Building and Environment 146 (2018) 258–267

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

A Bayesian method of evaluating discomfort due to glare: The effect of order bias from a large glare source

M.G. Kent[a,*], T. Cheung[b], S. Altomonte[c], S. Schiavon[d], A. Lipczyńska[b]

[a] Department of Architecture and Built Environment, University of Nottingham, Nottingham, UK
[b] Berkeley Education Alliance for Research in Singapore, Singapore
[c] Architecture et Climat, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
[d] Centre for the Built Environment, University of California, Berkeley, CA, USA

A R T I C L E I N F O

Keywords:
Bayesian method
Discomfort glare
Experimental bias
Order bias
Statistical analysis

A B S T R A C T

Replicating scientific findings is a fundamental aspect of research. However, in studies of discomfort due to glare, it is difficult to make comparisons between the results of different experiments since the statistical tests usually reported do not allow independent findings to be directly compared to each other. Here we present an alternative Bayesian approach that can address this problem. To show how this approach works, we performed a laboratory test with 55 participants to validate the effect of order bias previously detected in a similar study evaluating discomfort due to glare but, this time, under a large luminous source. Using the luminance adjustment procedure, the glare source was varied to meet four sensations of discomfort due to glare. Adjustments were performed under three different order sequences: ascending, descending, and randomised. Test participants provided glare settings using a newly proposed evaluation scale. The effect of order bias detected in the original study was compared to the data obtained with the same methodological procedure in the new experiment using Bayesian inferential tests. The results showed a close replication, highlighting that the order bias effect found in the original study was also present in the new experiment. The wide application of Bayesian methods in the design and analysis of experimental studies may improve the accuracy and validity of glare models.

1. Introduction

Discomfort due to glare is one of the challenges of building façade design. While studies have found that visual discomfort is a significant problem in many conventional buildings, occupants have reported glare five times more often in green-rated buildings [1]. A study based on 2540 occupant responses, collected from 11 countries and 36 different "sustainable" buildings, has also shown that glare – particularly from daylit windows – remains a pertinent issue [2]. To minimise the risk of glare, various models have been developed to provide precise measures of discomfort from a visual scene, with the objective of quantifying the perceived levels of glare based on physical measurements [3]. However, these models often give a low prediction accuracy [4]. Among many models recommended in the literature and in international standards, Table 1 presents a selection of key experimental studies used to derive prediction models of discomfort glare, also illustrating the subjective criteria that observers used to evaluate the glare sources.

From a methodological perspective, the studies presented in Table 1 [5–7] – together with many others – relied invariably on frequentist approaches (e.g., null hypothesis significance testing (NHST)) to analyse the predictive performance of the models proposed. However, NHST testing has several limitations, such as:

- Statistical significance is dependent on both the size of the sample and the magnitude of the effect, which cannot be measured using NHST alone [8,9]. This implies that, when large samples are used, statistically significant findings can be detected even though the magnitude of the effect is not practically relevant. For example, Altomonte and Schiavon [10] showed that even the smallest variations in occupant satisfaction scores between LEED and non-LEED rated buildings produced highly significant differences ($p \leq 0.001$) due to a large amount of sample data available ($n = 21\,250$).
- NHST tests do not provide any evidence that two or more studies will produce similar findings (i.e., no reliable information about the replication of experimental findings). In fact, when replicating an effect across studies with fixed sample sizes, but with different observers, statistical significance levels (p-values) can vary considerably [11]. Even small changes to the means, correlation

* Corresponding author.
E-mail address: michael.kent2@nottingham.ac.uk (M.G. Kent).

https://doi.org/10.1016/j.buildenv.2018.10.005
Received 13 June 2018; Received in revised form 1 October 2018; Accepted 3 October 2018
Available online 05 October 2018

Table 1
Key studies of discomfort due to glare.

Study	Prediction Model	Evaluation Criteria
Petherbridge and Hopkinson (1950)	Glare Constant/IES-GI	Multiple-Criterion Scale
Hopkinson and Bradley (1960)	Daylight Glare Index (DGI)	Multiple-Criterion Scale
Wienold and Christoffersen (2006)	Daylight Glare Probability (DGP)	4-Point Glare Scale

coefficients, or regression coefficients can lead to large variations in the calculated *p*-values, and therefore on the conclusions that are drawn from the data [9]. This can be problematic when comparisons are made between significant and non-significant results [12].

- Since differences in statistical significance (*p*-values) are not always statistically significant themselves, the comparisons made when using NHST analyses can often be misleading [9,13].

These conflicts arise also in discomfort glare research, whereas the strength of the significant relationships detected between evaluations of visual discomfort and calculated glare index values can vary considerably across different studies, even when the same prediction model has been used (e.g. [14–17]). For example, Tuaycharoen and Tregenza [18] showed that the correlation coefficients ($r = 0.72–0.86$) measuring the relationship between calculated Daylight Glare Index (DGI) values and the evaluations of discomfort due to glare reported by observers on Hopkinson's multiple-criterion scale were statistically significant ($p \leq 0.01$). Conversely, similar studies [19,20], which also used the DGI and the multiple-criterion scale, reported smaller correlation coefficients ($r = 0.28–0.56$) and did not show a statistically significant ($p > 0.05$) relationship between the same variables. Since results from separate studies often lead to inconsistent conclusions, we believe that NHST should not be used as the sole analysis method to support the statistical inferences derived from discomfort glare experiments.

The use of different statistical tools that can build on the work of previous studies may lead to a more robust and reliable characterisation of discomfort due to glare. An alternative approach to NHST is to use a Bayesian framework, whereby information from previous studies can be used to inform the analysis of data obtained in a new experiment [21]. Bayes' rule describes the probability of the occurrence of an outcome based on the conditions that might be related to it [22], positing how a degree of belief from previous knowledge should change once accounting for new evidence [23]. Bayesian inference treats data as fixed and model parameters as random variables [24]. A Bayesian approach is, thus, distinctly different from frequentist statistics, since it assumes that each unknown parameter has a *posterior* probability distribution that describes the uncertainty about population parameter values. The aim of the analysis is to estimate the posterior distribution given the data. The posterior density is the normalised product of a *prior* distribution, reflecting initial beliefs, and the *likelihood* from the data [25]. Once new data are collected, the prior is combined with the likelihood to produce a posterior distribution. In so doing, Bayesian models of analysis can deal with the complexity of real-world settings and overcome some of the limitations of controlled laboratory studies [26]. Clearly, since the Bayesian approach relies on knowledge from previous research, for it to be applied to inform new experimental studies there is a need to make data publicly available along with the original study findings. For example, in a recent article, Bayesian inferences were applied to analyse the effect of personalised control systems on the levels of visual satisfaction in daylit offices [27]. Using this analytical approach, previous knowledge of human visual preferences was combined with newly collected information to develop personalised visual satisfaction profiles within private workstations. Another important application of a Bayesian approach is to examine whether a new experiment can successfully replicate the results found in earlier research [13].

One further critical aspect of any scientific investigation is to verify whether the conclusions drawn from an original study were not the result of an experimental or analytical bias (i.e., a random error). In the context of discomfort glare research, we previously identified a substantive effect of order bias (i.e., the sequence in which the magnitude of discomfort glare was evaluated using a multiple-criterion scale and a luminance adjustment task) in the procedure used by Petherbridge and Hopkinson [6] to obtain the Glare Constant, which is at the basis of many successive glare models [28]. To ensure that our previous conclusions – based on an experimental setup using a small glare source – were not the result of a random error, a new experiment was designed. We applied a Bayesian approach to validate the previously detected effect of order bias, using a Hopkinson-like luminance adjustment task but under slightly modified experimental conditions. Informing the alternative hypothesis with the data from the earlier experiment [28], we used the same procedure in a new experimental setting with a large artificial window and a different sample of test participants. On this basis, in this paper we aim to: (1) demonstrate how a Bayesian approach can be used as a suitable alternative to NHST when analysing experimental findings derived from independent glare studies; (2) replicate the detection of the order bias effect when using a luminance adjustment procedure to evaluate the subjective degrees of discomfort due to glare from a large glare source. Therefore, while the effect of order bias is of relevant interest to this study, it was used primarily to illustrate how a Bayesian approach can reinforce the experimental conclusions derived from independent discomfort glare studies.

2. Method

2.1. Experimental setting

The new experiment was carried out in a test room located at the Berkeley Education Alliance for Research in Singapore (BEARS) centre, within the SinBerBEST testbed (Fig. 1). The room contained an artificial window (known as Daylight Emulator (DLE)), backlit by an array of cool and warm LEDs, capable of emitting light with a spectral composition approximate to daylight. The DLE provided a variable luminance in the range between 952 and 10 005 cd/m^2.

The artificial window featured three separate panes of glass, each of 0.98 × 2.15 m^2, surrounded by metal sill frames of 0.08 m in width and depth. Behind each glass pane, a fabric sheet membrane was mounted in front of the DLE. The membrane had uniform transmission properties allowing direct light from the DLE to be evenly distributed across the area of the window. Each window pane was equipped with a fabric roller blind, which remained fully retracted during the experiment. The room surfaces had reflectance properties of: $\rho_{wall} = 0.56$, $\rho_{floor} = 0.72$, $\rho_{ceiling} = 0.72$ (these were estimated using the Munsell system). A workstation (chair, desk, and desktop computer) was placed inside the room at a position parallel to the window. This arrangement was informed by the study from Osterhaus and Bailey [29] and was preferred over an internal spatial organisation where the workstation was orthogonal to the artificial window. Since a parallel arrangement produced higher glare index values, we believed this would have increased the likelihood of detecting an order bias effect. The surface of the desk had a reflectance of $\rho_{desk} = 0.56$, dimensions of 1.80 × 0.75 m^2, and a height of 0.74 m from the floor. We used a flat 24” liquid crystal display computer screen (hp zdisplay z24i, mean self-luminance = 150 cd/m^2) to present visual tasks to test participants. The screen was mounted on the desk top.

2.2. Photometric measurements

We used a Charged Coupled Device (CCD) Canon EOS 7OD camera with a 4.5 mm f/2.5 EX DX GSM 180° sigma fish-eye-lens, a luminance