

# A Perceptual Color-Matching Method for Examining Color Blending in Augmented Reality Head-Up Display Graphics

Joseph L. Gabbard<sup>1</sup>, Missie Smith<sup>1</sup>, Coleman Merenda, Gary Burnett, David R. Large<sup>1</sup>  
*<sup>1</sup>Member, IEEE*

**Abstract**— Augmented reality (AR) offers new ways to visualize information on-the-go. As noted in related work, AR graphics presented via optical see-through AR displays are particularly prone to color blending, whereby intended graphic colors may be perceptually altered by real-world backgrounds, ultimately degrading usability. This work adds to this body of knowledge by presenting a methodology for assessing AR interface color *robustness*, as quantitatively measured via shifts in the CIE color space, and qualitatively assessed in terms of users' perceived color name. We conducted a human factors study where twelve participants examined eight AR colors atop three real-world backgrounds as viewed through an in-vehicle AR head-up display (HUD); a type of optical see-through display used to project driving-related information atop the forward-looking road scene. Participants completed visual search tasks, matched the perceived AR HUD color against the WCS color palette, and verbally named the perceived color. We present analysis that suggests blue, green, and yellow AR colors are relatively robust, while red and brown are not, and discuss the impact of chromaticity shift and dispersion on outdoor AR interface design. While this work presents a case study in transportation, the methodology is applicable to a wide range of AR displays in many application domains and settings.

**Index Terms**—D.2.14.a User interfaces, H.1.2.a Human factors, H.5.1.b Artificial, augmented, and virtual realities, I.4.8.a Color

## 1 INTRODUCTION

DESPITE its growing marketability, augmented reality's (AR) implementation in day-to-day mobile usage settings remains hampered by continuing challenges. Current usability issues include unstable or shifting user interface colors and poor text legibility caused by fluctuations in real-world lighting levels and color blending [1]. These distortions resulting from luminance washout and chromaticity shifts can impact both the meaning and salience of AR graphics as perceived by users, render them useless or even distracting [2-4], and may compound issues with depth perception [3, 5, 6]. However, dynamically changing environments and high levels of lighting variability are frequent in outdoor environments and must inevitably be accounted for in AR interface design. Although most technology needed to generate conformal AR interfaces is readily available [2], the technology needs to undergo further refinement and integration into users' environments before it can realize its full potential [7]. Be-

cause of its status as a novel technology, many people are inexperienced with using AR head-up displays (HUDs), which is one manifestation of an optical see-through AR display (head worn AR displays such as Microsoft HoloLens are another). In fact, new users are likely to experience some perceptual biases from the mere presence of an AR interface in their environment [8], making successful design more difficult until AR is established as a mainstream technology. Additionally, since the market for augmented reality is still relatively new, standard guidelines and regulations for AR interface design largely do not exist. Despite this lack of guidance, designers are still expected to contend with the frequent constraints and standard practices common to industrial environments [9, 10].

Even with the continuing technical and perceptual obstacles facing AR, interest in optical see-through AR displays continues to grow and numerous applications are nearing market introduction. The benefits of using AR technology over traditional visual displays include overlaying contextual real-time graphics atop real world referents, resulting in faster and larger information gains than would be available through previous approaches, particularly when conveying spatial information [11, 12]. Although the benefits of AR are being realized in emerging novel head-mounted applications (e.g., as delivered via Microsoft HoloLens and Magic Leap One), its commercial use is largely limited by these technical and perceptual challenges, particularly in the automotive industry where there are likely promising gains to user (i.e., driver) performance. Specifically, when delivered via in-vehicle HUDs, AR graphics can increase drivers' forward situational awareness and cognition while minimizing distraction [13, 14]. AR HUDs allow drivers to attend to the display while

- J.L. Gabbard is with the Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061. E-mail: jgabbard@vt.edu.
- M. Smith is with the Industrial and Systems Engineering Department, Oakland University, Rochester, MI 48309. E-mail: mismith@oakland.edu.
- C. Merenda is with the Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061. E-mail: cjm120@vt.edu.
- G. Burnett is with the Human Factors Research Group, University of Nottingham, Nottingham, NG7 2RD, UK. E-mail: Gary.Burnett@nottingham.ac.uk.
- D.R. Large is with the Human Factors Research Group, University of Nottingham, Nottingham, NG7 2RD, UK. E-mail: david.r.large@nottingham.ac.uk.

still keeping their eyes focused in the direction of the road, resulting in improved divided and selective visual attention as compared to traditional head-down displays [7]. In some use cases, such as navigation, AR HUDs can afford more efficient divided attention as compared to traditional center stack navigation visual displays without negatively impacting driving performance [15].

Nonetheless, there is still much work needed to better understand how AR interface design affects users' perception of AR interface elements, and its resulting effect on user performance; be it in HUD-based driving or in head-worn applications domains. In this paper, we present an approach for assessing different AR interface colors in outdoor settings that examines the degree to which a given interface element can retain its perceived color name under changing environmental conditions. The approach allows for further exploration of individual differences in color perception and presumably resulting performance while using an AR display. While this paper is about optical see-through displays, we believe the method could be applied to other forms of AR (e.g. spatial augmented reality, video pass-through, head-worn AR).

We hope that this method will be replicated by others that study AR interface design, and ultimately be used to help generate design guidelines for AR interface color selection. Further, this approach, along with associated measured color data, could be used by real-time color correction systems to predict whether or not an AR color's shift will be perceived as a different color (i.e., be named a different color), and further to establish acceptable ranges of color blending and shifting for which the color correction system to manage and/or intervene.

## 2 RELATED WORK

Among outdoor optical see-through AR's most significant human factors problems are perceptual issues such as lighting, text legibility, and color blending. Lighting issues in AR are already well known; for example, early attempts to produce AR graphics in an outdoor environment resulted in certain color subsets like magenta and cyan appearing translucent and difficult to see in sunlight due to washout [16, 17]. Generally, users' experiences with AR are subject to change based on the optical elements present, display technology used, and the user's visual context [17]. Under frequent changes in lighting and real-world backgrounds, perceived colors and lighting of created AR graphics can become distorted and result in poor visual presentation for users [18]. Seo, Kang and Park [19] proposed a method to solve visualization incompatibility problems in AR by altering the rendering style of virtual objects as based on real outdoor lighting color data and real weather attributes (weather focused AR correction). These efforts strive to increase the visual fidelity of the AR graphics with respect to the environments (e.g., matching lighting, shadows, specular highlights). While this work could undoubtedly help produce more realistic looking 3D AR models, it is unclear if such methods could identify robust AR color sets for use in AR interfaces, where we define "robust" colors as ones that retain color naming when viewed

atop different background by different users.

Another common visual obstacle in AR is text legibility, which often results from poor lighting and color presentation. With either of these issues, AR text can easily become difficult or impossible to interpret. It is already known that real-world background colors and textures impact AR text legibility within proposed industrial uses of AR. For instance, background changes can reduce the visual performance of 3D textures by distorting and producing masking effects on displayed text [9]. Methods to correct for and maximize text legibility have also been examined. For example, text readability can be improved by maximizing the contrast between AR text and background in optical see-through displays by using saturation and contrast correction and calibration [20]. However, correction techniques for both contrast and polarity depend on both the type of AR being considered and ambient lighting conditions present, especially when evaluating the limits of text readability within work environments. For example, optical see-through AR technology is much more severely impacted by background luminance as compared to video-based AR when used in industrial application settings [10]. Consequently, correction methods must be tailored to, and limited by, each AR application setting which may not be a feasible or conventional solution to these aforementioned obstacles.

Be it text, symbology or 3D graphics, color blending produces additional challenges to AR interface designers and is characterized by the path that light goes through to reach the users' field of view, including emission, reflection, combination with the display, and final product [21]. Color blending and distortion can also impact the effectiveness of AR graphics. Previous work has identified two types of color distortion; render distortion (i.e. the accuracy at which a display can render color) and material distortion (i.e., the extent to which real-world background colors are changed by the display material) [22]. Research on color blending and distortion in AR technology has often focused on exact replication of the desired color, frequently by having participants match colors overlaid onto various backgrounds [23]. Other research that examines color blending superficially in optical see-through AR offers several different classes of approaches to the problem including: empirical user observations, real-time correction algorithms and mathematical modeling of the phenomenon, and measurement-based color studies. We present related works in each of these areas below, followed by relevant work in color perception.

### 2.1. Empirical User Observations

Early work by Thomas et al., [24] aimed to identify a set of colors that were effective in outdoor (i.e., bright) environments. Their method involved gathering subjective human feedback regarding the visibility and opaqueness of AR UI elements of nine prescribed colors and four intensities (36 total combinations) against four different outdoor backgrounds. Gabbard, Swan, Hix and collaborators present a method for empirically examining the legibility of different AR text colors against prescribed backgrounds [25, 26]. Several empirical studies were conducted at Politecnico di Bari on the effect of

color blending on text legibility in industrial AR environments; specifically examining color and style [27], contrast polarity [10], and text luminance [28] with optical see through AR displays, and text color and background surface texture when using projection-based spatial augmented reality [9]. These works collectively represent examples empirical approaches that aim to understand color blending by exposing participants to AR interface colors in various contexts and documenting participant feedback and task performance. However, these approaches do not provide the data needed to understand how a particular color may undergo color blending and how the resulting color may be perceived by different users.

With the gaining appreciation of dark mode UIs, [29] conducted an empirical user study using a HoloLens optical see through AR HWD to examine the effect of dark mode color schemes on user acuity, fatigue and usability. Users completed text reading tasks and landolt C visual acuity tests with graphics rendered in positive and negative contrast against three backgrounds (uniform, lightness distortions, and chromatic distortions) under two lighting conditions (low and high relative to indoor levels). Results suggest dark mode could be an effective approach to reduce fatigue and increase effectiveness of AR graphics, especially with indoor lighting levels and perhaps with video pass-through AR. However, more work is needed to test whether their results would hold under outdoor lighting levels.

## 2.2. Real-time Adaptive Approaches

Gabbard et al. [26] presents one of the earliest color correction approaches that employed a method to examined different real-time correction algorithms aimed at increasing the legibility of text on common outdoor backgrounds. This method applied algorithms adapted from other domains in an effort to maximize luminance contrast. This work aimed to identify what color characteristics were most important to modify in future, real-time adaptive interface algorithms.

Sridharan et al. [22] used binned profiles to describe an accurate color blending model and produce color correction in optical see-through displays, and subsequent work resulted in "SmartColor" which used computer algorithms to employ correction, contrast, and show-up contrast (when natural contrast is too low) management strategies to mitigate color blending and loss of legibility [30]. Still, correcting for color blending remains a complex issue in AR, especially since a successful calibration strategy for one AR display type may not translate well to other displays with differing specifications.

In 2008, the groundwork for a more sophisticated approach to real-time adaptive AR UIs was laid by Grundhöfer and Bimber [31]. In this work, they present a real-time adaptive radiometric compensation technique to support the projection of images onto colored and textured surfaces; a color blending problem in spatial augmented reality that is very much similar to the color blending phenomenon that occurs in optical see-through AR head-worn display (HWD). Their approach creates a compensation image from a per-pixel image of the projection surface which is then used to minimize

geometric distortions and color blending caused by the specific real-world background (in this case, a projection surface). The authors performed a preliminary user study where the adaptive algorithm was preferred over a static approach when viewing images and videos projected onto a stone background.

In 2016, Langlotz, Cook and Regenbrecht [32], presented a method for mitigating the effect of color-blending in optical see-through HWDs by also using real-time radiometric compensation. As with other similar approaches, the method assumes that a camera is able to capture the users' view in real-time to establish a compensation image. The work explores the use of three different algorithms to compensate for color blending, each of which the authors argue have various strengths and weaknesses. The work demonstrates that per-pixel corrected images are significantly better than uncorrected images (as evidenced by images captured through the AR HWD). However, the results of their user study suggest that while the approach is promising the perceptual improvements in image quality were marginal.

Silva et al. [33] presented an approach to mathematically model color correction in AR and included a user study with a color matching task conceptually similar to our method presented herein that varied AR colors against differing backgrounds and foregrounds. The authors concluded with a statement of how difficult and complex the issue of AR color correction is. Thus, methods that can potentially help inform real-time color correction methods could be beneficial to the field. Specifically, we posit that our method could be useful in defining more flexible boundaries in which a real-time correction algorithm may need to operate (yet still retain usability and color semantics).

In other approaches to mathematically modelling the problem, Itoh et al. [34] present a color calibration method for optical see-through AR that is based on a model of AR display optics and human perception as experienced through an AR HWD. The goal of the work is to create a calibration routine that affords pre-processing of AR source images such that the image produced by an AR HWD maintains its originally colored appearance. The authors frame the problem as a semi-parametric model, separating non-linear color distortions from linear color shifts. Results are quite promising, and demonstrate how properly calibrated input to an HWD can result in rendered images that, when measured by "an industrial camera" (presumably a form of colorimeter), are more closely aligned to the intended rendered color than an image that is not pre-processed. However, the method assumes that an image of the scene exists, so as to form the basis by which corrections are made. Further, the study does not examine individual differences in color perception, however, the authors suggest conducting user studies as possible next steps.

Fukiage, Oishi and Ikeuchi [35], describe a method to accurately and linearly predict the visibility of the graphics overlaid onto background textures, and then optimize a blending parameter so as to enhance the visibility of the blended AR content. The authors compare two blending

methods: the first locally optimizes a blending parameter such that the visibility of the blended object achieves a specified visibility level, while the second adaptively (and locally) ensures the visibility of an AR graphics specifically for optical see-through AR displays. The authors note the challenges of such approaches, especially when light from real-world scenes is brighter than the maximum brightness capability of the AR display; which is often the case with today's state-of-the-art AR HWDs in outdoor daytime usage contexts.

Mori et al. present *BrightView* [36], a creatively different approach to mitigating some of the challenges associated with color blending in bright outdoor environments; namely the phenomenon termed "washout due to luminance" in [4]. In their work, the authors attach a liquid crystal filter to an AR HWD to attenuate the ambient lighting in real-time and thus, increase the perceived brightness of AR graphics without altering users' perceived brightness of the real-world (although the actual brightness is attenuated). The authors detail an AR HWD prototype and evaluate its performance on users' brightness perception across three scenes. Interestingly, participants perceived changes introduced by the real-time BrightView as increased brightness in virtual objects as opposed to decreased brightness in the real-world scene. Their work focuses exclusively on the luminance of the display relative the scene, and does not address chromaticity shifts nor color perception per se.

Itoh et al., 2019, presents a completely novel approach to adaptively addressing color blending by developing a prototype AR optical see-through display with light attenuation capabilities that spatially remove (i.e., filter) colored light from the real-world scene on a per-pixel basis [37]. The authors do not employ a user study in this early work, however benchmark tests provided evidence that the display can indeed successfully modify background colors in effective and deterministic ways and in some cases can enhance colors. Since the display operates on a per-pixel basis, it is conceivable that this approach could assist users with color vision deficiencies, and once calibrated to an individual's perceptual system, could generate alternate, discernible UI colors when needed.

### 2.3. Measurement-based Color Studies

Gabbard [4] employed a strictly measurement-based approach to quantifying the effects of color blending by placing a colorimeter in front of an AR display, whilst presenting 27 full-screen colors on five backgrounds made from physical materials, four colored poster backgrounds, and one white poster background. The results classify the measured effects of color blending into four categories: washout due to chromaticity, washout mostly due to luminance, washout due to both chromaticity and luminance, and, linear shift in chromaticity.

In the work presented herein, we focus explicitly on chromaticity and luminance as independent aspects of colored light. Chromaticity is defined as the quality of a color independent of its luminance, and in, for example the HSV color model, would be defined by hue and saturation. Luminance is the property of light that describes "brightness" of light,

independent of its chromaticity. Note that these terms are not to be confused with chrominance, a term often used in luminance-chrominance models whereby a color is defined by a combination of both chrominance (color) and luminance (brightness), and importantly whereby any attenuation to the luminance will have a proportional effect on the resulting chrominance [38].

Regardless of the method, it is the consensus of researchers in this space that correcting for color blending remains a complex issue in AR, especially since a successful calibration strategy for one AR display type may not necessarily translate well to other displays with differing specifications. Thus, there is a need for more psychophysical, user-based methods that investigate the relationship between physical stimuli (e.g., AR graphics and real-world backgrounds and lighting) and the sensations and perceptions they produce (e.g., subjective color judgements). In this area, yet despite the application of various methods described above, published work that employs actual users to understand the effects of background and ambient lighting on AR color perception are scant. Indeed, a systematic review of 10 years of AR usability studies between 2005 to 2014 [39], reveals just six user studies on color perception in AR [17, 20, 25, 26, 40, 41] out of 369 user studies reported in 291 papers.

### 2.4. Color Perception

Colorimetry, or the science of color perception, measurement and reproduction, is a complete field of science in itself and a full treatment of even the fundamental tenants is beyond the scope of this work. Instead, we cover just a papers that are directly related to the work described here. For a comprehensive entry point into the field consider the following [42, 43], and for a brief overview of color models commonly used in computer graphics applications, such as AR, see [44, 45].

Some of the most relevant seminal early work in color focused on identifying quantitative links between physical pure colors (i.e. wavelengths) and how colors are perceived in the human visual system (HVS). In the late 1920s, in separate but similar intellectual endeavor, William David Wright and John Guild performed color matching experiments (e.g., [46] and [47] respectively), where participants viewed a circular split screen (or bipartite field) comprised of two halves: a target color and an adjustable color. Using a method of adjustment, participants independently altered the luminance of three primary light sources (red, green and blue) until the adjustable color perceptually matched the target color. Interestingly, Wright's study employed 10 participants and Guild's only 7, and both recruited participants "whose colour vision was found to be free from any marked abnormality". Wright and Guild's work generated data that formed the basis of the *Commission Internationale de l'éclairage* (CIE) RGB color space, from which the CIE XYZ color space was later developed. To date, the 1931 CIE XYZ color space (and its subsequent modifications over the years) has been a foundational component of color science for nearly 80 years allowing researchers to mathematically explore the relationship between color (be it from electronic devices, dedicated lamps or even paint pigments) and human perception.

In the 1940's, MacAdam employed a method of adjustment similar (but distinctly different) than that of Wright and Guild, to present 25 target colors to a set of nine participants, and asked participants to match the target color to the best of their ability [48]. MacAdam then plotted individual responses to each test color and studied their distribution in the 1931 standard chromaticity diagram (what we refer to herein as the CIE x-y plane). MacAdam's results found that for any specific test color, the set of user responses fell into ellipsoids of varying shape and orientation and that each ellipsoid represented a set of colors that are indistinguishable to the average observer. In some ways what MacAdam found is unsurprising, however, for the purposes our work, it is interesting that the size and orientation of each ellipsoid was different depending on the target color, providing evidence that not all colors are perceptually equivalent. Said another way, some colors may be more easily distinguished from its neighbors than others. As for the application of MacAdam ellipses in this work, we can consider two cases. First, we can envision an AR user interface color that undergoes color blending, but the blending is such that the resulting color is still within the source color's ellipse. If we place that AR color atop different backgrounds, and find similar results, then we could argue that this color is rather robust since most participants will not be able to distinguish the source color from the resulting blended color. Second, more broadly, we can consider a variation of MacAdam ellipses which are likely larger than the original by which the average observer would name the color the same as the source/target color. It is this perspective which we embrace for this work, as we expect color blending will result in shifts outside the original MacAdam ellipses but still may not result in said color being named differently. This perspective is not unlike that of Berlin and Kay's work developing the World Color Survey [49] (described briefly below). Lastly, while MacAdam noted that color differences are occasionally different across a given observer's right and left eye (even those that pass standard eye tests), like Wright and Guild's work, MacAdam also recruited subjects that had "normal": color vision as verified through "all the usual tests" at that time.

More recently, across a number of fields including human factors and human-computer interaction, there has been increasing interest in understanding and supporting individual differences in perception, cognition and capabilities of various kinds, including color perception. While most readers are likely familiar with common color vision deficiencies (CVDs) such as red-green color blindness (the most common affecting ~8% of males and less than 1% of females worldwide) and blue-yellow color blindness, work by Flatla and others' work have provided valuable insight into the factors that affect an individuals' ability to differentiate one color from another. Specifically, Reineke, Flatla and Brooks [50], demonstrate how a user population's ability to differentiate color may be measured and modeled by UI designers. They describe an open-source color differentiation test (WebCDT) that addresses perception of computer-generated light as viewed under varying lighting conditions, as well as a design tool,

ColorCheck, that, given a source image will predict specifically which portions of the image likely contain indistinguishable colors. A key finding from this work is the fact that color discrimination is an individual experience, that of course is affected by common CVDs (such as those methods above) but also by other internal factors such as age, gender, fatigue, as well as, external factors such as monitor brightness and environmental lighting. And moreover, that ellipsoids of indistinguishable colors are not only larger than previously assumed, but larger for men than women (controlling for likelihood of inherited CVDs), as well as larger for outdoor environments as compared to indoor environments. The latter finding being especially relevant for promising AR settings such as outdoors and even in driving as the present study examines

Flatla and Gutwin also produced a method for identifying individual models of color differentiation that require no a priori knowledge of user's color vision and is sensitive to real-time contexts such as lighting and user fatigue [51, 52]. As mentioned in our section on lessons learned, we envision ways to incorporate these models into the future iterations of our proposed method.

Our work presented herein does not aim to determine what set of colors may be indistinguishable from each other, but instead help AR UI designers identify colors that are more likely than not to be perceived *by name* as designers' intended given both color blending phenomenon and individual differences in color perception.

### 3 METHOD OBJECTIVES & DESCRIPTION

For this work, we posit that perfectly controlled color rendering may not be necessary to adequately convey an intended color-coded message. Therefore, determining which colors are most effective for text readability and symbol recognition could offer AR designers a foundational set of guidelines to assist in creating future AR applications. General principles and patterns that moderate interactions between an AR graphic's chromaticity, luminance, and real-world backgrounds are still not well understood, and this knowledge could help designers understand and consequently limit subsequent negative impact on users' color perception and recognition. In addition to exploring the relationship between common rendering distortions (such as color shift and wash-out), further work is needed to quantify how these distortions impact user behavior and color recognition performance. The proposed method is one such approach for addressing these issues.

As detailed color correction techniques already exist, the purpose of this work is not to identify exact or subtle interactions between AR displays and backgrounds nor to predict text legibility or resulting color rendering. Rather, the work presented herein aims to present a method for assessing the robustness of candidate colors when presented via optical see-through AR displays, and to identify patterns in user performance and color perception when using AR HUDs in different driving-relevant contexts. Moreover, to make effective

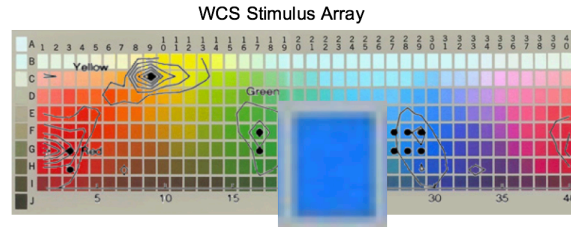
use of AR interfaces, we seek to understand which color subsets are most robust when used in real-world outdoor environments as a function of both visual perception (i.e., acuity and visual search performance) and verbal identification (i.e., color naming).

Thus, we developed a systematic method to afford perceptual selection of AR color, and then connect these selections to set of responses in a color space in which we can measure and perform calculations previously established for visual color perception. The method affords capturing and analyzing individual differences in color perception, which are inherently at play when a variety of different people use an AR interface (and is especially important to understand when the AR interface is subject to color blending). Our method employs five steps as outlined below.

**Step 1: Choose target color(s) to study.** While each study may have different specific aims, we assume that there are a set number of user interface colors that can be used to render interface elements. Identify the number of distinct colors needed to adequately encode the user interface, and then the set of colors by name (e.g., red, blue, yellow). Next, chose a specific color chip from the World Color Survey (WCS) stimulus array [49], keeping in mind that chips designated as the color naming centroids (for the language in which the study takes place) are likely to be the best representative color. For example, in Figure 1, the dots represent the best examples of English color terms based on [53]).

For context, we briefly describe the WCS and the color palette used for this work. The WCS leveraged a global network of “linguist-missionaries” to access speakers of 110 unwritten

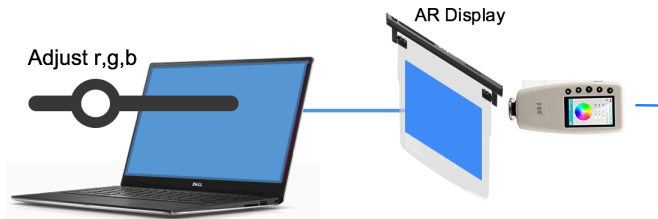
**STEP 1: Choose target color(s) from WCS stimulus array in which to study.** For this example, we chose the color blue.



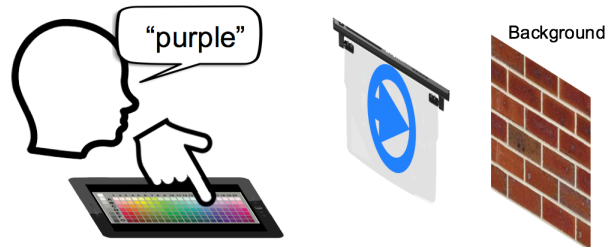
**STEP 2: Render the WCS Stimulus Array on a tablet.** We used a truncated WCS palette due to resolution and size constraints. Measure the CIE xyY color of each target color using a colorimeter.



**STEP 3: Tune the AR Source Colors.** Render a color image to AR display and measure the CIE xyY values. Tune the presentation of color sent to AR display (e.g., r,g,b values) until target CIE xyY measurements are achieved.



**STEP 4: Participants Name and Match Perceived AR Colors.** Render r,g,b colors as AR graphics atop real-world backgrounds. Have participants name perceived color aloud and match to WCS palette.



**STEP 5: Map Participant Palette Responses to CIE Color Space & Analyze.** Map tablet-rendered swatches to CIE color space via colorimeter measurement.

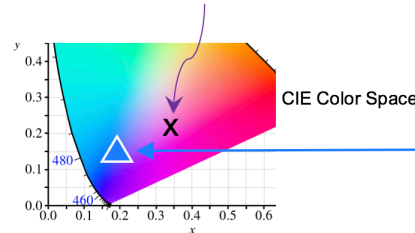


Fig. 1. Overview of proposed method to assist in quantifying the perceived effects of color blending in optical see-through AR.

languages from non-industrialized societies representing forty-five language families [54]. Participants were shown, one by one, each of the 330 color chips contained in the established Munsell Chip set [55, 56] and then asked to name the color in their native language. The Munsell Chip Set (hereafter referred to as the WCS Palette), contains color chips systematically generated using 40 gradations of hue crossed with 8 levels of lightness (Munsell value) at maximum saturation (Munsell chroma) resulting in 320 chips. The WCS palette further contains 10 grayscale “colors” ranging from black to white modified using 10 different levels of lightness. For each of the color categories named, participants were asked to choose one chip (from the 330 presented on a single palette) that represented the “best example” of each color category. Results of the work yielded defined clusters of 11 basic colors that are universally named, as well as color centroids for each of the 11 color categories.

**Step 2: Render the WCS Stimulus Array on a Tablet & Measure.** To support participants’ selections of “closest match” colors, we recommend using a tablet or touch-based computer to render a set of color chips. Specifically, we suggest using the specific colors and arrangement of color chips used in the World Color Survey (WCS) Palette [49]. Depending upon the display device used, it may not be possible to render all 330 WCS palette colors, in which case we recommend a subset that is representative of the color gamut of interest and contains the relevant color naming centroids.

To assist in ensuring tablet-rendered colors match AR source colors, measure each of the Munsell color chips as rendered on the tablet paying particular attention to the measured target colors of interest. Use the CIE color space to define chromaticity and luminance [49] for each of the target colors you wish to study (as opposed to a technology-based color model such as RGB). The 1931 CIE color space was created by the *Commission Internationale de l’Eclairage*, in a collective effort to generate a color space that adequately represents human perception and preserves relationships within the space that approximates relationships in perceptual differences present in human color judgements. The 1931 CIE color space is defined by the tuple:  $x, y, Y$ . Chromaticity is defined in 2-space using  $x$  and  $y$  (and generates the commonly used 1931 CIE color plots). Luminance is represented using the remaining  $Y$  parameter. While the 1931 CIE space has many potential uses, we felt that it was a well-suited for analysis, but not necessarily so for participant selection since it visually represented in 2 dimensions as a continuous color space (see Figure 1, Step 5 for an example of part of the 1931 CIE color space). Conversely, the WCS palette is well-suited for selecting a closest match, but not for perceptually based post-hoc analysis.

Most research-quality colorimeters will support measurement using the CIE color model (i.e., output  $xyY$  values). In Figure 1, one would measure all 148 colors (we used a truncated WCS color palette) and would specifically note the  $xyY$  values for the chip associated with the “blue” naming centroid chip.

**Step 3: Tune the AR Source Colors.** Since each visual display is likely to have a different color gamut and color rendering

properties, we must account for the fact that simply rendering a color’s  $xyY$  value measured during Step 2 to an optical see-through display does not guarantee that the color rendered by said optical see-through display would actually result in the same  $xyY$  measurements. Thus, for each target color, we must tune the actual color presented to the AR display (i.e., the *source color*) such that the resulting measured CIE  $xyY$  matches each target color. To achieve this, first render a full-sized (and adjustable) source color image to an optical see-through AR display. For the initial color condition, convert the CIE  $xyY$  of the target color to a computer graphics color model (e.g.,  $rgb$ ,  $hsv$ , etc.) taking care to choose the correct standard illuminant (D65 in mid-day outdoor conditions), and standard observer (10 degrees provide the best average spectral response in human observer especially at close distances). Then, measure the CIE  $xyY$  color of the presented color using a colorimeter in a dark room with no real-world backgrounds. Adjust the source color provided to the AR display until the measured  $xyY$  values are equal to those measured in Step 2. It is important to note that visual displays (including AR displays) have per pixel differences in color. However, users still need to use these displays despite the limitations. To mitigate some of these issues, we recommend sampling the presented color at several different screen locations during this step. Do this for each of the target colors you wish to study, and note the associated source color values (e.g.,  $r, g, b$  or other) for each target color.

**Step 4: Participants Name and Match Perceived AR Colors.** Select an experimental testing location where the AR graphics can be rendered atop real-world backgrounds of interest. For each target color, systematically render the documented  $r, g, b$  source colors to an AR display in usage contexts with real-world backgrounds. Use contexts are very flexible and could include interface stimuli such as text (e.g., to examine legibility, menuing, notifications), symbols (e.g. dashboard indicator lights), and 3D models (e.g., conformal AR graphics such as avatars, virtual objects, etc.). If desired, have participants complete some task that requires the use of the AR information (e.g., visual search task as presented in our exemplar user study below). Time on task and accuracy data may be useful depending on the goals of the assessment. Further, to facilitate assessment of the perceived AR color (given the size, shape, saliency of the chosen AR graphics), have participants verbally name the perceived AR color as well as match the perceived color to the closest swatch on the tablet-rendered WCS palette.

**Step 5: Map Participant Palette Responses to CIE Color Space & Analyze.** For each participant response choice using the tablet-rendered WCS palette, document the CIE  $xyY$  color measured during Step 2 associated with each participant color chip choice. With the original target colors noted (from Step 2) as well as the corresponding participant responses in CIE  $xyY$  color space, we can now turn to a series of various analysis techniques to better understand the effects of tested backgrounds on tested AR colors. We provide a sample of analysis techniques below as a starting point. We hope that this method inspires other analysis techniques to help the

community better understand the perceived interactions between AR display color and real-world backgrounds.

For analyzing response data, we propose two quantitative measures: shift and dispersion. In simple terms, shift can be thought of as a measure of accuracy (resulting primarily from color blending), and dispersion a measure of precision (resulting primarily from individual differences in color perception). More specifically:

- **Shift** we consider a measure of *how far a response color gamut has moved* from the original AR source color location in the space.
- **Dispersion** we consider a measure of the *size of a response color gamut* (or footprint in the 1930 CIE color space) created by the set of participants' perceived colors.

Specifically, *shift* is a measure of total deviation of participants' color matching tablet responses from the original target color. We consider shift in the 1931 CIE chromaticity x-y plane (hereafter referred to as the CIE x-y plane) as chromaticity shift independent of luminance changes (although luminance shift can also be calculated using CIE Y values). Shift measurements allow us to better understand how real-world backgrounds and lighting blend with source AR color and subsequently affect users' perception of that interface color. And when examined in tandem with color naming accuracy data, can provide insight into which AR interface colors are more and less susceptible to significant x-y shifting that may result in participants' ability to perceive and correctly name an interface target color.

*Dispersion* is as a measure of similarity between the set of CIE x,y values of each participants' color matching tablet responses for an experimental condition independent of how similar those responses are to the target source color. Thus, dispersion can provide insight on to what degree a given source color will be more (or less) likely to be perceived as different colors across different individuals. And in turn, help identify AR interface colors that may be less susceptible to individual differences in color perception.

Both shift and dispersion are calculated as a normalized quantitative difference using the root mean square distance method (RMSD), as shown in Equation (1), which has been used to identify the degree of similarity and alignment between varying structures [57].

$$RMSD = \sqrt{\frac{\sum_i d_i^2}{n}} \quad (1)$$

For shift,  $d_i$  represents the distance between a specific user-specified response color and the respective target color for a given experimental condition (e.g., blue text, brick background, symbol task). For dispersion,  $d_i$  represents the distance between a specific user-specified response color and the average of all user-specified response colors for a given experimental condition. In both cases,  $n$  represents the total number of participant responses per experimental condition, and  $i$  denotes a specific participant response.

Equation (2) shows how the term  $d$  is calculated. Chromaticity shift and dispersion analysis values for  $x$  and  $y$  denote points on the CIE x-y plane of the 1931 CIE color space. While

$x_i$  and  $y_i$  represent a point corresponding to a specific participant response,  $x$  and  $y$  represent the CIE x-y value of the target color for shift and average x-y value of participants' responses for dispersion.

$$d = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (2)$$

Additionally, response color naming data can be qualitatively analyzed by, for example, looking at summary statistics of accuracy. Together, the dispersion, shift, and color name help to collectively define the *robustness* of an AR color in specific real-world usage contexts. That is, to what extent a color retains its utility on a particular display for different users across dynamic contexts with changing backgrounds and lighting.

## 4 EXEMPLAR USER STUDY

While the method presented in Section 3 could be used for various types of optical see-through AR displays in different application domains, the user study presented herein applies our method to an optical see-through HUD in a transportation setting. When color is used to encode important information, particularly in safety-critical situations like driving, it is vital that the color robustly maintain its perceived color across a variety of backgrounds. For example, a red AR graphic signaling a driver to stop should clearly be perceived as red despite any changes as a result of color blending with the environment. If users consistently perceive the graphic to be red regardless of background, then we consider it to be a robust color. This user study includes an initial analysis to identify robust (and not robust) colors in the driving domain. Participants and Study Design

We engaged twelve participants ranging from 21-44 years old (six male: mean age =28, SD = 8; six female: mean age =27, SD = 6.8) with at least one year of previous driving experience and self-reported perfect or corrected-to-perfect vision.

During each experimental session, participants conducted both text- and symbol-based visual search tasks using an AR HUD (Figure 2). We rendered AR stimuli in one of eight different colors atop one of three real-world backgrounds. After



Fig. 2. Experiment setup. The AR HUD (right) was fixed to the dashboard and adjusted to the user's line of sight. A color palette was also displayed using a tablet mounted at eye level (left) of the driver, simulating a large tablet-style in-vehicle display.



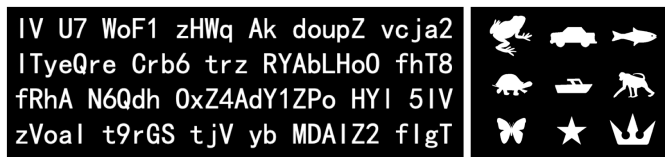


Fig. 3. Participants completed two types of visual search task, structured visual search tasks using pseudo-text (left) and semi-structured visual search tasks using symbols (right).

each visual search trial, participants also performed a color naming and color matching task. Each participant experienced four repetitions of each condition. Thus the experiment was an 8x3 within-subjects repeated measures design. Details of the independent variable levels and tasks follow.

**Independent Variable: AR Source Color:** In choosing the AR source colors to study, we began by constraining the selection to colors present on a tablet-rendered, truncated WCS palette to ensure that a color match existed for each trial (albeit under optimal or perfect viewing conditions). We initially adopted the 11 basic color categories which are most salient for human color recognition [49, 53], but excluded black, white and gray from our selection due to poor rendering on the AR HUD, leaving eight basic colors: blue, brown, green, orange, pink, purple, red, and yellow. For each of the eight colors, we chose the specific Munsell color chip from our truncated WCS palette historically shown to be the naming centroid [53] and applied the method described in Section 3 to ensure the colors presented on the tablet were adequately rendered to the HUD. Specifically, we color matched (in the dark) the projected light through the HUD for each of the 8 colors to the xyY values viewed on the tablet to within 0.01 chromaticity distance in the CIE color space and 10 cd/m<sup>2</sup> for luminance.

**Independent Variable: Background:** Each participant completed visual search and color naming/matching tasks using the AR HUD against three different backgrounds; brick, pavement, and grass. We intentionally selected very different colors including red (brick), green (grass), grey (pavement) in order to examine blending across a range of colors. We further wanted backgrounds with a high tendency to appear in vehicle operations, based on previous studies (e.g., [15]). While these backgrounds do not represent a comprehensive set of potential backgrounds for drivers, they represent some possible use-cases. These backgrounds also allow study of a variety of colors that could be reasonably experienced in the real world and are somewhat homogeneous in color within the HUD field of view: a property that we explicitly chose to control for possible confounds. Complicated backgrounds (e.g., advertising billboards) would be more difficult to study, but certainly of interest for future work.

**Visual Search Task:** We assessed two types of visual search tasks for this study. We used a *text-based task* to examine structured visual search tasks where users employ known strategies (i.e., left-to-right, top-to-bottom) to visually examine a set of stimuli in order to complete the task (Figure 3, left). Throughout the text task, participants were instructed to find and record the number of times a target letter occurred in four

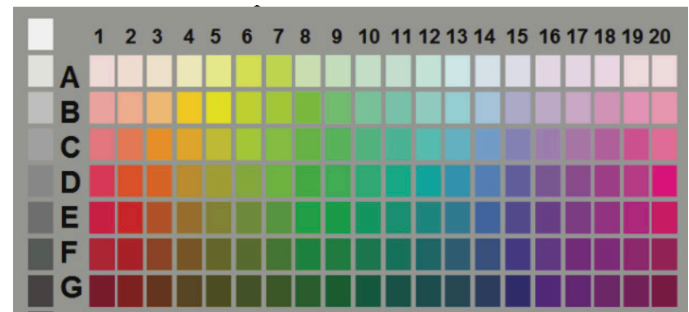


Fig. 4. We rendered a truncated World Color Survey palette on a tablet. Participants selected the color that most closely matched the perceived color as viewed through the AR HUD.

lines of pseudo text built from a randomized sequence of letters, both capitalized and lowercase, spaced similarly to written words but following ISO 9241-3, and with characters subtending approximately 1.0 degree in height. We used a sans serif mono spaced SimHei font, rendered in its most simple, plain stroke form. For this work we did not employ more visually complex approaches to rendering fonts such as those used in [58, 59]. While these approaches are known to increase the legibility of fonts in 2D GUIs, we were specifically interested in color perception of the UI element, and adding more visual complexity to the font could have confounded our results (i.e., adding transparency or juxtaposing other colored pixels as a font outline could affect participants' color perception).

The target letter occurred at least once and no more than nine times in the four lines of pseudo text.

We also used a *symbol-based task* as a semi-structured visual search tasks that required participants to identify a target symbol within a grid of 3 by 3 symbols randomized from a standardized library of commonly known symbols with similar levels of visual recognition and complexity [60] (Figure 3, right). The average symbol height was approximately 1 degree (min symbol height of 0.067 degree and max height of 1.653 degree). Following completion of each grid search task, participants likewise recorded their response time via mouse button press and verbally denoted the location of the target symbol. Grid location was defined simply, using the terms left-middle-right, and top-middle-bottom for horizontal and vertical dimensions respectively.

**Color Matching and Naming Tasks:** During both the text and symbol tasks, participants were given a color palette displayed on a mounted tablet showing a variety of colors ordered by hue (left to right) and brightness (top to bottom, Figure 4). Following completion of each visual search task trial, participants used this palette to indicate the tablet color that most closely matched the AR color in which they perceived through the HUD. Participants verbally designated the closest color match using letters A-G for row position and numbers 1-20 for column position. We selected 170 colors to present on the tablet from the World Color Survey (WCS) stimulus palette (originally consisting of 330 Munsell color chips [20]). We used a subset of the 330 WCS colors for two reasons: first the smaller palette was easier to visually scan given the size of our tablet; and secondly, we did not want the color

matching task to be overly demanding especially given the fact that some participants might not be able to easily differentiate adjacent colors on the full 330 palette due to either physiological or inherited differences in their visual system [51]. To assist in differentiating colors, our truncated palette used a proportional number of the original selection by equidistant spacing.

### 4.1. Experimental Procedure

Before testing, participants sat in the vehicle driver seat, adjusted the seat to a position that kept viewing position consistent across participants, and familiarized themselves with the AR HUD, tablet, and environment in which they completed tasks.

Each participant completed a set of text- and symbol-based visual search tasks. We limited our choice of visual elements to text and symbols because these are very common visual elements used in interface design. After completing each visual search task, participants verbally identified the color of the text or symbol they perceived through the AR HUD, and then chose a chip from the truncated WCS palette rendered on the tablet (Figure 4) that most closely matched the color they perceived through the AR HUD.

Since the study employed a within-subjects design, participants experienced 192 trials in total (8 AR source colors \* 3 real-world backgrounds\* 4 repetitions). We counterbalanced task type and background order, and randomized AR source color and repetitions. Throughout the experiment, all trials were self-paced, allowing participants to take a short rest interval between any two trials if needed.

After each experimental background condition concluded and participants exited the vehicle, an experimenter navigated the car to the next background location while a second experimenter escorted participants to the next location. The three background sites were located within five minutes of walking distance of each other.

### 4.2. Ambient Light

Since we conducted trials in an outdoor testbed, there were some inevitable changes in ambient lighting that occurred due to varying weather. We documented fluctuations in ambient lighting during experimental trials using a CEM DT 8820 environment meter equipped with a lux sensor. All trials occurred in ambient lighting bounded between 500-2000 lux, corresponding to lighting conditions ranging from a dark overcast day and light overcast day [61]. Ambient light was measured before the commencement of each experiment and when ambient lighting changed significantly during the experiment (e.g. over 1000 lux from overcast to light overcast), the experiment was paused, and lighting measurements were taken again to ensure that it remained within acceptable bounds before resuming trials. In practice, most weather consisted of overcast skies which limited the resulting variability in ambient lighting.

### 4.3. Materials

All experimental sessions took place in a parked car in front of three different outdoor locations, one for each designated

background. We used Microsoft PowerPoint and Matlab software to generate visual cues presented via a Pioneer laser scanning HUD with a field of view of 17.1° x 5.7° and a virtual image distance of approximately 2 meters placed approximately where a driver’s resting gaze would be located (Figure 2). Both the PowerPoint and Matlab programs were embedded with time tracking, and systematically controlled images displayed on the HUD to participants, while simultaneously capturing response time data. For each trial, participants mapped their perceived color on an Android tablet mounted on the dashboard of the car.

## 5 RESULTS AND ANALYSES

Our results focus first on reporting task time and errors for both text and symbol tasks. The remaining analyses focus on participants’ verbal color naming performance, specifically exploring the interactions between participants’ ability to accurately perceive and name the perceived AR graphics’ colors and the degree of chromaticity shift and dispersion associated with the set of participants’ color matching responses.

### 5.1. Response Time, Errors & Naming Accuracy

We measured response times and error for both text and symbol visual search tasks. Using Analysis of Variance (ANOVA), we found higher error rates for the text task as compared to the symbol task across all color conditions, and no significant differences across response times by color. We also saw, as expected, text tasks ( $\mu=12.266$ ) took significantly longer to complete than symbol tasks ( $\mu=2.712$ ,  $p<.0001$ ) as shown in Figure 5.

We also measured the accuracy at which each an AR source color could be verbally identified. Because participants were unconstrained when identifying colors, we categorized participants’ verbal responses with a binary hit (success) or miss (failure) based on whether or not the color name was used in the description. For example, a source color of yellow named “yellow-orange” by a participant would be a hit, but “cream-orange” would be a miss since it did not include the term “yellow.” Pearson chi-square results show statistically significant differences between all AR source color naming (Figure 6) with blue, green and yellow associated with higher naming accuracy (above 80%) and all other AR colors associated with relatively poor naming accuracy (0-

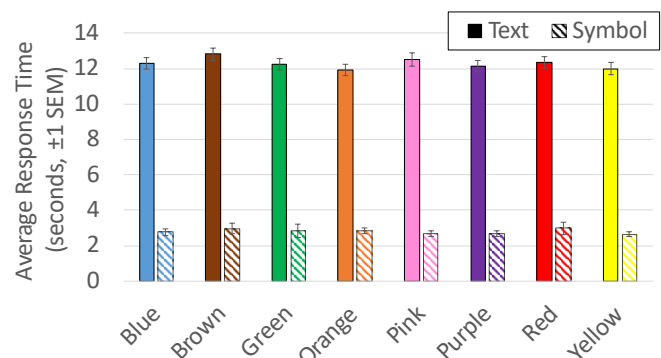


Fig. 5. Response time by color for both text and symbol tasks.

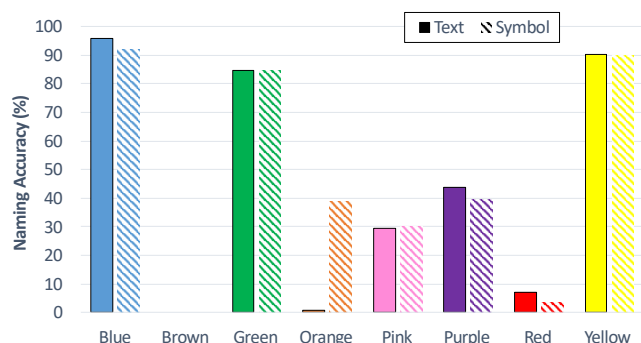


Fig. 6. Naming accuracy by color for both text and symbol tasks. Every color significantly different with  $p < 0.01$  or smaller. Letters denote significant differences between AR source colors.

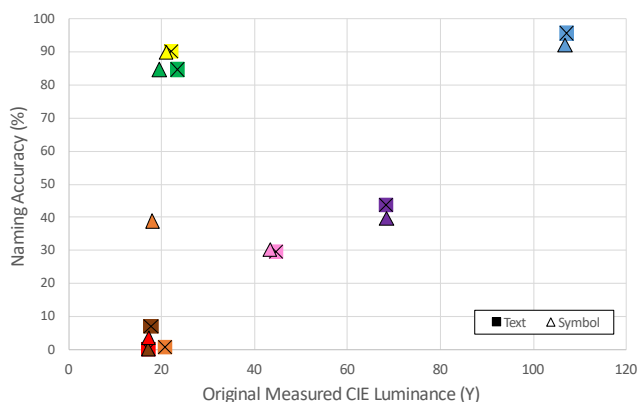


Fig. 7. The original measured display luminance of a color contains a weak positive correlation with the resulting naming accuracy of that color when rendered through an AR HUD. Note that brighter colors are not universally associated with higher naming accuracy.

40%). Participants were notably unable to correctly name brown and often reported it as a dark green or yellow instead.

## 5.2. Luminance and Naming Accuracy

One issue meriting further exploration was whether the inherent luminance of each source color impacted its resulting naming accuracy, meaning for example, that perhaps brighter AR source colors would be more likely to be perceived “correctly” as compared to darker AR source colors. To analyze, we plotted the original measured luminance of each AR source color against naming accuracy associated with that color (Figure 7) and found a weak positive correlation between source luminance and naming accuracy when modeled linearly. Specifically, luminance accounts for less than 25% of variation for either task type (symbol task:  $R^2=0.1559$ , text task:  $R^2=0.2394$ ). As a result, the remainder of our analysis will focus on chromaticity shift and dispersion.

Moreover, note that with the exception of the color orange, where there were more correct answers associated with the symbol task as compared to text ( $X=65.759$ ,  $p<0.0001$ ), there are no significant differences in color naming accuracy between text and symbol. Since we are interested in understanding how chromaticity shifts and dispersion are related to naming accuracy, and since there appears to be no con-

sistent difference in naming accuracy across task type, the remainder of our analysis reports AR source color using combined text and symbol results.

## 5.3. Chromaticity and Naming Accuracy

For shift and dispersion analyses, we quantified the color that each participant perceived using the measured CIE  $xyY$  values associated with each participant’s selected color chip (i.e., using the method presented herein, for each response noted in step 4, we used the measured CIE values obtained in step 2).

Using CIE  $xy$  values, and equations 1 and 2 above, ANOVA revealed several significant differences in perceived  $x-y$  shift across AR source colors ( $p<0.0001$ ), with green, orange and blue associated with the least amount of perceived shift, and pink, red and brown AR source colors associated with the most perceived shift (especially brown, Figure 8, left). Post-hoc analysis revealed significant differences between most colors as denoted with letters in Figure 8, left).

When we examine the relationship between  $x-y$  shift and naming accuracy, we see an expected strong negative correlation ( $R^2 = 0.69958$ ) where lower amounts of shift are associated with higher naming accuracy and conversely, colors with high  $x-y$  shift are associated with low naming accuracy (Figure 8, right).

Additional ANOVA revealed several significant differences in perceived  $x-y$  dispersion across AR source colors ( $p<0.0001$ ) but no differences across backgrounds. Results show that brown and pink are associated with greater dispersion than all other colors, and blue, purple and yellow are associated with the least amount of  $x-y$  dispersion as compared to other AR source colors (Figure 9, left). Post-hoc analysis revealed significant differences across groups of colors as denoted with letters in Figure 9, right).

When we examine the relationship between  $x-y$  dispersion and naming accuracy, we see an a moderate-to-weak negative correlation ( $R^2 = 0.40708$ ) where lower amounts of dispersion are somewhat associated with higher naming accuracy and conversely, colors with high  $x-y$  dispersion are loosely associated with low naming accuracy (Figure 9, right).

Lastly, to evaluate the relationship between shift and dispersion in the CIE  $x-y$  color space, we graphed both variables by AR color and annotated each data point with its corresponding naming accuracy (Figure 10). In the chromaticity space, significant decreases in naming accuracy occur as both shift and dispersion increase on a generally largely uniform scale. Further studies designed specifically to identify the relative weighting of chromaticity and dispersion shifts would be needed to fully understand the relationship between these two parameters and user perception and performance.

## 6 DISCUSSION

Our analyses allow us to characterize the relationships and interactions between (1) task performance (response time and accuracy), (2) naming accuracy, and, (3) the chromaticity shift and dispersion derived from participant tablet responses – ultimately attempting to examine which AR UI colors we can

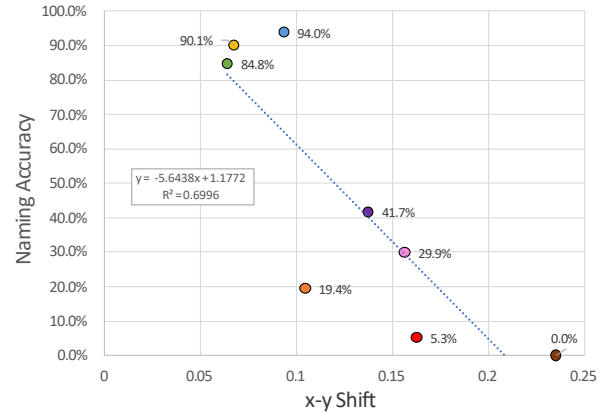
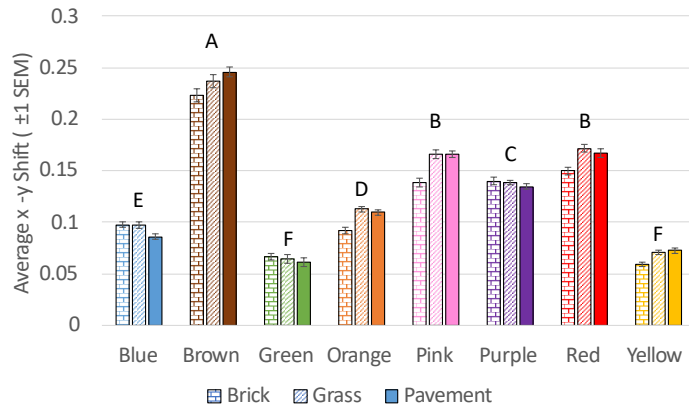


Fig. 8. (left) Average x-y shift by color and background. Shifts in the x-y plane can be attributable to both the effects of color blending and, to a less extent, individual difference in color perception. (right) Average x-y shift by naming accuracy shows a strong negative correlation.

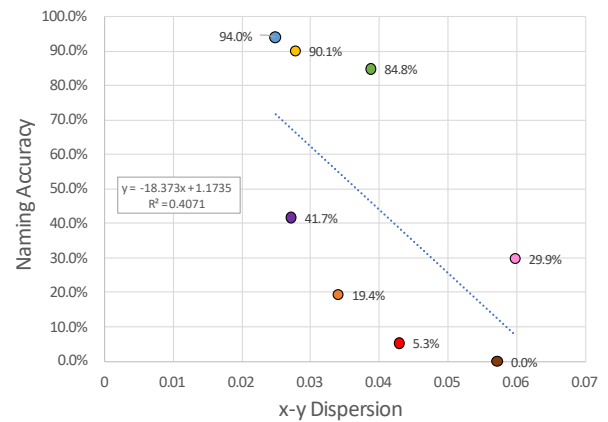
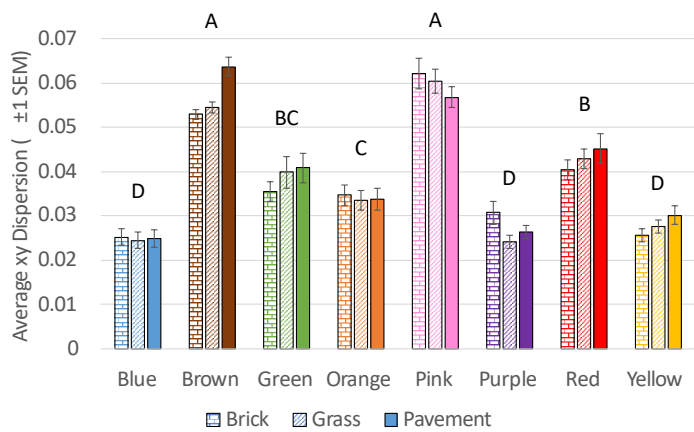


Fig. 9. (left) Average x-y dispersion by color and background (both tasks), where differences in dispersion levels are related to how likely a color is to be perceived differently by different users; a reflection of individual differences in color perception. (right) Average x-y dispersion by naming accuracy where a regression analysis reveals a moderate-to-weak negative correlation.

consider robust against the testbed real-world backgrounds.

In order to identify robust AR colors, we first examined task response time and task accuracy. We found no differences in task response time or accuracy due to color, although the task itself (text or symbol) did have a significant impact, with text resulting in lower accuracy and longer response times. This outcome can be expected because the text task included more characters to parse through. And in essence was a more difficult task as compared to the symbol task.

Intuitively, we might expect that the robustness of a color may depend on the nature of how the color used in an AR visual element (e.g., as a thin line vs. a filled rectangle). Interestingly, Figure 6, shows that color naming accuracy was fairly consistent across most colors, with the exception of orange (and possibly red and purple). In this case, it may be that there is some threshold by which a color becomes robust, but under that threshold its perceived name can be more easily affected by the number of pixels colored. For example, orange symbols are more accurately named than orange text. But confusingly, red and purple text are more accurately named than their symbol counterparts. Clearly, more work is needed to further understand the impact of visual element

size/footprint and color robustness, but that would require a more systematic experimental design and analysis of visual stimuli shape and size which is out of scope for the current work.

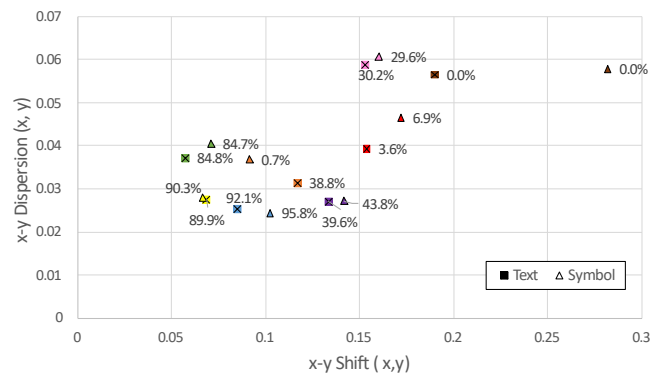


Fig. 10. Average chromaticity shift plotted against average chromaticity dispersion, for each color and task type (symbol  $\Delta$ , text  $\times$ ). Both chromaticity shift and dispersion concurrently impact the accuracy of the color naming (shown numerically as a percentage next to each color), and potentially in turn the perceived semantics of colored symbols and/or text.

There were, however, significant differences in color naming accuracy with blue, green, and yellow resulting in the most consistently correct color naming. A possible explanation for higher naming accuracy associated with these colors might be that these colors had inherently higher source luminance, because luminance is often seen as a solution to mitigating color blending issues in AR [62]. A complementary reason may be that the human visual system (HVS) contains three types of specialized cone that are particularly responsive to specific short, medium and long wavelengths of light. In particular, the HVS' tristimulus response is most sensitive to blue, green and red colored light (i.e., 420nm, 524nm, and 564nm wavelengths detected by S, M, and L cones respectively). While the specifics of our HVS may help explain the relative robustness of blue and green seen in our results, we must consider the relatively high amount of shift and dispersion associated with red to understand why participants found red more difficult to accurately perceive and name. Specifically, we posit that while the HVS is particularly sensitive to red light, the actual color that reached users eyes was likely not red and, indeed, this fact could be explained by its relatively low luminance.

Therefore, we examined naming accuracy as a result of luminance in Figure 7. However, this finding showed that luminance only accounts for a small part of color naming accuracy. That is, blue exhibited high luminance and was more accurately named than other colors, as were green and yellow despite exhibiting low luminance. Moreover, orange and purple symbols were associated with similar luminance levels but very different naming accuracy.

This may indicate that the brightness of rendered colors does not directly drive the resulting accuracy to which a color is identified since yellow and green exhibited much lower luminance levels as compared to blue (which had the highest luminance of all colors measured) but nonetheless were named with similar accuracy to blue. This result provides evidence against a commonly held idea that a solution to a specific color correction is that "brighter is better".

This initial analysis examining luminance and color naming accuracy provided some potential direction for AR display designers. Specifically, that blue, green, and yellow seem to have potential to be more robust UI colors across varying backgrounds. However, this initial analysis does not help us understand the reason behind this finding. For that reason, we examined in more detail participants perceived colors' chromaticity shift and dispersion from the source color.

### 6.1. Chromaticity x-y shift and color naming accuracy

When considering the concept of chromaticity x-y shift, in this work, we first should unpack what we are measuring with respect to both object and subject phenomenon. Specifically, we have to first consider that there exists an objective, physical change in light that reaches users' eyes via color blending which has been shown to be measurable via colorimeter [1, 4]. That is to say that the AR graphics may have undergone a change that could result in some x-y shifting even before we ask users to perceive them. And it is the case that analysis of such colorimeter data could predict, to some

degree, what colors may undergo less x-y shift than others which arguably would suggest these colors would be more accurately named (on average) as compared to those colors that exhibit quantifiably more x-y shift. However, this measurement-based approach does not take into account individual differences in color perception, which are ultimately at play when using AR.

Next, we must consider the second phenomenon that occurs while using this method, whereby we ask participants to perceive the color as viewed through the AR display and attribute a name to that color as well as select a matching color

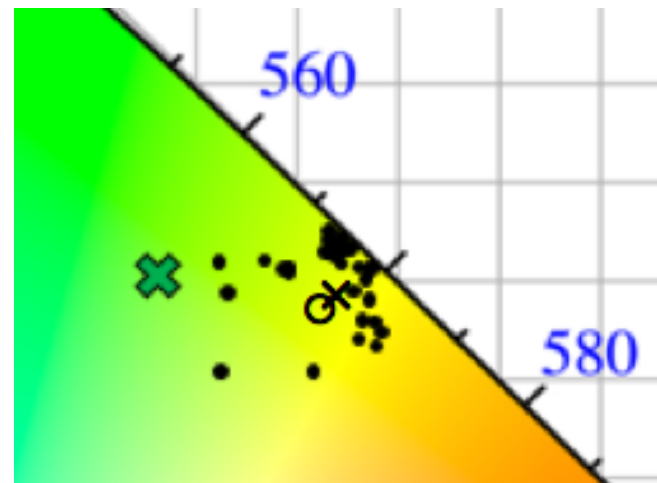


Fig. 11. The average perceived x-y shift (o) of a green AR color (\*) represents the actual x-y shift as a result of color blending (x) as well as, to a lesser extent, the effect of individual color perception differences (•).

swatch on experimental tablet. While we acknowledge the fact that there is likely some noise in the perceptual color matching subtask associated with this step, for now, we are mostly focused on the fact that such a matching task allows for capturing individual differences in *perceived* x-y shift including those related to color vision deficiencies.

Thus, we consider perceived x-y shift (which we examine as an average across all participants), to encompass both the result of objective, measurable x-y shifts associated with color blending as well as individual differences in color perception (but arguably to a lesser degree than actual x-y shift). This is illustrated in Figure 11 showing a set of participant responses (•) for a green AR source color (\*), as well as the average perceived x-y shift (o), and our estimate of the actual x-y shift (x) shown for illustration purposes. Note that the distance between the actual x-y shift and average perceived x-y shift is due to individual differences in color perception and is smaller than the distance between the distance between the source color and actual x-y shift. If all participants had theoretically perfect perception, then we'd see no differences across participants' responses and the resulting average perceived x-y shift would indeed be the same as the actual x-y shift due to color blending. However, humans are not perfect replicas of each other nor perfect in color perception! Thus a contribution of this method is that it takes into account

the fact that we have imperfect users perceiving colors on the fly, and this method may assist the community in understanding which colors are associated with less perceived shift (and potentially more accurate color naming) as a worthwhile endeavor.

As mentioned in Section 5.3, we found a strong negative correlation between x-y shift and naming accuracy, with lower amounts of shift associated with higher naming accuracy (Fig 8, right). Of note is the fact that green, yellow, and blue were associated with relatively little shift and accurate naming accuracy as compared to other colors. Since we measured blue to be significantly more luminous than green and yellow, we must consider the possibility that some colors undergo less actual and perceived x-y shift as compared to others independent of source color brightness. However, what we do not know is whether these smaller x-y shifts are due to actual x-y shift or perceived x-y shift. The answer may lie in the fact that other colors measured as similarly low in luminance (e.g., brown and red) were associated with large amounts of perceived x-y shift; suggesting that these differences are perceptual as opposed to the effect to measurable color blending. This is important to note, as our method can uncover these differences and identify sets of colors as “promisingly robust” despite their apparent lack of brightness.

## 6.2. Chromaticity x-y dispersion and color naming

When employing measurement-based approaches, such as those that use a colorimeter on a benchtop, we can only measure dispersion as a result of using different lighting levels or backgrounds or colored AR light (holding 2 of those other three constant). What we cannot do in these cases is capture the perceptual differences that exist between potential AR users. Using a colorimeter, we can observe that an AR color has been blended to a different location on the CIE x-y plane, but we cannot know if that change will result in has significant perceptual differences between users. That is, we cannot know which resulting colors would likely be associated with greater x-y dispersion and as a result be more likely to be named differently by different users. In short, benchtop measurement alone is insufficient.

By employing the method presented herein as well as the measure of chromaticity x-y dispersion, we believe we can hone directly into individual differences. Indeed, unlike x-y shift, the measure of x-y dispersion is solely a function of differences between individual responses to a presented color. It is calculated based solely on the x-y position of all participants’ responses, regardless of the source color’s position or even the color-blended position as measurable via colorimeter). Thus, x-y dispersion is a more direct measure of the extent to which a particular color is likely to be misperceived by users. Due to inherent color vision differences, both physiological and cultural, that exist, some colors are more susceptible to be misperceived than others. Moreover, there exists a set of colors adjacent to any given color that are likely to be perceived as that same given color (see MacAdam’s work that defined a set of ellipsoids in the CIE color space whereby observers are unlikely to distinguish between colors that lie

within a given ellipsoid when viewed at the same luminance [48]).

Thus, the smaller the x-y dispersion, the more likely that a set of users will be to name the perceived color the same name. This is not to say that small x-y dispersions will result in accurate color naming, as it is the *position* of the set of responses that define the average x-y dispersion relative to the source color that plays an equally compelling role in naming accuracy. We can nonetheless envision small average x-y dispersions that plot atop of a target color in the x-y plane, which would likely result in not only many users agreeing on the color name, but also that name being correct. But a small x-y dispersion that is significantly displaced from the target color could result in very low naming accuracy. This concept is a bit asymmetric however, in that large average x-y dispersions would almost certainly always increase the chances that: (1) users will perceive the color very differently from each other, and, (2) that those color perceptions would be farther away on the x-y plane from the target color (and thus be likely to be misnamed).

Recall that we calculate x-y dispersion based on the position of the average color response, not the target color position nor even the measurable position of the color produced as a result of color blending. It could be argued that perhaps a more meaningful measure of dispersion could be calculated relative to the actual color rendered to users’ eyes as a result of color blending (the  $\times$  in Figure 11 as opposed to the  $\circ$ ). However, since we are interested in a human-centered method that can be easily conducted in the field, we will assume that the relative difference between the actual measurable color and average perceived color is relatively small as compared to the difference between the average perceived color and individual observations (this assumption can be visually explored in Figure 11).

At this point, it is prudent to note that since x-y dispersion calculations are exclusively weighted to individual differences in color perception (as compared to x-y shift which is more heavily weighted by color blending effects), there is likely some noise introduced in to the data collection method since not only do users need to perceive and the color of the AR graphic but also the color of the tablet swatches. Thus, researchers must take care when choosing tablet swatches as well as performing calibration and measurement of tablet swatches to match candidate AR UI colors. Specifically, it might behoove researchers interested in more fully exploring individual color perception differences to use a large tablet with more swatches than used in the study presented herein.

Circling back to our evolving notion of “robustness”, we can posit that that colors associated with less dispersion are more likely to be named correctly than those with high x-y dispersion. Conversely, we would likely consider colors with high dispersion *not robust* since they are more likely to be associated with significant individual color perception differences.

However, the role of x-y dispersion in robustness is a little more complicated. Our study suggests that two colors with similar dispersion may not necessarily be equally robust. For

example, green and red show relatively similar levels of x-y dispersion (0.038 and 0.043 respectively) but very different naming accuracy (with green correctly name 84.8% of the time, and red only 5.3%).

Also, consider the colors pink and red as shown in Figure 8. Both exhibit similar x-y shift (and were not statistically different in our testing) but are associated with different levels of x-y dispersion (0.060 and 0.043 respectively) and very different naming accuracies (29.9% and 5.3% respectively). The end result is that red has similar shift, less dispersion but is still less accurately named as compared to pink. Thus, while x-y dispersion can help us understand individual differences in color perception, it is insufficient alone to predict robustness.

## 7 DESIGN IMPLICATIONS

Examining both x-y shift and dispersion in tandem (Figure 10) suggests that distortion in both x-y shift and dispersion may each play an active role in determining how accurately a color can be perceived and thus named. Colors that exhibit high naming recognition were associated with low levels of either chromaticity dispersion or chromaticity shift but were not necessarily associated with low levels of *both* chromaticity dispersion *and* shift. For example, blue averaged more x-y shift than green but was associated with lower levels of x-y dispersion than green, and thus both colors could still be verbally identified across different backgrounds. It is possible that an AR color's propensity to be perceived differently by different users under varying viewing conditions may in fact be compensated by lower amounts of x-y shifting; and vice versa.

Rather than developing a precise or mathematical color correction strategy for optical see-through AR displays (such as HUDs), this work instead contributes a novel method for exploring and understanding the general principles that govern users' recognition of AR rendered colors after undergoing both objective color blending and subjective individual perception.

In an ideal world, we could color-correct to minimize x-y shift and personalize to account for individual differences and minimize x-y dispersion. For example, through the use of adaptive color correction techniques, we could minimize effects of color blending resulting in rendered colors that more closely align with designers' intent. Further, we could add a personalized adaptive component that takes into account individual differences in color perception.

Until it is perfectly possible to account for individual differences in perception (i.e., x-y dispersion), our method could help identify colors that are likely to be associated with lower levels of dispersion across a population (i.e., colors that are more likely to be named the same by many different users).

To increase the odds of consistent user experience with regards to color, AR UI designers could select colors with high inherent source luminance (e.g. blue, green, and yellow). And while it appears that high luminance does not guarantee accurate color identification, luminance may play a signifi-

cant role in visualization acuity, resulting in better user performance (e.g., reading small text). Thus, the luminance of an AR source color likely functions as a foundational "starting point" for that color and may moderate its color blending sensitivity.

However, in some cases where color is used to convey important messages (e.g., red) and where that color in its purest form exhibits low luminance on a given AR display, designers should instead consider other color(s) that exhibit higher luminance but are still meaningfully interpretable (i.e., will be interpreted as a "warning" color more so than specifically a "red" color). Indeed, our method could be useful in choosing this set of potentially more robust candidate adjacent colors.

It is also interesting to note that brown, as the color with the lowest source luminance of the tested colors, was also never named correctly by participants. This supports the idea that, while increasing luminance may not be a solution to color recognition itself, AR graphics may require a minimum inherent brightness threshold or suffer becoming unrecognizable. Given this constraint, designers may be required to consider certain colors inherently unfit for use in AR applications if intending to render those at their correct CIE values.

Designers should be aware that issues in color recognition may not always be solved simply by increasing a color's projected brightness (if, for example, that color is still prone to heavy x-y shifts and/or x-y dispersion in that specific environment). Instead, choosing colors associated with lower levels of dispersion may provide another avenue for correcting color perception issues in AR displays. In any event, identifying and selecting robust colors that users can correctly identify in situ is a key to successful color-coded UI designs.

Lastly, it is important to continue exploring design tradeoffs in AR UI design which balance photorealistic and practical color rendering strategies, as maintaining both will be critical to their future adoption and effectiveness.

## 8 LIMITATIONS AND FUTURE WORK

We recognize several limitations associated with the nature of this study.

While we had participants self-report CVDs, we can envision an improvement to our method would be to pre-screen participants using a suite of CVD tests. For example, by administering ColorCheck [50], or perhaps better yet a situation-specific model of color differentiation such as ICD-2 [52], researchers could ensure that the set of tablet choices are more likely than not to be distinguishable from each other and modify the set of tablet choices as needed per participants' needs. Using ICD-2 could also be used in analysis to assess the degree to which participant responses may be attributable to CVD, and possibly even posit (using the participant-named color) whether the inability to differentiate colors affected the AR color judgment or the tablet response.

As discussed in the exemplar user study (section 4.2), the use of an outdoor testbed resulted in periodic changes in ambient lighting. Though the experimenters limited data collection to a small range of lighting conditions, some variation in light may have still impacted users' perception of rendered

graphic color, and this method does not include a way to account for lighting variations (e.g. screen reflections or ambient changes) in the color mapping process. We expect that repeating a similar design under more controlled lighting comparable to the outdoor levels studied herein could provide more consistent results and may reduce some of the dispersion observed during data collection. Future versions of this method may include analysis of ambient lighting to better understand the interaction between ambient lighting and results. Running a study with several different but constant lighting levels may also be a good strategy to further explore the impact of luminance and washout on color recognition.

We also note that some AR display designs would not allow easy measurement of the actual color which might limit the usability of this method for some display types. Therefore, our method is a first step in further understanding the perceptual experience of color through AR displays. Additionally, the equipment used for this study (e.g. the Pioneer HUD) have likely improved both in terms of fidelity and color rendering capability, given the current state of rapid growth in this area of technology. Future studies could apply our method to other ARs. We also recognize that only eight color categories were tested out of a much larger range of possible choices. Future work could take a broader approach to by expanding the selection of colors. Nonetheless, it is likely that the eight colors chosen for this study could encode the most critical elements of an AR user interface.

Further, our user study only examined text and symbol visual UI elements. The fact that we see differences in response time performance across task type suggests that the display element type (coupled with task) could produce different findings. Indeed, using the method proposed herein, future studies could examine other visual elements of AR UIs such as transparency, size, shape, etc. and their effect on AR color perception.

Future research with larger sets of backgrounds, including those that are dynamic or exhibit more complex spatial, specular and/or color properties may identify additional characteristics of robust colors. Similar extensions could further examine the robustness of UI colors across other AR hardware such as video pass-through AR and spatial augmented reality, both of which we believe the method presented herein could be applied with minimal adaptations.

Follow on studies could also identify the acceptable chromaticity and luminance values bounding a given color name centroid within which a presented AR source color retains its color naming. Such a study would help set boundaries of acceptable parameters for real-time color correction systems that have constraints imposed by the environments. While our work does not explicitly draw the expected boundaries, it proves that these relationships between dispersion and chromaticity shift exist, and that understanding them is important to achieve color perception as desired by AR display designers. Continuing this work could help us to define the thresholds for color-naming sensitivity, or the boundaries within which an AR source color should be presented for the most successful rendering.

Lastly, future work could use color measurement techniques (described in section 2) to first quantifiably measure color blending and the associated x-y shifts, and then compare that data to perceived shift & dispersion obtained via the method presented herein to develop a model. Given the environmental lighting and backgrounds as well as the AR source color, this model could first predict the actual x-y shift (that creates differently color light that reaches users' visual system) as well the subsequent set of possible perceived x-y shift & dispersion responses across a user population.

## 9 CONCLUSION

Our psychophysical method of evaluating AR color perception across a variety of backgrounds can help us better understand color-recognition boundaries and the resulting impact on task performance. This method could be a foundation for standardizing the way that we test color AR displays. The luminance and chromaticity output, while important, are only inputs into the human perception of color. Our method extends the work done with the World Color Survey (WCS) for application in the AR domain. By applying our method in future data collection, we could build a data set that shows the "robustness" of AR colors much like the WCS stimulus array in Figure 4. Further, we propose that this method is a way to categorize and draw boundaries around the colors that a particular display could provide. This method can be used in any outdoor setting and with most AR displays to understand the capabilities of an individual display.

## 10 ACKNOWLEDGMENT

The authors wish to thank Nicole Sanderlin, Jack Lesko, Valerie Kane, Catherine Harvey, Elizabeth Crundall, Ayse Eren, and Claire Merenda. This work was supported in part by a grants from the NSF (awards No. 1261162 & 1816721).

## REFERENCES

- [1] J. L. Gabbard, J. Zedlitz, J. E. Swan II, and W. W. Winchester, "More Than Meets the Eye: An Engineering Study to Empirically Examine the Blending of Real and Virtual Color Spaces," in *IEEE conference on Virtual Reality*, Waltham, MA, 2010, pp. 79-86.
- [2] J. L. Gabbard, G. M. Fitch, and H. Kim, "Behind the Glass: Driver Challenges and Opportunities for AR Automotive Applications," *Proceedings of the IEEE*, vol. 102, pp. 124-136, 2014.
- [3] E. Kruijff, J. E. Swan II, and S. Feiner, "Perceptual Issues in Augmented Reality Revisited," in *Technical Papers, IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2010)*, Seoul, Korea, 2010, pp. 3-12.
- [4] J. L. Gabbard, J. E. Swan II, and A. Zarger. (2019, Color Blending in Outdoor Optical See-through AR: The Effect of Real-world Backgrounds on User Interface Color. *arXiv:1908.09348 [cs.HC]*.
- [5] D. Van Krevelen and R. Poelman, "A Survey of Augmented Reality Technologies, Applications and Limitations," *International Journal of Virtual Reality*, vol. 9, p. 1, 2010.
- [6] L. Lisle, C. Merenda, K. Tanous, H. Kim, J. L. Gabbard, and D. A. Bowman, "Effects of Volumetric Augmented Reality Displays on Human Depth Judgments: Implications for Heads-Up Displays in Transportation," *International Journal of Mobile Human Computer Interaction (IJMHCI)*, vol. 11, pp. 1-18, 2019.



- [7] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Müller, J. Wieghardt, et al., "Augmented reality navigation systems," *Universal Access in the Information Society*, vol. 4, pp. 177-187, 2006.
- [8] J. A. Jones, J. E. Swan, G. Singh, E. Kolstad, and S. R. Ellis, "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception," in *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, 2008, pp. 9-14.
- [9] M. Di Donato, M. Fiorentino, A. E. Uva, M. Gattullo, and G. Monno, "Text legibility for projected augmented reality on industrial workbenches," *Computers in Industry*, vol. 70, pp. 70-78, 2015.
- [10] M. Gattullo, A. E. Uva, M. Fiorentino, and G. Monno, "Effect of Text Outline and Contrast Polarity on AR Text Readability in Industrial Lighting," *IEEE transactions on visualization and computer graphics*, vol. 21, pp. 638-651, 2015.
- [11] A. Tang, C. Owen, F. Biocca, and W. Mou, "Comparative effectiveness of augmented reality in object assembly," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2003, pp. 73-80.
- [12] S. J. Henderson and S. Feiner, "Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret," in *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, 2009, pp. 135-144.
- [13] J.-W. Lee, C.-R. Yoon, J. Kang, B.-J. Park, and K.-H. Kim, "Development of lane-level guidance service in vehicle augmented reality system," in *Advanced Communication Technology (ICACT), 2015 17th International Conference on*, 2015, pp. 263-266.
- [14] B.-J. Park, C. Yoon, J.-W. Lee, and K.-H. Kim, "Augmented reality based on driving situation awareness in vehicle," in *Advanced Communication Technology (ICACT), 2015 17th International Conference on*, 2015, pp. 593-595.
- [15] Z. Medenica, A. L. Kun, T. Paek, and O. Palinko, "Augmented reality vs. street views: a driving simulator study comparing two emerging navigation aids," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, 2011, pp. 265-274.
- [16] B. Thomas, B. Close, J. Donoghue, J. Squires, P. D. Bondi, and W. Piekarski, "First person indoor/outdoor augmented reality application: ARQuake," *Personal and Ubiquitous Computing*, vol. 6, pp. 75-86, 2002.
- [17] M. A. Livingston, J. H. Barrow, and C. M. Sibley, "Quantification of contrast sensitivity and color perception using head-worn augmented reality displays," in *Virtual Reality Conference, 2009. VR 2009. IEEE*, 2009, pp. 115-122.
- [18] E. Kruijff, J. E. Swan, and S. Feiner, "Perceptual issues in augmented reality revisited," in *ISMAR*, 2010, pp. 3-12.
- [19] S. Seo, D. Kang, and S. Park, "Real-time adaptable and coherent rendering for outdoor augmented reality," *EURASIP Journal on Image and Video Processing*, vol. 2018, p. 118, 2018.
- [20] J. L. Gabbard, J. E. Swan, D. Hix, R. S. Schulman, J. Lucas, and D. Gupta, "An empirical user-based study of text drawing styles and outdoor background textures for augmented reality," in *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*, 2005, pp. 11-18.
- [21] 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 *Virtual Reality Conference (VR), 2010 IEEE*, 2010, pp. 79-86.
- [22] S. K. Sridharan, J. D. Hincapié-Ramos, D. R. Flatla, and P. Irani, "Color correction for optical see-through displays using display color profiles," in *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, 2013, pp. 231-240.
- [23] L. Zhang and M. J. Murdoch, "Color matching criteria in augmented reality," in *Color and Imaging Conference*, 2018, pp. 102-109.
- [24] B. Thomas, B. Close, J. Donoghue, J. Squires, P. De Bondi, M. Morris, et al., "ARQuake: An outdoor/indoor augmented reality first person application," in *Digest of Papers. Fourth International Symposium on Wearable Computers*, 2000, pp. 139-146.
- [25] J. L. Gabbard, J. E. Swan II, and D. Hix, "The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 15, pp. 16-32, 2006.
- [26] J. L. Gabbard, J. E. Swan II, D. Hix, S. J. Kim, and G. Fitch, "Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications," in *IEEE conference on Virtual Reality*, Charlotte, North Carolina, USA, 2007, pp. 35-42.
- [27] S. Debernardis, M. Fiorentino, M. Gattullo, G. Monno, and A. E. Uva, "Text readability in head-worn displays: Color and style optimization in video versus optical see-through devices," *IEEE transactions on visualization and computer graphics*, vol. 20, pp. 125-139, 2013.
- [28] M. Gattullo, A. E. Uva, M. Fiorentino, and J. L. Gabbard, "Legibility in Industrial AR: Text Style, Color Coding, and Illuminance," *Computer Graphics and Applications, IEEE*, vol. 35, pp. 52-61, 2015.
- [29] K. Kim, A. Erickson, A. Lambert, G. Bruder, and G. Welch, "Effects of dark mode on visual fatigue and acuity in optical see-through head-mounted displays," in *Symposium on Spatial User Interaction*, 2019, pp. 1-9.
- [30] J. D. Hincapié-Ramos, L. Ivanchuk, S. K. Sridharan, and P. P. Irani, "SmartColor: Real-Time Color and Contrast Correction for Optical See-Through Head-Mounted Displays," *IEEE transactions on visualization and computer graphics*, vol. 21, pp. 1336-1348, 2015.
- [31] A. Grundhofer and O. Bimber, "Real-time adaptive radiometric compensation," *IEEE transactions on visualization and computer graphics*, vol. 14, pp. 97-108, 2007.
- [32] 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*, vol. 22, pp. 2385-2394, 2016.
- [33] S. R. Sanches, D. M. Tokunaga, V. F. Silva, A. C. Sementille, and R. Tori, "Mutual occlusion between real and virtual elements in augmented reality based on fiducial markers," in *2012 IEEE Workshop on the Applications of Computer Vision (WACV)*, 2012, pp. 49-54.
- [34] 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*, vol. 21, pp. 1269-1278, 2015.
- [35] T. Fukiage, T. Oishi, and K. Ikeuchi, "Visibility-based blending for real-time applications," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2014, pp. 63-72.
- [36] S. Mori, S. Ikeda, A. Plopski, and C. Sandor, "Brightview: Increasing perceived brightness of optical see-through head-mounted displays through unnoticeable incident light reduction," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 251-258.
- [37] Y. Itoh, T. Langlotz, D. Iwai, K. Kiyokawa, and T. Amano, "Light attenuation display: subtractive see-through near-eye display via spatial color filtering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, pp. 1951-1960, 2019.
- [38] D. A. Kerr, "Chromaticity and Chrominance in color definition," *INTERNET: [http://doug.kerr.home.att.net/pumpkin/Chromaticity\\_Chrominance\\_DAK.pdf](http://doug.kerr.home.att.net/pumpkin/Chromaticity_Chrominance_DAK.pdf)*, 2003.
- [39] A. Dey, M. Billinghamurst, R. W. Lindeman, and J. Swan, "A systematic review of 10 years of augmented reality usability studies: 2005 to 2014," *Frontiers in Robotics and AI*, vol. 5, p. 37, 2018.
- [40] M. A. Livingston, "Quantification of visual capabilities using augmented reality displays," in *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, 2006, pp. 3-12.
- [41] E. E. Veas, E. Mendez, S. K. Feiner, and D. Schmalstieg, "Directing attention and influencing memory with visual saliency modulation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 1471-1480.
- [42] N. Ohta and A. Robertson, *Colorimetry: fundamentals and applications*: John Wiley & Sons, 2006.
- [43] J. D. Mollon, "The origins of modern color science," *The science of color*, vol. 2, pp. 1-39, 2003.
- [44] J. Foley, R. Phillips, J. Hughes, A. van Dam, and S. Feiner, *Introduction to Computer Graphics*: Addison-Wesley Longman Publishing Co., Inc., 1994.
- [45] G. H. Joblove and D. Greenberg, "Color spaces for computer graphics," in *Proceedings of the 5th annual conference on Computer graphics and interactive techniques*, 1978, pp. 20-25.
- [46] W. D. Wright, "A re-determination of the trichromatic coefficients of the spectral colours," *Transactions of the Optical Society*, vol. 30, p. 141, 1929.
- [47] J. Guild, "The colorimetric properties of the spectrum," *Philosophical Transactions of the Royal Society of London. Series A. Containing Papers of a Mathematical or Physical Character*, vol. 230, pp. 149-187, 1931.
- [48] D. L. MacAdam, "Visual sensitivities to color differences in daylight," *Josa*, vol. 32, pp. 247-274, 1942.

- [49] B. Berlin and P. Kay, *Basic color terms: Their universality and evolution*: Univ of California Press, 1991.
- [50] K. Reinecke, D. R. Flatla, and C. Brooks, "Enabling designers to foresee which colors users cannot see," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 2693-2704.
- [51] D. R. Flatla and C. Gutwin, "Individual models of color differentiation to improve interpretability of information visualization," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 2563-2572.
- [52] D. R. Flatla and C. Gutwin, "Improving calibration time and accuracy for situation-specific models of color differentiation," in *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, 2011, pp. 195-202.
- [53] J. Sturges and T. Whitfield, "Locating basic colours in the Munsell space," *Color Research & Application*, vol. 20, pp. 364-376, 1995.
- [54] R. S. Cook, P. Kay, and T. Regier, "The world color survey database," in *Handbook of categorization in cognitive science*, ed: Elsevier, 2005, pp. 223-241.
- [55] D. Nickerson, "History of the Munsell color system and its scientific application," *JOSA*, vol. 30, pp. 575-586, 1940.
- [56] E. H. Lenneberg and J. M. Roberts, "The language of experience: A study in methodology," 1956.
- [57] M. Tkalcic and J. F. Tasic, "Colour spaces: perceptual, historical and applicational background," in *Eurocon*, 2003.
- [58] W. B. Paley, "Designing better transparent overlays by applying illustration techniques and vision findings," in *Proceedings of the 16th annual ACM symposium on User interface software and technology, Supplement*, 2003, pp. 57-58.
- [59] B. L. Harrison and K. J. Vicente, "An experimental evaluation of transparent menu usage," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1996, pp. 391-398.
- [60] D. Large, G. Burnett, E. Loon, and L. Skrypchuk, "Grouping icons to ease the visual burden of search and select pointing tasks during interactions with in-vehicle touchscreen HMIs," presented at the Human Factors in Complex Systems, Nottingham, UK, 2016.
- [61] P. R. Boyce, "Illumination," *Handbook of Human Factors and Ergonomics, Fourth Edition*, pp. 673-698, 1997.
- [62] M. VIGIER, T. PILLOIX, B. DUPONT, and G. MORITZ, "Very High Brightness, High Resolution CMOS Driving Circuit for Microdisplay in Augmented Reality," in *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2020, pp. 876-879.

**Joseph L. Gabbard** is director of the COGNITIVE Engineering for Novel Technologies (COGENT) Lab and Associate Professor of Human Factors at Virginia Tech's Grado Department of Industrial & Systems Engineering. Dr. Gabbard is also an executive committee member of Virginia Tech's Center for Human-Computer Interaction. Dr. Gabbard received his PhD and MS in computer science from Virginia Tech. Dr. Gabbard's research focuses on the connections between user interface design and human performance; and specifically the development of techniques to design and evaluate user interfaces for augmented reality and virtual environments. Gabbard has been a pioneer in usability engineering with respect to applying to, and creating methods for, new interactive systems for more than 20 years. With funding from a variety of sources, he has developed several methods for designing complex interactive systems and assessing their usability and impact on human performance, and disseminated this work in over 100 publications.

**Missie Smith** earned her B.S. and M.S. from Mississippi State University in 2010 and 2012, respectively. She earned her Ph.D. from Virginia Tech in 2018. Dr. Smith is currently an assistant professor at Oakland University, researching the impact of technology on users' perception, performance, and behaviors.

**Coleman Merenda** earned his B.S. in Industrial and Systems Engineering from Virginia Tech. He is currently a graduate research assistant pursuing his Ph.D.

**Gary Burnett** has a BEng from Warwick University and MSc and PhD qualifications from Loughborough University in the UK. He is presently a Professor in Transport Human Factors in the Faculty of Engineering at

the University of Nottingham - where he leads a team of researchers and PhD students focused on human factors and HCI issues for future vehicles. In Prof Burnett's career, he has published over 160 papers in peer-reviewed locations, including approximately 60 journal papers, and has been awarded approximately over £20M in funding.

**David R. Large** holds a BSc in Engineering from Durham University (1993) and an MSc in Information Technology (2007), followed by a PhD in Human Factors (2013), both from the University of Nottingham. He has spent time working within the manufacturing industry and is now a Senior Research Fellow with the Human Factors Research Group at the University of Nottingham. David has been involved in a broad range of industry, UK and internationally funded projects primarily concerning the design, evaluation and acceptance of novel and emerging in-vehicle human-machine interfaces (HMIs) and systems for both road and rail transport.