. *Landscape Research*, *27*(3), 271–295. https://doi.org/10.1080/01426390220149520

Skjervheim, Ø., Høgestøl, A., & Bjørnebekk, O. (2018). *Norwegian Citizen Panel 2018, twelfth wave methodology report* (Technical report). Ideas2Evidence and University of Bergen.

Søgaard, G., Allen, M., Astrup, R., Belbo, H., Bergseng, E., Blom, H. H., Bergseng, E., Blom, H. H., Bright, R., Dalsgaard, L., Fernandez, C. A., Gjerde, I., Granhus, A., Hanssen, K. H., Kjønaas, O. J., Nygaard, P. H., Stokland, J., & Sætersdal, M. (2019). *Effects of planting forests on new areas*. Significance for climate.

Straka, M., Hajič, J., & Straková, J. (2016). UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In *Proceedings of the tenth international conference on language resources and evaluation (LREC'16)*, (pp. 4290–4297). Portorož, Slovenia: European Language Resources Association (ELRA), Portorož, Slovenia.

Strumse, E. (1996). Demographic differences in the visual preferences for agrarian landscapes in western Norway. *Journal of Environmental Psychology*, *16*(1), 17–31. https://doi.org/10.1006/jevp.1996.0002

Tempesta, T. (2010). The perception of agrarian historical landscapes: A study of the Veneto plain in Italy. *Landscape and Urban Planning*, *97*(4), 258–272. https://doi.org/10.1016/j.landurbplan.2010.06.010

Tveit, M., Ode, Å., & Fry, G. (2006). Key concepts in a framework for analysing visual landscape character. *Landscape Research*, *31*(3), 229–255. https://doi.org/10.1080/01426390600783269

Tveit, M. S. (2009). Indicators of visual scale as predictors of landscape preference; a comparison between groups. *Journal of Environmental Management*, *90*(9), 2882–2888. https://doi.org/10.1016/j.jenvman.2007.12.021

Tvinnereim, E., & Fløttum, K. (2015). Explaining topic prevalence in answers to open-ended survey questions about climate change. *Nature Climate Change*, *5*(8), 744. https://doi.org/10.1038/nclimate2663

van Zanten, B. T., Verburg, P. H., Scholte, S. S. K., & Tieskens, K. F. (2016, October). Using choice modeling to map aesthetic values at a landscape scale: Lessons from a Dutch case study. *Ecological Economics*, *130*, 221–231. https://doi.org/10.1016/j.ecolecon.2016.07.008

Velldal, E., Øvrelid, L., & Hohle, P. (2017). Joint UD parsing of Norwegian Bokmål and Nynorsk. In *Proceedings of the 21st Nordic conference of computational linguistics*, (pp. 10). Sweden: Gothenburg.

Zoderer, B. M., Tasser, E., Carver, S., & Tappeiner, U. (2019, November). An integrated method for the mapping of landscape preferences at the regional scale. *Ecological Indicators*, *106*, 105430. https://doi.org/10.1016/j.ecolind.2019.05.061

Zube, E. H., Sell, J. L., & Taylor, J. G. (1982). Landscape perception: Research, application and theory. *Landscape Planning*, *9*(1), 1–33. https://doi.org/10.1016/0304-3924(82)90009-0

Appendices

A1. Respondents' profile

A2. Model selection procedure and the best-fit model

Model selection was based on likelihood ratio tests. The detailed procedure is listed in Table A2. The model that allows the landscape photos to interact with region, gender and age produced the best fit (i.e., model 8 in Table A2). Adding demographic variables such as education, income and residence type did not improve the model fit. Likelihood ratio tests also suggested that accounting for randomly paired photos is sufficient to treat the aforementioned random effects (comparison of models 1 & 2 in Table A2).

A3. Estimates from ordinal logistic regression models

A4. Most frequent words

A5. Estimate effects by topics

Table A1. Respondents' profiles.

	Comment[a] (N = 820)	No comment (N = 628)	Total (N = 1448)
Region			
Eastern	168 (20.5%)	163 (26.0%)	331 (22.9%)
Northern	60 (7.3%)	37 (5.9%)	97 (6.7%)
Oslo/Askerhus	242 (29.5%)	190 (30.3%)	432 (29.8%)
Southern	37 (4.5%)	32 (5.1%)	69 (4.8%)
Trøndelag	68 (8.3%)	58 (9.2%)	126 (8.7%)
Western	245 (29.9%)	148 (23.6%)	393 (27.1%)
Gender			
Male	396 (48.3%)	329 (52.4%)	725 (50.1%)
Female	424 (51.7%)	299 (47.6%)	723 (49.9%)
Age category			
<30	69 (8.4%)	55 (8.8%)	124 (8.6%)
30–59	350 (42.7%)	332 (52.9%)	682 (47.1%)
60+	401 (48.9%)	241 (38.4%)	642 (44.3%)
Income (NOK)[b]			
N-Miss	74	72	146
0–150k	57 (7.6%)	38 (6.8%)	95 (7.3%)
150–300k	98 (13.1%)	69 (12.4%)	167 (12.8%)
300–400k	125 (16.8%)	73 (13.1%)	198 (15.2%)
400–500k	147 (19.7%)	101 (18.2%)	248 (19.0%)
500–600k	122 (16.4%)	95 (17.1%)	217 (16.7%)
600–700k	71 (9.5%)	58 (10.4%)	129 (9.9%)
700k-1000k	73 (9.8%)	70 (12.6%)	143 (11.0%)
1000k+	37 (5.0%)	35 (6.3%)	72 (5.5%)
Not answered	16 (2.1%)	17 (3.1%)	33 (2.5%)
Education level			
Elementary	51 (6.2%)	48 (7.6%)	99 (6.8%)
Upper secondary	223 (27.2%)	187 (29.8%)	410 (28.3%)
University	511 (62.3%)	372 (59.2%)	883 (61.0%)
Not answered	35 (4.3%)	21 (3.3%)	56 (3.9%)

[a]'Comment' and 'no comment' refer to respondents who answered and did not answer open-ended text questions respectively.
[b]1 NOK $\approx$ 0.10 USD.

Table A2. A brief summary of model selection procedure.

Model	Formula	Random	Test	Pr(>|z|)
0	Choice[a] ~ photo.a	NA		
1	Choice ~ photo.a	photo.b[b]	0 vs 1	0.0000
2	Choice ~ photo.a	ID[c], photo.b	1 vs 2	0.9971
3	Choice ~ photo.a *[d] region	photo.b	1 vs 3	0.0121
4	Choice ~ photo.a * age	photo.b	1 vs 4	0.0116
5	Choice ~ photo.a * gender	photo.b	1 vs 5	0.0295
6	Choice ~ photo.a * income	photo.b	1 vs 6	0.6618
7	Choice ~ photo.a * region + photo.a * gender	photo.b	2 vs 7	0.0299
			5 vs 7	0.0122
8	Choice ~ photo.a * region + photo.a * gender+ photo.a * age	photo.b	7 vs 8	0.0069

[a]Choice refers to a discrete choice (*dislike, neutral* and *like*) made for photo.a, or the photo on the left.
[b]Photo.b, photo on the right refers to the photo paired with photo.a.
[c]ID = respondent ID.
[d]All " * " in the table cover main effects and their interactions: e.g., Photo.a * region = Photo.a + region + photo.a $\times$ region.

Table A3. Ordinal regression models.

	Base model	Spruce model
photo.aHeathland	− 0.02	− 0.04
	(0.11)	(0.13)
photo.aOld mixed	− 0.38***	− 0.21
	(0.11)	(0.13)
photo.aOld planted	− 0.60***	− 0.41**
	(0.11)	(0.12)
photo.aYoung mixed	− 1.38***	− 1.23***
	(0.11)	(0.13)
photo.aYoung planted	− 2.67***	− 2.52***
	(0.13)	(0.14)
dislike|neutral	− 1.35***	− 1.27***
	(0.36)	(0.36)
neutral|like	− 0.21	− 0.13
	(0.36)	(0.36)
spruce.share		1.70
		(0.88)
photo.aHeathland:spruce.share		0.20
		(1.29)
photo.aOld mixed:spruce.share		− 3.14**
		(1.17)
photo.aOld planted:spruce.share		− 3.72**
		(1.19)
photo.aYoung mixed:spruce.share		− 2.73*
		(1.23)
photo.aYoung planted:spruce.share		− 3.03*
		(1.47)
Log Likelihood	−3760.39	−3750.00
AIC	7538.79	7529.99
BIC	7596.18	7625.64
Num. obs.	4344	4344
Groups (responseid)	1448	1448
Groups (photo.b)	6	6
Variance: responseid: (Intercept)	0.00	0.00
Variance: photo.b: (Intercept)	0.74	0.74

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A4. Top 30 most frequent words from open-ended text responses.

Rank	Norwegian	English	Frequency	Share
1	åpen	open	347	0.359
2	se	see	189	0.195
3	grønn	green	145	0.15
4	fin	pretty	135	0.14
5	frodig	lush	103	0.107
6	kulturlandskap	cultural landscape	98	0.101
7	frisk	fresh	86	0.089
8	natur	nature	85	0.088
9	lys	bright	78	0.081
10	liten	small	68	0.07
11	skog	forest	66	0.068
12	tre	tree	65	0.067
13	beite	graze	63	0.065
14	naturlig	natural	58	0.06
15	lyst	bright	55	0.057
16	gjengro	overgrow	54	0.056
17	mangfold	diversity	51	0.053
18	viktig	important	48	0.05
19	vakker	beautiful	45	0.047
20	tur	hike	44	0.046
21	gi	give	39	0.04
22	lett	easy	38	0.039
23	gro	grow	37	0.038
24	skogbunn	forst floor	37	0.038
25	variere	vary	37	0.038
26	dyr	animal	36	0.037
27	leve	live	36	0.037
28	fjell	mountain	36	0.037
29	farge	colour	36	0.037
30	flott	great	35	0.036

Table A5. Estimate the relationship between topic prevalence and covariates. The dependent variable is the topic proportion of each document assigned by the six-topic STM model[a].

Coefficients	Estimate	Std. Error	Pr(>|t|)	Sig.	Estimate	Std. Error	Pr(>|t|)	Sig.
Topic 1: Culture					*Topic 2: Biodiversity*			
(Intercept)	0.130	0.009	0.000	***	0.145	0.009	<2e-16	***
[a] coastal	0.002	0.010	0.855		−0.006	0.011	0.613	
[b]age60+	0.007	0.010	0.503		−0.006	0.009	0.517	
female	−0.022	0.008	0.004	**	−0.009	0.008	0.241	
coastal:age60+	0.010	0.013	0.452		0.044	0.016	0.005	**
Topic 3: Aesthetics					*Topic 4: Openness*			
(Intercept)	0.140	0.009	0.000	***	0.277	0.015	0.000	***
[b]coastal	0.002	0.010	0.813		−0.012	0.020	0.550	
[c]age60+	−0.010	0.009	0.276		−0.012	0.018	0.486	
genderfemale	0.026	0.008	0.001	**	0.068	0.013	0.000	***
coastal: age60+	0.001	0.013	0.917		−0.027	0.028	0.329	
Topic 5: Overgrown nature					*Topic 6: Forest*			
(Intercept)	0.167	0.008	0.000	***	0.142	0.007	0.000	***
coastal	0.002	0.011	0.820		0.011	0.009	0.249	
age60+	0.017	0.009	0.062	.	0.005	0.008	0.534	
female	−0.016	0.007	0.015	*	−0.047	0.007	0.000	***
coastal:age60+	−0.041	0.015	0.006	**	0.014	0.013	0.298	

Notes: Significance codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1.

[a]The model is estimated by function *estimateEffect* in STM. The estimation incorporates measurement uncertainty from the STM model (i.e. 'global').

[b]Region is aggregated into two levels with 'coastal' to represent western and northern Norway and 'non-coastal' for other regions.

[c]Age has two levels: "60 + " if aged 60 and above, and "60 − ".