Evaluating the effects of colour blending on optical-see-through displays for ubiquitous visualizations

Charles-Olivier Dufresne-Camaro* Yumiko Sakamoto† Pourang Irani‡

Department of Computer Science, University of British Columbia - Okanagan

Abstract

Optical-see-through (OST) augmented reality headsets offer users the flexibility to access relevant data visualizations anytime and anywhere. However, the appearance of content displayed on OST displays varies in colour and transparency depending on the environment they are viewed in, potentially leading to interpretation challenges. We present the findings of a psychophysical study ($N = 24$), aimed at assessing the impact of two environmental factors – lighting intensity and background colour – on user performance and colour perception accuracy in a visualization and colour-matching task using an OST headset. Our results suggest the effect of background colour on visualization interpretation is notable only under bright lighting conditions. Interestingly, participants perceived low-colour-contrast scenarios as more challenging, although their performance did not decline. Additionally, visualization colours were perceptibly and distinctly mismatched, but did not blend with the background colours. Finally, we discuss visual comfort and colour coding in the context of designing ubiquitous visualizations on OST displays, highlighting open challenges.

Index Terms: Human-centered computing—Visualization—Visualization application domains—Information visualization; Human-centered computing—Visualization—Empirical studies in visualization; Human-centered computing—Ubiquitous and mobile computing—Empirical studies in ubiquitous and mobile computing.—

1 Introduction

Ubiquitous visualizations were recently defined as data visualizations available anytime and anywhere [45], affording users with the ability to view relevant information *in-situ*, and respond to *in-the-moment* queries. Augmented reality (AR) optical-see-through (OST) head-mounted displays (HMDs) can support such applications by augmenting the user's physical environment. For example, a user at a mall could concurrently view the price history of a product they are purchasing and compare prices between stores, while keeping track of their surroundings behind the AR content.

OST displays are unique in the way they augment the user's environment: virtual content is rendered on a semi-transparent display, directly overlaying the physical environment. Dark colours (e.g., black) can thus be impossible to perceive if the physical environment is too bright. This design introduces the problem of colour blending: reductions in display contrast due to the virtual elements blending in with the environment's colours [16]. For example, a red virtual object may appear greener when viewed over grass. Colour blending could induce significant usability issues in ubiquitous visualization applications, as colour is commonly used to encode information. Users may take longer to interpret a chart due to lower foreground-background contrast, or arrive to an incorrect interpretation due to inaccurately perceiving a colour encoding [15]. Contrast issues on recent OST-HMDs are further exacerbated by environmental lighting conditions. For example, the Microsoft Hololens 2 has roughly 25% luminance contrast in lighting intensities equivalent to an overcast day [14, 44]. Although the effects of colour blending have been quantified through metrological experiments [16, 23], we find a general lack of perceptual research focusing on evaluating its effects on user performance in ubiquitous visualization settings.

Countless guidelines for designing colour-coded visualizations on traditional opaque displays have been proposed and validated [36, 47]. However, few have been examined on OST displays, where they may not generalize due to unique viewing conditions. Recent work in colour science have notably suggested that colour perception on OST displays may be unique [10, 20, 37, 53], further blurring the extent to which traditional guidelines may generalize to this medium. We thus identify a need to understand how colour blending affects visualization interpretation on OST displays to inform on the design of ubiquitous visualizations, and the necessity of correcting its effects. We note existing corrective methods [9, 25, 54] do not preserve colour encodings, limiting their uses to support visualization applications.

For ubiquitous visualizations to support users *in-situ* and provide engaging experiences [45], we must understand how the environment itself affects users (e.g., through colour blending). In this work, we simultaneously investigate the effects of two visual factors of physical environments – *lighting intensity* and *background colour* – on visualization interpretation performance and colour perception accuracy on an OST-HMD. To this effect, we present the results of a psychophysical study ($N = 24$) which involved interpreting and colour-matching 2D scatter charts situated within visually-distinct indoor environments on a Microsoft Hololens 2. Our results show that participants' interpretation performance was only affected by more visually complex backgrounds under brighter lighting. Although low-contrast conditions (e.g., green visualization on green background) were found more difficult by participants, we found no significant differences in performance. Additionally, participants perceptibly mismatched visualization colours in luminance and chromaticity, but no clear patterns emerged which would indicate colour blending as the main cause. Similar to prior work [15], colours like blue and yellow were more accurately matched than red, providing insights for colour map design on OST displays. Finally, we discuss our findings in the context of visual comfort and colour coding for ubiquitous visualizations on OST displays, and outline open challenges. Study materials and data are available at: `https://osf.io/qbh32/`.

*e-mail: chcamaro@student.ubc.ca
†e-mail: yumiko.sakamoto@ubc.ca
‡e-mail: pourang.irani@ubc.ca

Our contributions include:

- C1: empirical support for the effects of lighting intensity and background colour on the interpretation of coloured visualizations on OST-HMDs;
- C2: empirical evidence of the impact of colour blending on colour perception accuracy on OST-HMDs, contrasting with earlier research;
- C3: a discussion on colour coding and the importance of visual comfort in ubiquitous visualizations on OST displays.

2 Related work

2.1 Colour perception on OST displays

OST displays are semi-transparent, resulting in a *see-through* appearance for the rendered augmented reality (AR) elements. Due to their additive nature, the visible gamut of OST displays vary based on lighting conditions, with darker colours disappearing under brighter lighting [26]. Current OST-HMDs offer limited brightness, resulting in poor contrast in outdoor conditions [14]. Transparency has been notably shown to affect user performance on visual-search tasks on OST-HMDs when reaching critical levels [24].

Metelli [33] framed transparency perception as a spectrum where viewers may see two coloured patches – a semi-transparent overlaying an opaque one – as two separate layers (i.e., colour scission), or as a single patch of the mixed colours (i.e., colour fusion). Complete colour scission appears ideal from a usability perspective for OST displays, as the perception of the colour of virtual elements would not be affected by the underlying background. However, colour perception on OST displays appear closer to incomplete colour scission [10, 15]: users may identify two layers, but incorrectly separate the colours.

Colour blending is defined on OST displays as the mixing of virtual content's colours with the physical environment underlying them [16]. Metrological experiments showed that under natural light, the colours' appearance were perceptibly distorted towards the background colours [16]. Follow-up user studies revealed contrasting effects of background from large distinct distortions [10] to negligible differences [15] between backgrounds. The distortions also notably differ from those measured in metrological studies [10]. Correlated colour temperature (CCT) of illuminants has additionally been shown to affect perceived colour appearances on OST displays [6, 20]. We note multiple colour-corrections approaches have been proposed to minimize colour blending [9, 28, 29, 54]. However, they do not preserve colour encodings, making them unusable on colour-coded visualizations.

Colours on OST displays are not perceived as equally distorted in colour matching tasks [15, 32, 51]. For example, matching errors for red were ≥ 2 times larger than for blue and yellow on an OST heads-up display [15]. Furthermore, colour appearance models, such as the CAM16 [30], do not accurately predict the perception of AR elements [20, 37, 53]. Colour perception on OST displays may thus be unique, and rely on other contextual cues to assess colour [10, 20], luminance [37, 53], and transparency [52]. Colour-scission models have been proposed for luminance [37] and colour [10, 19], but have yet to reliably predict colour appearance. Finally, colour perception may be affected by the display's characteristics and imperfections [46, 51]. Thus, modeling colour perception on OST displays remains an open problem, which further complicates the design of ubiquitous visualizations.

Although colour perception has been explored on OST displays, the impact of colour blending on user colour perception in ubiquitous visualization tasks remains an open question. Prior studies have additionally focused primarily on the effects of single attributes of the physical world, limiting our understanding in more complex and visually diverse settings. Furthermore, the tasks did not account for visual factors critical in visualizations (e.g., content density, mark shape and sizes). We thus conduct a psychophysical study involving distinct physical environments and lighting conditions to identify their combined effects on colour perception accuracy on an OST-HMD.

2.2 Designing ubiquitous visualizations on OST displays

Ubiquitous visualization has been recently defined as data visualizations that can be viewed anywhere, anytime [45]. Ubiquitous visualizations can be situated [3, 50], and facilitate a broad range of applications from supporting in-the-moment queries to more complex analyses (e.g., immersive analytics applications [1, 11]). Although ubiquitous visualizations have not yet been extensively studied, their design can be inspired from neighbouring research areas: For example, where to position content within a user's field of view [31], and situate visualizations in the world [38]. We additionally find research in related areas has focused primarily on specific applications and settings rather than more fundamental challenges (e.g., visual perception [13]), which limits our understanding of designing applications more generally.

Countless guidelines for designing data visualizations on traditional displays have been proposed over the last few decades based on human visual perception [8, 36, 47]. Colour is an effective visual channel to encode information, and has received extensive attention. Guidelines focus notably on designing effective colour maps [21, 49], with colour differences large enough to correctly interpret the codes [41, 42, 43]. Furthermore, colour codes may be affected by simultaneous contrast effects, where the surrounding (e.g., background) of a visual mark will change its perceived appearance [4, 22, 48] (e.g., a coloured mark surrounded by a darker colour will appear lighter). However, we find most guidelines have not been formally studied for OST displays, where display characteristics and viewing conditions can differ greatly.

Finally, we note the medium on which a visualization is displayed constrains its design (e.g., size, content density, colours). OST displays impose stricter constraints than traditional displays, due to hardware limitations and complexity [26]. OST-HMDs have been shown to be spatially non-uniform, producing perceptible colour distortions across their surface [23, 26, 51]. Additionally, binocular OST displays may produce slightly different stimuli for each eye, further distorting the appearance of virtual elements [46].

In summary, we find a lack of grounded and actionable design guidelines surrounding the use of colour in ubiquitous visualizations on OST displays. We address this by conducting a psychophysical study to identify the main effects of two core visual factors of physical environments (lighting intensity and background colour) on user performance and colour perception in a general visualization task.

3 Study Methodology

Colour blending represents the mixing of the light emitted by the OST display with the light reflected from the environment towards the display [16]. The reflected light's intensity modulates the transparency of the rendered AR elements [14]. The light's colour, which varies notably based on the illuminant's correlated colour temperature (CCT) and the colour of the surface from which it is reflected, may in turn distort the elements' appearance [16, 10, 20].

We thus conducted an in-person psychophysical study with controlled physical environments to evaluate the effects of colour blending on user performance interpreting ubiquitous visualizations on OST-HMDs. Specifically, we studied the effects of three factors – background (colour), lighting intensity, and visual mark colour – on user visualization interpretation performance, colour perception accuracy (i.e., how close one's perception of a colour is to the target), and perceived workload. We focused on background (surface) colour to represent the reflected light's colour as this property is more likely to vary in real-life environments and produce noticeable changes in viewing conditions. Participants completed two tasks requiring them to 1) interpret simple 2D scatter charts, and 2) identify the colour of the visual marks rendered on an OST-HMD (Microsoft Hololens 2).

Our study design is grounded in established psychophysical experimental designs used to assess colour perception on OST displays. Such methods involve participants matching an AR stimulus colour to a target on a different medium (e.g., opaque display) through adjustments [20, 51] or by selecting from a fixed set [10, 15]. In this work, we specifically adapted Gabbard *et al.*'s methodology [15] to support exploring colour perception for ubiquitous visualization research, and to control and mitigate the effects of extraneous variables based on hardware (e.g., spatial distortions on OST-HMDs [26, 46, 51]), visual perception (e.g., mark size in visualizations [43]) and colour-vision deficiencies. We selected this approach over others because it has notably been used to evaluate the effects of colour in practical tasks [15].

3.1 Research Questions & Hypotheses

Our study was designed to answer the following research questions with respect to users' performance interpreting visualizations, and their colour perception of visual marks on an OST-HMD:

RQ1. **What are the effects of lighting intensity?** We hypothesize brighter lighting intensities will lead to worse interpretation performance due to decreased OST display contrast [14] (**H1.1**), and less accurate colour perception as colour blending will be more intense [16] (**H1.2**).

RQ2. **What are the effects of the background?** We hypothesize backgrounds will not affect interpretation performance [40] (**H2.1**), but will decrease colour perception accuracy of the marks due to colour blending [16] (**H2.2**).

RQ3. **What are the effects of visual marks' colour?** We hypothesize interpretation performance will not be affected by the visual marks' colour [15] (**H3.1**). However, colour perception accuracy will vary [15] (**H3.2**).

RQ4. **What are the interaction effects between the background and the visual marks' colour?** We hypothesize that when the difference between the background's colour and mark colour is small (i.e., low contrast), interpretation performance will decrease (**H4.1**), but colour perception will be more accurate as the appearance of the blended colour will be closer to the mark's colour [16] (**H4.2**).

3.2 Factors

This study focuses on three within-subject factors summarized in Fig. 1: *Background*, *Lighting intensity*, and *Mark colour* (i.e., the colour of the marks in a visualization).

Background *Background* represents the colour of the illuminated surface from which light is reflected towards the OST-HMD. We considered three backgrounds with distinct surface colours (Fig. 1): brick (brown and grey), paint (green), and wallpaper (blue-grey). The textures were selected to reflect commonly-encountered indoor environments, and to produce different contrast conditions when viewed under data visualizations. We further used non-reflective materials over prints or digital renders to generate realistic viewing conditions where physical lighting interacts with the environment. Although each background has a relatively different texture complexity (e.g., brick having a two-colour pattern compared to paint having only a faint depth pattern), we expected these differences to not affect visualization interpretation performance based on prior work [40]. Finally, all backgrounds were illuminated by the same 3000K illuminant (warm light) to control the effects of CCT on colour perception [6, 20].

Lighting intensity We defined *Lighting intensity* as the amount of light reflected from the background's surface towards the OST-HMD (Hololens 2), before it passes through the display. This factor directly relates to the transparency of the AR elements rendered on the OST display[14]. We specifically selected two levels corresponding to typical indoor lighting conditions [44]: 100 cd/m^2 (hallways) and 350 cd/m^2 (office spaces). Furthermore, these levels exemplify two distinct display contrast levels on the Hololens 2: 70% and 42% luminance contrast [34] respectively[1].

Mark colours Five mark colours were selected based on their frequency of use in traditional information visualizations [47], and current OST-HMDs limitations [26]: blue, green, red, white (light grey), and yellow. The specific colour values were chosen to fit within the Hololens 2's colour gamut. Colours were also grouped by lightness with blue, green, and red having medium lightness ($L^* \approx 50$ in $L^*a^*b^*$ space), and white and yellow having high lightness ($L^* \approx 80$). This allowed us to explore whether colour lightness may affect user performance [15].

3.3 Apparatus

Sessions were conducted in a closed indoor space with controllable halogen light sources (3000K) as illuminants (Fig. 2). Backgrounds were created using 1.2m× 1.1m physical panels hung on a stand. Participants sat on a chair facing the stand, 1.25m away. Panels were designed to be large enough to cover the participant's field of view ($52^\circ \times 46^\circ$), such that the visualizations would be positioned entirely within a panel. Light sources were 1.65m high, facing the stand at an angle to produce uniform lighting across the surface underlying the visualizations without *hotspots*. Light source intensities were adjusted for each *background* to ensure the light reflected towards the participant corresponded to the two intensity levels. We used an Extech LT45 light-meter to measure the intensities.

Visualizations were rendered on a Microsoft Hololens 2 OST-HMD with maximum display brightness. The headset was eye calibrated for each participant. Visualizations were anchored on the physical panels, centered, at eye-level, and directly facing the participant. The visualization application was built in Unity using MRTK[2], UXF [5], and u2vis [39]. Visualizations consisted of 2D scatter charts with non-overlapping disk-shaped marks distributed semi-randomly (Fig. 2; see Sect. 3.5.1 for details). Visualizations were $25^\circ \times 25^\circ$ in visual angles, fitting fully inside the Hololens

[1]We calculated these values using Erickson *et al.*'s approach [14] with a halogen light source (3000K CCT).

[2]https://github.com/microsoft/MixedRealityToolkit-Unity

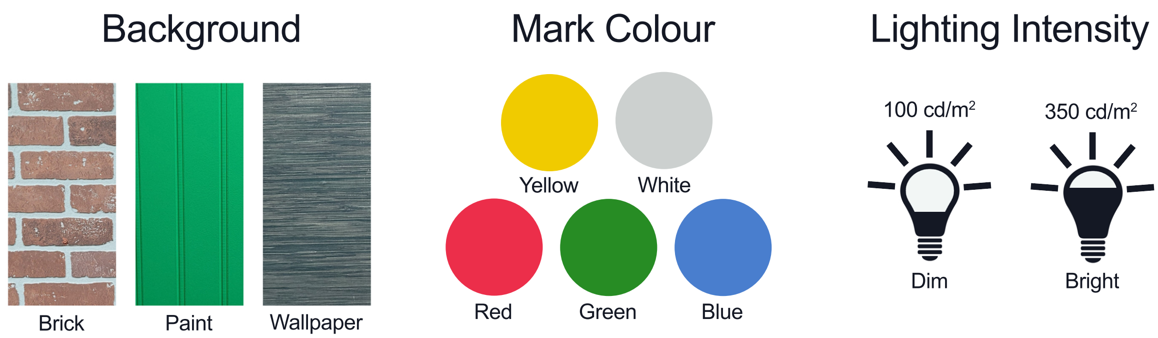

Figure 1: Background conditions, visual mark colours and lighting intensities used in our study. Colours are included for illustrative purposes, and may not reflect how they appeared in the study, due to display format and calibration. Colour values for the background and visual mark colours are provided in the supplementary materials.

2's field of view ($43^\circ \times 29^\circ$). Disk mark diameters were designed to be 1.17° in order for colour perception to be more sensitive to colour contrast than brightness contrast[3] [35].

We collected responses using a custom Android application deployed on a 12.4" Samsung S7 FE tablet at maximum brightness on a table to the right of the participants. Colour responses were collected through a discrete colour palette, similar to the one used by Gabbard *et al.* [15] (Fig. 2). We opted for a fixed palette rather than free adjustments (e.g., in HSV space [51]) to reduce the time taken to match colours. The colour palette consisted of 148 colours uniformly sampled from the $L^*a^*b^*$ colour space with $L^* \in [30, 90]$ [4]. Each row corresponded to a different L^* value, and columns to a different hue. Colours were adjusted to fit the colour gamut of the tablet as needed. Neighbouring columns had an average colour distance of $\Delta a^*b^* = 19.1$, and neighbouring rows a lightness difference of $\Delta L^* = 9.89$. *Mark colour* levels were included in the palette, and were tuned, in the CIE 1931 xyY colour space, in complete darkness on the Hololens 2 to be within 2cd/m^2 (equivalent to $Y \leq .05$) and .01 chromaticity (xy) distance from the palette's, using an off-the-shelf display colorimeter adapted for this purpose. This ensured participant responses reflected the effects of colour blending on colour perception, and minimized the effects of the display's colour profile. [15]. The tuning method is described in supplementary materials.

3.4 Participants

The study was conducted at a North American post-secondary institution, and ethics approval was obtained from their research ethics board. Recruited participants completed a physical 14-plate Ishihara test [7] to verify whether they can perceive red-green colour differences. We identified participants as outliers if they responded incorrectly to at least three plates, and did not use their data in our analysis.

Twenty-six participants were recruited. While all participants reported having normal or corrected-to-normal vision, and not suffering from any colour vision deficiency, two participants' data were excluded as their Ishihara test results indicated some colour deficiency. This left us with twenty-four participants (11F, 13M; $M_{age} = 25.3$ yo., $SD_{age} = 7.4$, range: 19 - 47). Participants were not so familiar with AR in general ($M = 3.38$, $SD = 1.56$), with only three (12.5%) using OST-HMDs at least a few times per month. Furthermore, participants were strongly familiar with simple 2D visualizations ($M = 6.08$, $SD = 1.06$), and were confident in their ability to correctly interpret them ($M = 5.63$, $SD = 1.17$). All responses are on a 7-point Likert scale, where a higher rating indicated higher familiarity or confidence.

3.5 Procedure

After signing the consent form, participants were fitted with the Hololens 2, and were introduced to the *visualization interpretation* and the *colour matching* tasks. Each participant completed four practice trials to familiarize themselves with the study. Note that during the practice trials, conditions were different from the studied factor levels to avoid any potential learning effects. In each trial, participants were shown a 2D scatter chart on the Hololens 2, and asked to successively complete the visualization interpretation and colour matching tasks. We included a 2.5 second pause with no visual stimulus between trials to reduce afterimages, and allow participants to prepare for the next trial. The specific pause time was set based on results of a pilot study. Participants completed a total of 60 trials (2 *lighting intensities* $\times$ 3 *backgrounds* $\times$ 5 *mark colours* $\times$ 2 repetitions).

Trials were evenly distributed across six blocks: one per *background* $\times$ *lighting intensity*. At the end of each block, participants completed an 11-point-scale NASA-TLX questionnaire [18] focusing on the interpretation task in the specific environment, and were offered a two minute break. At the end of the study, participants completed a questionnaire on their perception of each factor's impact on their interpretation performance. The study lasted on average 75 minutes, and participants received a $20 CAD Amazon gift card.

For each trial, the following dependent variables were collected: interpretation (completion) time and response for the visualization interpretation task, and colour response for the colour matching task. The backgrounds' order was counterbalanced using a 6×6 Latin square, so that the same background appeared in two consecutive blocks (one per lighting intensity). Lighting intensities were set to appear in *bright-dim* order for half the participants, and *dim-bright* for the others. Finally, mark colours were randomized within each block, so that each colour would appear two times per block.

[3] We translated the 0.5cycle/deg guideline of Mullen [35] to a minimum mark size of 1 degree by posing 1 "cycle" to correspond to two times the diameter of a disk mark (half ON, half OFF), similar to a square wave signal.

[4] Blue hues were sampled at a coarser rate, since they appeared too similar once rendered on the tablet.

Figure 2: **Left**: Study space set up for a bright lighting and paint background condition during a trial. The user is currently matching the colour of the marks using the colour palette on the right. **Right**: Example scatter chart with *white* coloured marks shown during a trial (**Top**). Note that this image is only used for illustrative purposes, and does not reflect the visualization's appearance on the Hololens 2 (i.e., colour, and transparency). Participants were asked to identify which of the four groups contained the most marks, followed by matching the mark's colour as it appeared to them with the colour palette rendered on the tablet (**Bottom**).

3.5.1 Visualization interpretation task

The 2D scatter charts consisted of four distinct groups of randomly distributed disk-shaped marks (32 in total), delimited by a 2D axis system with no tick marks (Fig. 2). For each chart, participants were asked to identify which of the four groups contained the most marks. Participants were told to focus on obtaining the correct answer using any technique, and that they had no time limit. They were also allowed to rotate their head to bring marks closer to the centre of the display to increase their opacity, due to the Hololens 2's distortions along its periphery [26, 51]. Participants were instructed to press the *Space* key on a keyboard as soon as they identified the correct answer, and then to press a button corresponding to the group on the tablet. The chart disappeared after pressing the *Space* key to avoid participants gaming the task. We defined interpretation time as the time taken for the participant to press the *Space* key after the chart first appeared. We used this definition, as participants of our pilot studies would take a variable amount of time to respond after identifying the correct answer. Finally, task accuracy was measured based on whether the participant selected the right group.

We used this discrimination task [12] for its simplicity: solutions involved deriving the number of marks in each group, and comparing them to find the max value [2]. We further selected scatter charts, because it allowed us to control the effects of mark size on colour perception accuracy [42, 43], without affecting the *fidelity* of the task. The difference between the two most populous groups was randomly set to be one or two marks, based on the results of a pilot study, to reduce potential learning effects. The number of marks per visualization was also optimized to ensure the task can be completed in $\leq$ 20s, and avoid participants subitizing [27] (i.e., obtaining the solution without *counting* the marks).

3.5.2 Colour matching task

Upon completing the interpretation task, the scatter chart reappeared and participants were asked to match the appearance of the marks' colour as they perceived it using the colour palette. Participants were instructed to centre their field of view on the middle of the chart, and focus only on the marks closest to the centre. Participants would then flip open the Hololens 2 display, look directly at the colour palette rendered on the tablet to their right, and select the colour patch that most closely matched their perception of the mark's colour (Fig. 2). These steps were included to avoid biasing the user's colour perception due to the Hololens 2 display's non-uniformities [23, 26, 51]. Furthermore, displaying the colour palette on an external opaque display allowed us to measure potential effects of colour blending on perception accuracy. We additionally used a fixed, discrete colour palette to keep trials' duration short.

4 Results

Three-way Repeated Measures (RM) ANOVA (*Lighting intensity* $\times$ *Background* $\times$ *Mark colour*) was conducted on all measures (averaged over repetitions), correcting for sphericity where necessary using Hyunh-Feldt correction (if $\epsilon > 0.75$) or Greenhouse-Geisser [17], unless stated otherwise. We additionally conducted pairwise comparisons where necessary using paired t-tests with Bonferroni correction.

4.1 Visualization interpretation

4.1.1 Interpretation performance

We evaluated performance through interpretation time and accuracy. No interaction effects or main effects were found on interpretation accuracy: The average accuracy across all conditions was 99.3% (SD = .3%), indicating that regardless of the conditions, the visualization interpretation task

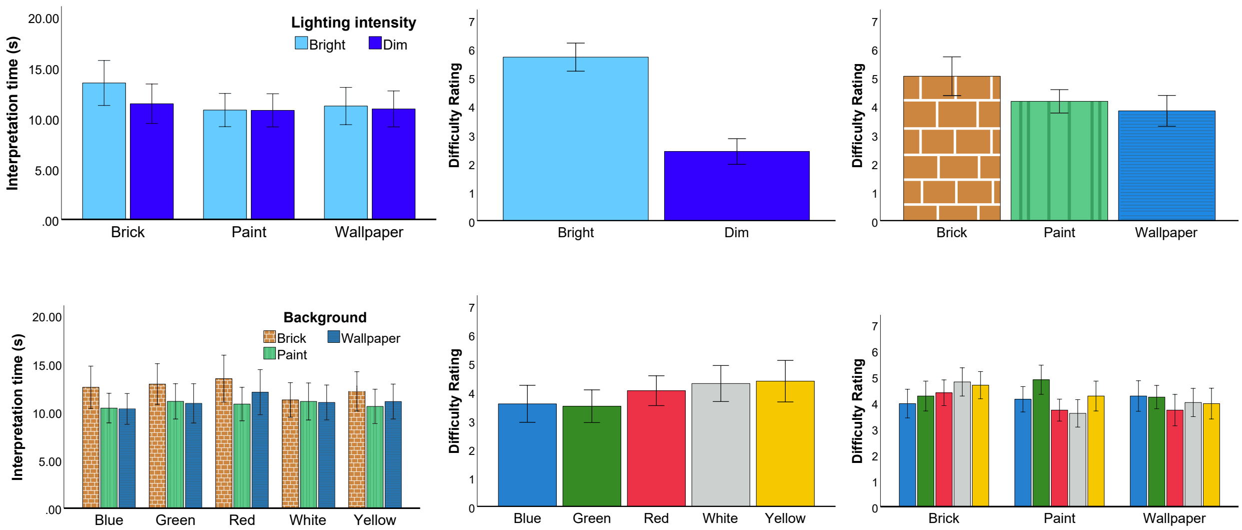

Figure 3: **Left:** Mean interpretation times in the visualization interpretation task ($N = 24$) between *Background* and *Lighting intensity* (Top), and between *Mark colour* and *Background* (Bottom). **Middle & Right:** Mean perceived difficulty rating responses ($N = 24$; 7-point scale in post-study questionnaire) for the visualization interpretation task for *Lighting intensity* (Top-middle), *Background* (Top-right), *Mark colour* (Bottom-middle), and *Background* $\times$ *Mark colour* (Bottom-right). Error bars represent 95% confidence intervals.

was performed accurately. To explore potential learning effects, we compared the mean interpretation times of the first and last three trials of the session. The difference was not significant, but had a medium effect size ($p = .14$, $\eta^2 = .09$).

We observed a large interaction effect between *Background* and *Lighting intensity* on interpretation time, $F(2, 46) = 4.46, p = .02, \eta_p^2 = .16$ (see Fig. 3 for mean interpretation times). Participants spent a longer time interpreting visualizations against the brick background when the lighting was bright in comparison to when the lighting was dim ($p \leq .01$). Furthermore, only under bright lighting, participants spent a longer amount of time with the brick background than the paint and wallpaper backgrounds ($ps \leq .01$). This indicates that when participants performed the task under dim lighting, the background might not have had any significant effect on the interpretation time. Additionally, a marginal interaction effect was found between *Mark colour* and *Background*: $\epsilon = .98$, $F(7.85, 180)$, $p = .097$, $\eta_p^2 = .07$. Pairwise comparisons revealed a significant *Mark colour* effect only with the brick background ($p = .02$): participants spent less time when the marks were white, compared to when they were red. Taking the accuracy and interpretation time data together, it appears that participants were able to perform the task accurately as long as they had sufficient time to interpret the visualization.

4.1.2 NASA-TLX

We performed two-way RM-ANOVA (*Lighting intensity* $\times$ *Background*) to compare the responses for each of the six NASA-TLX dimensions [18] (see Fig. 4 for mean response distributions). Three medium to large interaction effects were found for the following dimensions: Mental demand; $\epsilon = .88$, $F(1.75, 40.3) = 4.34, p = .02, \eta_p^2 = .16$, Physical demand; $\epsilon = .90$, $F(1.80, 41.4) = 4.53, p = .02, \eta_p^2 = .17$, and Frustration; $\epsilon = .99$, $F(1.99, 45.8) = 3.46, p = .04, \eta_p^2 = .13$. Under bright lighting, participants perceived the interpretation task as more mentally and physically demanding when viewed over the brick background than the wallpaper background ($p_m = .04$, $p_p \leq .01$). Brick also led to higher physical demand under bright lighting, compared to paint ($p \leq .01$). Additionally, completing the task on the brick and paint backgrounds was found to be more mentally demanding under bright than under dim lighting ($ps \leq .02$). Lastly, participants found the task more physically demanding and frustrating under bright lighting, compared to dim lighting, for every background ($ps \leq .04$).

Furthermore, we observed two main effects of *Lighting intensity*: Performance ;$F(1, 23) = 6.46, p = .02, \eta_p^2 = .22$, and Effort; $F(1, 23) = 11.6, p \leq .01, \eta_p^2 = .33$. Participants were more confident about their performance, and felt interpretation required less effort when completed under dim lighting than under bright lighting. Finally, *Background* had a large effect on Effort; $\epsilon = .84$, $F(1.69, 38.8) = 4.16, p = .03, \eta_p^2 = .15$. Two comparisons were close to marginal significance but none were statistically significant (brick vs. wallpaper, $p = .10$; brick vs. paint, $p = .11$).

4.1.3 Perceived difficulty

In the post-study questionnaire, participants rated their perceived task difficulty based on *Lighting intensity*, *Background*, *Mark colour*, as well as *Background* $\times$ *Mark colour*. Participants completed four questions using 7-point Likert scales (e.g., *In general, this mark colour made the task...* with a scale; 1 = *much easier* to 7 =*much harder*), and justified their ratings through open-ended responses. See Fig. 3 for the mean ratings of each question.

Lighting intensity Participants felt the task was more difficult under bright lighting $F(1, 23) = 67.23, p \leq .01, \eta^2 = .75$, than dim lighting. Almost all the participants (i.e., 23/24 or 96%) further described the task as more difficult under bright lighting (e.g., P10: "*Brighter light makes the holographic display appear dimmer and more transparent*").

Background The task was found more difficult when performed in front of the brick background, $F(2, 46) = 5.29, p \leq .01, \eta^2 = .19$,instead of the wallpaper background ($p = .02$). Participants noted the brick's more complex texture and colours (e.g., P08: "*brick was the hardest because it was a*

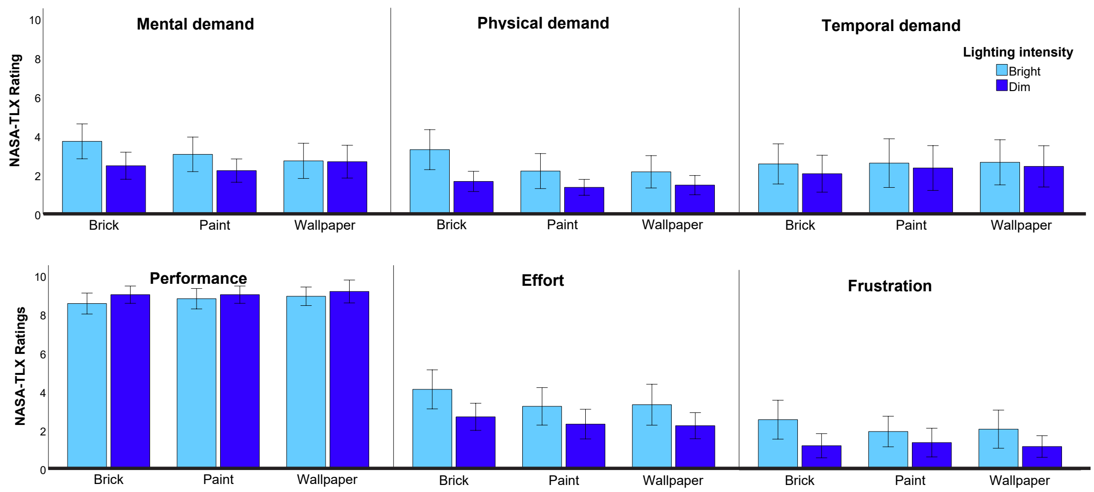

Figure 4: Mean NASA-TLX [18] responses ($N = 24$) between *Background* and *Lighting intensity*. Error bars represent 95% confidence intervals.

textured background that made it difficult to focus on the [marks]."; P06 "*the white lines interferred [sic] with the [marks], made them appear lighter.*"

Mark colour Perceived task difficulty did not vary significantly between *Mark colours*, $\epsilon = .96$, $F(3.84, 88.5) = 1.88, p = .12, \eta^2 = .08$. Fourteen participants (58%) noted certain colours appeared more distorted without mentioning lighting or background (e.g., P07: "*especialy [sic] for yellow, where it looked peach, lime and yellow*"; P09: "*The white looked a bit purplish and pink*"). However, no clear trends emerged from their justifications.

Background × Mark colour We found a large interaction effect, $\epsilon = .94$, $F(7.53, 173) = 5.26, p \leq .01, \eta^2 = .19$. Participants found the task more difficult on the paint background when the marks were green, compared to red ($p = .048$) or white ($p = .03$). Similarly, participants felt interpreting: 1) green marks over paint to be more difficult than over wallpaper ($p \leq .01$); 2) red and yellow marks over brick to be more difficult than over wallpaper ($ps \leq .04$); and 3) white marks over brick to be more difficult than over paint ($p \leq .01$). Fifteen participants (63%) further noted that low-contrast conditions made the visualization more difficult on average (e.g., P20: "*If the colour matched with the background it was harder to identify it*").

4.2 Colour matching

We define colour perception accuracy as the difference in xyY colour space between the participants' colour responses and the target *Mark colours* in terms of *luminance* (i.e., Y) and *chromaticity* (i.e., xy). Following Gabbard *et al.*'s approach [15], we calculated luminance and chromaticity *shifts*, as well as *dispersions*. Note *shift* indicates the degree to which a colour's appearance is distorted, which we quantified using *the Euclidean distance between a colour response and the target mark colour*. *Dispersion* represents how much colour responses vary between participants, and is calculated for each condition separately as *the Euclidean distance between a participants' colour response and the mean colour response*. We also considered the distributions of colour responses in xy space to examine whether colour blending pulled participants' responses towards the appearances of the backgrounds [16, 10]. Note that we only report differences larger than our luminance and chromaticity tuning tolerances (i.e., 2cd/m^2 and .01). Fig. 5 presents the mean colour responses for each condition.

4.2.1 Luminance matching

Luminance shift We performed three-way RM-ANOVA (*Lighting intensity* × *Background* × *Mark colour*) on signed luminance shifts ($GM = 7.22$ lux, $SD = 1.1$ lux), which revealed no significant interaction effects. A large *Mark colour* effect was found; $\epsilon = .82$, $F(3.27, 75.2) = 38.2, p \leq .01, \eta_p^2 = .62$ (see Fig. 6). Pairwise comparisons revealed participants were inclined to overestimate the brightness of green marks much more than every other mark colour ($ps \leq .01$), and were more accurate for yellow marks than all other colours ($ps \leq .01$). Furthermore, *Background* strongly affected participants' accuracy in matching luminances; $\epsilon = .92$, $F(1, 83, 42.1) = 12.2, p \leq .01, \eta_p^2 = .35$. Participants overestimated colour luminance more when viewed over the paint background than the brick and wallpaper backgrounds ($ps \leq .01$), indicating the higher saturation of the paint background might have biased participant's perception towards lighter colours.

Luminance dispersion We found no interaction effects between any of the factors, with a grand mean (GM) of 6.14 lux ($SD = .50$ lux). Only *Mark colour* had a large main effect; $\epsilon = .66$, $F(2.65, 61.0) = 8.56, p \leq .01, \eta_p^2 = .27$ (see Fig. 6 for mean responses), with red having less dispersed luminance responses than blue, green, and white ($ps \leq .02$), and yellow being less dispersed than green and white ($ps \leq .05$).

4.2.2 Chromaticity matching

Chromaticity shift Shifts were calculated using the Euclidean distance with (x, y) of the xyY colour space. We then performed three-way RM-ANOVA (*Lighting intensity* × *Background* × *Mark colour*), which revealed no significant interaction effects. Mean shift values across conditions are shown in Fig. 6, with the grand mean equal to .067 ($SD = .005$). *Mark colour* had a large main effect; $\epsilon = .65$, $F(2.62, 60.2) = 35.5, p \leq .01, \eta_p^2 = .61$, with red mark responses being less accurate than all other colours

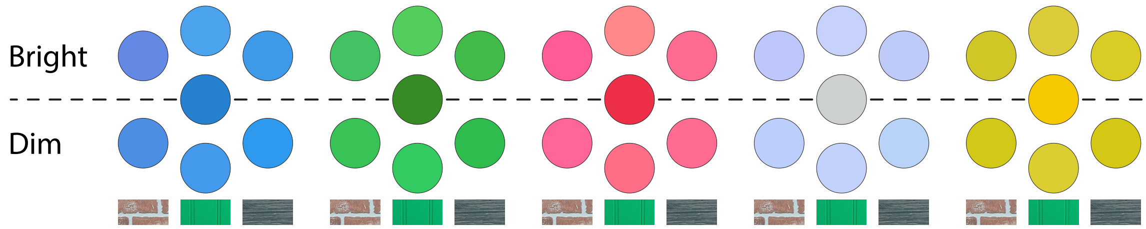

Figure 5: Mean colour matching responses ($N = 24$) for each *Background* (columns), *Lighting intensity* (rows) and *Mark colour* (charts) conditions. Ground-truth mark colours are shown in the middle of each chart. Note the colour appearances may differ due to display calibration.

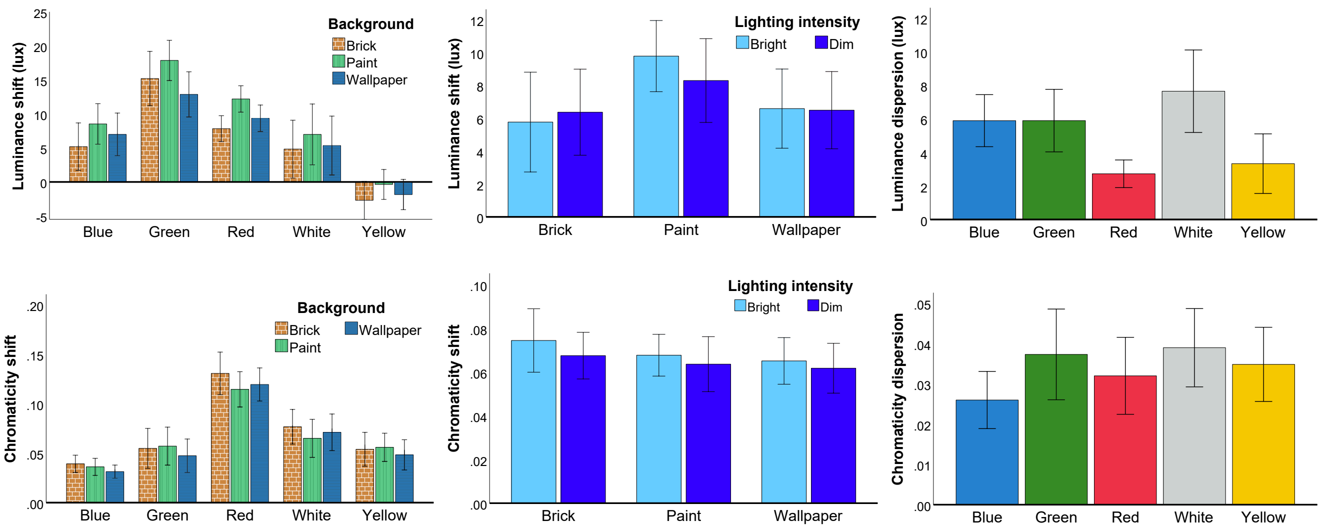

Figure 6: Mean luminance (**Top row**) and chromaticity (**Bottom row**) shifts and dispersions of colour responses ($N = 24$) in xyY colour space. Error bars represent 95% confidence intervals.

($ps \leq .01$). White mark responses were also significantly more distorted than blue marks ($p = .01$). Furthermore, we found a large main effect of *Background*; $\epsilon = .85$, $F(1.70, 39.1) = 4.23, p = .03, \eta_p^2 = .16$, but pairwise differences were all smaller than our tuning tolerance (.01).

Chromaticity dispersion Our tests revealed no significant interaction nor main effects across any of the factors. The grand mean across all conditions is .044 ($SD = .003$), indicating participants are equally likely to perceive colours as distorted across conditions [15]. See Fig. 6 for the mean dispersion values between mark colours.

Chromaticity response distributions *Shifts* quantify the intensity of the distortion, but not its direction (e.g., towards blue or greener hues) which could indicate colour blending affected participants. Similarly, *dispersions* quantify the variability of colour responses, but not how they are distributed in chromaticity. Fig. 7 shows the mean chromaticities and *prediction* ellipses of the colour responses in the xy plane. Shift directions varied between backgrounds for certain colours (e.g., red on paint vs. brick). However, we observed no trends across colours where mean shifts are oriented towards the backgrounds' colour, suggesting the chromaticity shifts may not be caused by colour blending. We additionally found prediction ellipses to vary in orientation and size based on *Mark colour*, similarly to dispersion measures: blue had on average 2.25 times smaller areas than the other colours. Furthermore, prediction ellipses of each *Mark colour* differed between *Lighting intensities*, with intersection-over-union (IoU) scores averaging .72 ($SD = .10$), suggesting participants' colour responses may vary based on lighting intensity. However, no clear trends emerged with respect to the background colours, indicating colour blending may not have been amplified by brighter lighting. Finally, we observed important overlap between the blue and white mark ellipses for the brick background, which show participants may have confused the two colours together.

5 Discussion

5.1 Effects on visualization interpretation

We hypothesized visualization interpretation performance would be affected by *Lighting intensity* (**H1.1**), and the interaction between *Background* and *Mark colour* (**H4.1**), but not by *Background* (**H2.1**) nor *Mark colour* (**H3.1**) independently. Considering participants had near perfect accuracy across conditions ($GM = 99.3\%$), and had unlimited time to complete the interpretation task, we operationalize performance in terms of interpretation time. We include workload and subjective measures to supplement our interpretations and situate our findings.

We observed an interaction effect between *Background* and *Lighting intensity*, where only the more complex (i.e., multi-coloured) brick background led to longer interpreta-

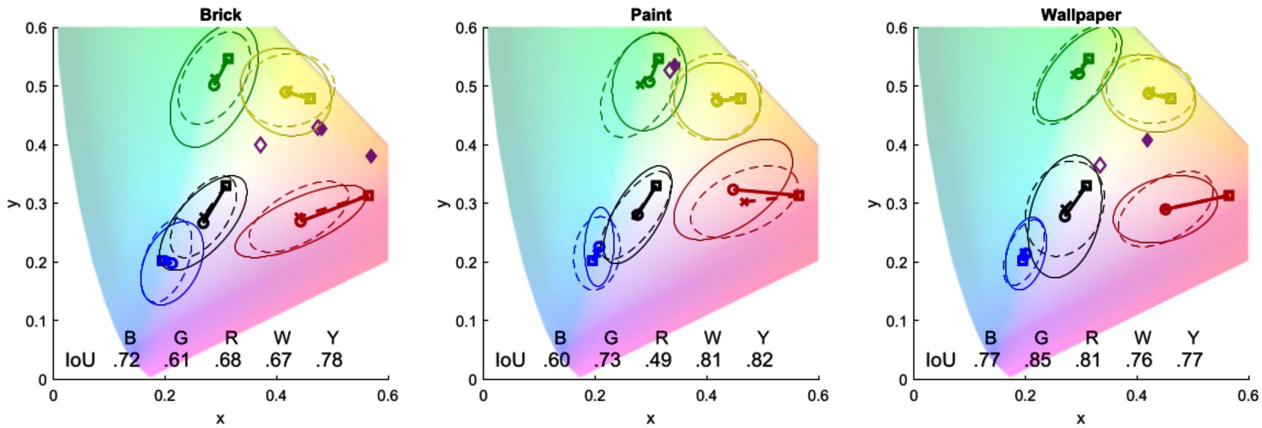

Figure 7: Mean chromaticities and 95% prediction ellipses of colour responses in xy plane ($N = 24$) for each *Mark colour* (coloured □, with white shown in black), *Background*, and *Lighting intensity* (dashed lines and cross marks for bright; plain lines and disk marks for dim) condition. Background chromaticities under bright and dim lighting are represented by purple ◊ and ♦ marks respectively. Brick chart contains two ◊/♦ to reflect the colours of the bricks and the mortar separately. Intersection-over-unions (IoU) between ellipses of the same mark colour under different lighting intensities are shown at the bottom of each chart.

tion times and higher workload when viewed under bright lighting. This suggests the effects of the background on user performance may only arise when the virtual elements are more transparent and blend in more with the physical environment. Our OST-HMD's luminance contrast notably decreased from 70% to 42% in our bright lighting condition. Participants additionally felt the task required more effort and was more frustrating under bright than dim lighting. Taken together, we have limited support towards brighter lighting intensities reducing user performance, and can only **partially retain H1.1**.

Our findings on the effects of *Background* contrast with prior work on background clutter, suggesting users' performance may be affected even in *single task* conditions [40]. We posit this may be caused by differences in the lighting intensities used in each experiment. Considering strictly the results under bright lighting where the brick background decreased user performance (Fig. 3), we note *discounting* the background [10] may be simpler to accomplish when it is more uniform (e.g., has fewer distinct colours). As we observed a condition where a more complex background led to a slower interpretation time, we **reject H2.1**.

Lastly, we hypothesized task performance would decrease in low-colour-contrast conditions, such as the green marks on the paint background. However, we found only a marginal interaction where interpretation time was shorter for white marks than red marks when viewed over brick. We posit this difference is caused primarily by the display's luminance contrast and not colour contrast, similar to the *Lighting intensity* and *Background* interaction: colours with lower lightness (i.e., blue, green, and red) which appear dimmer on the OST-HMD had 1.3s longer interpretation times ($SD = .4$s) on the brick background than the lighter colours white and yellow ($p \leq .01$, $\eta^2 = .35$; Fig. 3). We therefore **reject H4.1**. Additionally, we found no main effect of *Mark colour* on user performance, and thus **retain H3.1**.

5.1.1 Perspective on Visual Comfort

We highlight the value of considering visual comfort based on our results. We observed disparities between performance and *perceived difficulty* regarding the interaction effect of *Background* × *Mark colour*. We expected these measures to match to some extent, as longer interpretation times can indicate a visualization was more difficult to interpret.

Although performance did not vary between colour-contrast extremes, participants' perceived task difficulty did (Fig. 3): For example, green marks on paint was found more difficult than red marks on paint (red-green contrast), and yellow marks on brick to be more difficult than yellow marks on wallpaper (blue-yellow contrast). Participants' effort may have varied across contrast conditions, reducing differences in performance [40]. However, as the responses were collected at the end of the study, participants might have responded based more on their understanding of colour contrast than how they actually felt during the task. Nevertheless, colour-contrast enhancing methods for ubiquitous visualizations could be valuable to enhance user comfort as they move across different environments [9, 54], provided they can maintain coherent colour codes.

Furthermore, while we found limited support towards the effects of *Lighting intensity* and *Background* on interpretation performance, our results provide insights on the minimal viewing conditions required to affect the usability of OST displays for visualization applications. Specifically, lighting conditions reducing an OST display's luminance contrast to 42% (e.g., $\approx 350\text{cd/m}^2$ on the Hololens 2). As OST display technology improves, we believe lighting intensities causing this contrast will grow outside of typical use conditions, and reduce the need to consider lighting intensity as a constraint. However, we argue there is need to explore lighting-based and background-based adjustments on the aspect of comfort, particularly in more complex and dynamic environments: users may compensate for more complex viewing conditions by putting more effort (Fig. 4), which could in turn increase fatigue and reduce application use times. For example, a navigation application on an OST-HMD could display a map (screen-fixed) or coloured glyphs (world-fixed) to help tourists traverse through an area or exhibition. As they traverse different spaces, the legibility of the visualizations may fluctuate: a green arrow or line tracing the path to a destination could be easy to follow through hallways, but appear faded outdoors as the user traverses parks or sunlit areas. In such cases, context-aware techniques could modify the visualizations' colours to be more noticeable, or alter the elements' brightness to reduce eye strain in dimlit areas, or to increase contrast outdoors.

5.2 Effects on colour perception

5.2.1 Colour perception accuracy

Our hypotheses focused on the potential effects of colour blending [16] on colour perception accuracy through changes in *Lighting intensity* (**H1.2**), *Background* (**H2.2**), and colour contrast (**H4.2**). We additionally considered the effects of *Mark colour* independently (**H3.2**) based on prior work [15]. We discuss results of both luminance and chromaticity *shifts* (i.e., the mean perceived distortions).

Mark colours were on average perceptibly mismatched across conditions (Fig. 5), with $L^*a^*b^*$ perceptual distances ($\Delta L = 8.81$, $\Delta a = 9.61$, $\Delta b = 12.5$) ≤ 1.30 times larger than the just-noticeable differences (JNDs) for the $1.17°$ visual marks used in our study [43].

Distortions differed between *mark colours*. Similar to the results of Gabbard *et al.* [15], red had the largest chromaticity shift (Fig. 6), nearly twice as large as all other colours. Blue was the least distorted, making it a more robust colour to encode information. Furthermore, participants overestimated the lightness of all colours except yellow (Fig. 6), with green having the largest mean shift. Differences between colours may be due to their relative footprints in colour gamuts and our colour palette (Fig. 2)): participants had fewer *yellow-like* colours to pick from across lightness, compared to green. Additionally, participants may have struggled to match colours between the two displays, as they were not given any colour references. We thus **retain H3.2**.

We found no support for our hypothesis regarding *Background-Mark colour* contrast. Low-contrast (e.g., paint and green marks), and high-contrast (e.g., paint and red marks) responses did not significantly differ in chromaticity and luminance shifts, and so we **reject H4.2**. Furthermore, we observed no differences in shifts between *Lighting intensities*. Although prediction ellipses (Fig. 7) differed, their position and size do not suggest colour responses were more blended with the backgrounds. The OST-HMD's luminance contrasts (70% vs. 42%) may have been too large to accentuate colour blending or affect participants' ability to accurately scission the colours [10]. We therefore **reject H1.2**.

Lastly, the effects of *Background* were limited to differences in luminance matching, with larger shifts for the paint background. This may be due to the paint background having a lighter and more saturated colour than the others: participants may have struggled to accurately discount the more apparent colour from the marks [10, 37]. Moreover, although chromaticity prediction ellipses differed between backgrounds (Fig. 7), we did not observe any trends that would indicate colour blending to have affected participants' colour perception (e.g., colour responses being closer to the background's chromaticity). The viewing conditions (e.g., texture, number of colours) may have been too simple for colour blending to affect the participants. Additionally, participants did not have to interact with the background, potentially reducing its effects [40]. Nevertheless, we note the directions of the chromaticity shifts were likely affected by our illuminant's CCT (3000K) [6, 20]. A different CCT, especially under brighter lighting conditions, would likely alter shift directions by modifying the colour of the light reflected from the background. However, based on our results, we posit colour responses under a different illuminant would not lead to significantly different chromaticity shifts between *Backgrounds*. We thus only **partially retain H2.2**.

5.2.2 Variability in colour perception

Dispersion measures capture how colour responses vary between participants: the larger the dispersion, the more different the perceptions [15]. We found differences in luminance dispersions between *Mark colours*, with red and yellow having smaller dispersions, indicating a higher agreement between participants. In contrast to prior work conducted outdoors on an OST head-up display [15], we observed no main effect of *Mark colour* on chromaticity dispersion. We also found overlaps between prediction ellipses of blue and white (Fig. 7), but note the differences in luminance may be sufficient for the colours to not be misinterpreted for each other. Furthermore, we posit the prediction ellipses may vary when evaluated on multi-coloured visualizations, as participants' perception can be further affected by simultaneous contrast effects [22, 48] caused by the other mark colours as well as the background. Finally, beyond individual differences in colour vision, we posit dispersion levels may be partly due to the non-uniformities of the OST-HMD [26, 46, 51].

5.2.3 Perspective on Colour-Coding

Our results inform on the challenges in designing colour-coded visualizations on OST displays. We observed colours are perceived as distorted differently in terms of chromaticity and luminance, posing potential interpretation issues.

First, messages encoded through specific colours may not be accurately perceived by users, leading to information loss. For example, red or white may be perceived closer to pink and light blue (Fig. 5) when the visual marks are small. Including a reference colour (e.g., legend) on the OST display could mitigate this risk, but the colour may not appear equally distorted due to display spatial distortions [26, 51]. However, we note most mean colour matches (Fig. 5) appear as the same or similar hue as the original colour. Encodings could thus be preserved in spite of the original colours appearing different, provided colours with similar appearances are not used to encode distinct information. For example, small red glyphs on a vehicle's OST heads-up display alerting the user of an imminent danger outside may be interpreted correctly even if their appearance is shifted towards pink, unless pink hues are used for non-urgent notifications. Based on our results, we suggest relying on less chromatically distorted colours, such as blue, green, and yellow.

Secondly, variations in perception among users might result in different interpretations of colour maps, if the step size between colours is smaller than these perceptual differences. To design colour maps for visualizations on OST displays, dispersion measures and existing colour JND models for traditional displays [43] can be considered as initial references. We suggest using less *shifted* and *dispersed* colours, such as blue, as a foundation for monochromatic colour maps. However, we advocate for further investigation to determine whether their effectiveness generalizes across different lightness levels. For example, the use of lightness-based colour maps on OST displays may result in darker steps being more difficult to interpret than others. In the case of multi-coloured maps, priority should be given to colours with no overlap between prediction ellipses and substantial perceptual differences (e.g., blue and yellow).

Lastly, despite previous reports highlighting colour blending as a possible usability concern for OST displays [10, 16, 23], our findings show it did not noticeably impact participants' colour perception *within the specific environments examined in our study*. This refers to static, low-visual-complexity settings where the OST display maintains $\geq 42\%$ luminance contrast. We contend that additional research in more complex environments is essential to extend our understanding of colour blending to more *realistic* conditions.

5.3 Limitations

We acknowledge two main limitations in our approach. First, our *environments* were limited to only two lighting intensities and three backgrounds, restricting the depth of our evaluation. The chosen lighting intensities may not have been sufficient to capture the evolution of user performance across a broader range of display transparencies. Our illuminant (3000K) additionally represents only a small subset of light colours encountered in real-life conditions which can affect colour perception [6, 20]. Furthermore, our environments were static and relatively simple in complexity (e.g., lack of clutter, textures, and a limited selection of colours), failing to encompass the diverse array of environments users encounter in their daily lives. Lastly, while our task selection allowed for the control of critical visual factors (e.g., mark size, visualization type), it also constrained the generalizability of our findings beyond 2D scatter charts.

Secondly, our selection of OST display introduced limitations to the precision of our approach. In our task prototypes, we observed spatial non-uniformities in the Hololens 2, leading to perceptible distortions in colours and brightness outside the display's center. Although we structured our study to minimize the impact of these distortions on participants' performance, we cannot separate the effects of the display from those of the studied factors influencing our results. Finally, we acknowledge that our choice of colorimeter constrained the precision of our colour-tuning process between the Hololens 2 and the tablet, potentially contributing to some of the colour variations observed in our study.

5.4 Future work

Our upcoming efforts involve delving into the design of colour maps for visualizations on OST-HMDs and assessing their usability in relation to the user's environment for ubiquitous visualization applications. Furthermore, we intend to explore the disparities between user performance and the user's workload and comfort concerning visualization contrast in more varied, intricate and realistic environments. This research will contribute valuable insights into enhancing and sustaining user performance and experience, particularly as users navigate diverse environments, including approaches such as encoding-preserving colour-correction methods.

6 Conclusion

In this work, we investigated how lighting conditions and colours of physical environments may affect visualization interpretation and colour perception accuracy on OST displays to inform on the design of ubiquitous visualizations. We observed interpretation performance was only impacted by the more visually complex background, when the display had increased transparency due to brighter lighting conditions. Although participants found low-colour-contrast scenarios to be more challenging, their performance did not decline, suggesting a need to design for visual comfort. Furthermore, we found colours to be on average perceptibly and distinctively mismatched, suggesting blue and yellow are more robust colours for colour coding on OST displays than red. However, our results do not indicate colour blending to have noticeably affected participants' colour perception. Taken together, our findings inform on the visual contexts in which interpretation of colour codes of ubiquitous visualizations on OST displays may be affected, and highlight the need to design for both performance and visual comfort. We envision future work exploring colour map design and techniques enhancing visual comfort for ubiquitous visualization applications on OST displays in complex and realistic environments.

Acknowledgments

This project was supported by an NSERC Discovery Grant awarded to Pourang Irani. We thank Julia Petrie, and Kai Hagen for their help in creating the figures.

References

[1] F. Alallah, Y. Sakamoto, and P. Irani. Exploring the Need and Design for Situated Video Analytics. In *Symposium on Spatial User Interaction*, pp. 1–11. ACM, oct 2020. doi: 10.1145/3385959.3418458

[2] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*, pp. 111–117, 2005. doi: 10.1109/INFVIS.2005.1532136

[3] N. Bressa, H. Korsgaard, A. Tabard, S. Houben, and J. Vermeulen. What's the situation with situated visualization? a survey and perspectives on situatedness. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):107–117, 2022. doi: 10.1109/TVCG.2021.3114835

[4] C. A. Brewer. Prediction of simultaneous contrast between map colors with hunt's model of color appearance. *Color Research & Application*, 21(3):221–235, 1996. doi: 10.1002/(SICI)1520-6378(199606)21:3<221::AID-COL4>3.0.CO;2-U

[5] J. Brookes, M. Warburton, M. Alghadier, M. Mon-Williams, and F. Mushtaq. Studying human behavior with virtual reality: The unity experiment framework. *Behavior research methods*, 52(2):455–463, 2020.

[6] S. Chen and M. Wei. Real-world environment affects the color appearance of virtual stimuli produced by augmented reality. *Color and Imaging Conference*, 27(1):237–237, 2019. doi: 10.2352/issn.2169-2629.2019.27.42

[7] J. H. Clark. The ishihara test for color blindness. *American Journal of Physiological Optics*, 5:269–276, 1924.

[8] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[9] J. David Hincapie-Ramos, L. Ivanchuk, S. K. Sridharan, and P. Irani. SmartColor: Real-time color correction and contrast for optical see-through head-mounted displays. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 187–194. IEEE, sep 2014. doi: 10.1109/ISMAR.2014.6948426

[10] T. Downs and M. Murdoch. Color layer scissioning in see-through augmented reality. *Final Program and Proceedings - IS and T/SID Color Imaging Conference*, 2021-Novem:60–65, 2021. doi: 10.2352/issn.2169-2629.2021.29.60

[11] T. Dwyer, K. Marriott, T. Isenberg, K. Klein, N. Riche, F. Schreiber, W. Stuerzlinger, and B. H. Thomas. *Immersive Analytics: An Introduction*, pp. 1–23. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-030-01388-2_1

[12] M. A. Elliott, C. Nothelfer, C. Xiong, and D. A. Szafir. A design space of vision science methods for visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1117–1127, 2021. doi: 10.1109/TVCG.2020.3029413

[13] B. Ens, B. Bach, M. Cordeil, U. Engelke, M. Serrano, W. Willett, A. Prouzeau, C. Anthes, W. Büschel, C. Dunne, T. Dwyer, J. Grubert, J. H. Haga, N. Kirshenbaum, D. Kobayashi, T. Lin, M. Olaosebikan, F. Pointecker, D. Saffo, N. Saquib, D. Schmalstieg, D. A. Szafir, M. Whitlock, and Y. Yang. Grand Challenges in Immersive Analytics. In *Proceedings of the 2021 CHI Conference on Human Fac-*

tors in Computing Systems, pp. 1–17. ACM, may 2021. doi: 10.1145/3411764.3446866

[14] A. Erickson, K. Kim, G. Bruder, and G. F. Welch. Exploring the Limitations of Environment Lighting on Optical See-Through Head-Mounted Displays. *Proceedings - SUI 2020: ACM Symposium on Spatial User Interaction*, 2020. doi: 10.1145/3385959.3418445

[15] J. L. Gabbard, M. Smith, C. Merenda, G. Burnett, and D. R. Large. A Perceptual Color-Matching Method for Examining Color Blending in Augmented Reality Head-Up Display Graphics. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2020. doi: 10.1109/TVCG.2020.3044715

[16] J. L. Gabbard, J. E. Swan, J. Zedlitz, and W. W. Winchester. More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. In *2010 IEEE Virtual Reality Conference (VR)*, pp. 79–86. IEEE, mar 2010. doi: 10.1109/VR.2010.5444808

[17] E. R. Girden. *ANOVA: Repeated measures*. Number 84. sage, 1992.

[18] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988.

[19] N. Hassani. *Modeling color appearance in augmented reality*. Rochester Institute of Technology, 2019.

[20] N. Hassani and M. J. Murdoch. Investigating color appearance in optical see-through augmented reality. *Color Research and Application*, 44(4):492–507, aug 2019. doi: 10.1002/col.22380

[21] C. Healey. Choosing effective colours for data visualization. In *Proceedings of Seventh Annual IEEE Visualization '96*, pp. 263–270, 1996. doi: 10.1109/VISUAL.1996.568118

[22] E. G. Heinemann. Simultaneous brightness induction as a function of inducing-and test-field luminances. *Journal of experimental psychology*, 50(2):89, 1955. doi: doi/10.1037/h0040919

[23] D. Hirobe, Y. Uranishi, J. Orlosky, S. Shirai, P. Ratsamee, and H. Takemura. Characteristics of Background Color Shifts Caused by Optical See-Through Head-Mounted Displays. In H. Uchiyama and J.-M. Normand, eds., *ICAT-EGVE 2022 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2022. doi: 10.2312/egve.20221285

[24] M. Hussain and J. Park. Effect of transparency levels and real-world backgrounds on the user interface in augmented reality environments. *International Journal of Human–Computer Interaction*, 0(0):1–10, 2023. doi: 10.1080/10447318.2023.2212218

[25] Y. Itoh, M. Dzitsiuk, T. Amano, and G. Klinker. Semi-parametric color reproduction method for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 21(11):1269–1278, 2015. doi: 10.1109/TVCG.2015.2459892

[26] Y. Itoh, T. Langlotz, J. Sutton, and A. Plopski. Towards Indistinguishable Augmented Reality, jul 2021. doi: 10.1145/3453157

[27] E. L. Kaufman, M. W. Lord, T. W. Reese, and J. Volkmann. The discrimination of visual number. *The American Journal of Psychology*, 62(4):498–525, 1949.

[28] T. Langlotz, M. Cook, and H. Regenbrecht. Real-time radiometric compensation for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(11):2385–2394, 2016. doi: 10.1109/TVCG.2016.2593781

[29] K.-K. Lee, J.-W. Kim, J.-H. Ryu, and J.-O. Kim. Ambient light robust gamut mapping for optical see-through displays. *Optics Express*, 28(10):15392, may 2020. doi: 10.1364/oe.391447

[30] C. Li, Z. Li, Z. Wang, Y. Xu, M. R. Luo, G. Cui, M. Melgosa, M. H. Brill, and M. Pointer. Comprehensive color solutions: Cam16, cat16, and cam16-ucs. *Color Research & Application*, 42(6):703–718, 2017. doi: 10.1002/col.22131

[31] G. Lin, M. Haynes, S. Srinivas, P. Kotipalli, and T. Starner. Towards finding the optimum position in the visual field for a head worn display used for task guidance with non-registered graphics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(1), mar 2021. doi: 10.1145/3448091

[32] M. A. Livingston, J. H. Barrow, and C. M. Sibley. Quantification of contrast sensitivity and color perception using head-worn augmented reality displays. In *2009 IEEE Virtual Reality Conference*, pp. 115–122, 2009. doi: 10.1109/VR.2009.4811009

[33] F. Metelli. The Perception of Transparency. *Scientific American*, 230(4):90–99, 1974.

[34] A. Michelson. *Studies in Optics*. University of Chicago science series. University of Chicago Press, 1927.

[35] K. T. Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359(1):381–400, 1985. doi: 10.1113/jphysiol.1985.sp015591

[36] T. Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press, 2014.

[37] M. J. Murdoch. Brightness matching in optical see-through augmented reality. *J. Opt. Soc. Am. A*, 37(12):1927–1936, Dec 2020. doi: 10.1364/JOSAA.398931

[38] A. Prouzeau, Y. Wang, B. Ens, W. Willett, and T. Dwyer. Corsican twin: Authoring in situ augmented reality visualisations in virtual reality. In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3399715.3399743

[39] P. Reipschlager, T. Flemisch, and R. Dachselt. Personal augmented reality for information visualization on large interactive displays. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1182–1192, 2021. doi: 10.1109/TVCG.2020.3030460

[40] M. Satkowski and R. Dachselt. Investigating the impact of real-world environments on the perception of 2d visualizations in augmented reality. In *Conference on Human Factors in Computing Systems - Proceedings*, pp. 1–15. ACM, may 2021. doi: 10.1145/3411764.3445330

[41] S. Smart and D. A. Szafir. Measuring the separability of shape, size, and color in scatterplots. *Conference on Human Factors in Computing Systems - Proceedings*, pp. 1–14, 2019. doi: 10.1145/3290605.3300899

[42] M. Stone, D. A. Szafir, and V. Setlur. An engineering model for color difference as a function of size. *Final Program and Proceedings - IS and T/SID Color Imaging Conference*, 2014-Janua(January):253–258, 2014.

[43] D. A. Szafir. Modeling Color Difference for Visualization Design. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):392–401, jan 2018. doi: 10.1109/TVCG.2017.2744359

[44] E. ToolBox. Illuminance - recommended light level, 2004.

[45] J. Vermeulen, C. Collins, R. Dachselt, P. Irani, and A. Joshi. Reflections on ubiquitous visualization. In *Mobile Data Visualization*, pp. 263–316. Chapman and Hall/CRC, 2021.

[46] M. Wang, J. Ding, D. M. Levi, and E. A. Cooper. The effect of interocular contrast differences on the appearance of augmented reality imagery. *ACM Trans. Appl. Percept.*, aug 2023. Just Accepted. doi: 10.1145/3617684

[47] C. Ware. *Information Visualization: Perception for Design*. Elsevier Science & Technology, 2012.

[48] C. Ware and W. B. Cowan. Changes in perceived color due to chromatic interactions. *Vision Research*, 22(11):1353–1362, 1982. doi: 10.1016/0042-6989(82)90225-5

[49] C. Ware, T. L. Turton, R. Bujack, F. Samsel, P. Shrivastava, and D. H. Rogers. Measuring and modeling the feature detection threshold functions of colormaps. *IEEE Transactions on Visualization and Computer Graphics*, 25(9):2777–2790,

2019. doi: 10.1109/TVCG.2018.2855742

[50] W. Willett, Y. Jansen, and P. Dragicevic. Embedded Data Representations. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):461–470, jan 2017. doi: 10.1109/TVCG.2016.2598608

[51] L. Zhang and M. J. Murdoch. Color matching criteria in augmented reality. *Journal of Perceptual Imaging*, 1(1):010506–1–010506–1, 2018. doi: 10.2352/J.Percept.Imaging.2018.1.1.010506

[52] L. Zhang and M. J. Murdoch. Perceived transparency in optical see-through augmented reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 115–120, 2021. doi: 10.1109/ISMAR-Adjunct54149.2021.00033

[53] L. Zhang, M. J. Murdoch, and R. Bachy. Color appearance shift in augmented reality metameric matching. *Journal of the Optical Society of America A*, 38(5):701, 2021. doi: 10.1364/josaa.420395

[54] Y. Zhang, R. Wang, E. Y. Peng, W. Hua, and H. Bao. Color Contrast Enhanced Rendering for Optical See-through Head-mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2021. doi: 10.1109/TVCG.2021.3091686